

**NOVEL ACOUSTIC EMISSION SIGNAL PROCESSING
METHODS FOR BEARING CONDITION MONITORING**

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Yanhui Feng

Department of Engineering

University of Leicester

May 2008

**NOVEL ACOUSTIC EMISSION SIGNAL PROCESSING
METHODS FOR BEARING CONDITION MONITORING**

by

Yanhui Feng

Declaration of Originality:

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Engineering, University of Leicester, UK. All work presented in this thesis is original unless otherwise acknowledged in the text or by references. No part of it has been submitted for any other degree, either to the University of Leicester or to any other University.

Signed:_____

Date:_____

Yanhui Feng

To my daughter Ruixue,
who brings us eternal happiness and hopes.

To my wife Yingning and my parents,
who give me forever supports behind.

NOVEL ACOUSTIC EMISSION SIGNAL PROCESSING METHODS FOR BEARING CONDITION MONITORING

Yanhui Feng

Abstract

Rolling Element Bearing is one of the most common mechanical components to be found in critical industrial rotating machinery. Since the failure of bearings will cause the machine to malfunction and may quickly lead to catastrophic failure of the machinery, it is very important to detect any bearing deterioration at an early stage. In this thesis, novel signal processing methods based on Acoustic Emission measurement are developed for bearing condition monitoring. The effectiveness of the proposed methods is experimentally demonstrated to detect and diagnose localised defects and incipient faults of rolling element bearings on a class of industrial rotating machinery – the iGX dry vacuum pump. Based on the cyclostationary signal model and probability law governing the interval distribution, the thesis proposes a simple signal processing method named LocMax-Interval on Acoustic Emission signals to detect a localised bearing defect. By examining whether the interval distribution is regular, a localised defect can be detected without *a priori* knowledge of shaft speed and bearing geometry. The Un-decimated Discrete Wavelet Transform denoising is then introduced as a pre-processing approach to improve the effective parameter range and the diagnostic capability of the LocMax-Interval method. The thesis also introduces Wavelet Packet quantifiers as a new tool for bearing fault detection and diagnosis. The quantifiers construct a quantitative time-frequency analysis of Acoustic Emission signals. The Bayesian method is studied to analyse and evaluate the performance of the quantifiers. This quantitative study method is also performed to investigate the relationships between the performance of the quantifiers and the factors which are important in implementation, including the wavelet order, length of signal segment and dimensionality of diagnostic scheme. The results of study provide useful directions for real-time implementation.

Acknowledgements

I am very grateful to my supervisor Dr. Fernando Soares Schlindwein for his great support and encouragement during my whole Ph.D. study. The discussions with him always enlighten me and push my research forward. He carefully reads any piece of my work and patiently helps me to correct mistakes in grammar. It is his insights, knowledge and experiences that guide me to find the way out when I almost got lost in maze.

I really thankful to my family who provides loyal supports in finance and spirit that help me go through each difficulty. My family always gives me comfort when I feel frustrated and encourages me when I hesitate. Special thanks are given to my wife, Dr. Yingning Qiu, without whom this thesis even this study will not have a chance to start. I am indebted to my parents Jinzhang Feng and Yanping Qi who are always supporting me from the very beginning of my life to overseas study.

I am thankful to a number of research staff and colleagues for their invaluable advice, assistance and discussions: Dr. Suguna Thanagasundram, Dr. Christos Kitsos and especially my project team leader Dr. John Twiddle; undergraduate students Fanbo Huang and Ihab Abou Rayan who assisted the development of analogue instrumentation. Thanks are given to other staff in the Engineering department who provided assistance and advice: Mr. Dipak Raval, Mr. Bilal Haveliwala, Mr. Luigi Alessandro, and Mr. Tom Robotham. Thanks are also given to many of the friends or housemates in Leicester: Dr. Royan Ong, Mr. Huawei Su, Miss Lei Dong, Miss Huijuan Dai, Dr. Zhou Chen, Mrs. Lanna Boyd-Russell, Dr. Teera Phatrapornnant, Dr. Halim Alwi, Mr. Jose Manuel Andrade Da Silva, Mr. Peter Vidler, and Mr. Zemian Hughes.

I would also like to thank BOC Edwards for the supply of iGX dry vacuum pumps. Special gratitude should go to Jeremy Watson, Larry Marini and Nick Tyler from BOC Edwards for their collaboration. I am thankful to Nick Dowding from Barden Bearings, who provided us the faulty ceramic ball bearings on which the experimental investigations were carried out.

Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Figures	vi
List of Publications	ix
Abbreviations	xi
Chapter 1 Introduction	1
1.1 Research Motivation	1
1.2 Research Objectives	2
1.3 Original Contributions	4
1.4 Thesis Organisation	5
Chapter 2 Background	7
2.1 Rolling Element Bearings	7
2.2 Bearing Kinematics and Characteristic Defect Frequency	8
2.3 Premature Bearing Failures	10
2.4 Measurements for Bearing Condition Monitoring	11
2.5 Summary	12
Chapter 3 Experimental Setup	14
3.1 Introduction	14
3.2 Dry Vacuum Pump	15
3.3 Mounting Acoustic Emission Transducer	17
3.4 Bearing Faults for Investigation	18
3.5 Instrumentation and Data Acquisition.....	20
3.6 Summary	21
Chapter 4 Literature Review	22
4.1 Introduction	22
4.2 Time Domain Methods	25
4.3 Frequency Domain Methods	27
4.4 Envelope analysis	29
4.5 Cyclostationary Analysis.....	32
4.6 Wavelet Methods	37
4.6.1 Single-scale Feature Extraction	38
4.6.2 Multi-scale Denoising and Singularity Detection	40

4.7	Summary	41
Chapter 5	A New Method for Localised Defect Detection	45
5.1	Introduction	45
5.2	Signal Model	47
5.3	Methodology.....	47
5.3.1	An Algorithm for Finding out Local Maxima	47
5.3.2	Interval Distribution.....	50
5.4	Simulations	51
5.4.1	The Parameter of Algorithm.....	52
5.4.2	Noisy Cyclostationary Process	52
5.4.3	Denoising to Improve Effective Range of Parameter Values.....	54
5.4.4	White Noise Process and Zero Process	56
5.5	Results and Discussion.....	56
5.6	Summary	62
Chapter 6	Application of Un-decimated Discrete Wavelet Transform	64
6.1	Introduction	64
6.2	Signal Models	66
6.3	Methodology.....	67
6.3.1	UDWT Decomposition	67
6.3.2	Significance Measure.....	67
6.3.3	UDWT Denoising	68
6.4	Results and Discussion.....	70
6.5	Summary	78
Chapter 7	Application of Wavelet Packet Quantifiers and Bayesian Method ..	80
7.1	Introduction	80
7.2	Methodology.....	81
7.2.1	WP Quantifiers	81
7.2.2	Bayesian Method	83
7.3	Analysing AE signals via WP Quantifiers	85
7.4	Finding out the Best Quantifier via Bayesian Method	90
7.5	Performance Factor 1: Wavelet Order	92
7.6	Performance Factor 2: Length of Signal Segment	93
7.7	Performance Factor 3: Dimensionality of Diagnostic Scheme	94
7.8	Discussion	96

7.9	Summary	98
Chapter 8	Discussion and Future Work	99
8.1	Discussion	99
8.2	Future Work.....	101
Chapter 9	Conclusions.....	103
Bibliography	104
Appendix A	Wavelet Methods	112
A.1	Introduction to Wavelet Representation	113
A.2	Beyond the CWT: DWT, UDWT and DWPT	116
A.3	Multiresolution Analysis.....	119
A.4	Discrete Wavelet Transform	121
A.5	Un-decimated Discrete Wavelet Transform.....	124
A.6	Discrete Wavelet Packet Transform.....	127
A.7	Wavelet Basis.....	129
A.8	Wavelet Thresholding.....	131
A.9	Wavelet for Singularity Detection	134
A.10	Wavelet or WP Quantifiers: Relative Energy and Entropy	135
A.11	Summary	137
Appendix B	A Localised Defect Detection using ACF.....	139
B.1	Introduction	139
B.2	Autocorrelation Function.....	139
B.3	Detection Scheme.....	140
B.4	Experimental Setup and Signal Processing.....	141
B.5	Results and Summary	141
Appendix C	MATLAB Codes	143
C.1	Cyclostationary Signals Simulation	143
C.2	LocMax Algorithm and Interval Calculation.....	145
C.3	UDWT Denoising Algorithm	146
C.4	Wavelet Packet Quantifiers	149
C.5	One Dimensional Bayesian Classification	151
C.6	Two Dimensional Bayesian Classification	153

List of Figures

Figure 1-1	The BFDD system using acoustic and vibration measurements.	3
Figure 2-1	A single-row, deep-groove bearing having a shield.....	7
Figure 2-2	Left: Bearing structure; Right: Shaft and bearing system with rotating inner race and fixed outer race.	7
Figure 3-1	The BOC Edwards iGX dry vacuum pump is used as the test bed.....	15
Figure 3-2	The combined Roots and claw mechanism in dry vacuum pump (Harris 2001). 16	
Figure 3-3	The AE Transducer is firmly held at the surface of the pump housing.	17
Figure 3-4	The scale of the bearing used for investigation.....	18
Figure 3-5	Artificially made indent on inner race of bearing.....	19
Figure 3-6	The seeded indent is approximately 2 mm wide and 2 mm deep.....	19
Figure 3-7	Schematic diagram of the complete data acquisition system.	20
Figure 4-1	Diagram of basic processing used in envelope analysis of vibration signals. ...	30
Figure 4-2	Cyclic spectrum of a cyclostationary process (Antoni 2007).....	33
Figure 4-3	Relationship between cyclostationary analysis and envelope analysis (Randall et al. 2001).	34
Figure 5-1	Local maxima Detection using LocMax algorithm.	48
Figure 5-2	Original distribution for interval (left) and generated interval distribution (right).....	50
Figure 5-3	Comparison between the proposed LocMax-Interval method and envelope analysis method.....	51
Figure 5-4	Generated interval distribution from pure cyclostationary process using different parameter values.	53
Figure 5-5	Generated interval distribution from noisy cyclostationary process.	54
Figure 5-6	Generated interval distribution from denoised cyclostationary process.....	55
Figure 5-7	Generated interval distributions from white noise process (left), approximate zero process (middle) and zero process (right).....	56
Figure 5-8	Generated interval distributions from raw AE signals of localised defect.....	57
Figure 5-9	Generated interval distributions from AE Signals of localised defect using wavelet multi-scale denoising methods.....	58
Figure 5-10	Generated interval distributions from raw AE signals of fault free bearing.....	60
Figure 5-11	Generated interval distributions from AE Signals of fault free bearing using wavelet multi-scale denoising methods.....	61
Figure 6-1	The method based on UDWT decomposition and denoising for BFDD.....	70
Figure 6-2	Raw AE signal of localised defect (first row) and UDWT-SURE denoised coefficients at two scales.....	71
Figure 6-3	Clear timings of impact signals are revealed at two scales after the LocMax detection algorithm is performed on denoised coefficients.....	72

Figure 6-4	Interval distribution from AE signals of localised defect using UDWT-SURE denoising and LocMax-Interval method.	73
Figure 6-5	Interval distribution from AE signals of incipient contamination fault using UDWT-SURE denoising and LocMax-Interval method.	74
Figure 6-6	Interval distribution from AE signals of fault free bearing using UDWT-SURE denoising and LocMax-Interval method.	74
Figure 6-7	Interval distribution from AE signals of localised defect using UDWT-SURE denoising and LocMax-Interval method (the parameter of LocMax algorithm is set to 60, 80, 90, 100).	76
Figure 6-8	Comparison of interval distributions using UDWT and DWT denoising as pre-processing for LocMax-Interval Method.	77
Figure 7-1	2-level DWPT decomposition on AE signals (frequency ordering).	82
Figure 7-2	The histogram of RE of A band for three classes ($L = 2048, db10$).	85
Figure 7-3	The histogram of RE of B band for three classes ($L = 2048, db10$): Different classes are well separated.	86
Figure 7-4	The histogram of RE of C band for three classes ($L = 2048, db10$).	87
Figure 7-5	The histogram of TE for three classes ($L = 2048, db10$).	88
Figure 7-6	The histogram of NE in band C for three classes ($L = 2048, db10$).	89
Figure 7-7	The normalised WP coefficients in band C ($L = 2048, db10$).	90
Figure 7-8	Total classification error probability of the quantifiers using different Daubechies wavelet orders ($N = 1, 2, \dots, 10$) ($L = 2048$).	92
Figure 7-9	Total classification error probability increases when the length of signal segment decreases.	93
Figure 7-10	The histogram of quantifier F2 for three classes using Haar wavelet ($L = 2048$); the total classification error probability is very high at 27.6%.	94
Figure 7-11	Two-dimensional diagnostic map using two quantifiers F2 and F4 (Haar wavelet; $L = 2048$): Total classification error probability has greatly reduced. ...	95
Figure 7-12	Using two-dimensional diagnostic scheme reduces the classification error probability; improvement is significant for lower wavelet order ($L = 2048$). ...	96
Figure A-1	Modulated burst signals on carrier (left), Y-axis is amplitude of arbitrary unit, and Fourier spectrum (right).	114
Figure A-2	Elementary functions for STFT (left) and CWT (right) to analyse the signal (Vetterli and Kovacevic 1995).	115
Figure A-3	High frequency bursts (top) are highlighted in scalogram (bottom). Scale parameters are 1: 32.	116
Figure A-4	Elementary functions and frequency (or scale) representations for different transforms.	117
Figure A-5	Sampling of time-scale plane for DWT (left) and UDWT (right). Circle, square and triangle represent wavelet coefficients at dyadic scale 1, 2, 3, respectively. ...	118
Figure A-6	Relationship between different transform variations.	119

<i>Figure A-7 Divisions of frequency domain for approximation and detail spaces.</i>	<i>120</i>
<i>Figure A-8 Filter banks computation for DWT (3-level decomposition).....</i>	<i>123</i>
<i>Figure A-9 Filter banks computation for UDWT (3-level decomposition).....</i>	<i>126</i>
<i>Figure A-10 Decomposition structures of DWT (left) and DWPT (right).</i>	<i>128</i>
<i>Figure A-11 Examples of Daubechies wavelets db N (N is the number of vanishing moments).</i>	<i>130</i>
<i>Figure A-12 Biorthogonal scaling and wavelet functions of bior 6.8.....</i>	<i>131</i>
<i>Figure A-13 Signal denoising via wavelet multi-scale thresholding and reconstruction.</i>	<i>132</i>
<i>Figure B-1 The histogram of the positions found using the AE signals.....</i>	<i>142</i>
<i>Figure B-2 The histogram of the positions found using the vibration signals.....</i>	<i>142</i>

List of Tables

<i>Table 3-1 The bearing specifications for iGX dry vacuum pump.....</i>	<i>18</i>
<i>Table 4-1. Comparison of different time-frequency analysis methods (Peng and Chu 2004)..</i>	<i>37</i>
<i>Table 5-1 A recursive algorithm LocMax to detect local maxima.</i>	<i>49</i>
<i>Table 7-1. Classification error probability using different quantifiers (L = 2048 and db10). </i>	<i>91</i>

List of Publications

Main Publications

Y. Feng, F. S. Schlindwein, “Normalised Wavelet Packets Quantifiers for Condition Monitoring,” *Mechanical Systems and Signal Processing*, In Press, Corrected Proof, Available online 24 July 2008

<http://dx.doi.org/10.1016/j.ymssp.2008.07.002>.

Y. Feng, F. S. Schlindwein, “Discriminating Contamination Fault and Localised Bearing Defect using Acoustic Emission Signal,” Presentation on WCEAM-CM Conference, Harrogate, United Kingdom, June 2007a.

Y. Feng, F. S. Schlindwein, “Improving Detectability on Localised Bearing Defect using Acoustic Emission Signal,” Proceeding of WCEAM-CM Conference, Harrogate, United Kingdom, June 2007b.

Y. Feng, S. Thanagasundram, F. S. Schlindwein, “Discrete Wavelet-based Thresholding study on Acoustic Emission Signals to Detect Bearing Defect on A Rotating Machine,” Proceeding of ICSV13 Conference, Vienna, July 2006.

Y. Feng, F. S. Schlindwein, “Processing Acoustic emission signals from dry vacuum pumps for fault diagnostics using wavelets” Presentation on 2nd Festival of Postgraduate Research, Leicester, June 2006.

Joint Publications

Jiang, W, Spurgeon, S K, Twiddle, J A, Schlindwein, F S, **Feng, Y**, Thanagasundram S, “A Wavelet Cluster-based Band-pass Filtering and Envelope Demodulation Approach with Application to Fault Diagnosis in a Dry Vacuum Pump”, Proceedings of the I MECH E Part C Journal of Mechanical Engineering Science, Vol. 221, No. 11, pp.1279-1286(8), 2007.

S. Thanagasundram, Gurung K.R., **Y. Feng**, F. S. Schlindwein, “AR pole trajectory in condition monitoring studies”, Proceeding of ICSV13 Conference, Vienna, July 2006.

S. Thanagasundram, Gurung K.R., **Y. Feng**, F. S. Schlindwein, “Exploring pole positions for fault diagnostics from dry vacuum pumps”, Presentation on 2nd Festival of Postgraduate Research, Leicester, June 2006.

S. Thanagasundram, **Y. Feng**, I. S. M. Abou Rayan and F. S. Schlindwein, “A case study of Autoregressive Modelling and Order Selection for a Dry Vacuum Pump”, Proceeding of ICSV12 Conference, Lisbon, July 2005.

S. Thanagasundram, **Y. Feng**, F. S. Schlindwein, “A LabView based system for condition monitoring of a dry vacuum pump using AR modelling techniques”, Presentation on 1st Festival of Postgraduate Research, Leicester, June 2005.

Abbreviations

ACF	Autocorrelation Function
AE	Acoustic Emission
AI	Artificial Intelligent
ANN	Artificial Neural Networks
AR	Autoregressive
BFDD	Bearing Fault Detection and Diagnosis
BSF	Ball Spin Frequency
BPFI	Ball Pass Frequency of Inner race
BPFO	Ball Pass Frequency of Outer race
CDF	Characteristic Defect Frequency
CDI	Characteristic Defect Interval
CSD	Cone-Shape Distribution
CWD	Choi-Williams Distribution
CWT	Continuous Wavelet Transform
DWPT	Discrete Wavelet Packet Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiograph
EEG	Electroencephalogram
FFT	Fast Fourier Transform
FTF	Fundamental Train Frequency
FWT	Fast Wavelet Transform
HFRT	High Frequency Resonance Technique
IC	Internal Combustion

ICA	Independent Component Analysis
MLE	Maximum Likelihood Estimation
MRA	Multiresolution Analysis
NE	Node Entropy
RE	Relative Energy
<i>rms</i>	root mean square
SANC	Self-Adaptive Noise Canceller
SK	Spectral Kurtosis
SNR	Signal to Noise Ratio
STFT	Short Time Fourier Transform
SURE	Stein Unbiased Risk Estimate
SVD	Singular Value Decomposition
TE	Total Entropy
UDWT	Un-decimated Discrete Wavelet Transform
WMC	Wavelet Maxima Chain
WP	Wavelet Packet
WVD	Wigner-Ville Distribution

Chapter 1

Introduction

1.1 Research Motivation

The Rolling Element Bearing* is a class of bearing in which the main load is transferred through elements in rolling contact. It is found in broad applications from very simple commercial devices to complex engineering machinery. Particularly, bearing is one of the most common mechanical components to be found in the critical industrial rotating machinery, e.g. gas turbine engines, rocket engines, gearboxes and dry vacuum pumps. The safety and reliability of the critical industrial rotating machinery are vital in industry. Since premature bearing failures will cause machinery malfunction and may quickly lead to catastrophic failure of the machinery, it is important to monitor bearings on-line and predict when maintenance should be performed. The main motivation of this Ph.D. study is to investigate condition monitoring systems for predicting premature bearing failures at an early stage.

Previous research works in University of Leicester based on vibration measurements (accelerometers) have achieved success on the detection of a localised bearing defect, which were reported in detail by Thanagasundram (2007). Besides the localised indent defect on bearing inner race, as reported before, the incipient contamination fault of bearing will be included in this study. When contaminants have just entered a bearing, the symptom is hidden and the fault at this incipient stage is difficult to detect by using traditional vibration measurements. This Ph.D. study attempts to adopt Acoustic Emission (AE) as a measurement method to provide a solution to this class of detection problem.

* Rolling Element Bearing is abbreviated to bearing in this thesis.

1.2 Research Objectives

The subject of bearing condition monitoring can be divided up into three main areas, namely bearing fault detection, fault diagnosis and fault prognosis (Howard 1994). Detection concerns the question of whether a specific fault is present on the bearing; diagnosis is concerned with the determination of the type of fault; prognosis is concerned with the estimation of the remaining life of the faulty bearing.

This thesis is mainly about Bearing Fault Detection and Diagnosis (BFDD). Some background knowledge about bearings and BFDD will be given in the following chapters. A BFDD system based on signal processing usually follows three stages: sensing, signal processing, fault detection and diagnosis. Sensing describes the stage of setting up the measurement system, including the selection and installation of transducers, signal conditioning and data acquisition (Analogue to Digital conversion). The time of computation and the cost of implementation are major limitations for many advanced fault detection and diagnosis systems to be applied in practice. In many cases, the reduction of features and the selection of appropriate features are important engineering problems.

These problems could be alleviated by choosing suitable signal processing methods. The main role of signal processing is to perform the transform and algorithm to extract characteristic information from the acquired signals. The term “signal processing” in the thesis encompasses both the pre-processing for denoising signals and the analysis for extracting features from signals. This stage is essential to the whole system. Appropriate signal processing methods can reduce the data to be handled, highlight characteristic signals and provide the optimal feature(s), which are all important for a BFDD system. See Figure 1-1 for a type of BFDD method using acoustic and vibration measurements (including sound, vibration and Acoustic Emission measurement) which are the most popular measurements for BFDD because of the non-destructivity, high sensitivity and cost effectiveness (Tandon and Choudhury 1999). The challenges for this type of BFDD system are:

- First, the capability of the system to cope with noise. Such noise may derive from the mechanical components, from the instrumentation, from internal electronics and also include computational noise introduced by signal processing.
- Second, the capability of the system to operate with as little *a priori* knowledge as possible. Such knowledge may include the precise shaft speed and bearing geometry from which the Characteristic Defect Frequencies (CDFs) of bearings can be estimated.
- Third, the capability of the system to detect and diagnose bearing deterioration at an early stage (i.e., the incipient bearing faults, such as the incipient contamination fault discussed in this thesis).

To meet these challenges, it is necessary to apply appropriate signal processing methods. Some classical signal processing methods used for BFDD will be reviewed in Chapter 4.

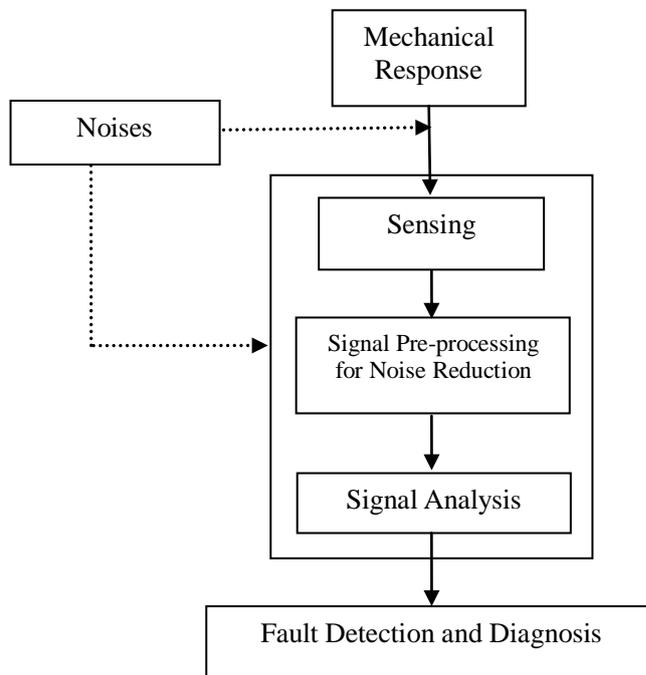


Figure 1-1 The BFDD system using acoustic and vibration measurements.

The main objectives of this thesis are:

- Study whether AE measurement is suitable for BFDD, particularly for the detection of localised defects and incipient faults.
- Develop suitable signal processing methods for BFDD based on AE signals.
- Demonstrate the effectiveness of the proposed BFDD systems on the iGX dry vacuum pump.
- Find an effective method to evaluate the performance of different signal processing approaches for BFDD and their parameters.

The rich information contained in the broadband high-frequency AE signals might provide new opportunities for the detection of bearing faults at different stages. This research is very promising for the future applications along with the reduction of AE transducer price. Two bearing faults of dry vacuum pump are studied to demonstrate the effectiveness of the proposed methods: the localised indent defect on bearing inner race which represents localised defects, and the incipient contamination fault which represents incipient faults.

1.3 Original Contributions

The thesis proposes novel signal processing methods on AE signals for BFDD:

1. Based on the cyclostationary signal model and probability law governing the interval distribution, the thesis proposes a simple signal processing method on AE signals to detect a localised bearing defect. The new method is composed of a new algorithm for finding out the timings of impact signals and the interval distribution for detection or diagnosis. By examining whether the interval distribution is regular, a localised defect can be detected without *a priori* knowledge of CDFs. The method, named LocMax-Interval, will be shown in chapter 5.
2. The thesis studies the application of Un-decimated Discrete Wavelet Transform (UDWT) in BFDD. The UDWT denoising is introduced as the pre-processing approach for AE signals. The proposed UDWT denoising

approach can improve the effective parameter range of the LocMax-Interval method. Also, it enables the method to detect incipient contamination bearing fault. This study will be shown in chapter 6.

3. The thesis introduces the first study on the application of Wavelet Packet (WP) quantifiers in BFDD. The WP quantifiers construct a quantitative time-frequency analysis of AE signals: Relative Energy (RE) measures the normalised energy of the WP node; Total Entropy (TE) measures how the normalised energies of the WP nodes are distributed in the frequency domain; Node Entropy (NE) describes the uncertainty (i.e. the degree of disorder) of the normalised coefficients of the WP node. When using on AE signals, these quantifiers are powerful and particularly useful for diagnosing multiple bearing faults. The Bayesian method is also studied to analyse and evaluate the performance of the quantifiers. The best quantifier can be found by means of comparing the Bayesian classification error probabilities. This study will be reported in chapter 7.
4. The Bayesian method also finds application on the implementation of a BFDD system using WP quantifiers. Chapter 7 will show the first investigation on relationships between the quantifiers and some important factors in implementation. The study provides useful directions for real-time implementation.

1.4 Thesis Organisation

Chapter 2 provides a broad overview of the research background, including some basic knowledge about bearings, premature bearing failures and measurements for BFDD. Chapter 3 introduces the details of the test bed (i.e. the iGX dry vacuum pump), transducer mounting, bearing faults, instrumentation and data acquisition. Chapter 4 reviews the classical applications of signal processing methods in BFDD, including the time domain methods, frequency domain methods and envelope analysis. The limitations of these methods are discussed in detail. The cyclostationary analysis and wavelet methods which have attracted increasing interest are also reviewed in detail.

Chapters 5, 6 and 7 detail the development of the new methods and present the results. Chapter 5 proposes a novel method called LocMax-Interval method to detect localised bearing defects. A difference between the proposed method and the traditional envelope analysis method is illustrated for comparison. Both simulations and experimental data are studied to test the new method. Finally, the Discrete Wavelet Transform (DWT) denoising method is implemented to improve the performance of the LocMax-Interval method. Chapter 6 introduces the UDWT decomposition and denoising as a pre-processing approach for the LocMax-Interval method. The signal models for the bearing faults are described in detail and the signal processing diagram is depicted. The applications of WP quantifiers and Bayesian method for BFDD are studied in chapter 7. The quantitative study method is also highlighted in this chapter.

Finally, discussion and recommendations for future research work are presented in chapter 8. The conclusions are made in chapter 9. Appendix A briefly introduces the principles of wavelet methods, the differences between the transform variations, the mathematical descriptions and filtering conventions. Sections A.8-A.10 review some important wavelet techniques. This chapter is the essential background reading for the techniques used in this thesis. Appendix B reports the study of detection of a localised bearing defect using autocorrelation function. Appendix C lists the MATLAB codes used for signal processing.

Chapter 2

Background

2.1 Rolling Element Bearings

A bearing generally consists of four parts: an inner ring, an outer ring, a set of rolling elements (balls or rollers) and a cage. The most popular types of bearing are ball and roller bearings (Harris 1966). Figure 2-1 shows a ball bearing with nine rolling elements and Figure 2-2 sketches the bearing structure.



Figure 2-1 A single-row, deep-groove bearing having a shield.

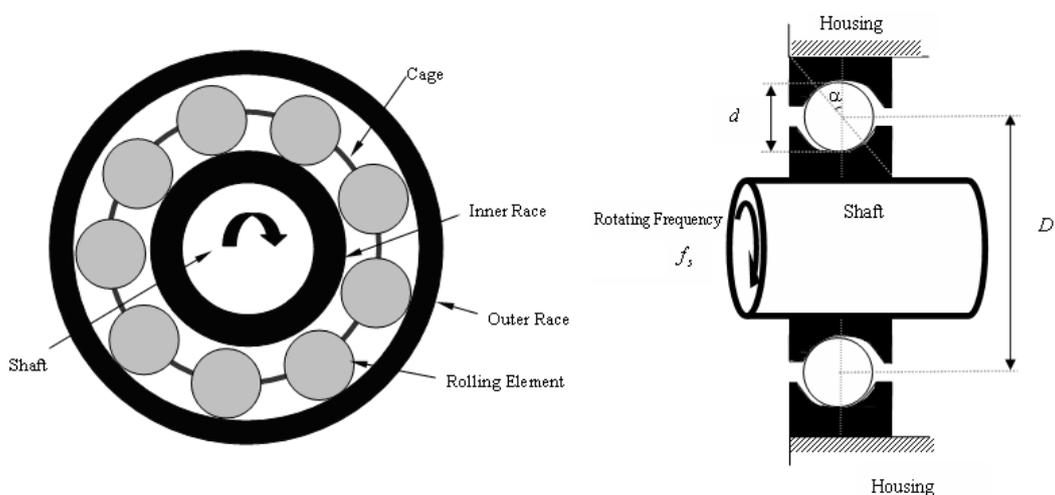


Figure 2-2 Left: Bearing structure; Right: Shaft and bearing system with rotating inner race and fixed outer race.

The rolling elements are used to maintain the motion between the fixed supporting structure (i.e. housing) and the rotating shaft. See Figure 2-2. The rolling elements are held in an angularly spaced relationship by the cage. The inner race is mounted on the rotating shaft, the outer race is fixed on the housing (Harris 1966).

2.2 Bearing Kinematics and Characteristic Defect Frequency

Bearing kinematics is essential for understanding localised defects detection because it determines the rotational velocities of the bearing elements with respect to each other and the theoretical Characteristic Defect Frequencies.

Howard (1994) gives the following formulas and related details. The race diameter of the bearing can be expressed in terms of the pitch circle diameter D , contact angle α and roller diameter d to give

$$D_i = D - d \cos \alpha, \quad (2-1)$$

and
$$D_o = D + d \cos \alpha, \quad (2-2)$$

where D_i is the inner race diameter and D_o is the outer race diameter.

The circumferential velocity of the cage V_c is the average of the velocity of inner race V_i and outer race V_o assuming no slip occurs:

$$V_c = \frac{V_i + V_o}{2}, \quad (2-3)$$

For a bearing with fixed outer race, the velocity of the cage is half of the velocity of the inner race:

$$V_c = \frac{V_i}{2} = \frac{\omega_i \frac{D_i}{2}}{2}. \quad (2-4)$$

The cage frequency is the circumferential velocity dividing by πD :

$$f_c = \frac{V_c}{\pi D} = \frac{\omega_i D_i}{4\pi D} = \frac{f_i D_i}{2D}. \quad (2-5)$$

By substituting equation (2-1) into (2-5), the cage frequency becomes:

$$f_c = \frac{f_i \left(1 - \frac{d}{D} \cos \alpha\right)}{2}. \quad (2-6)$$

When the inner race is fixed on the shaft, the inner race frequency is the same as the shaft frequency, the Cage Frequency (i.e. Fundamental Train Frequency, FTF) in equation (2-6) becomes:

$$f_c = \frac{f_s \left(1 - \frac{d}{D} \cos \alpha\right)}{2}. \quad (2-7)$$

The frequency of rotation of a single rolling element with respect to the cage frequency is:

$$f_{iz} = f_s - f_c = \frac{f_s \left(1 + \frac{d}{D} \cos \alpha\right)}{2}. \quad (2-8)$$

The frequency of rotation of Z rolling elements with respect to the cage frequency, named Ball Pass Frequency of Inner race (BPFI), is f_{iz} multiplied by Z:

$$f_{bpfi} = Z f_{iz} = \frac{Z f_s \left(1 + \frac{d}{D} \cos \alpha\right)}{2}. \quad (2-9)$$

Likewise, the Ball Pass Frequency of Outer race (BPFO) is f_c multiplied by Z:

$$f_{bpfo} = Z f_c = \frac{Z f_s \left(1 - \frac{d}{D} \cos \alpha\right)}{2}. \quad (2-10)$$

The Ball Spin Frequency (BSF) is given by f_{iz} multiplied by the ratio of inner race diameter to the ball diameter:

$$f_{bsf} = f_{iz} \frac{D_i}{d} = \frac{Df_s}{2d} \left(1 - \frac{d^2}{D^2} \cos^2 \alpha\right). \quad (2-11)$$

The f_c , f_{bphi} , f_{bphi} and f_{bsf} are the well known Characteristic Defect Frequencies (CDFs) which form the basis for the detection of localised bearing defects (Howard 1994, Li and Ma 1997, Tandon and Choudhury 1999, McInerny and Dai 2003, Randall 2004). If the shaft speed and bearing geometry are known, the CDFs can be theoretically estimated by the above equations. When the defective part of the bearing makes contact with the surface of its mating part under load, an impulse impact is generated and captured by the transducer at a nearby location on the casing of the machine. If the machine rotates at a certain steady speed, the repetitive impacts will be generated with an equal interval which characterises a specific type of localised defect. The envelope signal, which is the outline of the impact signal, can be achieved after the rectification and smoothing of the impact signal (Randall 2004). See Section 4.4 for more details.

The characteristic impact interval is termed Characteristic Defect Interval (CDI) in the time domain and the Characteristic Defect Frequency (CDF) in the frequency domain. The CDI (expressed as the number of discrete sampling points) is the sampling frequency f_w divided by CDF:

$$CDI = \frac{f_w}{CDF}, \quad (2-12)$$

2.3 Premature Bearing Failures

The most common failure modes of premature bearing failures are fatigue and wear (Howard 1994). The failures may be caused by plastic deformation, corrosion, brinelling, poor lubrication, contamination, improper installation and incorrect design. The premature bearing failures will quickly lead to machine malfunction, even catastrophic failure. As a result, it is very important for an advanced condition monitoring system to be capable to find out any bearing deterioration problem as early as possible.

Bearing defects are usually classified into distributed defects and localised defects (Tandon and Choudhury 1999). Surface roughness, waviness, misaligned races and off-size rolling elements are included in the class of distributed defects. This class of defects is caused by incorrect design, improper installation or wear. Cracks, pits, spalls and indents caused by fatigue on the rolling surfaces are included in the class of localised defects.

Solid contamination is one of the main failure root causes of premature bearing contact fatigue (Ville *et al.* 2006). About 20% of premature bearing failures are caused by solid contamination (FAG Bearings Corp.). Because of the harsh environments present in most industrial settings, contaminants such as dirt and foreign particles entering bearings through inadequate sealing commonly contaminate bearing lubrication and cause damage. The effects of contamination in rolling bearings can be varied, from abrasive wear, when a significant concentration of abrasive debris is present causing surface roughness, to the effects of a small number of particulates in relatively clean lubricant systems, causing localised surface defects such as dents, scuffing and melting (Maru *et al.* 2007). When contaminants have just entered a bearing, serious damage may not have developed. In comparison to localised defects, the incipient contamination problem is more difficult to detect because the symptom is hidden at this stage.

2.4 Measurements for Bearing Condition Monitoring

Many measurements have been proposed for bearing condition monitoring, such as acoustic and vibration measurement, temperature measurement and wear debris analysis. Acoustic and vibration (including sound, vibration and Acoustic Emission measurements) are the most popular measurements for bearing condition monitoring because of the non-destructivity, high sensitivity and cost effectiveness (Tandon and Choudhury 1999). Using this type of measurement, it is then possible to continuously monitor the bearing condition of an industrial rotating machine without the need to interrupt the machine operation and production process. Previous research works based on vibration measurements

(accelerometers with frequency range 0-10 kHz) were reported by Thanagasundram (2007). The signal acquired via vibration measurement is called vibration signal in this thesis.

Acoustic Emission (AE) transducers (Howard 1994, Choudhury and Tandon (2000), Holroyd 2003, Mba and Rao 2006) have received increasing attention for use in bearing condition monitoring. In the 1960s, Josef Kaiser discovered that the AE technique was capable of detecting elastic waves caused by microscopic deformation when material releases its energy (Holroyd 2003). The AE technique has been applied on many areas, including structural testing and surveillance, process monitoring and control, and materials characterisation (Scruby 1987). Holroyd (2003) introduced the general knowledge, condition monitoring applications and equipments for AE technique in his book. Recently, Mba and Rao (2006) reviewed the research and application of developing AE techniques to condition monitoring of industrial rotating machines and components, including bearings, pumps, gearboxes, engines and other rotating structures.

Different types of AE transducer covering different frequency ranges of interest are now available in the market. Study suggests a moderate frequency band 10-40 kHz to be used for bearing condition monitoring of a gear test rig (Howard 1994): The low frequency band 0-10 kHz is dominated by the gear mesh harmonics; the mid frequency range 10-20 kHz contains some gear mesh harmonics and a structural resonance; the high frequency range 20-40 kHz is dominated by structural resonances. Following that study, a PAC R3 α AE transducer will be used in this study to provide a cost-effective solution and the frequency range of interest is between 10 kHz and 40 kHz, which is beyond that used by traditional accelerometers for vibration measurement (typically 0-10 kHz as reported in Thanagasundram 2007). The acquired AE signals are unlikely to be affected by the low frequency noise generated by other mechanical components.

2.5 Summary

The bearing is one of the most critical mechanical components in industrial

rotating machinery. Section 2.1 introduced the bearing geometry. Section 2.2 introduced the bearing kinematics and Characteristic Defect Frequency (CDF) which are essential for understanding the localised bearing faults detection. Since premature bearing failure will quickly cause machine malfunction, even catastrophic failure, it is very important to detect any bearing deterioration at an early stage. Section 2.3 introduced the bearing faults. Finally, AE technique and Section 2.4 briefly reviewed AE transducers. Next chapter will introduce the details of experimental setup.

Chapter 3

Experimental Setup

3.1 Introduction

The acoustic and vibration measurements are highly sensitive, but these are external measurements in which the signal attenuation and the presence of interferences from other components are unavoidable. When the internal forces and stresses propagate through the structure of the machine, they will be attenuated at joints and interfaces, damped, modified by the frequency response of the system and then measured by the transducer. These transmission path effects mean that the acoustic and vibration signal which is finally measured by the transducer will be a poor response of the original forcing function, particularly in machinery with complex structure (Howard 1994).

In practice, the manner of the transducer mounting and the complexity of the machine structure will have impacts on the quality of acquired signals, and then the system used for detection and diagnosis. An effective fault detection and diagnosis system often depends on the nature of signal acquired from the transducer and it may vary from one class of machine to another. In some laboratory situations where the transmission is quite effective and the signals acquired from the bearing test rigs are almost free from interferences, the measurement setup is a simple task and the detection system is easier to design. However, most of the industrial rotating machines are not that simple.

The following sections will introduce the experimental platform, the mounting method for the Acoustic Emission transducer, the details of the bearing faults under investigation, and the instrumentation and data acquisition system.

3.2 Dry Vacuum Pump

The BOC Edwards iGX dry vacuum pump has been used as the test bed for studying bearing faults. See Figure 3-1. The iGX dry vacuum pump has found crucial application in the semiconductor industry. Most of the processes in the semiconductor industry have in common the need to admit one or more process gases into the vacuum chamber for the generation of plasma and the promotion of any required chemical reactions. Pumping systems for these processes may therefore have to accommodate process gases, gaseous reaction products and particulates, most of which may be explosive, flammable, aggressive, corrosive or toxic, while solids are likely to be abrasive. The oil-free (dry) design of vacuum pump greatly reduces the possible occurrences of oil related problems, such as partial contamination-sealing or lubricating fluids reacting with the materials in the processes or retaining them. Detailed knowledge on dry vacuum pump is given by Harris (2001).

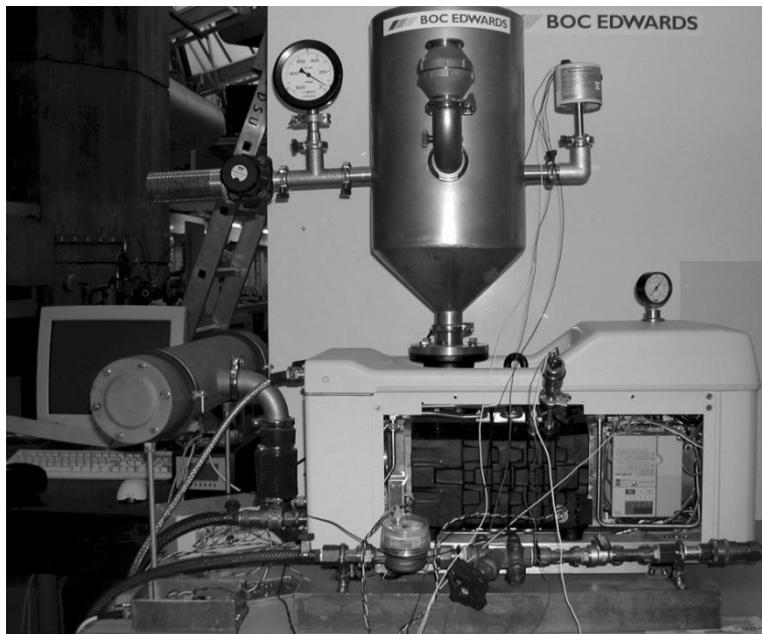


Figure 3-1 The BOC Edwards iGX dry vacuum pump is used as the test bed.

Any unexpected failure of the dry vacuum pump can be very costly for the

large-scale production process. A single failure incident might potentially cost over US \$100,000 in the semiconductor industry. Due to the important role played by the iGX dry vacuum pump in the semiconductor industry, the research outcomes are of important practical value.

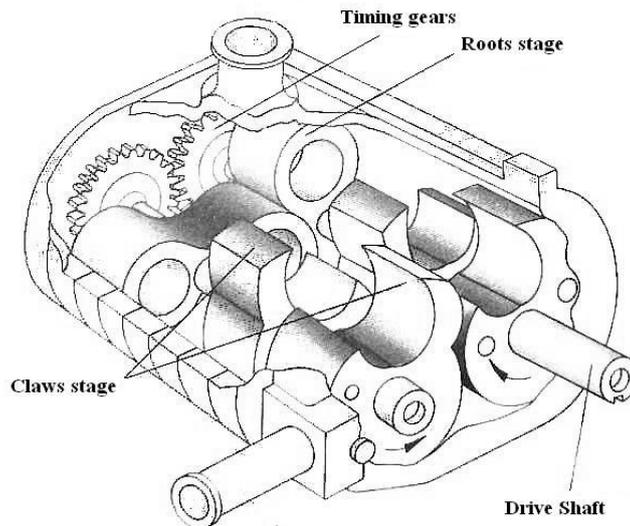


Figure 3-2 The combined Roots and claw mechanism in dry vacuum pump (Harris 2001).

The iGX dry vacuum pump consists of a pump housing, a pair of parallel shafts extending throughout the pump housing for supporting the rotary movement of rotors, timing gears for synchronising the movement of rotors, rotors arranged in complementary pairs, ball bearings at the inlet and outlet to support the shafts (Harris 2001). A combined Roots and claw mechanism is used in the pump to achieve optimal performance at both lower and higher pressures. See Figure 3-2 for the combined Roots and claw mechanism. The one stage Roots are devices which interlock, synchronise and rotate in opposite directions. The pair moves past each other and the stator wall with a small clearance. The true compressors are the following four stage claws. When the non-contact claw rotors rotate, gas is drawn in via an inlet slot, which matches the cavity in one of the rotors. Continued rotation closes the inlet while the claws compress the trapped volume

of gas until the cavity in the second rotor exposes the outlet or exhaust slot. In the iGX pump, the second claw stage of the pump is reversed so that the outlet of the first stage is directly in line with the inlet of the second stage. This creates a short path that allows direct transfer of gas.

3.3 Mounting Acoustic Emission Transducer

The suitable mounting point for a vibration transducer has been identified by the previous study (Thanagasundram 2007). A convenient means for obtaining signals of good quality is by holding the transducer at the surface of the pump housing in the radial direction of movement and as near as possible to the position of the bearing. The same position is used for mounting the Acoustic Emission Transducer to achieve good signal quality. See Figure 3-3. When the transducer is mounted at the appropriate position, the mechanical forces and stresses can be effectively transmitted from the bearing to the surface of the pump housing.

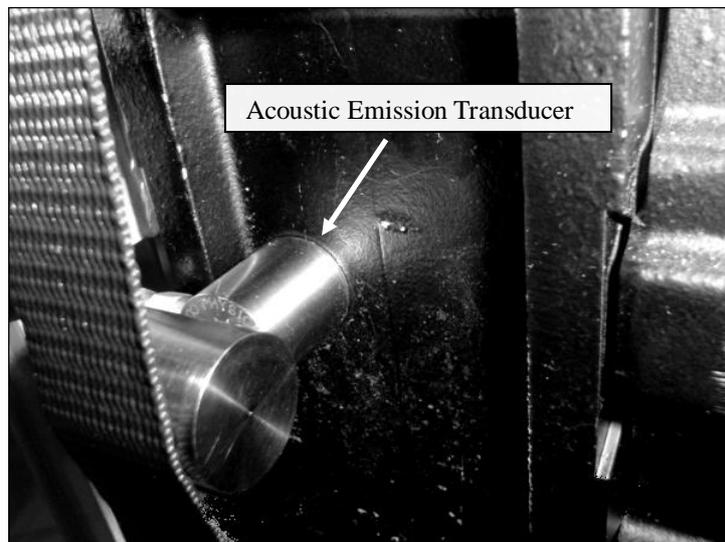


Figure 3-3 The AE Transducer is firmly held at the surface of the pump housing.

The high frequency stress waves are of small amplitudes and they will be attenuated at the surface of machine structure. Therefore, the transducer needs to be firmly held at the surface of the pump housing to ensure good enough signal quality. Because the repeated compressions can raise the temperature of the

surface of pump housing well above 100°C, the conventional mounting methods such as glue fixing or glue-magnetic fixing are not suitable. In this study, a strap is used to hold the transducer at the specific position with a steel cylinder transversely pressing on the transducer.

3.4 Bearing Faults for Investigation

A bearing mounted at the high vacuum side of iGX dry vacuum pump is seeded with different conditions. The scale of the bearing used for investigation is shown Figure 3-4. The bearing specifications are shown in Table 3-1.



Figure 3-4 The scale of the bearing used for investigation.

Table 3-1 The bearing specifications for iGX dry vacuum pump.

Number of balls Z	Pitch diameter D	Ball diameter d	Contact angle α
9	46.2 mm	9.5 mm	24.97°

The three bearing conditions are: fault free, localised indent defect which is artificially made on bearing's inner race, contamination fault with artificially introduced solid particles. The location and size of the localised defect are shown

in Figure 3-5 and Figure 3-6, respectively.



Figure 3-5 Artificially made indent on inner race of bearing.

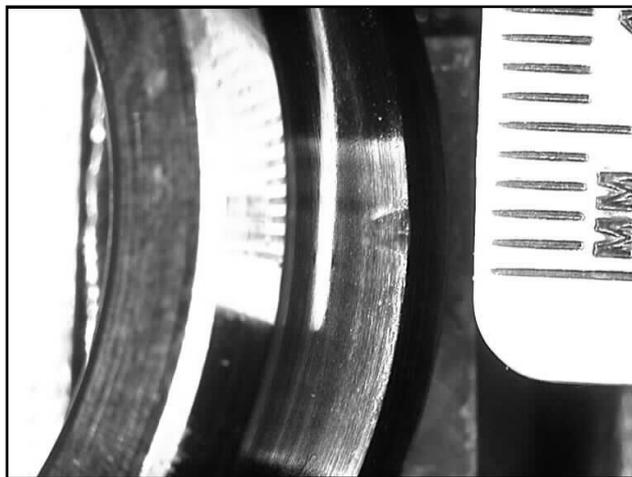


Figure 3-6 The seeded indent is approximately 2 mm wide and 2 mm deep.

These bearing faults can well represent two important deteriorative stages: foreign particles first enter and contaminate a good bearing, which is the incipient stage;

then the fatigue happens and a localised indent defect appears at the race, which is the localised fatigue stage. As the bearing faults are carefully identified by experienced engineers of BOC Edwards and introduced in test facilities by Barden Bearing specialists, they replicate the bearing damage and wear of the dry vacuum pump in natural semiconductor operating conditions.

3.5 Instrumentation and Data Acquisition

The iGX dry vacuum pump with empty load is used as test bed. The speed of the pump is set at 105 Hz ($6300 \text{ rev min}^{-1}$) and the inlet pressure is set at 0 mbar. The AE transducer (PAC R3 α) is firmly held at the surface of the pump house near the high vacuum end in the radial direction.

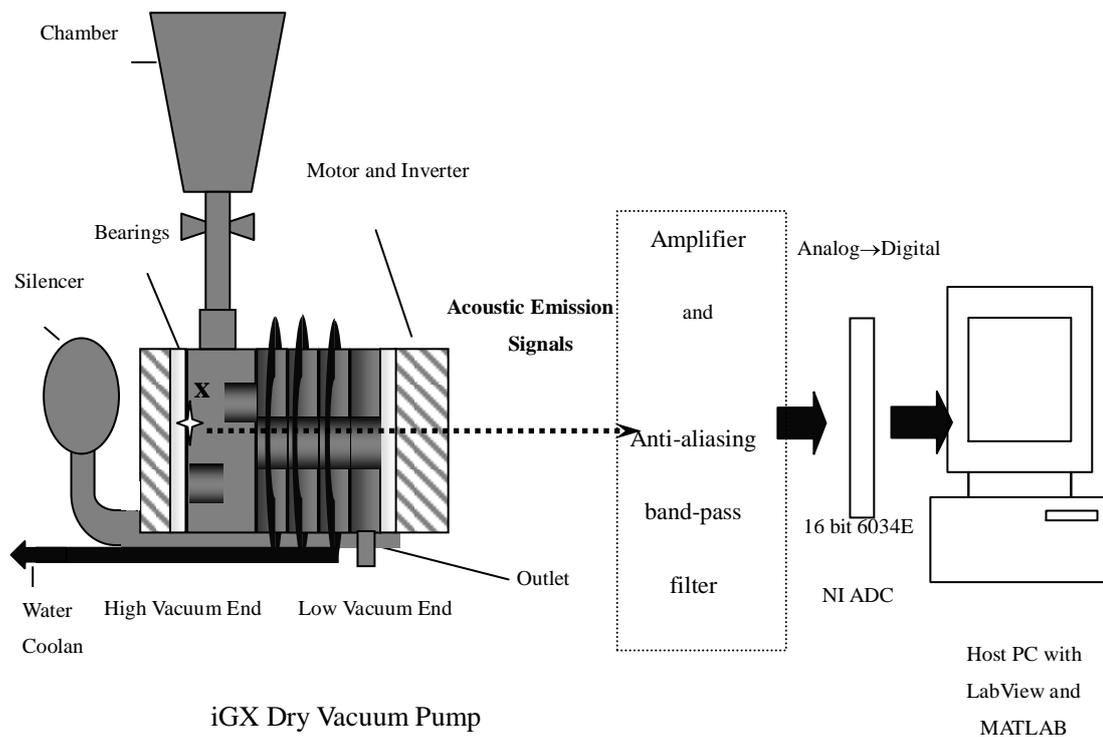


Figure 3-7 Schematic diagram of the complete data acquisition system.

The acquired signals are amplified with the gain of 1000 and filtered by an analogue second order Butterworth band-pass filter. Signals are digitised by a 16-bit NI Analogue to Digital Converter (ADC) with the sampling rate 200 kHz

and stored in a computer. The AE signals are down sampled to 80 kHz offline and further processed in MATLAB. See Figure 3-7 for the schematic diagram of the complete data acquisition system.

For the localised indent defect of inner race mentioned earlier, and the theoretical CDF is calculated as 560 Hz and the theoretical CDI is 143 samples (expressed as the number of discrete sampling points when sampling rate is 80 kHz) according to the equations (2-9) and (2-12) with the information on the shaft speed and the bearing specifications.

3.6 Summary

It is important to understand that many factors such as the mounting method of the transducer, transmission path and degree of interference will limit the success of a BFDD system. Section 3.2 briefly introduced the application and working principle of the iGX dry vacuum pump which serves as the test bed for this study. The mounting method of an AE transducer was explained in Section 3.3.

The details of bearing faults for investigation were given in Section 3.4. Two types of bearing faults are identified and introduced by the industrial specialists. These efforts are made to replicate bearing damage and wear of the dry vacuum pump in natural semiconductor operating conditions. The instrumentation and data acquisition were introduced in Section 3.5.

A reported advantage of AE technique is that it can be directly demodulated to reveal the localised bearing defects. Appendix B also shows that the detection performance on a localised defect can be improved by using the AE technique. The rich information contained in the broadband AE signals may also provide new opportunities for characterising different stages of bearing faults. However, the success of AE technique is limited, partly due to the difficulty in processing, interpreting and classifying the information from AE signals (Mba and Rao 2006).

Chapter 4

Literature Review

4.1 Introduction

Bearing Faults Detection and Diagnosis (BFDD) is essentially a problem of pattern recognition and pattern classification. It is a procedure of mapping the obtained features in the feature space to machine faults in the fault space. The main role of signal processing is to perform the transform and algorithm to extract features from the acquired signals. At this stage, useful information is extracted for further manual or automatic detection and diagnosis purpose. Some of the signal processing methods can provide the capability of denoising.

Traditionally, fault detection and diagnosis is done manually with aid of graphical tools such as parameter trending graph, power spectrum graph, envelope spectrum graph, AR spectrum graph, spectrogram, wavelet scalogram, etc. Manual fault detection and diagnosis may depend on some expertise and specific knowledge to interpret information in the specific application. Thus, a fault detection and diagnosis system which can operate with little *a priori* knowledge will be preferable in practice.

On the other hand, automatic fault detection and diagnosis is more desirable and it can be achieved by automatic classification based on the features extracted from the signals. Statistical, model-based and Artificial Intelligence (AI) are the main approaches used for machine fault detection and diagnosis. Statistical approaches such as Bayesian method, Support Vector Machine, nearest neighbour, Hidden Markov Model are applied on fault detection and diagnosis (Jardine *et al.* 2006). Widodo and Yang (2007) presented a review of condition monitoring and fault

diagnosis using Support Vector Machine. The model-based approaches basically consist of observer, parity space, and parameter estimation. Isermann (2005) gave an introduction on the model-based approaches.

Many Artificial Intelligence techniques have increasingly been applied for the fault detection and diagnosis, including Artificial Neural Networks (McCormick and Nandi 1996, Paya *et al.* 1997, Yang *et al.* 2002, Samanta and Al-Balushi 2003, Wang and Hope 2004, Yu *et al.* 2006, Rafiee *et al.* 2007) and fuzzy logic systems (Sugumaran and Ramachandran 2007). Artificial Neural Networks (ANN) consists of simple processing elements connected in a complex layer structure which enables the model to approximate a complex non-linear function with multi-input and multi-output. The training of an ANN is the process to learn the unknown function by adjusting its weights with observations of input and output. (Jardine *et al.* 2006). A main limitation of ANNs is that the results are difficult to interpret physically and the difficulty in the training process. The main difficulties of a fuzzy logic system stand in the fuzzy partitioning of the input and output spaces and in the establishment of the fuzzy rules, which may require a time-consuming trial-and-error process (Zio and Gola 2009). There has been much effort recently in making a fusion of fuzzy logic and neural networks for better performance in diagnostic systems, including fuzzy-neural network approach (Zhang *et al.* 2003) and neural-fuzzy approach (Zio and Gola 2009, Lou and Lopar 2004). Other AI techniques including Genetic Algorithm (Zhang *et al.* 2005), Expert System (Ebersbach and Peng 2008) or the combination of AI techniques (Lei *et al.* 2007, Saxena and Saad 2007) are also reported for fault diagnosis.

Many of the above methods have achieved a certain amount of success for fault detection and diagnosis, but there is no general guideline on how to choose the appropriate signal processing methods and the features being generated for diagnosis. The first reason for that is the understanding of fault mechanisms and interpretation of acquired signals are still limited. Secondly, the problems of fault detection and diagnosis often differ from one another. One method works well on

one type of machine but possibly will fail on another type of machine. The acquired signals also depend on the mounting method and the type of the transducer. In most applications of condition monitoring, particularly for those with emphasis on real-time detection and diagnosis, the amount of features cannot be too large. The number of features needs to be as small as possible; otherwise, the time of computation and the cost of implementation will increase considerably. The research advances on signal processing will bring many benefits for fault detection and diagnosis:

- Better understanding of fault mechanisms and interpretation of acquired signals;
- Correct selection of the appropriate signal processing methods and the number of features;
- Faster computation to allow on-line implementation.

Because the characteristics of AE signals are quite different from those of the vibration signals acquired by accelerometers, the signal processing methods for AE signals have to be reconsidered. The following sections will review some important signal processing methods, with emphasis on those used for BFDD. These methods, including the time domain methods, frequency domain methods, envelope analysis, cyclostationary analysis and wavelet methods will be introduced in Sections 4.2 - 4.6. The time domain methods are the simplest and use statistical parameters for trending bearing condition. The frequency domain methods use major frequency components and their amplitudes in the direct spectrum of the signal for trending and comparison. Both low frequency range and high frequency range of the spectrum are used for BFDD. The envelope analysis is the most popular method used for localised bearing defect detection. The envelope analysis aims to find out whether significant power exists at some characteristic frequencies in the envelope spectrum. The characteristic frequencies can be theoretically estimated by the equation of shaft speed and bearing geometry as *a priori* knowledge for the envelope analysis approach. In the

pre-processing stage of envelope analysis of vibration signals, a narrow band-pass filter needs to be used for isolating a resonance frequency band. The cyclostationary signal model and cyclostationary analysis on a localised bearing defect give valuable insight into the statistical properties of the impact signals. The main application of wavelet methods to date is to overcome the drawback of the envelope analysis in which determining the resonance frequency band is troublesome. Both single-scale and multi-scale wavelet methods have been investigated for pre-processing the impact signal. This chapter finishes with a summary in Section 4.7.

An early review on these signal processing methods can be found in the book edited by Braun *et al.* (1986). Howard (1994), Tandon and Choudhury (1999), Randall (2004a, b) updated the reviews. Recently, Sawalhi (2007) summarised some new developments in his thesis.

4.2 Time Domain Methods

The quantitative methods in the time domain for BFDD are based on trending some statistical parameters. The most common mentioned statistical parameters are: peak, root mean square (*rms*), crest factor, skewness and kurtosis (Braun *et al.* 1986, Howard 1994, Tandon and Choudhury 1999, Sawalhi 2007). The details are given as below. For a signal $x(n)$ with sample size N , such that

$$peak = (\max(x(n)) - \min(x(n))) / 2, \quad (4-1)$$

$$rms = \sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - \mu)^2} \quad (4-2)$$

$$crest\ factor = \frac{peak}{\sigma}, \quad (4-3)$$

$$kurtosis = \frac{\frac{1}{N} \sum_{n=1}^N (x(n) - \mu)^4}{\sigma^4}, \quad (4-4)$$

where μ is the mean value of the signal. The well known first, second and fourth statistical moments are the mean value μ , variance σ^2 , and kurtosis.

Statistical moments μ_k are defined as (E is the statistical expectation operator):

$$\mu_k = E\{(X - \mu)^k\}, \quad k = 2, 3, \dots \quad (4-5)$$

The third moment, also known as skewness, is defined as

$$skewness = \frac{\frac{1}{N} \sum_{n=1}^N (x(n) - \mu)^3}{\sigma^3}. \quad (4-6)$$

The denominators σ^k in the expressions are necessary for normalisation to make the statistical moments independent of scale (i.e., independent of the actual amplitude of signal). Peak and *rms* values are less sensitive to a localised fault. Kurtosis and crest factor are used to describe the “spikiness” of signals (Braun *et al.* 1986).

The lower moments are less sensitive to the “spikiness”, whereas the higher moments are over-sensitive to spurious interference and noise. The value of kurtosis is close to 3 for the fault free bearing. The value of kurtosis greater than 3 is used for an indication of bearing fault without any prior history (Tao *et al.*, 2007). However, the value may come down to the level for a fault free bearing (i.e., close to 3) as the damage is well advanced. Crest factor has the same problem as Kurtosis. These hinder the Kurtosis and crest factor becoming popular for BFDD in industry (Howard 1994, Tandon and Choudhury 1999). From the viewpoint of Rényi entropy, Tao *et al.* (2007) recently proposed the generalised statistical moments, of which the kurtosis is a special case. When relaxing the mean value property from an arithmetic mean to an exponential mean, the Rényi entropy becomes the well known Shannon entropy.

Two strategies are usually employed for time domain methods (Howard 1994). The first is to monitor the statistical parameters for the whole frequency range of signal. One main drawback of this approach is that the vibration signal from the

bearing usually mixes with that from other mechanical components. This makes the parameters less sensitive to certain modes of bearing failure, in which only some narrow frequency bands are influenced. To alleviate this problem, the second approach is to monitor the parameters over a number of frequency bands or a certain frequency band. Howard (1994) showed four typical frequency bands (0-40 kHz, 0-10 kHz, 10-20 kHz and 20-40 kHz) and the corresponding statistical parameters for bearing condition monitoring of a gear test rig.

Few time domain methods are reported for identifying incipient contamination bearing faults. Maru *et al.* (2007) proposed to trend the values of *rms* in the frequency band 600 Hz - 10 kHz for identifying the contamination bearing fault. The problem with this method is the values of *rms* in the low frequency band 600 Hz - 10 kHz will be sensitive to noise in some situations, e.g., Howard (1994) reported the frequency band 0-10 kHz was dominated by noises from gears on a gear test rig.

For the high-frequency AE signals, the most commonly used parameters in the time domain are ringdown counts and events counts of the signals (Choudhury and Tandon 2000, Miettinen and Andersson 2000, Mba and Rao 2006). The ringdown counts involve counting the number of times the amplitude exceeds a preset threshold level in a given time. An event consists of a group of ringdown counts and signifies a transient wave. In practice, it is difficult to specify threshold level for the counting methods. The accuracy of the threshold level being set relies on the empirical experiences.

4.3 Frequency Domain Methods

The frequency domain methods identify the major frequency components in the direct spectrum of the signal, and then use these components and their amplitudes for trending and comparison. Both low frequency range and high frequency range of the spectrum can be used for BFDD (Howard 1994, Tandon and Choudhury 1999, Randall 2004a). The spectrum (i.e., power spectral density) can be estimated via the well known Fast Fourier Transform (FFT) or parametric

techniques. A summary of the typical techniques for spectrum analysis is given by Kay and Marple (1981).

The approach using a low frequency range tries to find out if there is any significant power occurring at the spectral lines of the fault-related frequencies. For localised defects, the major frequencies of interest for spectrum trending are the Characteristic Defect Frequencies (CDF), their harmonics and sidebands. However, for some modes of failure such as lubrication starvation, general wear and contamination, the fault-related frequencies may not be readily apparent. This makes BFDD using low frequency range difficult or not appropriate for detecting those bearing faults (Howard 1994). The approach also has some limitations: factors like speed variations, bearing slip and bearing geometry need to be known *a priori* for setting up the baseline spectrum; the mechanical noises from other components often mask the major frequencies of interest and make the detection very difficult, especially if the defect is not sufficiently large.

For localised defects, the changes will often become apparent in the high frequency range where the impacts excite the structural resonances. To improve the detection sensitivity, some statistical parameters over a number of frequency bands in the high frequency range are suggested for monitoring. Recently, Antoni and Randall (2006) systematically studied the Spectral Kurtosis (SK) for BFDD using band-pass filtering. The computation of SK combines the classical time domain methods and the time-frequency analysis. SK computes the kurtosis at the outputs of the filter bank which correspond to the different frequency bands. The implementation of filter bank based on Short Time Fourier Transform (STFT) to compute SK was also proposed by Antoni (2006). In this thesis, Chapter 6 will show some new statistical quantifiers based on Discrete Wavelet Packet Transform (DWPT). The DWPT is computed by the fast filter bank algorithm, and the outputs of the filter bank are used for computing the statistical quantifiers. DWPT has some important advantages over STFT: faster computation; flexible basis to be chosen for practical implementation considerations; easier noise elimination. See also Appendix A for detailed technical introduction on DWPT.

4.4 Envelope analysis

Envelope analysis, also known as High Frequency Resonance Technique (HFRT) or demodulated resonance analysis, is one of the classical and most widely accepted signal processing techniques for localised bearing defect detection (McFadden and Smith 1984b). Envelope analysis is an amplitude demodulation process used to obtain the bearing defect harmonics from the spectrum for fault diagnosis purposes (Ho and Randall 2000). The basic sequence of processing used in envelope analysis of vibration signals is shown in Figure 4-1.

When a localised defect in a rolling element bearing makes contact with another mating surface in the bearing under cyclic load, a repetitive impact process is generated. The impact process is later recognised as the cyclostationary process by Antoni and Randall (2002, 2003), which will be introduced in next section. A localised bearing defect is then characterised by the regular impact interval of the impact process. The regular impact interval characterising the localised bearing defect is termed Characteristic Defect Interval (CDI) in the time domain and Characteristic Defect Frequency (CDF) in the frequency domain. As mentioned in Section 2.2, the CDFs can be estimated theoretically by equations of shaft speed, bearing geometry and defect location. The impact may excite resonances in the bearing and the machine which can be measured by a transducer mounted on the machine case near the bearing. With a narrowband filter around the resonance frequency band of vibration signals (typically several kHz), it is possible to exclude most of the interferences generated by other parts of the machine, enabling the impact energy from the bearing to be identified (McFadden and Smith 1984a). In contrast, the AE signals can be demodulated directly to reveal CDF without the need to find out the resonance frequency band.

Demodulation can be employed in many ways (Howard 1994). These methods are applicable for the demodulation of band-passed vibration signals and AE signals. The first method is to use full or half wave rectification followed by smoothing to obtain the envelope signal. The second method is to replace the rectification by a squaring operation and then smooth the band-passed signal. The third way of

extracting the envelope signal is to take the magnitude of the analytic signal whose imaginary part is the Hilbert transform of the real part. The real part is the band-passed signal in the time domain.

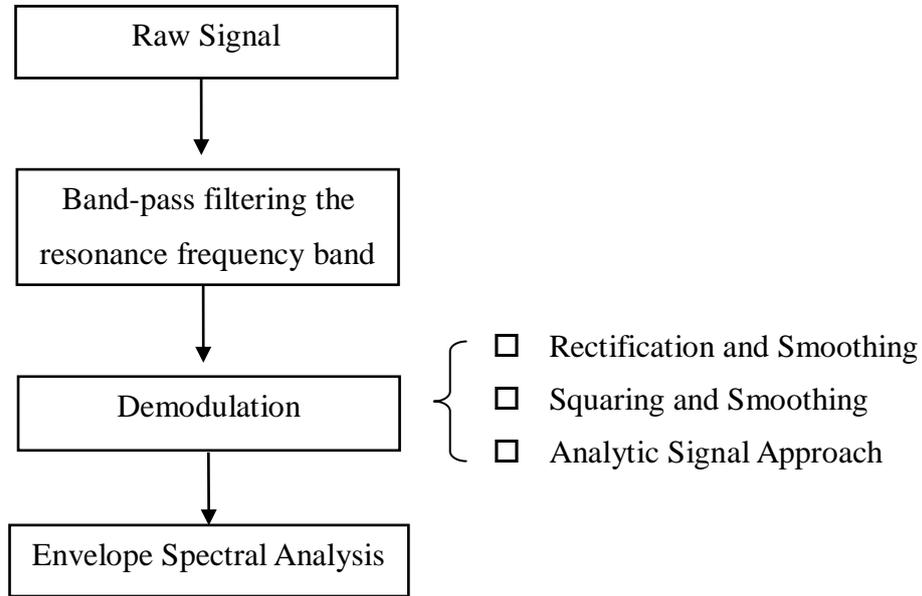


Figure 4-1 Diagram of basic processing used in envelope analysis of vibration signals.

The first comprehensive model for describing different sources of amplitude modulation of the bearing envelope signal was proposed by McFadden and Smith (1984a). This model attempts to explain the appearance of various spectral lines in the envelope spectrum and provides fundamental understanding of the bearing envelope signal. It incorporates the effects of bearing geometry, shaft speed, bearing load distribution, transfer function and the exponential decay of impact signal. The envelope signal $Y_r(t)$ is expressed as (McFadden and Smith 1984a, McFadden and Toozhy 2000)

$$Y_r(t) = (d(t) \cdot q(t) \cdot a(t)) \otimes h(t), \quad (4-7)$$

where $d(t)$ is the pattern of impacts which are generated by the action of the rolling elements under cyclic load, $q(t)$ is the amplitude modulation of the signal under the cyclic load, $a(t)$ is the amplitude variation of the transfer function at

the excited resonance frequency between the location of the defect when an impact occurs and the transducer location, $h(t)$ represents the impulse response of the low-pass filter in the envelope detector.

Improvement for envelope analysis can be achieved by adding a denoising operation between band-pass filtering and demodulation. Ho and Randall (2000) applied the Self-Adaptive Noise Canceller (SANC) to remove discrete frequency masking noise in the envelope spectrum. The main problem with applying SANC is that it may require an extremely long filter length and then become very slow to adapt if some parameters are not set properly (Antoni and Randall 2004a). A newer algorithm called frequency-domain algorithm, which is faster and simpler than SANC, is proposed by Antoni and Randall (2004b).

One of the difficulties with envelope analysis of vibration signals is how to determine the best frequency band to envelope (Howard 1994, Li and Ma 1997). Tse *et al.* (2001) pointed out that the technique itself had some drawbacks in practice: Impact tests are necessary to determine the excited resonance frequency of a particular bearing system; a skilful operator is needed to perform the impact tests. Some efforts are attempted using wavelet methods to overcome the difficulty on determining the resonance frequency band (See Section 4.6.1). To diagnose the bearing condition, envelope analysis needs to perform detection in the envelope spectrum to find out if significant power exists at one of the CDFs. Therefore, information about the shaft speed and bearing geometry should be known *a priori* for diagnosis, and this might not be possible in some situations. In other applications, it is only important to know whether the bearing has a localised defect rather than to know the exact type of localised defect (i.e., inner race defect, outer race defect or cage defect). It might be difficult for the envelope analysis to deal with these cases due to the highly complex envelope spectrum.

An easy way to avoid the problem to determine the resonance frequency band is to directly demodulate the high frequency Acoustic Emission (AE) signals. See Appendix B. Note that the proposed detection method using the AE signals and Autocorrelation Function is very similar to those in the envelope spectrum. The

power spectral density is essentially the Fourier transform of the autocorrelation function (Kay and Marple 1981). Usually, complicated processing operations need to be used for the envelope analysis, including pre-processing (no need for AE signals), demodulation, and the envelope spectral analysis which relies on *a priori* knowledge. A simple signal processing method for BFDD will be proposed in Chapter 5. The advantages of the proposed method are the detection is based on the time domain and it does not need to perform demodulation and envelope spectral analysis.

4.5 Cyclostationary Analysis

Increasing interest aroused on performing the cyclostationary analysis method for BFDD in recent years. The systematic research works on using the cyclostationary method for modelling and detecting localised bearing defects were pioneered by Randall and Antoni (Randall *et al.* 2001, Antoni and Randall 2002, Antoni and Randall 2003, Antoni 2007). Cyclostationarity is popular in nature and it has been broadly studied in many scientific and engineering fields. A survey of the literature review was recently undertaken by Gardner *et al.* (2006). Some important definitions of the method and the results for BFDD are reviewed as below. The details of the following definitions were given by Gardner and Spooner (1992).

A zero-mean process is said to be cyclostationary (in the wide sense) if its autocorrelation function $R_{xx}(t, \tau)$ varies periodically with time. Since $R_{xx}(t, \tau)$ is periodic, it admits a Fourier series representation

$$R_{xx}(t, \tau) = \sum_{\alpha} R_{xx}(\alpha, \tau) e^{j2\pi\alpha t}, \quad (4-8)$$

where the Fourier coefficients $R_{xx}(\alpha, \tau)$ are given by

$$R_{xx}(\alpha, \tau) = F(R_{xx}(t, \tau)) = \int_{-\infty}^{\infty} R_{xx}(t, \tau) e^{-j2\pi\alpha t} dt. \quad (4-9)$$

$R_{xx}(\alpha, \tau)$ is referred to as the cyclic autocorrelation; the set $\{\alpha : R_{xx}(\alpha, \tau) \neq 0\}$

is referred to as the set of cyclic frequencies. The Fourier transform of $R_{xx}(\alpha, \tau)$ is

$$S_{xx}(\alpha, f) = F(R_{xx}(\alpha, \tau)) = \int_{-\infty}^{\infty} R_{xx}(\alpha, \tau) e^{-j2\pi f\tau} d\tau, \quad (4-10)$$

or the double Fourier transform of $R_{xx}(t, \tau)$ is

$$S_{xx}(\alpha, f) = F_{t \rightarrow \alpha} \{ R_{xx}(t, \tau) \}, \quad (4-11)$$

$$\tau \rightarrow f$$

$$S_{xx}(\alpha, f) = \lim_{T_0/2 \rightarrow \infty} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} R_{xx}(t, \tau) e^{-j2\pi\alpha t} e^{-j2\pi f\tau} dt d\tau, \quad (4-12)$$

which is called the cyclic spectral density or cyclic spectrum. The cyclic spectrum is a continuous function over f . Meanwhile, it is a discrete function over α with non-zero values only at those frequencies related to some hidden frequencies in the cyclostationary process. Figure 4-2 shows the cyclic spectrum of a cyclostationary process related to cyclic frequencies i/T (Antoni 2007). Note that the cyclic spectrum shows a discrete structure in the α direction.

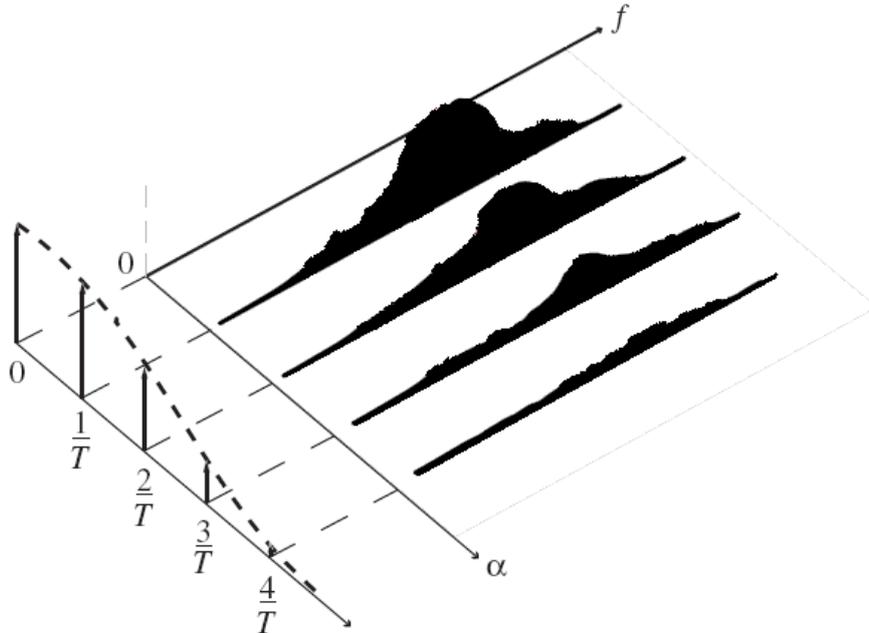


Figure 4-2 Cyclic spectrum of a cyclostationary process (Antoni 2007).

For stationary random signals, only the spectrum at zero cyclic frequency ($\alpha = 0$) has non-zero values in the f direction, which is actually the power spectral density. For those periodic signals, there will be discrete components in both f and α directions. Therefore, a simple test for evidence of cyclostationarity is to check for continuity of the cyclic spectrum for some values of $\alpha \neq 0$ (Antoni and Randall 2002).

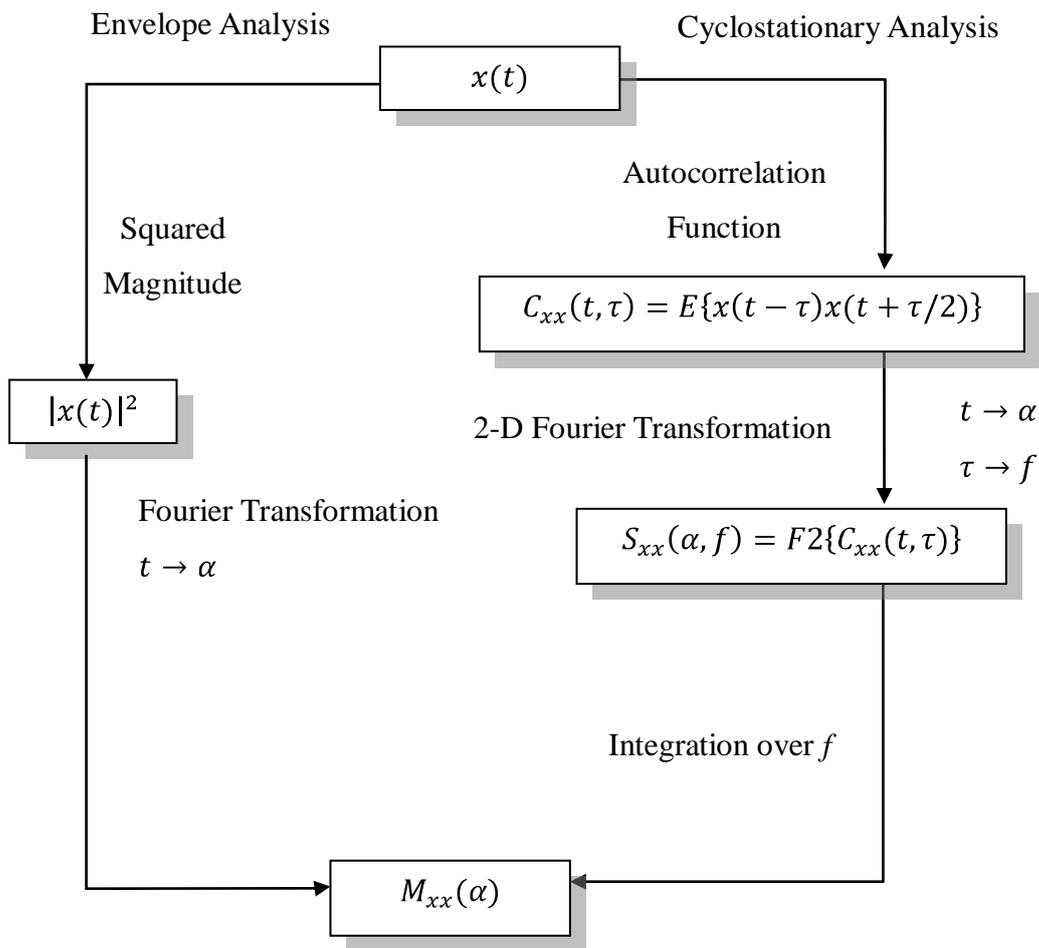


Figure 4-1 Relationship between cyclostationary analysis and envelope analysis (Randall et al. 2001).

Randall et al. (2001) introduced a statistical model which is composed of three parts: a deterministic part, a purely stochastic part and an additive background noise. The deterministic part is the periodic frequency components which are

phase-locked to the various shaft speeds and their harmonics, e.g., gear mesh frequencies. The stochastic part is the localised bearing defect signal which is also tied to shaft speeds. But this part is not strictly phase-locked because there is random slip between the bearing components. The reason is the actual load for individual rolling elements depend on their positions in the bearing, the rolling elements are trying to roll at different speeds. But the physical cage forces them to maintain an equal separation, which causes the slippage. As a result, this signal is not strictly periodic as the deterministic part but cyclostationary. When a defect excites some high frequency modes of the bearing, the energy of the stochastic part may be significantly higher than that of the deterministic part. Therefore, the contribution from the deterministic part is negligible if the analysis spans over sufficiently high frequencies, where the stochastic part may be simply obtained by high pass filtering the signal. This is particularly true for such techniques as envelope analysis. Randall *et al.* (2001) pointed out the integrated cyclic spectrum over f was equivalent to the envelope spectrum (see Figure 4-1).

Antoni and Randall (2002, 2003) modelled the impact process as a 2nd order cyclostationary process. This process is generated by the repetition of impact forces when a defect in one surface strikes a mating surface. Denote the impact process as

$$F(t) = \sum_{i=0}^{\infty} A_i \delta(t - t_i), \quad (4-13)$$

where A_i is the amplitude of the i^{th} impact on the defect, accounting for amplitude modulation. $\delta(t)$ is the Dirac delta function for describing each impact. t_i is the timing of the impact occurrence.

McFadden and Smith (1984a) developed the first model to explain the amplitude modulation of the envelope signal and the corresponding spectral lines of the envelope spectrum.

The uniform point process simply assumes all impacts have equal magnitudes (Antoni and Randall 2003)

$$F(t) = \sum_{i=0}^{\infty} \delta(t - t_i), \quad (4-14)$$

The interval time between two adjacent impacts, $\Delta T = t_i - t_{i-1}$, is the independent and identically distributed random variable. If the impact signals are cyclostationary, the interval time is regular. The term “regular” means that the interval standard deviation is a small fraction of the mean value; thus being regular is characterised by a narrow peak in the probability density function of interval distribution. The probability law governing the interval time between adjacent impacts was approximated by the Normal distribution (Antoni and Randall 2003). In the cyclostationary study, the probability law governing the intervals of a cyclostationary process was also approximated by the Gamma distribution (Gestri and Piram 1975, Landini and Verrazzani 1990). The probability law can also be well approximated by other distributions converging to the central limit, e.g. the Cauchy distribution. This will be studied in Chapter 5.

Antoni (2007) explained the reason why classical spectral analysis often failed to detect bearing faults in practice. Very slight random slippage on the interval time can completely destroy the harmonic structure and cause Dirac pulses to quickly vanish in the frequency domain. The amplitude modulation by shaft or cage rotation smears the raw spectrum. When background noise has very large power in the lower frequency range, it can easily mask the harmonic components produced by localised bearing defects.

Since the cyclic spectrum is the double Fourier transform of the autocorrelation function of the signal, the computation of the cyclostationary analysis will be even more complicated than that of the classical envelope analysis. This might hinder the application of the cyclostationary analysis. However, the above studies on the cyclostationary signal model can give us valuable insight into the statistical properties of the impact signals. The cyclostationary model of the impact signals and the probability law governing intervals provide the theoretical foundation for proposing a new BFDD method. If the timings of impacts can be detected by an algorithm in the time domain, the intervals between adjacent impacts will be

easily achieved and the distribution of the intervals can be directly used for BFDD. A localised defect can then be detected if the interval distribution is found to be regular. See Chapter 5 for the details on LocMax algorithm.

4.6 Wavelet Methods

Wavelet methods are particularly interesting for BFDD. The theory of wavelet methods will be given in Appendix A. One of the main reasons is that the localised bearing defect usually produces a non-stationary signal from which the timing information cannot be easily revealed by conventional spectral analysis. A comparison of the performances of different time-frequency analysis methods such as Short Time Fourier Transform (STFT), Wigner-Ville Distribution (WVD), Choi-Williams Distribution (CWD), Cone-Shape Distribution (CSD), is given in Peng and Chu 2004 (See Table 4-1).

Table 4-1. Comparison of different time-frequency analysis methods (Peng and Chu 2004).

Methods	Resolution	Interference term	Speed
CWT	Good frequency resolution and low time resolution for low-frequency components; low frequency resolution and good time resolution for high-frequency components	No	Fast
STFT	Dependent on window function, good time or frequency resolution	No	Slower than CWT
WVD	Good time and frequency resolution	Severe interference terms	Slower than STFT
CWD	Good time and frequency resolution	Less interference terms than WVD	Very slow
CSD	Good time and frequency resolution	Less interference terms than CWD	Very slow

The interference terms on the time-frequency plane will affect the signal analysis. Continuous Wavelet Transform (CWT) and STFT are the most appropriate methods because they have no interference terms. The difference between STFT

and CWT lies in the elementary function to analyse the signal. The wavelet basis is very compressed for high frequency, which allows CWT a better localisation in time than STFT. This property is particularly useful for the detection of localised bearing defects. Wavelet methods are a set of related processing and analysis techniques based on the idea of wavelet transform, including CWT and other wavelet transform variations, such as Un-decimated Discrete Wavelet Transform (UDWT), Discrete Wavelet Transform (DWT) and Discrete Wavelet Packet Transform (DWPT). The literature on the applications of wavelet methods in BFDD is reviewed below. The detailed introduction is given in the next chapter.

4.6.1 Single-scale Feature Extraction

The main application of wavelet methods in BFDD to date is to assist the envelope analysis of vibration signals in which determining the resonance frequency band is troublesome. These methods are usually employed as a pre-processing operation of the traditional envelope analysis for localised defect detection.

The first scheme for localised defect detection based on CWT was presented by Li and Ma (1997). The method does not require the prediction of the resonance frequencies. By finding out whether the magnitudes of wavelet coefficients at certain dilation vary at the rate of a certain characteristic defect frequency, it can be concluded that periodic structural response due to repetitive force impulses, generated when the passing of each rolling element on the defect, is present – and the localised defect appears. To detect the periodicity, the scheme calculates the autocorrelation of the absolute value of wavelet coefficients and examines the values at time lags around the characteristic frequencies. Once a local peak at one of those lags is found, the bearing is regarded as damaged.

As stated by Lin and Qu (2000), the Morlet wavelet is used for BFDD because of its similarity to the individual impact signal which characterises the localised defect. Based on similarity of the Morlet wavelet to the impact signal, the values of parameters of Morlet wavelet can be optimised. Lin and Qu (2000) proposed a

method based on Morlet CWT. The optimal value of parameter β of the Morlet wavelet is obtained by minimising the wavelet entropy. Here β controls the shape of Morlet wavelet. Qiu *et al.* (2006) proposed a two-step process to select optimal values of parameters β and a for the Morlet wavelet, where a is the scaling parameter). After β is optimised using the method given by Lin and Qu (2000), the optimal value of a scale parameter a is selected based on the Singular Value Decomposition (SVD). Lin *et al.* (2004) employed the Morlet wavelet to match the impact signal (the value of parameter β is set to 1), and then applied the thresholding rule based on the Maximum Likelihood Estimation (MLE) principle by approximating the probability density of impact. Shi *et al.* (2004) proposed an approach based on the fusion of Morlet wavelet transform and envelope spectrum. The approach selects the optimal scale a based on Shannon entropy.

The above Morlet wavelet filter-based method is based on the idea of detecting the “similar” impact components from the noisy signal by optimising the values of parameters of the Morlet wavelet. The method requires a good approximation to the impact signal as *a priori* information for the optimisation procedure, which is difficult to obtain in practice. When the amplitude modulation is considered, the shape of individual impact signals may vary widely. The amplitude variation and the noise from other mechanical components will also influence the shape of individual impact signals. On the other hand, the methods based on the optimisation of Morlet wavelets also cost large computation time. All of the above factors will hinder the application of Morlet wavelet filter-based method.

DWPT is the more flexible version of Discrete Wavelet Transform (DWT) to achieve small frequency separations. The DWPT was also used to overcome the limitations with regard to determining the frequency bands of interest. Altmann and Mathew (2001) presented a method which can automatically extract the Wavelet Packet (WP) node via an adaptive network-based fuzzy inference system. A signal is then reconstructed from the node and used in conjunction with the autoregressive (AR) modelling for spectral analysis. Significant gains in the

signal-to-noise ratio are evident compared to the high pass filtering. Nikolaou and Antoniadis (2002) proposed a method using DWPT as the systematic tool to decompose signals and select the “best” WP node with most energy concentration. A signal is then reconstructed from the node and its envelope spectrum is inspected for the presence of characteristic defect frequencies. In the methods using DWPT, a certain WP node usually is selected as the “best” or “optimal” component for characterising the impact signal. In principle, these methods are similar to the envelope analysis of vibration signals which were filtered around the resonance frequency band.

4.6.2 Multi-scale Denoising and Singularity Detection

In the aforementioned methods, a certain wavelet scale (or a certain WP node) usually is selected as the “best” or “optimal” component for characterising the impacts from vibration signals. However, note that the impact signal is a typical type of transient signal. It has wide-spread frequency distribution which may cross over several wavelet scales and excite several resonances. Therefore, the information at different scales (inter-scale information) should be jointly considered for characterising the impact signals well. The large wavelet coefficients at different scales maintain the characteristics of the impact signals. Recently, many research works in BFDD have linked to these methods.

Shao and Nezu (2005) proposed a mixture denoising techniques consisting of an Adaptive Noise Cancelling (ANC) filter and a wavelet based denoising estimator. A reference signal is necessary for ANC and the SquareTwoLog threshold estimator is used for wavelet denoising. Hong and Liang (2007) proposed a kurtosis-based hybrid thresholding method, K-hybrid, for wavelet denoising the mechanical fault signals. The threshold is determined based on kurtosis. The hybrid thresholding rule divides the wavelet coefficients into four zones associated with different denoising processes. Zuo *et al.* (2005) proposed a method using the wavelet transform as the pre-processor for Independent Component Analysis (ICA). The wavelet coefficients at different scales are modelled as the mixture observation containing the independent sources (transient

signal and noise) and used for ICA separation. The principle behind the ICA method is the IC coefficients of the transient and noise have different sparseness properties. Therefore, non-Gaussianity is used as the separation rule and the separated components with least Gaussianness are found.

Although the denoised signal can provide a clear image for visual inspection on the bearing condition, the quantities for automatic BFDD have to be generated by other means. Conventionally, the denoised signal is also demodulated and analysed with the envelope analysis. Thus the defect detection method is still based on the analysis of the envelope spectrum of the denoised signal. As pointed out in Section 4.4, however, this detection method would rely on *a priori* information, such as the geometry of the bearing and the rotational speed. An alternative method is to directly utilise the detected timings of impact signals to achieve interval distribution for defect detection. This will allow a localised bearing defect to be detected without *a priori* knowledge of CDFs. See Chapter 5.

Sun and Tang (2002) performed the singularity analysis based on Continuous Wavelet Transform (CWT). The method finds out some long ridgelines by tracing along the ridge of the wavelet transform modulus. The periodic patterns of these ridgelines can reveal the relevant information about the health of machine components. The ridgelines that do not propagate to large scales are due to noise. The method can effectively capture the timings of the impact signals. Finally, the method performs a level-dependent thresholding of wavelet coefficients and reconstructs the denoised signals with improved SNR.

Although Sun and Tang (2002) have shown in their work to reveal these timings based on CWT, the CWT for analysis is highly redundant and computational intensive. As an important simplified version of CWT, UDWT not only inherits the translation-invariant property of CWT, but also has much reduced computation requirements. Chapter 6 will present a UDWT denoising method.

4.7 Summary

Although many manual and automatic methods have been successfully applied for

fault detection and diagnosis, there is no general guideline on how to choose the appropriate signal processing methods and the features being generated for diagnosis. The research in signal processing will also allow better understanding and interpretation of acquired signals. Signal processing also studies the possible ways of fast computation which allows on-line implementation of a fault detection and diagnosis system.

This chapter reviewed some important signal processing methods used for BFDD. Sections 4.2 and 4.3 showed the time domain methods and frequency domain methods. The time domain methods are the simplest methods, which use statistical parameters for trending the bearing condition. When the statistical parameters are monitored for the whole frequency range of the signal, they are less sensitive to certain modes of bearing failure, in which only some narrow frequency bands are influenced. In the frequency domain methods, both low frequency range and high frequency range of the spectrum can be used for BFDD. For some modes of failure, the fault-related frequencies may not be readily apparent in the low frequency range. The mechanical noise from other components may also easily mask the major frequencies of interest and make the detection very difficult if the localised defect is not sufficiently large. To cope with noises and improve detection sensitivity, studies suggest the statistical parameters can be monitored over a number of frequency bands in the high frequency range. This method inherits the merits from both the time domain and frequency domain approaches. The Spectral Kurtosis is such a combined method that it computes the kurtosis at the outputs of the STFT. In Chapter 7, some novel statistical quantifiers based on the DWPT will be studied for diagnosing three bearing conditions.

Section 4.4 reviewed the classical envelope analysis, which is established as one of the most widely accepted signal processing techniques for localised bearing defect detection. In the pre-processing stage of envelope analysis of vibration signals, a narrow band-pass filter needs to be used for isolating a resonance frequency band. This pre-processing is essentially an empirical signal denoising approach that improves SNR. One of the difficulties with envelope analysis is

how to determine the ‘best’ frequency band for the envelope. To diagnose the bearing conditions, the envelope analysis also needs to perform detection in the envelope spectrum to find out if significant power exists at one of the CDFs. Therefore, the information about the shaft speed and bearing geometry should be known *a priori* for diagnosis.

Section 4.5 reviewed the cyclostationary analysis using the cyclic spectral density or cyclic spectrum for BFDD. For a cyclostationary process, the cyclic spectrum is a continuous function over f and a discrete function over α with non-zero values only at those frequencies related to some hidden frequencies. The impact signals are not strictly periodic but cyclostationary. If the impact signals are cyclostationary, the interval time is regular. The studies in cyclostationary analysis provide a theoretical base for understanding the statistical properties of the impact signals. Based on the cyclostationary model of impact signals, a new algorithm termed LocMax for detecting the timings of the impact signals will be proposed. The details will be given in Chapter 5.

Section 4.5 reviewed the wavelet methods used for BFDD. The main application of wavelet methods is to assist the envelope analysis of vibration signals in which determining the resonance frequency band is problematic. For pre-processing the impact signals which may cross over several wavelet scales and excite several resonances, however, the multi-scale wavelet denoising and singularity detection are more suitable. Sun and Tang (2002) suggested denoising could be directly performed on the CWT wavelet coefficients across the whole scale range. However, the translation-invariant CWT analysis is highly redundant and computationally intensive. Chapter 6 will introduce the UDWT denoising method which has an advantage over CWT denoising method in computation.

When dealing with the incipient contamination bearing fault, the problems with the reported time domain methods are either sensitivity to the noise or difficulty to automatically specify the threshold level for counting. Because no obvious repetitive impact signals can be observed at the early stage, some signal analysis methods for localised bearing defects may not be effective for the incipient

contamination fault. Chapter 6 will show a new solution – after the AE signals are applied with the automatic UDWT denoising and LocMax-Interval method, the incipient contamination fault can be successful diagnosed. Chapter 7 will provide an alternative solution using novel statistical quantifiers based on the DWPT.

Chapter 5

A New Method for Localised Defect Detection

5.1 Introduction

This chapter introduces a new method to detect a localised bearing defect. The new method includes an algorithm named LocMax for finding out the timings of impact signals and the interval distribution for defect detection. By examining whether the interval distribution is regular, a localised defect can be detected without *a priori* knowledge on Characteristic Defect Frequencies (CDFs). This method is named LocMax-Interval method. This study demonstrates the method using both simulations and experiments. The signal denoising method using DWT thresholding is then investigated to improve the effective range of the parameter values used in the LocMax algorithm.

The impact signals caused by a localised bearing defect can be modelled as a 2nd order cyclostationary process. Since the interval time is regular for a cyclostationary process, a localised bearing defect can then be characterised by the regular interval time of impact signals. To detect a localised bearing defect, the prevalent envelope analysis needs to perform detection in the envelope spectrum to find out if significant power exists at one of the CDFs. Then the information about the shaft speed and bearing geometry should be known *a priori* for defect detection. Refer to Chapter 4 for more details.

However, the information on the shaft speed and bearing geometry for estimating CDFs might not always be available. The cyclostationary method would benefit from a new method for localised defect detection without *a priori* knowledge about the shaft speed and bearing geometry. A localised bearing defect is detected

by examining if the interval distribution of the process is regular (i.e. whether it converges to a narrow peak). The position of the narrow peak of the regular distribution, which corresponds to the bearing Characteristic Defect Interval (CDI), is not necessary information for the diagnostic decision.

On the other hand, denoising techniques have been broadly studied in Bearing Fault Detection and Diagnosis (BFDD) as a pre-processing approach to reduce noise. The details were reviewed in Section 3.3 and 3.5. Denoised signals, which have simpler representation of impact signals, usually provide a clear image for visual inspection on localised defect. But they do not directly provide information for automatic fault diagnosis. To generate the information suitable for automatic BFDD, the denoised signals are conventionally demodulated and envelope spectral analysis is performed. Thus, the detection method is based on the envelope spectrum of the denoised signals.

In this chapter, a new LocMax-Interval method is proposed to find out the timings of impact signals in the time domain, estimate the intervals between adjacent impact signals, and then detect localised defects from the interval distribution. Compared to the envelope analysis, the LocMax-Interval method is also simpler for processing the denoised signals in the time domain. The Discrete Wavelet Transform (DWT) denoising is also investigated as a pre-processing approach for the new LocMax-Interval method.

This chapter is organised as follows: Section 5.2 provides a brief description of the signal model of the impact process. Section 5.3 overviews the methodology, including a new LocMax algorithm for finding out the timings of impact signals and the interval distribution method for defect detection. The simulations in Section 5.4 aim to investigate the effective range of the parameter values used by the algorithm for different cyclostationary processes. The interval distributions achieved for the white noise process and zero process are also shown. Section 5.5 shows the experimental results and a discussion of applying the new method on AE signals for optimal detection. Section 5.6 summarises the chapter.

5.2 Signal Model

The impact process $Y(t)$ for a localised defect can be modelled as

$$Y(t) = Y_r(t) + w_0(t). \quad (5-1)$$

The major signal of interest, $Y_r(t)$, is generated by the repetition of impact forces when a defect in one surface strikes a mating surface. The term $w_0(t)$ represents white noise. Without considering the amplitude variation, $Y_r(t)$ can be ideally modelled as the cyclostationary uniform point process signal assuming all impacts have equal magnitudes (Antoni and Randall 2003). That is:

$$Y_r(t) = \sum_{i=0}^{\infty} \delta(t - t_i), \quad (5-2)$$

where $\delta(t)$ is the Dirac delta function and t_i is its exact timing.

5.3 Methodology

5.3.1 An Algorithm for Finding out Local Maxima

From the above introduction, the problem of localised bearing defect detection is essentially to estimate the interval time ΔT between two adjacent impacts:

$$\Delta T = t_i - t_{i-1}. \quad (5-3)$$

The estimated interval time $\hat{\Delta T}$ is obtained from the timing of adjacent impacts:

$$E(\Delta T) = E(t_i - t_{i-1}) = E(t_i) - E(t_{i-1}), \quad (5-4)$$

$$\hat{\Delta T} = \hat{t}_i - \hat{t}_{i-1}, \quad (5-5)$$

where $i = 1, 2, 3, \dots$

For a pure cyclostationary process, the impact signals are quasi-periodic and their amplitude distribution shows a local maximum. That is, the estimated \hat{t}_i is the

timing on which there is a local maximum \hat{N}_i :

$$t_i = \hat{t}_i | \hat{N}_i \quad (5-6)$$

which is a conditional estimation. Then a signal detection method in the time domain can be used to find out the timings of impact signals. Note the timing mentioned below will be expressed as the index of the discrete sampling point for convenience.

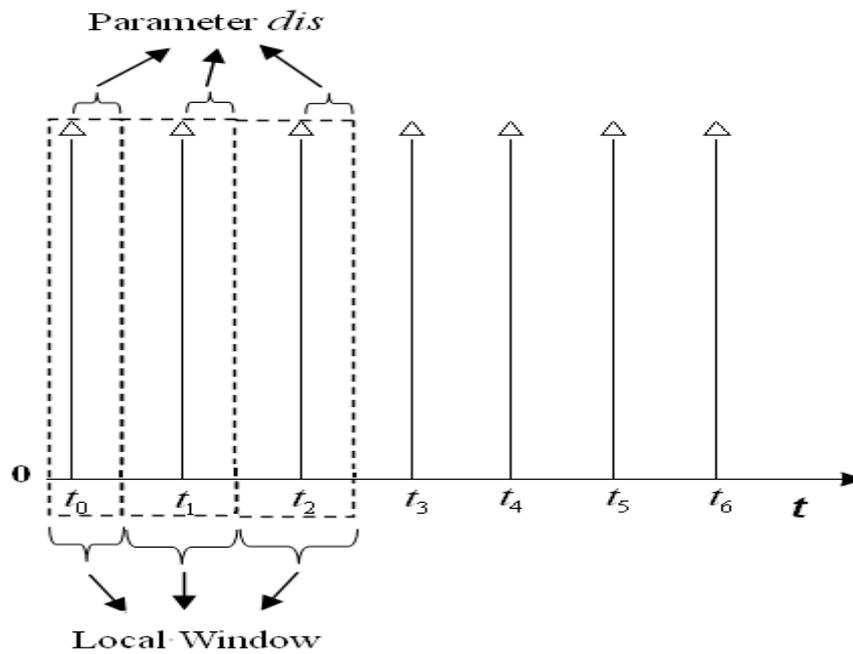


Figure 5-1 Local maxima Detection using LocMax algorithm.

Here a simple recursive algorithm is proposed to find out these local maxima: Take the coefficient with the largest magnitude within a local window A_i . That is, a local maximum \hat{N}_i is defined in a local window:

$$\hat{N}_i = \text{Max}(m_{ob} \in A_i), \quad (5-7)$$

where $i = 1, 2, 3, \dots$ and m_{ob} are all the observations within the local window.

The algorithm starts from the first index of the original observation ($t = 0$), takes its value as temporal value and the index as temporal timing. Then it searches for

another value greater than this temporal value. Once it finds the greater value, it stores the new value and timing, and discards the previous temporal record. If it does not find a greater value, it keeps the temporal record and takes the record at the next index for comparison. This search continues until the distance from an index to the temporal value is greater than an input value of parameter *dis*. The parameter *dis* is defined as the distance (expressed as the number of discrete sampling points) between a local maximum and the end of the local window. This is the end of a local search. The temporal value is confirmed to be a local maximum and its timing is permanently stored. The selection of the parameter *dis* will be studied in Section 5.4.

Table 5-1. A recursive algorithm *LocMax* to detect local maxima.

```

%% m(:) stores the observation;
%% position(:) stores the index of m(:) in the original observation;
%% index(:) stores the timing indices of the local maxima found.
l = length( position( : ) ); %% Take the length of position( : ) array;
t = position( 1 ); %% Initialise t ;
j = 1;
for i = 2 : l
    if position( i ) - t > dis %% The condition to terminate a local search;
        index( j ) = t;
        j = j+1;
        t = position( i );
    elseif abs(m( position( i ) )) > abs(m( t ))
        t = position( i );
    end
end
end

```

The next local search starts from the index next to the end of the previous search, and performs the same searching procedure. The algorithm repeats the same procedure and finds out all the local maxima together with their timings. The

algorithm can be implemented on-line for continuous detection without the need to separate the signal into segments. The proposed algorithm is named LocMax algorithm. See Table 5-1 and Figure 5-1.

5.3.2 Interval Distribution

Estimate the interval by subtracting the adjacent timing of local maxima (expressed as the number of discrete sampling points):

$$\Delta\hat{T} = (\hat{t}_i | \hat{N}_i) - (\hat{t}_{i-1} | \hat{N}_{i-1}). \quad (5-8)$$

As the subtraction is quite an easy computation, the diagnostic information is quickly generated.

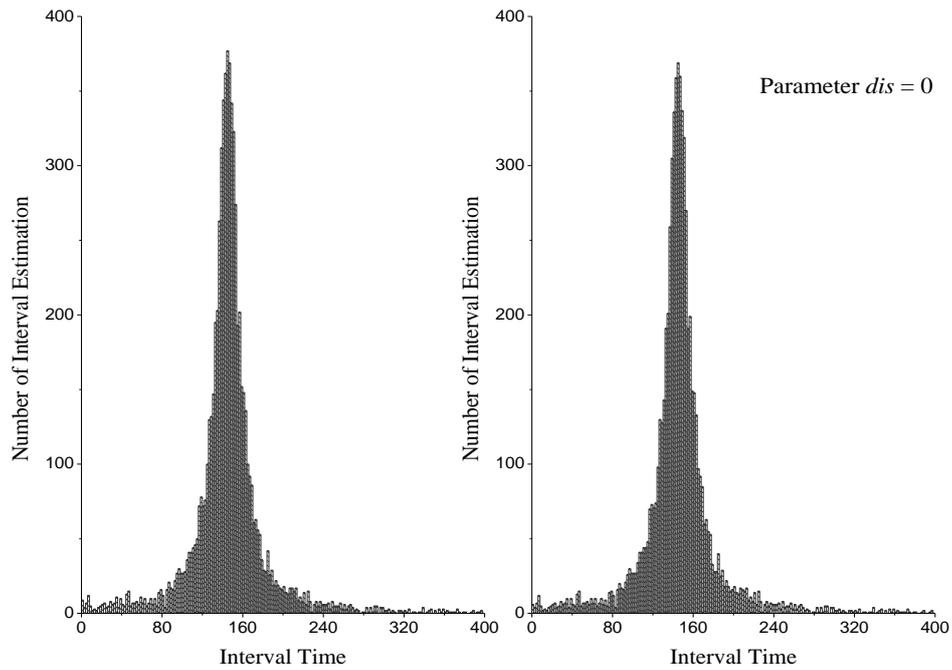


Figure 5-2 Original distribution for interval (left) and generated interval distribution (right).

Figure 5-2 shows a pure cyclostationary process (without any noise involvement) with interval of Cauchy distribution and the interval distribution achieved after applying the LocMax algorithm (parameter value $dis=0$). Here the algorithm

completely recovers the original distribution. It is possible to judge whether the generated interval distribution is regular (i.e. whether it converges to a narrow peak) to infer whether the original process is cyclostationary.

See Figure 5-3 for the comparison between the proposed method and the traditional envelope analysis method for localised bearing defect detection. One advantage of applying interval distribution is that the defect can be detected even without *a priori* knowledge about the shaft speed and bearing geometry. The position of the narrow peak of the regular interval distribution will correspond to the bearing CDI. This fact can be later used for validating the method.

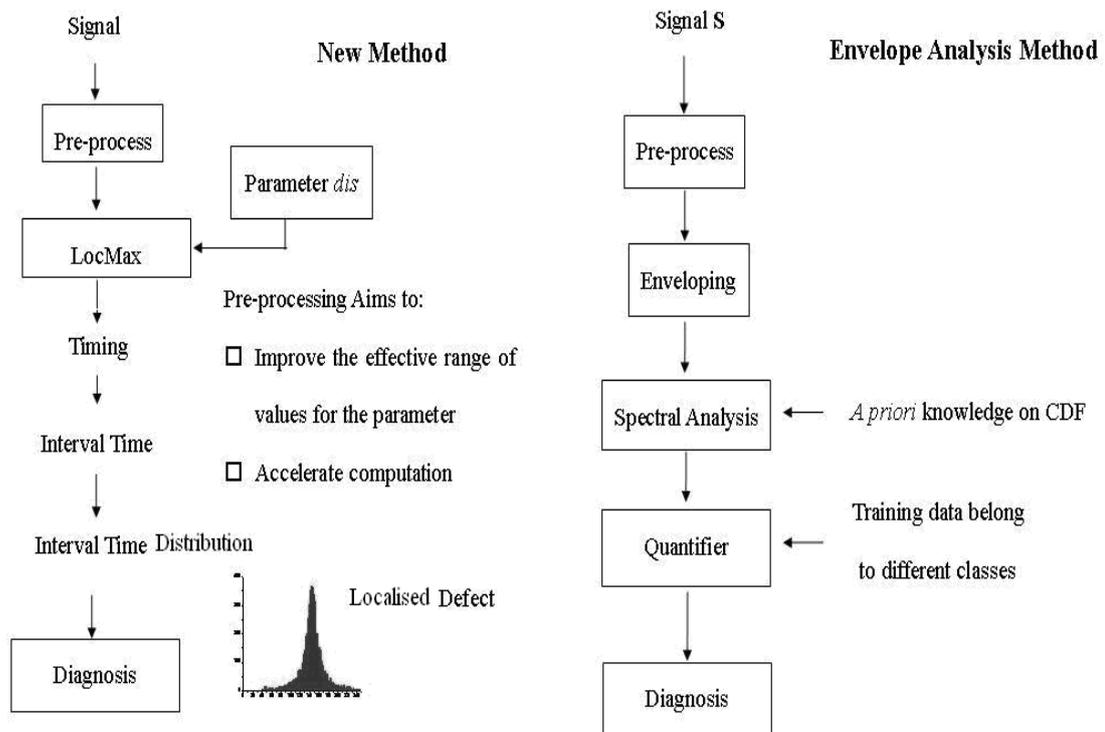


Figure 5-3 Comparison between the proposed LocMax-Interval method and envelope analysis method.

5.4 Simulations

The simulations below aim to investigate the effective range of values valid for the parameter for a pure cyclostationary process, a noisy cyclostationary process

and a denoised cyclostationary process. The performances of the LocMax-Interval method on the white noise process and zero process are also investigated.

5.4.1 The Parameter of Algorithm

Suppose a pure cyclostationary process with interval of Cauchy Distribution (mean interval is 143 samples), the LocMax algorithm is performed on the process to achieve the interval distribution. A complete regular interval distribution can be easily achieved, even taking a very small parameter value $dis = 0$ as shown in Figure 5-2 (right). When increasing the parameter value, the truncated distributions are achieved where the distribution part less than dis is removed. See Figure 5-4 top left and top right. Also note the truncated distribution maintains its regularity and the peak converges to the mean, which means the algorithm is adaptive to a wide range of the parameter values when the process is noise free. If the parameter value is set too large, a weaker peak will be seen at twice the value of the mean interval (see Figure 5-4 bottom).

5.4.2 Noisy Cyclostationary Process

The situation will be different for noisy cyclostationary processes. Suppose a cyclostationary point process with uniform amplitude, local SNR (defined to be the ratio between the impulse power and Root Mean Square of white noise) is 64.

When the value of the parameter is small, either the exponential distribution (see Figure 5-5 top left) or the bi-modal distribution (see Figure 5-5 top right) is achieved. Gradually increasing the parameter values, again the truncated regular distributions are achieved. See Figure 5-5 bottom left and bottom right. Note the effective range of the parameter values to achieve truncated regular distributions is narrower for the noisy process than the pure cyclostationary process. For a cyclostationary process containing noise, it can be seen that usually larger parameter values are chosen to achieve truncated regular distributions.

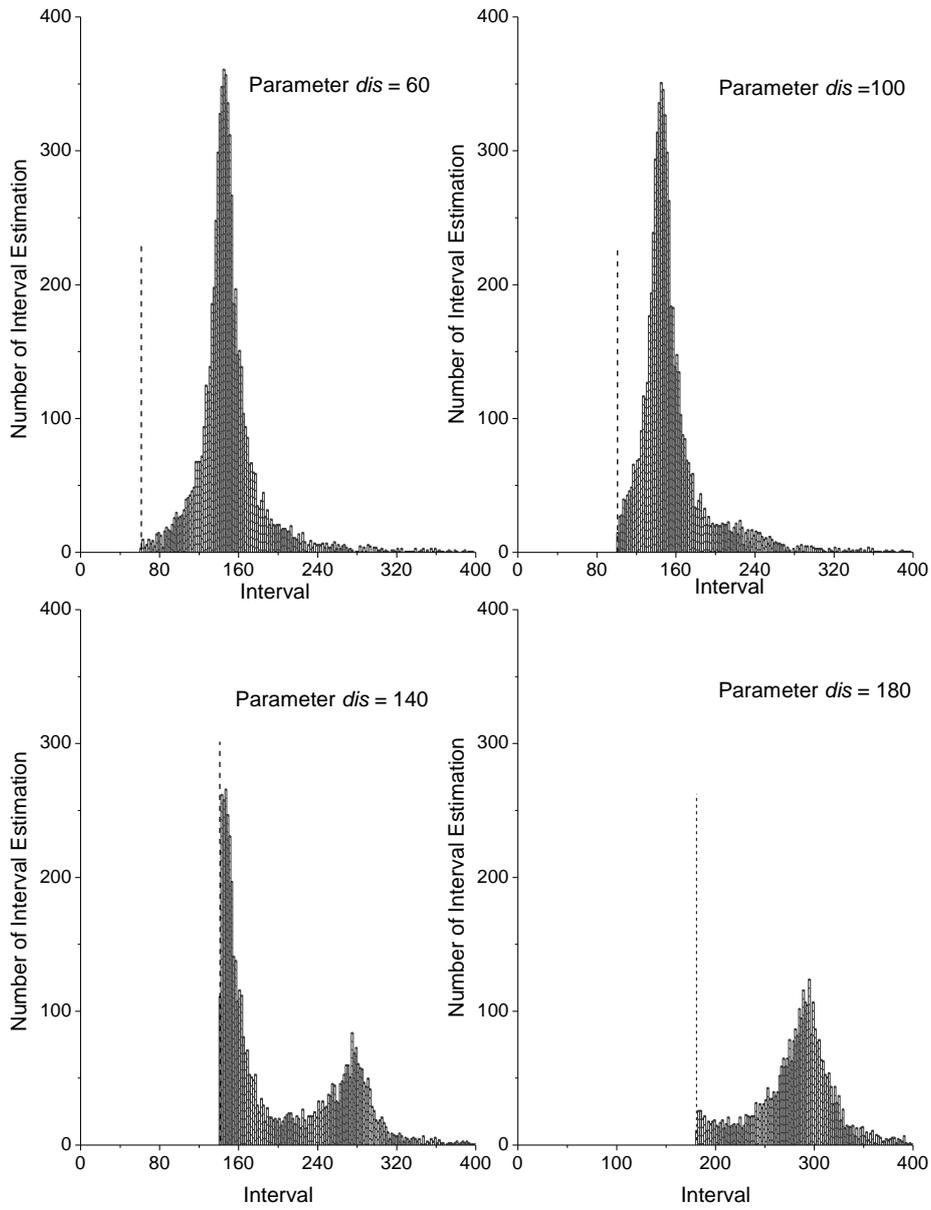


Figure 5-4 Generated interval distribution from pure cyclostationary process using different parameter values.

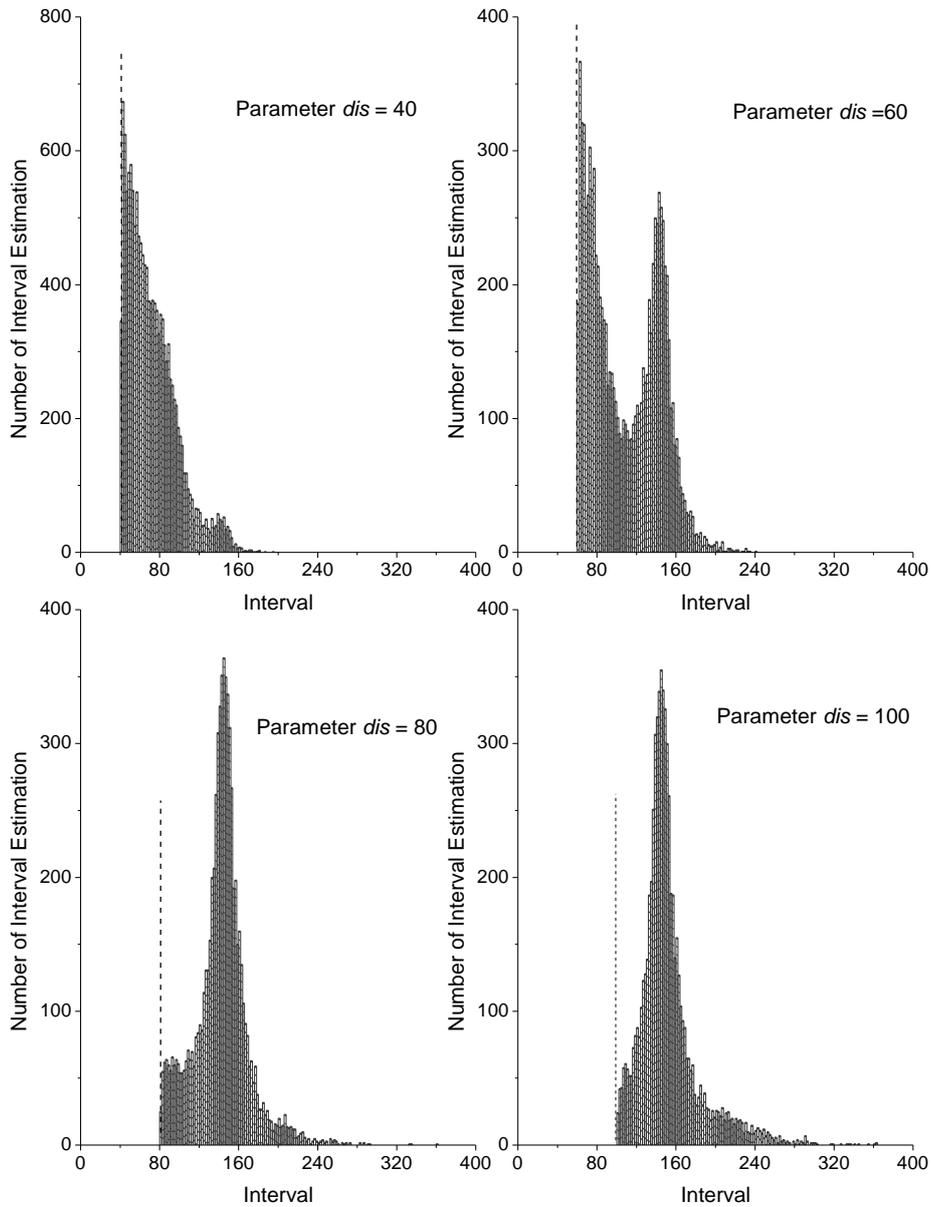


Figure 5-5 Generated interval distribution from noisy cyclostationary process.

5.4.3 Denoising to Improve Effective Range of Parameter Values

From the practical point of view, the effective range of the parameter values should be as wide as possible. Denoising can be employed to improve the effective range of the parameter values. Suppose a cyclostationary point process with uniform amplitude and the local SNR of 64, and single level denoising is

performed on this noisy cyclostationary process. Note that denoising greatly improved the effective range of the allowed parameter values (see Figure 5-6). Compared to the results shown in Figure 5-5, the interval distributions show regularity even using small parameter values, such as $dis = 0$ and $dis = 40$. This is important for the detection without *a priori* knowledge because the algorithm may have to start checking the distribution using small parameter values.

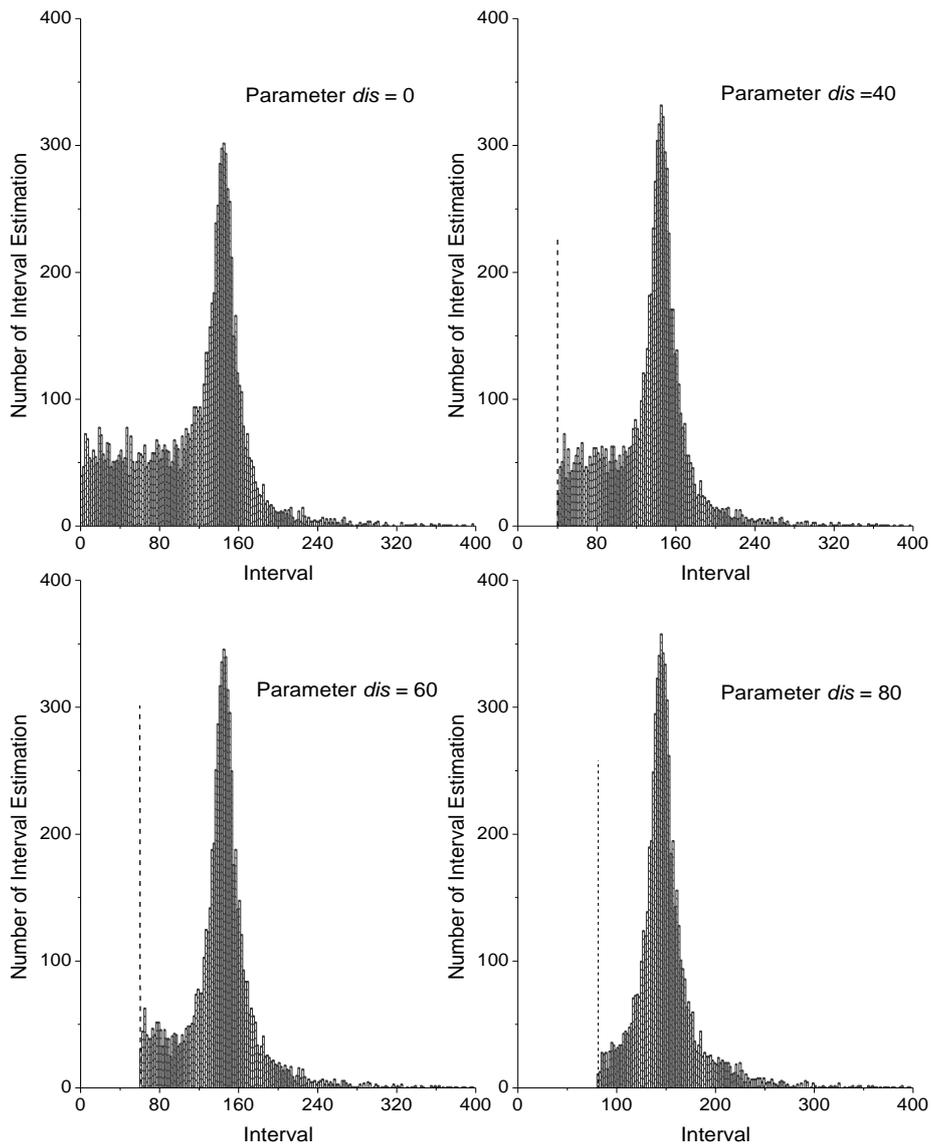


Figure 5-6 Generated interval distribution from denoised cyclostationary process.

5.4.4 White Noise Process and Zero Process

The performance of the algorithm on the white noise process is shown in Figure 5-7 left. For a white noise process, the detected intervals do not show a regular distribution but an exponential distribution. For a pure zero process (i.e. no signal appears), the algorithm cannot find a “local maximum” and the local window will extend infinitely. In this case, the interval distribution cannot actually be produced. (see Figure 5-7 right). In practice, an approximate zero process rather than a pure zero process is usually met. See Figure 5-7 middle for the generated interval distribution from an approximated zero process. The detected intervals show a uniform distribution. Both the pure zero process and the approximate zero process can be achieved when denoising is applied to the white noise process.

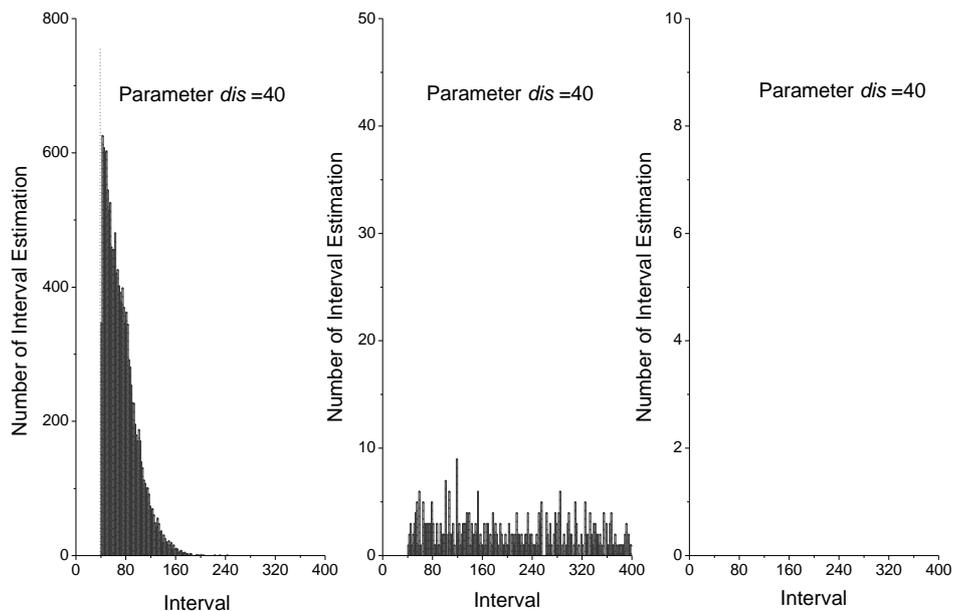


Figure 5-7 Generated interval distributions from white noise process (left), approximate zero process (middle) and zero process (right).

5.5 Results and Discussion

For the localised indent defect of inner race, the theoretical CDI is calculated as

143 samples (expressed as the number of discrete sampling points when sampling rate is 80 kHz) using the equations (2-9) and (2-12). The shaft speed is 105 Hz and the Barden bearing specifications are used in the equations for calculation. The acquired AE signals from localised defect bearing are divided into many 4096-point segments without overlap and subsequently processed by the proposed LocMax algorithm.

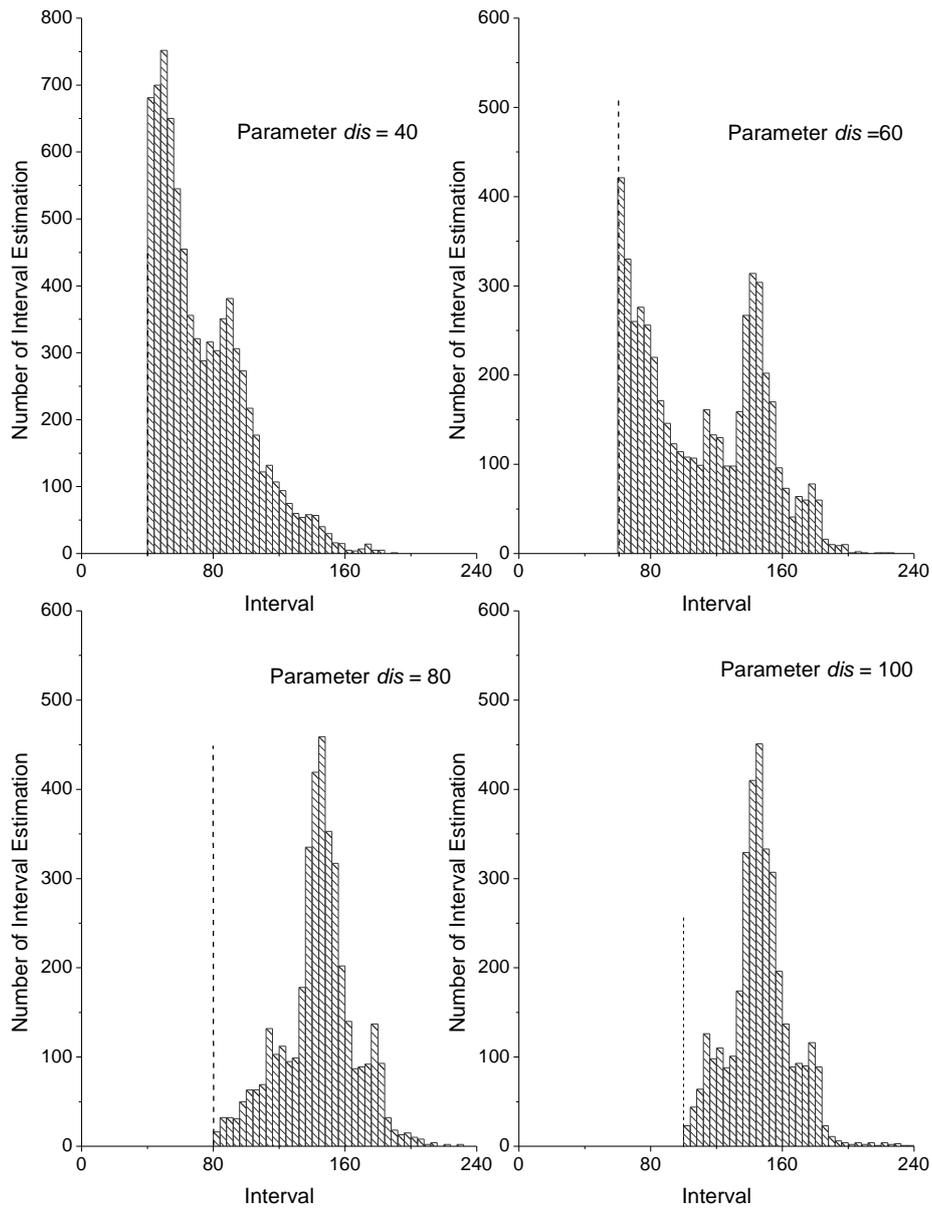


Figure 5-8 Generated interval distributions from raw AE signals of localised defect.

Figure 5-8 shows the interval distributions when using parameter values $dis=40$, $dis=60$, $dis=80$ and $dis=100$. The distributions gradually show regularity when using large parameter values, such as $dis=80$ and $dis=100$. The peaks of these two latter cases are at interval=144 samples, which is very close to the theoretical CDI.

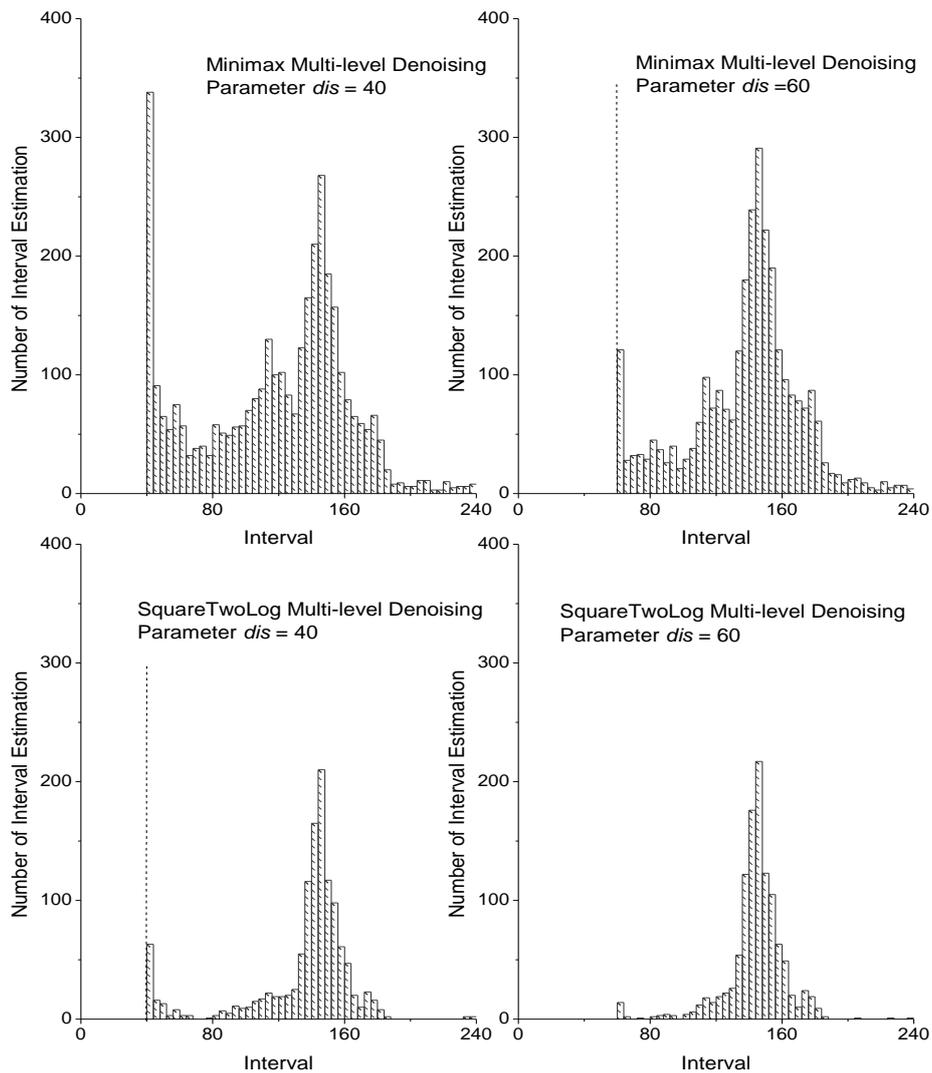


Figure 5-9 Generated interval distributions from AE Signals of localised defect using wavelet multi-scale denoising methods.

To improve the effective range of the parameter values and allow the method to work with small parameter values, the raw signals can be pre-processed by the wavelet denoising methods. The acquired AE signal segments are pre-processed by the 2-level DWT denoising (see Section A.8) and then the denoised signals are applied with the proposed LocMax-Interval method. The non-parametric DWT thresholding methods are performed and thresholds are automatically estimated. Figure 5-9 shows the interval distributions when using Minimax and SquareTwoLog, with parameter values $dis = 40$ and $dis = 60$, respectively. For small parameter values, the regularities of denoised signals are clearer than for the raw signals. This is evident when using parameter value $dis = 40$, where the distribution from raw signals fails to show any regularity in Figure 5-8 top left and the one from denoised signals highlights the regularity in Figure 5-9 left. The SquareTwoLog thresholding is stricter and eliminate more variations than the Minimax thresholding. When small parameter values are used, increasing the thresholds such as changing from Minimax to SquareTwoLog thresholding can lead to better visibility of the regularity. The first peaks of these cases are also at $interval = 144$ samples, which is very close to the theoretical CDI.

The acquired raw AE signals from a fault free bearing are also divided and processed by the LocMax algorithm. See Figure 5-10 for the interval distributions when using different parameter values, where they all show the exponential distributions. It means the bearing conditions can be diagnosed by testing the generated interval distribution directly from raw signals when using large parameter values, such as $dis = 80$ and $dis = 100$. It is evident that Figure 5-10 bottom show exponential distributions, while Figure 5-8 bottom show very clear regular distributions.

The AE signals from fault free bearing are also pre-processed by DWT denoising and then processed by the LocMax algorithm. See Figure 5-11 for the generated interval distributions. The intervals from Minimax denoising show exponential distributions (top), while the interval distributions cannot be produced after SquareTwoLog denoising (bottom). The latter cases are the distributions detected

from the zero processes. It means the bearing conditions can also be diagnosed by testing the distribution from the denoised signals when using small parameter values, such as $dis = 40$ and $dis = 60$. Comparing Figure 5-9 and Figure 5-11, the denoising can help to highlight the regularity of interval distribution for the bearing with localised defects (see Figure 5-9), and generate the exponential distribution (Figure 5-11 top) or fail to generate any distribution for the fault free bearing, corresponding to a zero process (Figure 5-11 bottom).

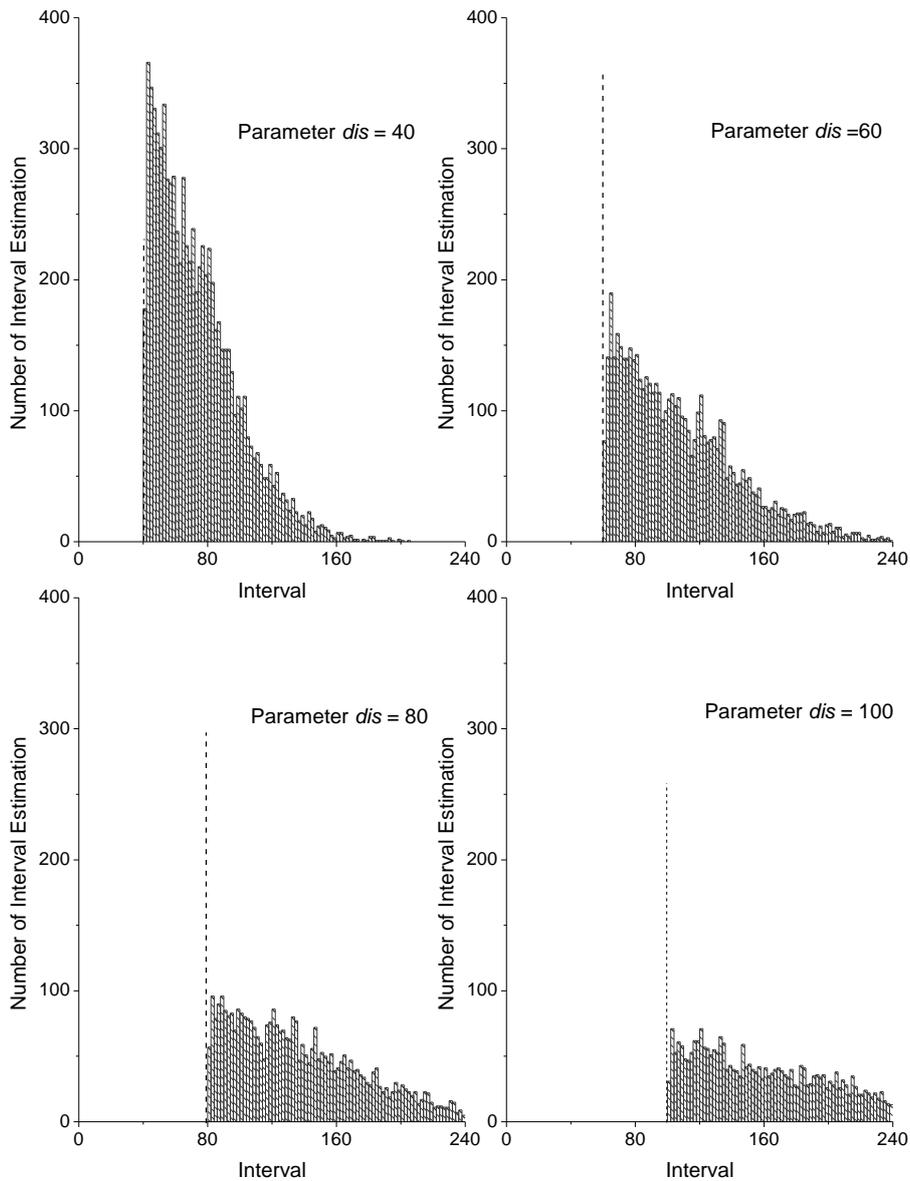


Figure 5-10 Generated interval distributions from raw AE signals of fault free bearing.

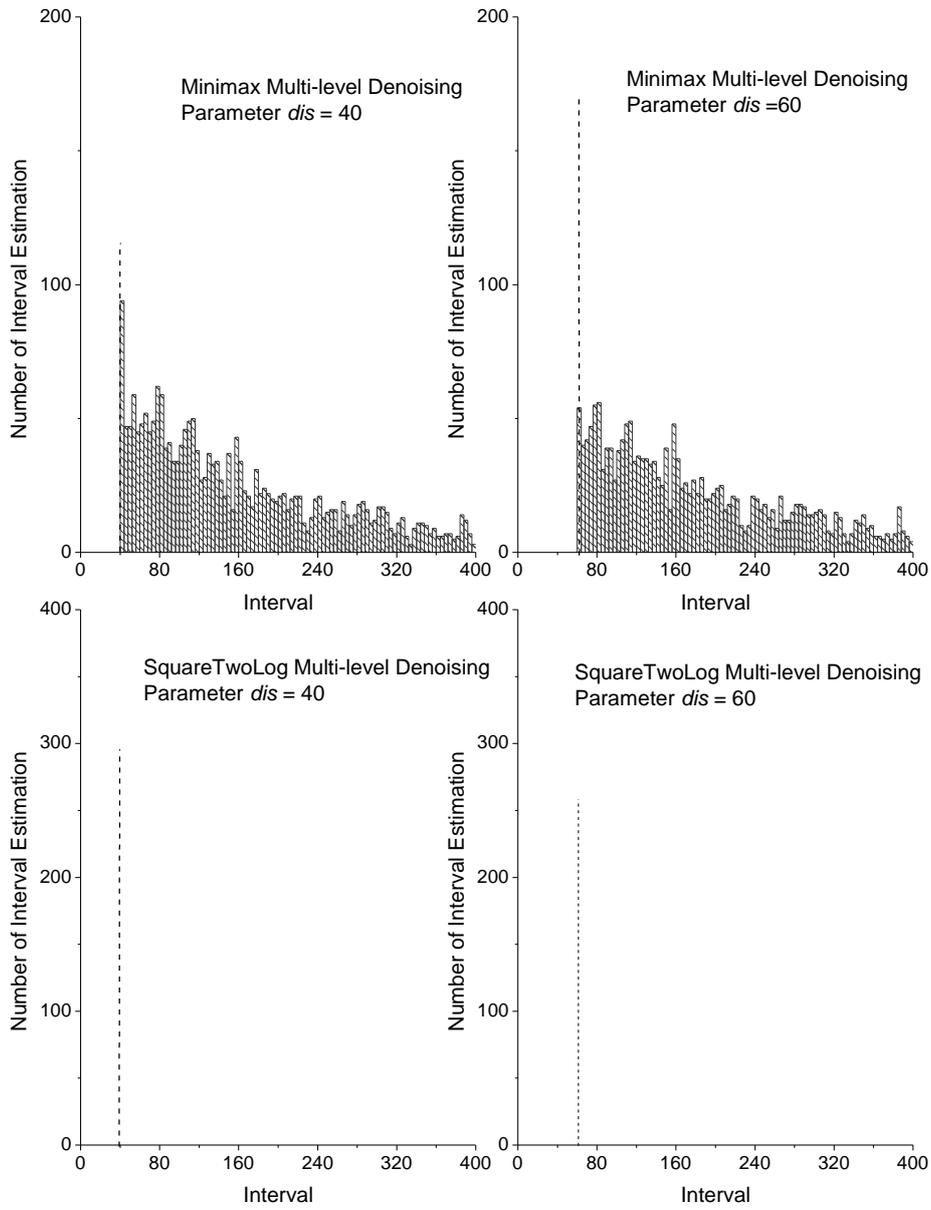


Figure 5-11 Generated interval distributions from AE Signals of fault free bearing using wavelet multi-scale denoising methods.

5.6 Summary

Many prevalent signal processing methods in the frequency-domain find out if significant power exists at one of the CDFs which should be obtained from the shaft speed and bearing geometry as *a priori* knowledge for defect detection. However, this knowledge might not be always available for some applications.

In this chapter, a new LocMax-Interval method was proposed to detect localised bearing defects. The new method first applies a new algorithm, LocMax, for finding out the timings of impact signals in the time domain, and the intervals between adjacent impact signals are estimated, the interval distribution is then used for defect detection. By examining whether the interval distribution of the process is regular, a localised bearing defect can then be detected.

The simulations show a wide range of values for the choice of the parameter of the algorithm when the process is noise free. By increasing the value of the parameter, the truncated distributions which maintain the regularity and the peaks which converge to the mean will be achieved. However, the involvement of noise in a cyclostationary process will limit the effective range of the parameter values. Denoising is shown to improve the effective range of the parameter values, which allows the proposed method to work with small values of the parameter.

The experimental results also show that a localised defect can be successfully detected using the proposed method. When using large parameter values, the bearing conditions can be diagnosed by testing the interval distribution directly from raw signals. The intervals from a localised defect bearing show very clearly a regular distribution; while the intervals from a fault free bearing show exponential distribution. The regular distribution converges to the peak corresponding to the theoretical CDI. It is shown that denoising methods can improve the effective range of the parameter values, thus allowing the proposed method to work with a wider range of the parameter values, including small values. Denoising can also help to highlight the regularity of interval distribution for bearings with localised defects, and generate the exponential distribution or

the distribution corresponding to a zero process for fault free bearings. Another benefit brought by denoising is that the computation of the algorithm can be greatly reduced, which is important for real-time implementation.

The proposed method shows the feasibility to detect a localised bearing defect when *a priori* knowledge on the shaft speed and bearing geometry is unavailable. The LocMax algorithm can also be implemented on-line for continuous detection, without the need to separate the signal into segments. The diagnostic information can be quickly generated after the timings are detected by the algorithm. This method is suitable for those detection applications which do not need to know the exact type of localised bearing defect. However, if the type of localised bearing defect needs to be known, the position of the narrow peak of the regular interval distribution is also capable to provide the necessary information.

Chapter 6

Application of Un-decimated Discrete Wavelet Transform

6.1 Introduction

This chapter investigates the application of the Un-decimated Discrete Wavelet Transform (UDWT) in Bearing Fault Detection and Diagnosis (BFDD). In the last chapter, the LocMax-Interval method including a new algorithm LocMax and the interval distribution was proposed for localised defect detection. In this chapter, UDWT decomposition and denoising will be introduced to process the acquired AE signals. UDWT representation allows convenient and direct detection on the wavelet coefficients. The local maxima of UDWT wavelet coefficients can highlight the timing information. The UDWT denoising is introduced as a pre-processing approach to apply with the LocMax-Interval method. The signal models for different bearing conditions are also described.

The LocMax-Interval method was proposed in the last chapter, where it was used for finding out the timings of impact signals. The idea of the LocMax algorithm is to take the coefficient with the largest magnitude within a local window. The local window should be able to extend infinitely and it has the minimal size controlled by a parameter. The algorithm detects the local maximum and the corresponding timing in the time domain. All the intervals calculated from the timings of the adjacent local maxima then form a distribution used for diagnosis.

For a cyclostationary process, the local maxima found will show strong periodicity. The detected interval distribution shows strong regularity - it

converges to a narrow peak. For a white noise process, the “local maxima” found will be distributed quite randomly. The intervals do not show a regular distribution but an exponential distribution. For a zero process, the algorithm will not find a “local maximum” and the local window will extend infinitely. An interval distribution cannot actually be produced. In practice, an approximately zero process rather than a pure zero process is usually met. The detected intervals show a uniform distribution for an approximately zero process.

UDWT has been introduced as a powerful tool used for transient analysis (see Section 4.6 and Appendix A.5 for details): UDWT not only inherits the translation-invariant property of CWT, but also has much reduced computation requirements; UDWT is translation invariant and suitable for the transient signal analysis which requires precise time localisation; UDWT representation allows convenient and direct detection on the wavelet coefficients; local maxima of UDWT wavelet coefficients can highlight the timing information. For the traditional signal denoising approaches, signal reconstruction is usually the necessary operation to achieve the denoised signal which will be used for fault detection in the next stage. In the last chapter, DWT denoising was performed to improve the effective range of the parameter values dis , thus allowing the proposed method to work with a wider range of the parameter values. After being processed by the DWT denoising, the AE signal was reconstructed as the denoised signal which was then processed with the LocMax-Interval method. However, the reconstruction process usually needs much computation.

In this chapter, UDWT denoising, which combines Wavelet Maxima Chain (WMC) search and thresholding, will be introduced as a pre-processing approach for the LocMax-Interval method. Since the UDWT-denoised wavelet coefficients have the same time resolution as the original signal, the timings of impacts can also be detected by the LocMax algorithm from the denoised UDWT wavelet coefficients. After being processed by the UDWT denoising, the denoised UDWT wavelet coefficients can be directly applied with the LocMax-Interval method without the need for reconstruction. Once timings are found by the LocMax

algorithm, the interval distribution can be easily generated for fault diagnosis.

This chapter is organised as follows: Section 6.2 provides a brief description of signal models. Section 6.3 overviews the methodology, including the description of the UDWT decomposition and denoising method. In Section 6.4, the application of the UDWT will be demonstrated with experimental results. The summary is in Section 6.5.

6.2 Signal Models

For a localised defect, the impact process $Y(t)$ is modelled as

$$Y(t) = Y_r(t) + Y_l(t) + Y_c(t) + w_0(t). \quad (6-1)$$

The major signal of interest, $Y_r(t)$, is generated by the repetition of impact forces when a defect in one surface strikes a mating surface. $Y_l(t)$ represents the weak harmonic component located in the lower frequency range (Antoni 2007). $Y_c(t)$ represents the coherent component of the signal $Y_r(t)$. It may show similar time-frequency behaviour as $Y_r(t)$ but with lower amplitudes. The terms $Y_c(t)$ and $Y_l(t)$ are essentially the interferences. The $w_0(t)$ term represents white noise with zero mean. In practice, the timing and amplitude of impact signals are usually difficult to estimate directly from the noisy observation $Y(t)$. The amplitude of signal $Y_r(t)$ was modelled by McFadden and Smith (1984a) (see Section 3.3 for details). Without considering the above amplitude variation, $Y_r(t)$ can also be ideally modelled as the cyclostationary point process signal with uniform amplitudes.

For an incipient contamination fault, the signal process $YC(t)$ is modelled as:

$$YC(t) = Y_l(t) + Z_h(t) + w_0(t), \quad (6-2)$$

which is a process containing the weak harmonic component $Y_l(t)$ located in the lower frequency range, the random non-cyclostationary component $Z_h(t)$ at high

frequencies and the white noise term $w_0(t)$. Since no obvious cyclostationary impact signals can be observed at an early stage, the prevalent envelope analysis method broadly used for localised defect detection will be ineffective to detect the incipient contamination fault.

For a fault free bearing, the signal process $YF(t)$ is modelled as

$$YF(t) = Y_l(t) + w_0(t), \quad (6-3)$$

which is a process containing the weak harmonic component $Y_l(t)$ located in the lower frequency range and the white noise $w_0(t)$. For both the fault free and incipient contamination fault models, the impact signal $Y_r(t)$ and the coherent component $Y_c(t)$ in the impact process $Y_r(t)$ no longer exist. The difference between these two processes is the random non-cyclostationary component $Z_h(t)$ at high frequencies.

6.3 Methodology

6.3.1 UDWT Decomposition

The UDWT decomposes the impact process $Y(t)$ as

$$W_j Y(t) = W_j y_r(t) + W_j y_l(t) + W_j y_c(t) + W_j w(t), j = 1, 2, \dots, J, \quad (6-4)$$

where $W_j y_r(t)$, $W_j y_l(t)$, $W_j y_c(t)$ and $W_j w(t)$ denote the UDWT wavelet coefficients of signal components $Y_r(t)$, $Y_l(t)$, $Y_c(t)$ and $w(t)$ at the dyadic scale j and time t , each scale having the same number of coefficients. For the incipient contamination fault process (6-2) and fault free process (6-3), the UDWT wavelet coefficients are denoted as $W_j YF(t)$ and $W_j YC(t)$, respectively.

6.3.2 Significance Measure

The impact signals generated by localised defects are typically transient signals

with wide spread frequency distribution and relatively high local SNR at high frequencies. The detection of the transient signal was reported in literature as singularity detection (see Section A.9 for details). After the Wavelet Maxima Chains (WMCs) are found, a significance measure m_k needs to be defined for detection in the time domain. For example, one can define m_k as the local regularity estimation, coefficient magnitude, or the amount of inter-scale correlation. The significance measure is essentially a new signal computed from the observed wavelet coefficients and the selection of the position is the detection. The significance measure allows the LocMax algorithm to detect failures in the time domain. In the following approach, the thresholded WMCs will be defined as the new significance measure. The non-parametric thresholding methods (i.e., SquareTwoLog, SURE and Minimax) with hard thresholding are also used here for investigation.

6.3.3 UDWT Denoising

Let the decomposition level be $J = 2$, the UDWT wavelet coefficients of impact signal at two dyadic scales $W_1Y(t)$ and $W_2Y(t)$ are:

$$W_1Y(t) = W_1y_r(t) + W_1y_l(t) + W_1y_c(t) + W_1w(t), \quad (6-5)$$

$$W_2Y(t) = W_2y_r(t) + W_2y_l(t) + W_2y_c(t) + W_2w(t). \quad (6-6)$$

The UDWT scaling coefficients which are viewed as the low frequency carrier representing the weak harmonic component $Y_l(t)$ in the model will be discarded as the interference components. After the UDWT denoising, which combines the WMC search and thresholding, most of the noise components can be eliminated. The remaining coefficients are supposed to be:

$$M_1Y(t) = M_1y_r(t) + M_1y_c(t), \quad (6-7)$$

$$M_2Y(t) = M_2y_r(t) + M_2y_c(t), \quad (6-8)$$

Choose the modulus magnitudes of the denoised UDWT wavelet coefficients at

the finest scale ($j = 1$) to perform timing detection using the LocMax algorithm. These denoised UDWT wavelet coefficients will be called as the UDWT denoised signals in the following description for simplicity. That is, the significance measure m_k is:

$$m_k = |M_1 Y(t)| = |M_1 y_r(t) + M_1 y_c(t)|. \quad (6-9)$$

The detection at the single scale means that there is no need to reconstruct signals. In this way, the local maximum \hat{N}_i is defined by:

$$\hat{N}_i = \text{Max}(m_k \in A_i), \quad (6-10)$$

where $i = 1, 2, \dots, L$ and m_k is the significance measure within the local window A_i . The detection and diagnosis can then be obtained using the LocMax-Interval method. See Figure 6-1 for the processing diagram of the proposed method. The interval distribution is supposed to show regularity for the impact process.

On the other hand, the signal processes in (6-2) and (6-3) at the finest scale for incipient contamination fault and fault free bearing are also supposed to have two simple forms after the UDWT denoising:

$$M_1 YC(t) = M_1 Z_h(t), \quad (6-11)$$

$$M_1 YF(t) = 0. \quad (6-12)$$

These are a random non-cyclostationary component and zero, respectively. For the signal from a fault free bearing, the UDWT denoising completely eliminates the noise and the process becomes a zero process. In an ideal situation, an interval distribution cannot be actually generated for this case. Since the incipient contamination fault maintains some random non-cyclostationary components at high frequencies which will not be completely eliminated by the denoising, the generated intervals are supposed to show an exponential distribution.

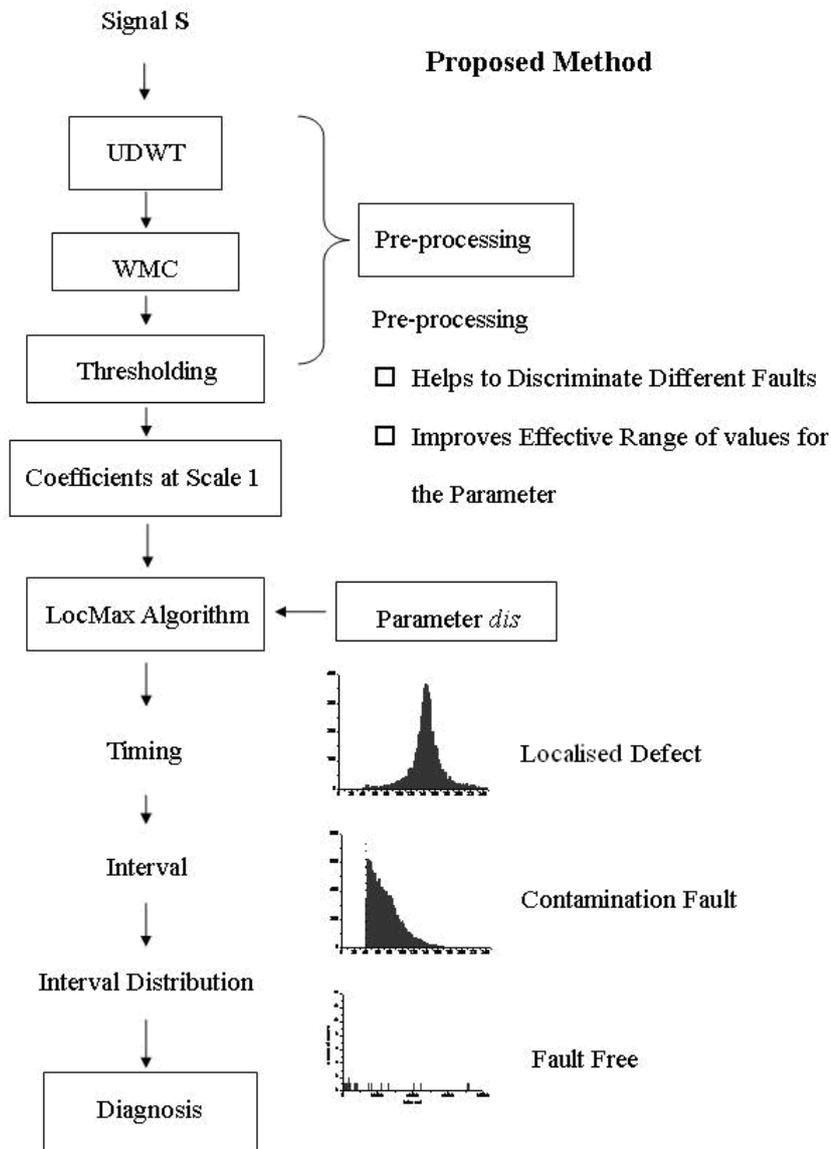


Figure 6-1 The method based on UDWT decomposition and denoising for BFDD.

6.4 Results and Discussion

The acquired AE signals are divided into segments without overlap and subsequently decomposed by the two-level UDWT using the biorthogonal wavelet bior 6.8 (see Figure 4-12 for the function of bior 6.8 wavelet). The scaling coefficients are discarded as the interference components and the wavelet

coefficients at two dyadic detail scales are processed with the UDWT denoising. The thresholds at different scales are automatically estimated according to the non-parametric thresholding methods (SURE, Minimax or SquareTwoLog) on a segment by segment basis.

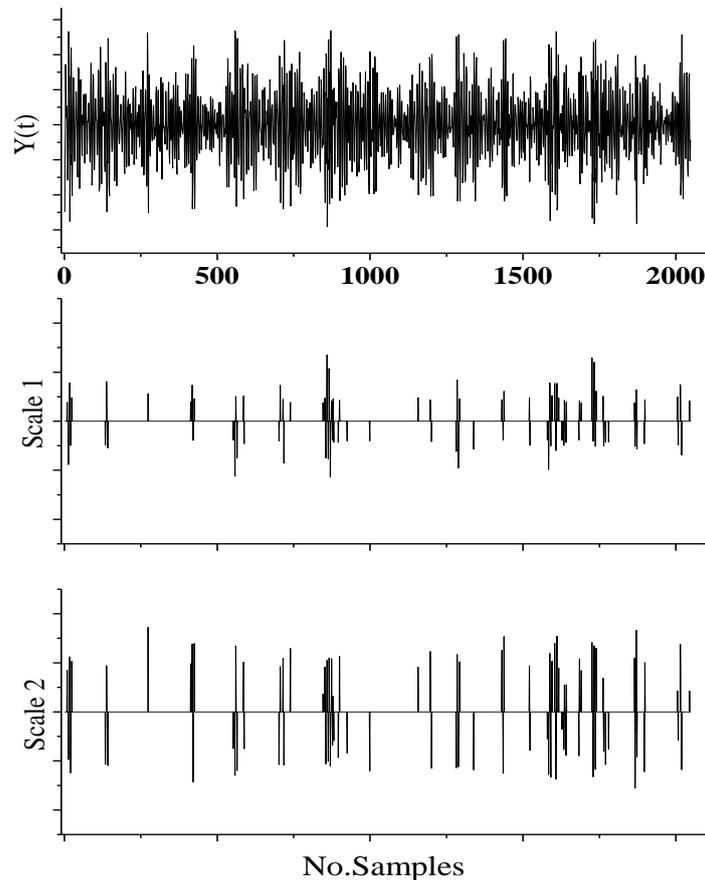


Figure 6-2 Raw AE signal of localised defect (first row) and UDWT-SURE denoised coefficients at two scales.

Figure 6-2 shows a 2048-point segment of the acquired AE signal from the localised defect bearing and the corresponding UDWT-SURE thresholded WMCs obtained from the UDWT wavelet coefficients. Scale 1 is the finest detail scale corresponding to frequency band [20 kHz, 40 kHz]; Scale 2 corresponds to frequency band [10 kHz, 20 kHz]. The WMCs at both scales always have the same sign and line up at the same positions. Note that the WMCs keep the

important signal features at sharp singularities and most of the noise fluctuation has been suppressed. This is because the signal feature shows persistence across scales with larger amplitude, while the noise attenuates very quickly when the scale increases.

Hence the local maxima on those UDWT-SURE denoised wavelet coefficients at the finest scale are detected using the LocMax algorithm. Figure 6-3 shows the acquired AE signal and the clear timings of impact signals, where the regular intervals between adjacent impacts are clearly visible at both scales.

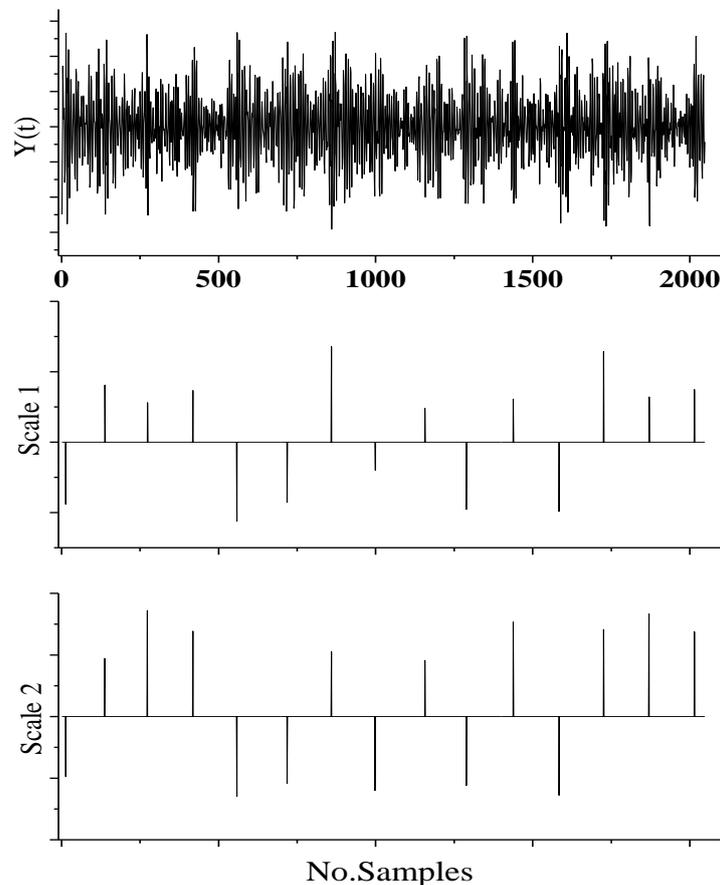


Figure 6-3 Clear timings of impact signals are revealed at two scales after the LocMax detection algorithm is performed on denoised coefficients.

The intervals calculated from the timings of adjacent local maxima then form the distribution for diagnosis. This interval distribution is plotted to study the

diagnostic performance of the signals after the UDWT-SURE denoising. Figure 6-4, Figure 6-5 and Figure 6-6 show the interval distributions from a localised defect, an incipient contamination fault and a fault free bearing, respectively. Clearly, the interval distribution from localised defect signals in Figure 6-4 shows a Cauchy distribution. The distribution in Figure 6-4 has a very high and narrow peak around the mean value. This peak converges to a mean value $\mu = 142$ samples, which is very close to the theoretical estimation $CDI = 143$ samples. This illustrates the performance of our detection method.

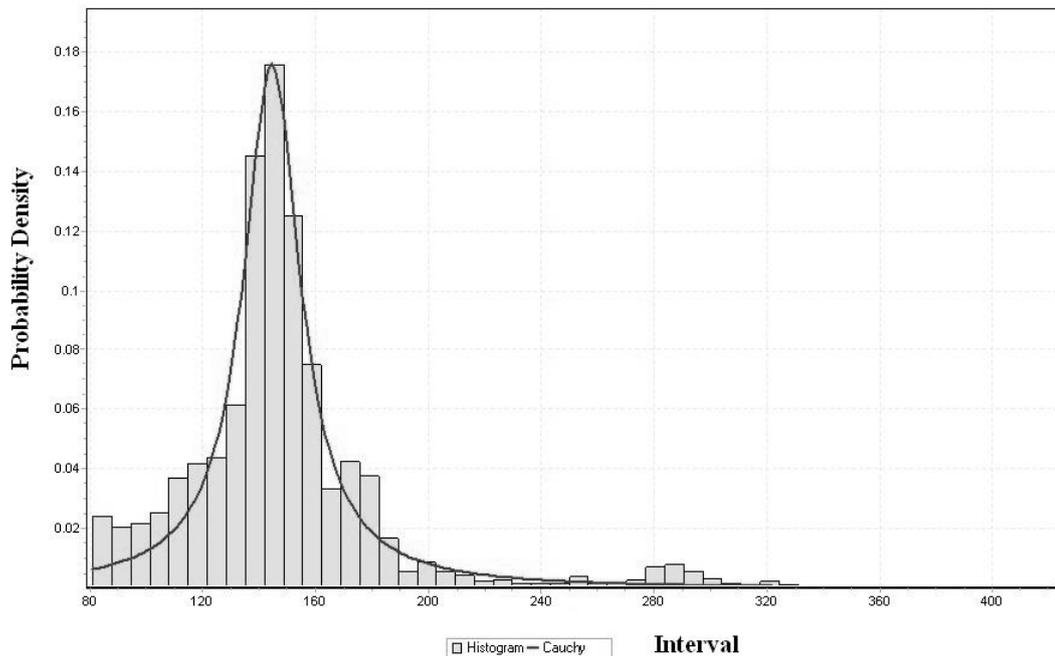


Figure 6-4 Interval distribution from AE signals of localised defect using UDWT-SURE denoising and LocMax-Interval method.

The interval distribution from an incipient contamination fault bearing signal in Figure 6-5 shows an exponential distribution; the distribution from a fault free bearing in Figure 6-6 shows that a rare event distribution is found. The bearing conditions can then be diagnosed from these interval distributions.

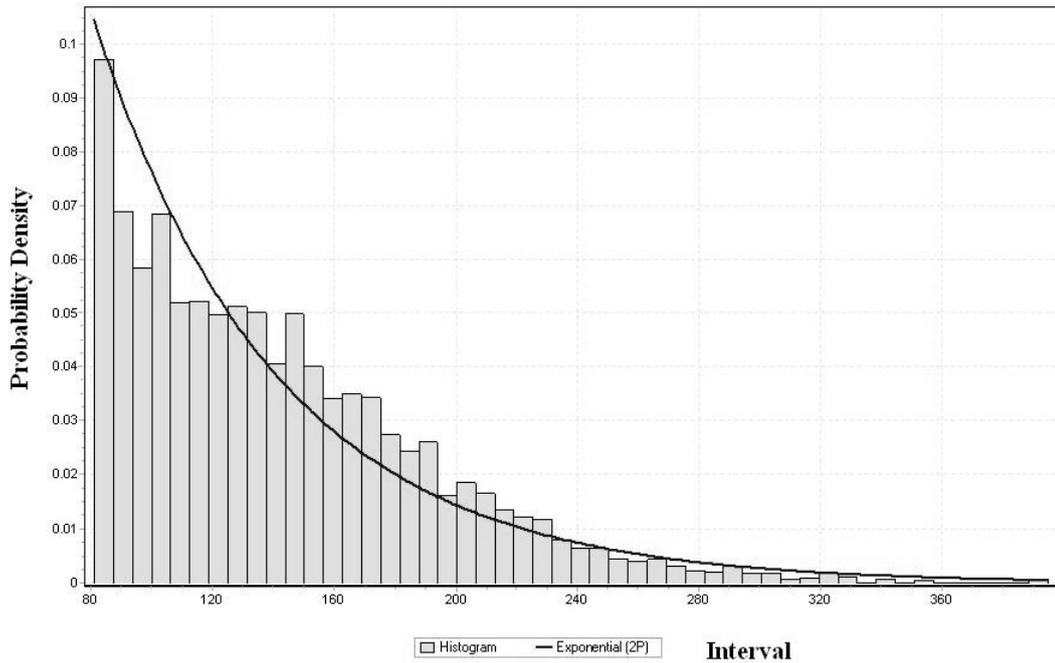


Figure 6-5 Interval distribution from AE signals of incipient contamination fault using UDWT-SURE denoising and LocMax-Interval method.

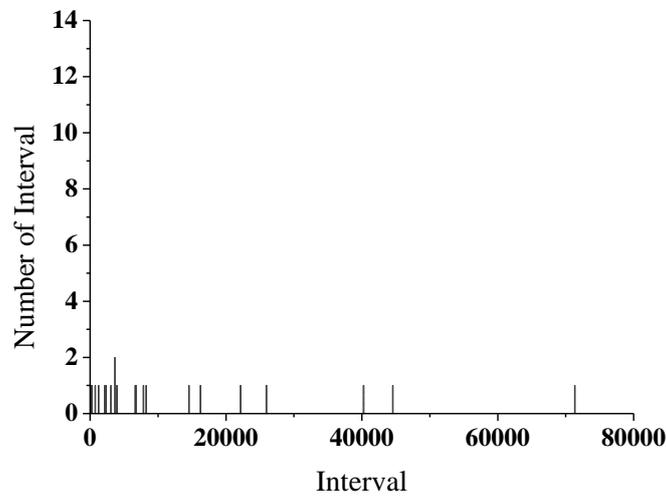


Figure 6-6 Interval distribution from AE signals of fault free bearing using UDWT-SURE denoising and LocMax-Interval method.

The advantage of the UDWT denoising is it can automatically adjust the denoising processing for the AE signals acquired from three different bearing conditions. When there is a localised defect on the bearing, cyclostationary

impacts are generated and the energy of these impacts extends to the high frequency band [20 kHz, 40 kHz]. The detected intervals will be distributed quite regularly. As a result, the signal process is characterised by the narrow peak in interval distribution as shown in Figure 6-4. For the signals from the bearing with an incipient contamination fault, there are some random non-cyclostationary components at high frequencies and the UDWT denoising does not completely eliminate these. Therefore, the intervals show an exponential distribution as shown in Figure 6-5. For the signal from the fault free bearing, the noise has been nearly eliminated and the signal process becomes an approximately zero process. The interval distribution is shown in Figure 6-6. The tests show the interval distributions generated from the bearing conditions of an incipient contamination fault and fault free can consistently exhibit exponential and rare event distributions accordingly, despite different thresholding estimators used in UDWT denoising (i.e. SURE, Minimax or SquareTwoLog).

When performed with the LocMax algorithm on the denoised signals, the parameter *dis* is used to control local windows. Figure 6-7 shows the interval distributions from a localised defect bearing when the parameter is 60, 80, 90 and 100, respectively. For these four distributions, the Chi-square goodness-of-fit Tests of Cauchy distributions give the statistical results 1021, 78, 106 and 127, respectively. The smaller the statistical result, the better the interval distribution fits the Cauchy distribution. Out of these four situations, *dis* = 80 is the best value of the parameter for diagnosing a localised bearing defect. For a specific application, it is then possible to optimise the value of parameter *dis* for the LocMax algorithm towards automatic detection and diagnosis when combined with the test statistics.

In Figure 6-7 top left, note that some wrong detections accumulate at the lower part of the distribution when a relatively small value of parameter is used (*dis* = 60). If a smaller value of parameter is used (e.g., *dis* = 10), this bi-modal distribution will develop into an exponential distribution because more and more wrong detections occur. The effect of increasing the threshold based on DWT

denoising for reducing this was studied in the last chapter. The purpose of increasing the threshold is to improve the effective range of the parameter values, which allows the LocMax-Interval method to work well with a wider range of parameters. For example, changing the threshold estimator from SURE to the estimators with higher threshold estimation, such as Minimax and SquareTwoLog, can improve the effective range of the parameter values.

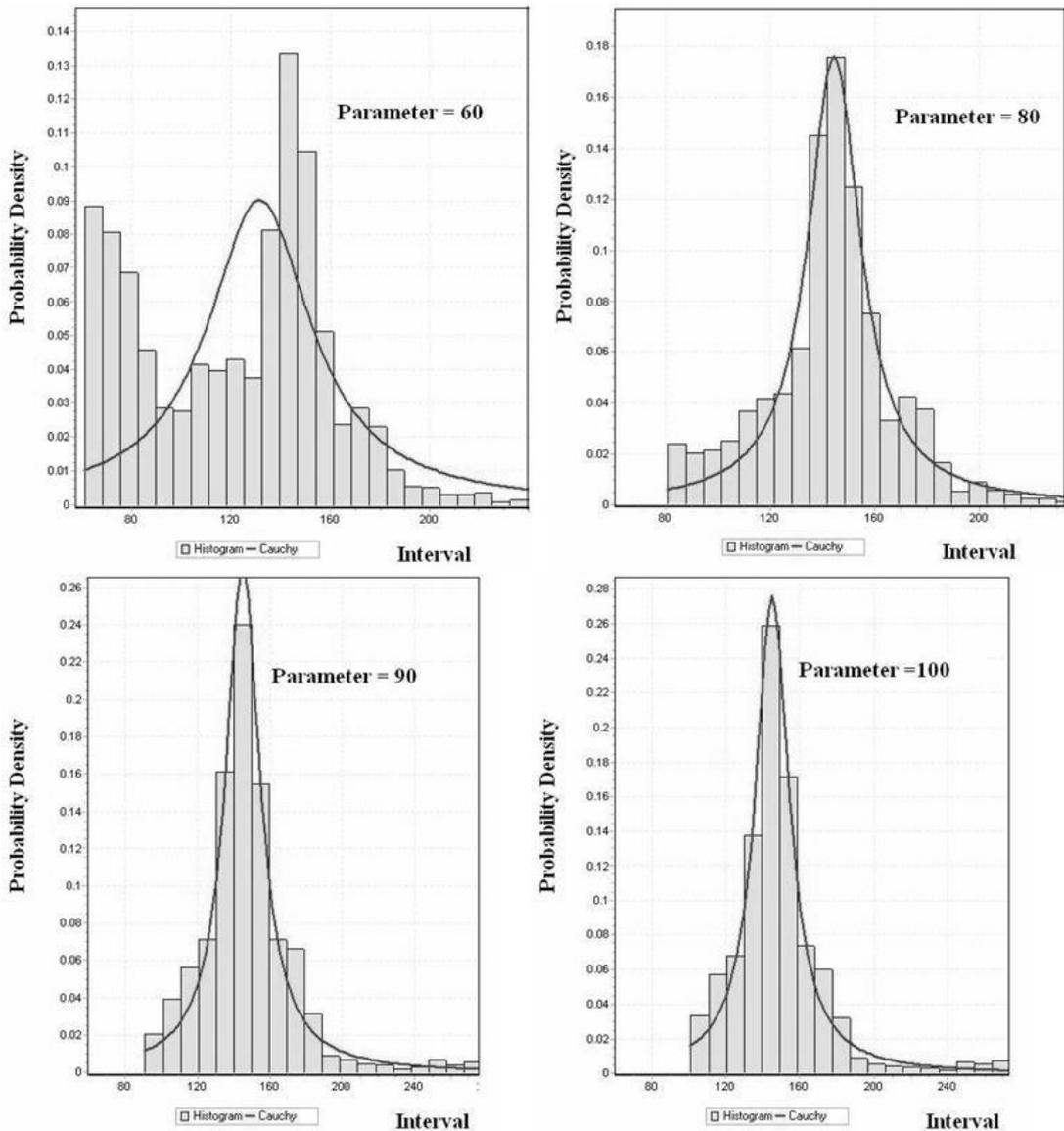


Figure 6-7 Interval distribution from AE signals of localised defect using UDWT-SURE denoising and LocMax-Interval method (the parameter of LocMax algorithm is set to 60, 80, 90, 100).

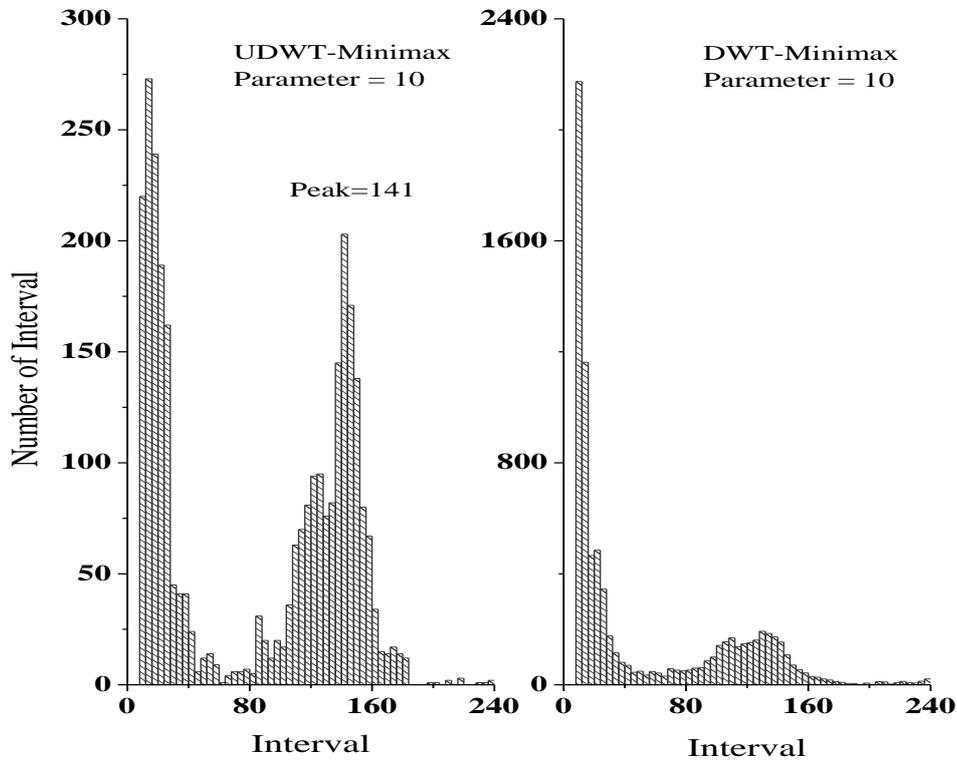


Figure 6-8 Comparison of interval distributions using UDWT and DWT denoising as pre-processing for LocMax-Interval Method.

A similar strategy can also be applied for the UDWT denoising to improve the effective range of the parameter values. When there is no *a priori* information about the bearing, a possible diagnostic scheme is to examine the interval distribution using some small value of parameter (e.g., $dis = 10$). In this case, UDWT-Minimax denoising which is stricter than UDWT-SURE denoising is suggested for pre-processing. See Figure 6-8 (left) for the generated interval distributions using the UDWT-Minimax denoising. Compared with the result from DWT denoising on the right side of Figure 6-8, the proposed UDWT denoising method can show better performance when a small value of parameter $dis = 10$ is used. The regularity is still visible for the UDWT denoised signal, while it is hardly seen for the DWT denoised signal. The UDWT-Minimax denoising

improves the diagnostic capability of the LocMax-Interval method using a small parameter. In our test, the interval distribution from a fault free bearing will show a rare event distribution like Figure 6-6; the interval distribution from a localised defect bearing will reveal the regularity as shown in Figure 6-8 left; the interval distribution from the bearing with an incipient contamination fault will maintain an exponential distribution. With UDWT-Minimax denoising as a pre-processing approach, LocMax-Interval method will work with small parameter values and diagnose three different bearing conditions.

6.5 Summary

This chapter studied a new AE signal processing approach based on UDWT representation and denoising. The method allows direct fault detection on UDWT denoised wavelet coefficients without the need to reconstruct the signal. It is also found that the UDWT denoising can effectively improve the performance of the LocMax algorithm and the diagnostic capability of the LocMax-Interval method.

The method is demonstrated on the bearings of a dry vacuum pump seeded with different conditions: localised defect, incipient contamination fault and fault free. The results show that UDWT denoising, which combines Wavelet Maxima Chain search and thresholding, can improve the diagnostic capability of the LocMax-Interval method. After the UDWT denoising, three different bearing conditions can be diagnosed by examining their generated interval distributions. When there is a localised defect on the bearing, the detected interval distribution shows a Cauchy distribution which is characterised by the narrow peak at the theoretical CDI. When there is incipient contamination fault on the bearing, the detected intervals show an exponential distribution. This is because there are some random non-cyclostationary components at high frequencies for the incipient contamination fault and the UDWT denoising does not completely eliminate these components. When the bearing is fault free, the noise can be completely eliminated by the UDWT denoising and the signal process becomes an approximately zero process. The interval distribution then shows a rare event

distribution. It is also possible to optimise the value for the parameter of the LocMax algorithm towards automatic detection and diagnosis when combined with a statistics test. If small values of the parameter need to be used in the LocMax-Interval method to diagnose three bearing conditions, the UDWT denoising outperforms the DWT denoising, and the UDWT-Minimax denoising is recommended for use as a pre-processing approach.

Chapter 7

Application of Wavelet Packet Quantifiers and Bayesian Method

7.1 Introduction

This chapter studies the application of Wavelet Packet (WP) quantifiers and the Bayesian method in Bearing Fault Detection and Diagnosis (BFDD). Appendix A.10 will introduce the concept and statistical explanation of WP quantifiers. The WP quantifiers which are obtained from WP coefficients can construct a quantitative time-frequency analysis. The materials presented in this chapter have been published in different form by Feng and Schlindwein (2008).

In the field of BFDD, wavelets and WPs to date were most commonly used as a complementary tool of envelope analysis to overcome the limitations with regard to determining the frequency bands of interest (see Section 4.6.1 for details). This is because the low frequency range of the commonly used vibration signals is usually dominated by noise. In this chapter, the WP quantifiers are applied to analyse the AE signals and extract signal characteristics from bearings in three conditions: fault free, with a localised defect and with an incipient contamination fault. The study is to find out whether the WP quantifiers are suitable for diagnosing those different bearing conditions and whether there exists a “best” quantifier.

To find out the “best” quantifier, the Bayesian method is introduced in this chapter to analyse and evaluate quantitatively the performance of the WP quantifiers. This quantitative study method also has many important applications in BFDD.

Quantitative studies can be performed to investigate the relationships between the performance of the quantifiers and some consideration factors in implementation. These factors include the wavelet order, length of signal segment and dimensionality of diagnostic scheme.

This chapter is organised as follows: Section 7.2 provides an overview of the methodology, including the introduction to WP quantifiers and the performance evaluation method. Section 7.3 shows the results of applying WP quantifiers on AE signals analysis. Section 7.4 evaluates the performance of WP quantifiers and finds out the “best” quantifier by comparing the Bayesian classification error probability. For implementation, wavelet order, length of signal segment and dimensionality of the diagnostic scheme are often the important consideration factors. Sections 7.5 - 7.7 study the relationships between the performance of WP quantifiers and these consideration factors via the Bayesian method. The results are discussed in Section 7.8 and finally, the summary is reported in Section 7.9.

7.2 Methodology

7.2.1 WP Quantifiers

By decomposing the details of DWT coefficients into dyadic frequency bands, the DWPT can yield a STFT-like decomposition but with some important advantages. The wavelet basis of the DWPT can be flexibly chosen for practical implementation considerations; the contaminating noises concentrating in some frequency bands can be easily eliminated; the DWPT can be efficiently computed via the filter bank algorithm.

Recall that the WP node (j, k) corresponds to the vector $\mathbf{P}_{j,k}$ of WP coefficients. The collection of nodes forming the indices of the WP table will be denoted by $T \equiv \{(j, k) : j = 1, \dots, J; k = 0, \dots, 2^j - 1\}$. The coefficients vector $\mathbf{P}_{j,k}$ is associated with the frequency band of signal on $[f_s \cdot k / 2^{j+1}, f_s \cdot (k + 1) / 2^{j+1}]$. The terminal WP nodes can preserve the energy of the signal. The new quantifiers construct the quantitative time-frequency analysis based on WP coefficients:

Relative Energy (RE) measures the normalised energy of the WP node; Total Entropy (TE) measures how the normalised energies of the WP nodes are distributed in the frequency domain; Node Entropy (NE) describes the uncertainty of the normalised coefficients of the WP node.

In the following study, these WP quantifiers are used to process the AE signals. AE signals acquired from three different bearing conditions (fault free, incipient contamination fault, localised defect) are divided into L -point signal segments without overlap. Each segment S is then decomposed at $j=2$ levels by DWPT using the Daubechies wavelet as base because Daubechies wavelets are compactly supported for a given number of vanishing moments (see Appendix A.7). The terminal WP nodes $(2,0)$, $(2,1)$, $(2,2)$, $(2,3)$ correspond to frequency bands $[0, 10 \text{ kHz}]$, $[10 \text{ kHz}, 20 \text{ kHz}]$, $[20 \text{ kHz}, 30 \text{ kHz}]$ and $[30 \text{ kHz}, 40 \text{ kHz}]$, respectively. The last three bands are labelled as A, B and C, see Figure 7-1.

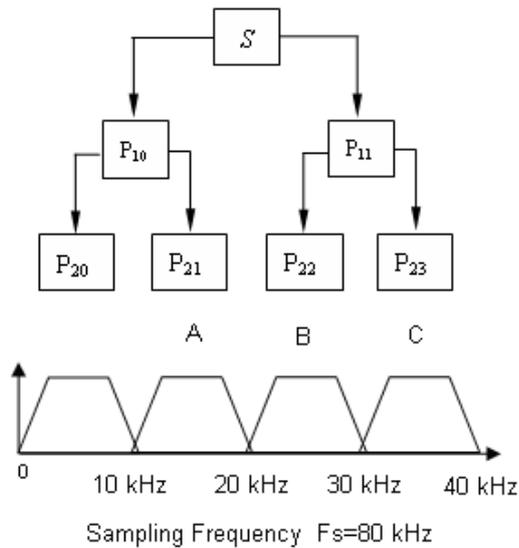


Figure 7-1 2-level DWPT decomposition on AE signals (frequency ordering).

Because the interesting information in AE signals is above 10 kHz, the low frequency band of AE signals $[0, 10 \text{ kHz}]$ is not used in the analysis and is discarded. WP nodes A $[10 \text{ kHz}, 20 \text{ kHz}]$, B $[20 \text{ kHz}, 30 \text{ kHz}]$ and C $[30 \text{ kHz}, 40 \text{ kHz}]$ are analysed using the proposed quantifiers. Specifically, the energy of the

denoised signal \hat{S} is the total energy of the signal on A, B and C band

$$E_{tot} = \sum_{n=A,B,C} E_n, \quad (7-1)$$

where $E_n = \sum_{t=0}^{N/4-1} P_{n,t}$ for $n = A, B, C$ band.

The RE of band n is

$$RE_n(f) = E_n / E_{tot}, \quad (7-2)$$

and the NE in band n is

$$NE_n(t) = - \sum_{t=0}^{N/4-1} P_{n,t} \cdot \log_2 P_{n,t}, \quad (7-3)$$

and the TE of a signal segment is

$$TE(f) = - \sum_{n=A,B,C} RE_n(f) \cdot \log_2 RE_n(f). \quad (7-4)$$

7.2.2 Bayesian Method

BFDD can be systematically studied in the context of pattern recognition (Theodoridis and Koutroumbas, 2003). The performance of the WP quantifiers, also known as the features, will be evaluated by comparing their classification error probability. The adopted classification method based on Bayes decision theory is optimal with respect to minimising the classification error probability or the total risk.

Consider a classification task of M classes, ω_i , $i = 1, 2, \dots, M$, the *a posteriori* probability of an unknown pattern, which is represented by a feature vector \mathbf{X} , is denoted by $P(\omega_i | \mathbf{X})$. It represents the probability that the unknown pattern belongs to the respective class ω_i , given the corresponding feature vector takes \mathbf{X} . For $\mathbf{X} \in R_i$, the *a posteriori* probability $P(\omega_i | \mathbf{X})$ is the correct classification

probability; otherwise, it is the classification error probability.

The risk associated with the class ω_k is

$$r_k = \sum_{i=1}^M \lambda_{ki} \int_{\mathbf{X} \in R_i} p(\mathbf{X}|\omega_k) d\mathbf{X}. \quad (7-5)$$

The total risk for the Bayesian classification is

$$r = \sum_{k=1}^M r_k P(\omega_k) = \sum_{k=1}^M \sum_{i=1}^M \lambda_{ki} \int_{\mathbf{X} \in R_i} p(\mathbf{X}|\omega_k) P(\omega_k) d\mathbf{X}, \quad (7-6)$$

and this can be written in the form of *a posteriori* probability as

$$r = \sum_{k=1}^M \sum_{i=1}^M \lambda_{ki} P(\omega_k | \mathbf{X} \in R_i). \quad (7-7)$$

In equation (7-5), the probability density function $P(\mathbf{X}|\omega_i)$ describes the distribution of the feature vectors in each class. The integral in the region R_i of the class ω_i represents the classification error probability of ω_k caused by misclassification, i.e. the values of a feature belonging to class ω_k being misclassified as belonging to class ω_i .

The penalty term λ_{ki} is zero for $k=i$ because no penalty is given to correct classification. When there is no practical risk evaluation available, assume λ_{ki} is 1 for any $k \neq i$. That is, any misclassification has equal penalty. The above total risk becomes the total classification error probability

$$E = \sum_{k=1}^M \sum_{i=1}^M P(\omega_k | \mathbf{X} \in R_i). \quad (7-8)$$

The less the probability densities of different classes with respect to a quantifier are overlapped, the better their ability to discriminate and diagnose different bearing conditions.

7.3 Analysing AE signals via WP Quantifiers

In this analysis, AE signals are first divided into 292 sets of 2048-point signal segments and decomposed by the 2-level DWPT using db10 as the wavelet basis. The histograms of the quantities of quantifier RE of A band [10 kHz, 20 kHz], B band [20 kHz, 30 kHz], C band [30 kHz, 40 kHz] for three different bearing conditions are shown in Figure 7-2, Figure 7-3 and Figure 7-4, respectively. The distributions are well fitted with a Gaussian distribution (profile in solid curves), likewise for the following figures. In Figure 7-2, most energy of the signals is concentrated in the A band [10 kHz, 20 kHz]: over 65% of the energy for the incipient contamination fault, over 80% for fault free and over 85% for the localised defect.

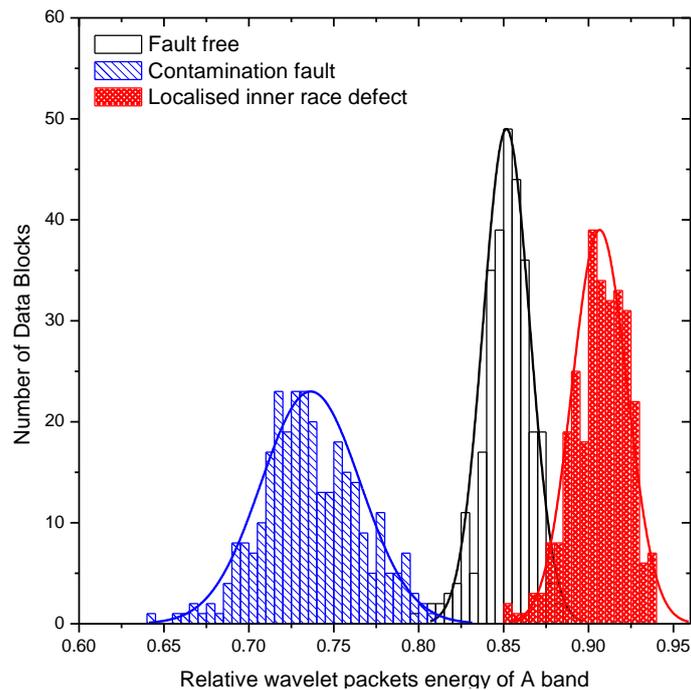


Figure 7-2 The histogram of RE of A band for three classes ($L = 2048$, db10).

The energy of B band [20 kHz, 30 kHz] is the second largest. The energy for an incipient contamination fault is over 18%, for fault free, about 10% to 18% and for a localised defect, no more than 8%, as shown in Figure 7-3. Notice that the

three classes here are well separated except for a slight overlap between fault free and incipient contamination fault. Compared to A and B band, C band [30 kHz, 40 kHz] contains the lowest energies as shown in Figure 7-4. Only 2% to 8% energies are for the localised defect and less than 2% of the energy for the other two classes.

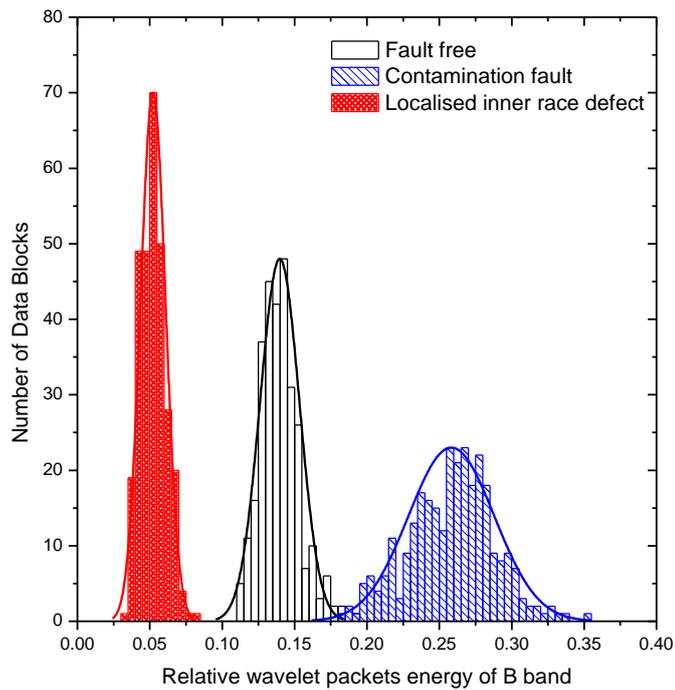


Figure 7-3 The histogram of RE of B band for three classes ($L = 2048$, $db10$): Different classes are well separated.

When an incipient contamination fault happens, it corresponds to the lowest energy portion in band A compared to the other two bearing conditions. However, relatively high energies tend to spread to band B and almost all the energies attenuate in band C. In contrast, the localised defect has very different characteristics. It shows relatively high energies in bands A and C, but lower energy in band B. In all three bands, the energy probability distributions for the fault free condition are located in the middle of the other two.

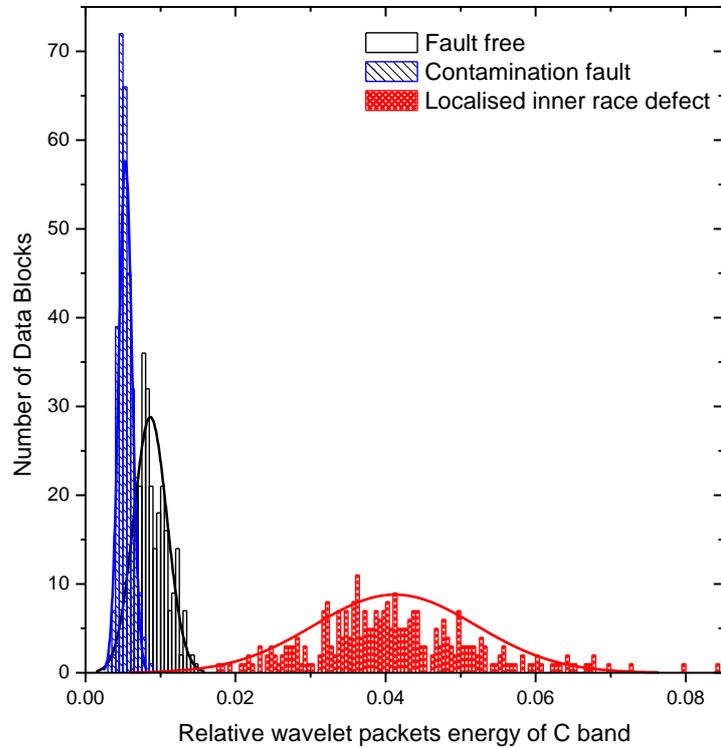


Figure 7-4 The histogram of RE of C band for three classes ($L = 2048$, $db10$).

To further investigate these energy distributions in the three bands, we obtained the histogram of the quantities of the TE for three bearing conditions (Figure 7-5). This quantifier describes the degree of the disorder of signals from the point of view of frequency energy distribution. Narrowband-like signals are more ordered than wideband-like signals. If the total energies of signals concentrate in one band, the quantities will then have relatively small values. From the previous graph, it is shown that localised defects produce more narrowband-like signals, where over 85% of the energy concentrates in band A, while bands B and C have together no more than 15% of the energy. The incipient contamination fault signal produces more broadband-like signals, where 65% to 80% of the energy is in band A and 18% to 35% of the energy is in band B. Broadband-like signals are more disordered than narrowband-like signals. The energy of wideband like signals tend to spread over the bands and the quantifier therefore takes larger values.

According to the above analysis, it is expected that the relationship of the quantities of the quantifier TE for the three bearing conditions would be: TE (incipient contamination fault) > TE (fault free) > TE (localised defect). Figure 7-5 shows a good agreement with this expectation. Again notice that the overlaps between the incipient contamination fault and the other two conditions are very small, which implies that this quantifier is good at discriminating incipient contamination faults. The quantitative analysis for the overlaps will be given in Table 7-1.

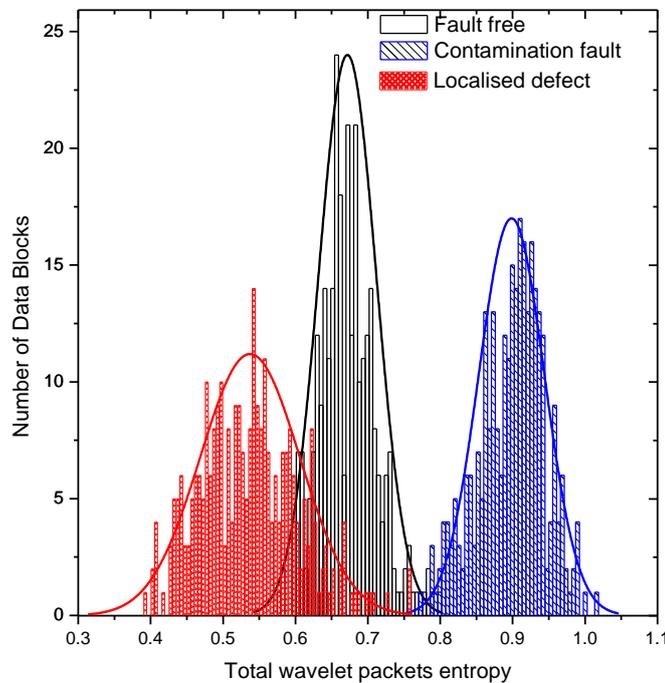


Figure 7-5 The histogram of TE for three classes ($L = 2048, db10$).

As mentioned earlier, the quantifier NE is used to describe how disordered the normalised coefficients are in a frequency band. The quantities of the quantifier will take small values if coefficients show ordered behaviour, and vice-versa. Figure 7-6 shows the histogram of the quantities of the quantifier NE in band C.

In Figure 7-6, the quantities for the localised defect are much lower than the other two, which implies that the coefficients might be very ordered and the processes

concentrate energy in a few large coefficients and have stronger peaks fluctuation. Figure 7.7 displays the normalised coefficients in band C (from top: localised defect, incipient contamination fault, fault-free) and shows good agreement with the above analysis. For the fault-free and the incipient contamination fault, the processes are relatively disordered and their energies are distributed more evenly. Consequently, the quantities for them take larger values. The overlaps between localised defect and the other two conditions are very small, which indicates this quantifier is good at discriminating localised defects.

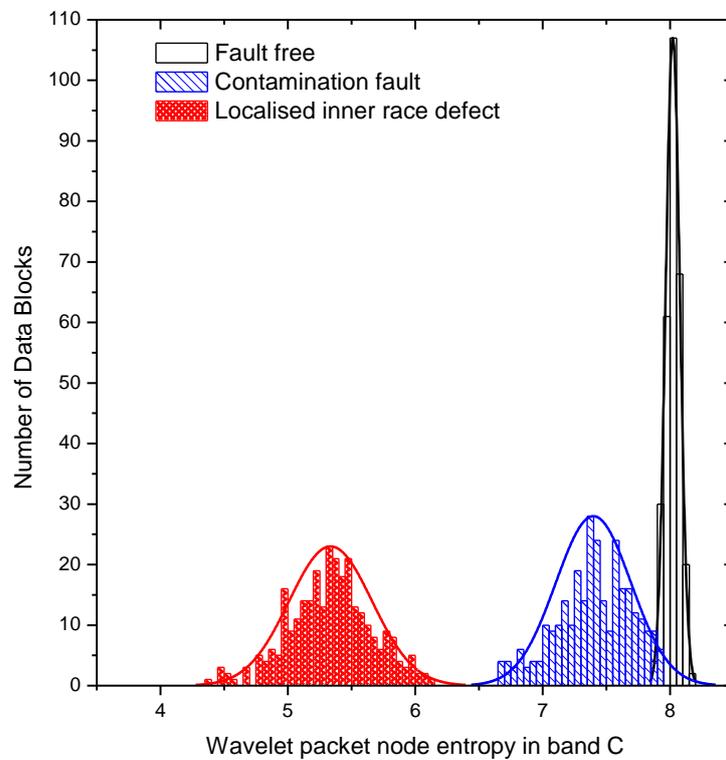


Figure 7-6 The histogram of NE in band C for three classes ($L = 2048, db10$).

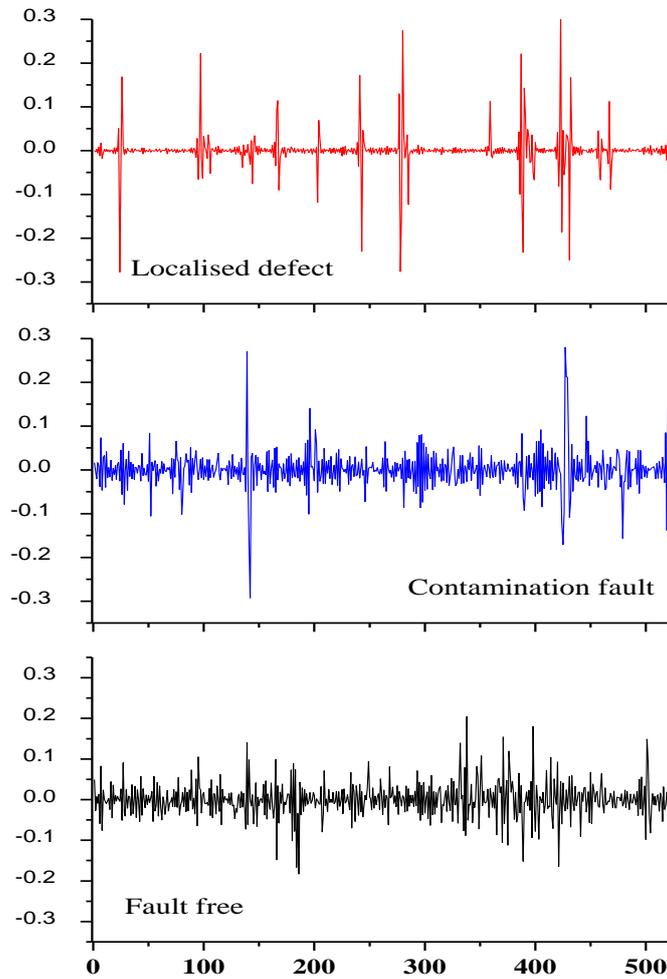


Figure 7-7 The normalised WP coefficients in band C ($L = 2048$, $db10$).

7.4 Finding out the Best Quantifier via Bayesian Method

To quantitatively evaluate the performance of the proposed quantifiers, the Bayesian classifier is used to compute the classification error probability, where the probability density functions are estimated by Gaussian distributions. The classifier is trained with 30% of the data and tested with the remainder. In the following context, the quantifier RE of A, B, C band are denoted as F1, F2 and F3; the quantifier TE is denoted as F4; the quantifier NE of A, B, C band is denoted as F5, F6 and F7. Table 7-1 shows the classification error probability using the quantifiers F1~F7.

The first row in Table 7-1 is the classification error probability for the fault-free condition, that is, the patterns belong to fault free but are misclassified to the other two classes. The second row and the third row are the classification error probability for the incipient contamination fault and localised defect, respectively. The fourth row is the total classification error probability from the sums of all three types of classification error probability. Significant low classification error probability (less than 5%) is highlighted in red colour.

Table 7-1. Classification error probability using different quantifiers ($L = 2048$ and $db10$).

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>	<i>F7</i>
<i>Fault free</i>	0.048	0.003	0.182	0.151	0.699	0.702	0.175
<i>Contamination</i>	0.003	0.003	0.092	0.010	0.387	0.469	0.592
<i>Localised defect</i>	0.031	0	0	0.140	0.397	0.188	0
<i>Total</i>	0.081	0.006	0.273	0.300	1.482	1.359	0.768

NB: since the total error includes false positives and false negatives it might be larger than 1.

The “best” diagnosis performance is from F2, where total error probability is only 0.6%. The second “best” performance is from F1, where total error probability is 8.1%. The results can also be explained from the corresponding Figure 7-2 and Figure 7-3, both showing small overlaps among different classes. In Figure 7-3, all three classes are well separated except slight overlaps between fault free and incipient contamination fault. The results in Table 7-1 show the small overlaps cause 0.3% classification error probability for each class. Because the distribution of the localised defect has no overlap with the other two classes, the classification error probability for it is zero. In Figure 7-2, more overlaps between the localised defect and fault-free classes appear, which cause the increasing classification error probability 4.8% and 3.1% for each class. The other quantifiers have not attained satisfactory results, with their total error probabilities over 25%. Although their total performance are not satisfactory, these quantifiers have very low classification error probability for the classes relevant to some faults (e.g. F3 and

F7 for localised defect, F4 for incipient contamination fault), which means they have good specificity for certain types of bearing fault.

7.5 Performance Factor 1: Wavelet Order

In the above results, the Daubechies wavelet dbN with large order $N = 10$ was used as the WP basis. This Section investigates how the change of order N influences the performance of the quantifiers. Figure 7-8 shows the total classification error probability of the proposed quantifiers when using different Daubechies wavelet orders ($N = 1, 2, \dots, 10$).

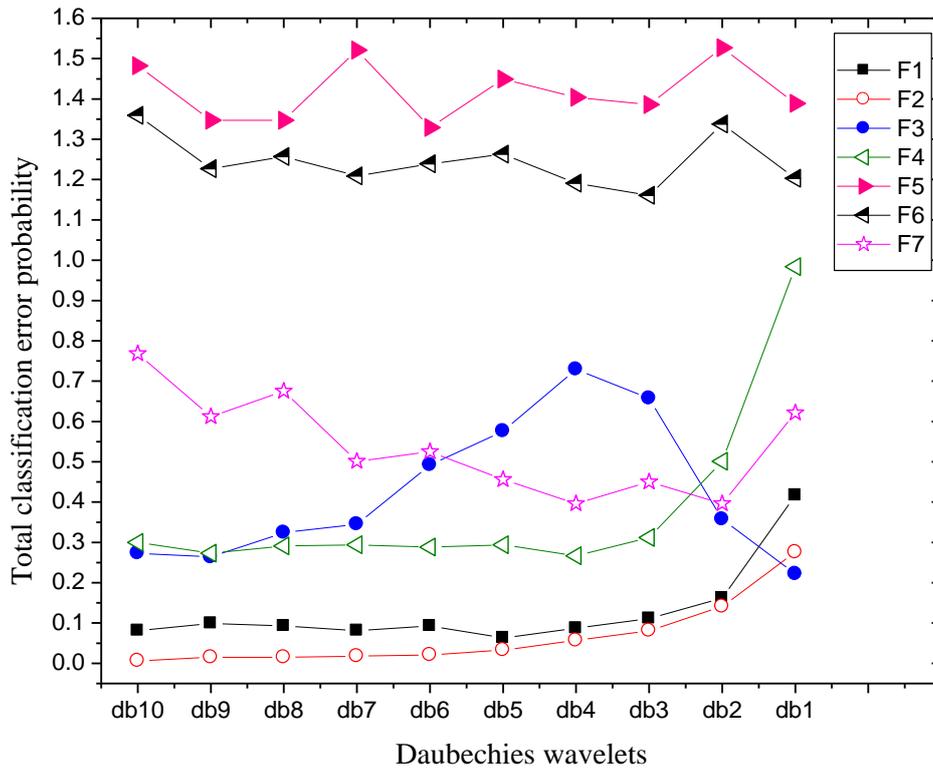


Figure 7-8 Total classification error probability of the quantifiers using different Daubechies wavelet orders ($N = 1, 2, \dots, 10$) ($L = 2048$).

The graph shows F5 and F6 both have a poor performance, with unacceptable

classification error probability. F4 and F7 also perform badly, with over 30% classification error probability for all the wavelet orders. F3 performs best when db1 (haar) basis is used. F2 has the best performance for the larger wavelet orders but the performance deteriorates when wavelet order goes lower. F1 is the second best quantifier overall, but the classification error probability also increases when wavelet order decreases.

7.6 Performance Factor 2: Length of Signal Segment

To study how the length of signal segment influences the performance of the quantifiers, the AE signals are divided into signal segments of different sizes. For different lengths, only 292 sets of signal segments are kept for Bayesian classification.

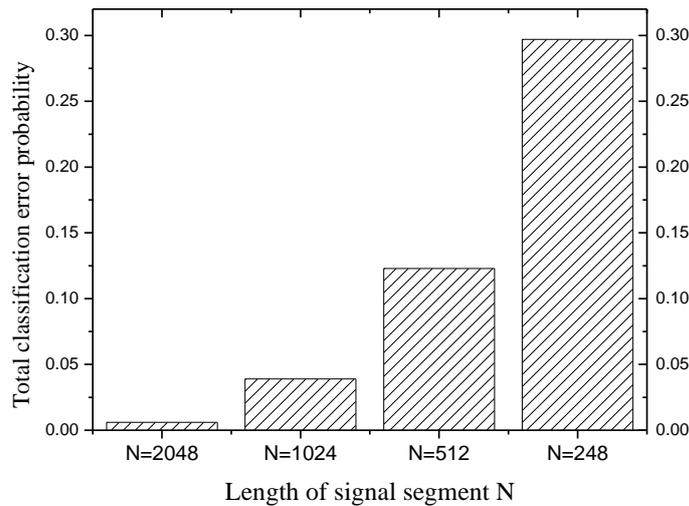


Figure 7-9 Total classification error probability increases when the length of signal segment decreases.

Figure 7-9 displays the total classification error probability when using different lengths of signal segment for the quantifier F2 and the Daubechies wavelet dbN with order $N = 10$ (Note: The previous analyses have showed the quantifier F2 and db10 as the wavelet basis which can achieve the best performance of all

amongst the options tested). Figure 7-9 shows that the total classification error probability increases significantly when the length of the signal segment decreases.

7.7 Performance Factor 3: Dimensionality of Diagnostic Scheme

Figure 7-8 showed how the performance of the “best” quantifier F2 deteriorates when wavelet order decreases. When very low order db1 (Haar) is used, the total classification error probability increases to 27.6%, where two classes - fault free and incipient contamination fault - overlap heavily. Figure 7-10 shows the histogram for this situation. A possible approach to improve the classification performance is to use a two-dimensional diagnostic scheme, i.e. combine a pair of quantifiers to diagnose bearing faults.

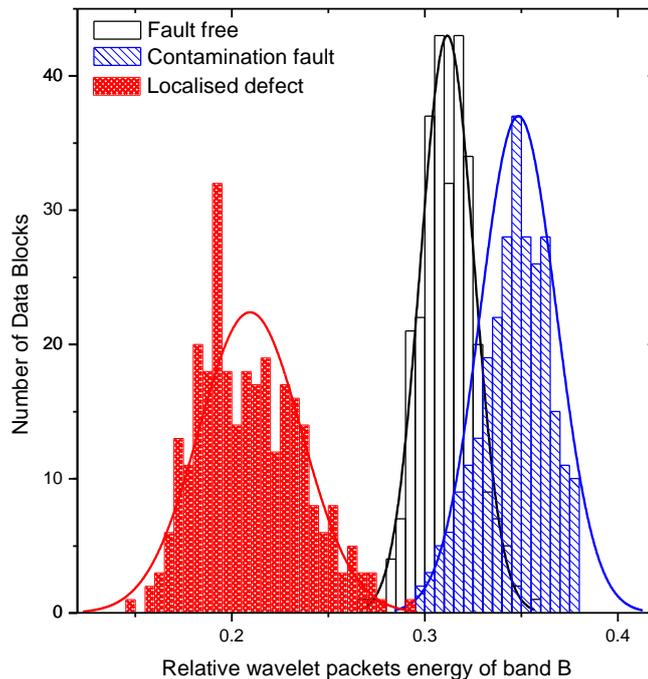


Figure 7-10 The histogram of quantifier F2 for three classes using Haar wavelet ($L = 2048$); the total classification error probability is very high at 27.6%.

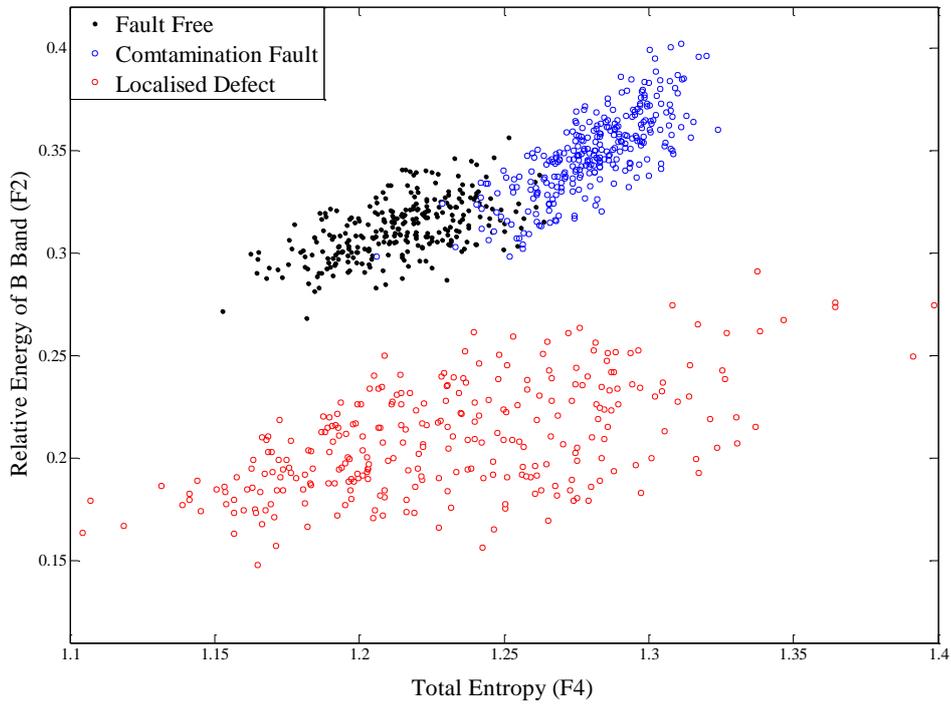


Figure 7-11 Two-dimensional diagnostic map using two quantifiers F2 and F4 (Haar wavelet; $L = 2048$): Total classification error probability has greatly reduced.

Figure 7-11 shows a two-dimensional diagnostic map using two quantifiers F2 and F4. The quantifier F4 is used to help for discriminating the incipient contamination fault and fault free cases, where the total classification error probability reduces from 27.6% to 9.6%.

Figure 7-12 shows the comparison of classification error probability between the “best” single quantifier F2 and the “best” combination pair of quantifiers. It shows that a two-dimensional diagnostic scheme can help to improve the diagnostic performance. The improvements are more significant when lower wavelet orders are used.

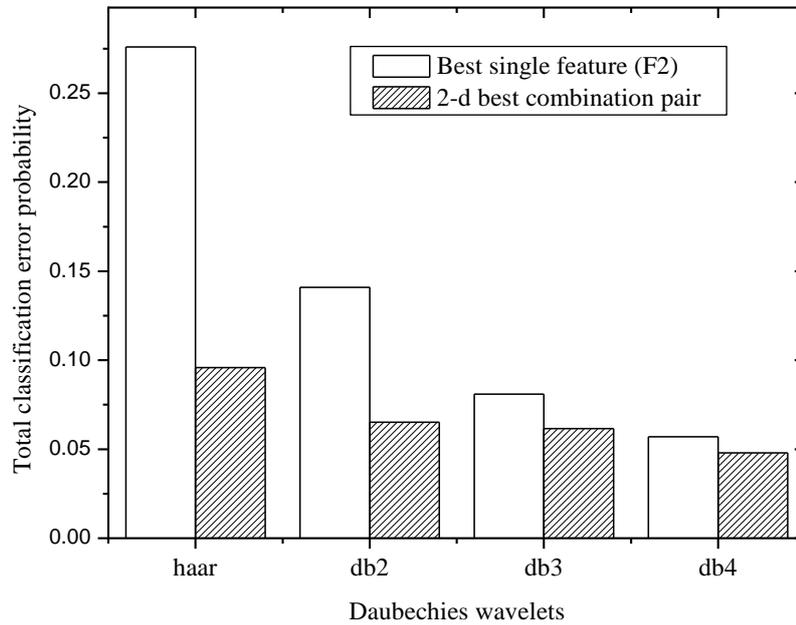


Figure 7-12 Using two-dimensional diagnostic scheme reduces the classification error probability; improvement is significant for lower wavelet order ($L = 2048$).

7.8 Discussion

For mechanical vibration and high frequency AE, the loss of energy due to attenuation increases with the frequency. This is evident in studying the RE, where for all the three bearing conditions the energies decay along the A, B, and C bands. Although the energies in band B [20 kHz, 30 kHz] are far less than for band A [10 kHz, 20 kHz], band B provides better information for the diagnosis. All three classes shown in the histograms are well separated, which means the quantifier RE of band B has the greater power for diagnosis. The corresponding total classification error probability of the Bayesian classifier (only 0.6%) proves this point. Band C [30 kHz, 40 kHz] contains the lowest energies, but it is a very good band for detecting localised defects by using RE or NE quantifiers. The quantifier TE is good at detecting whether the signals are broadband and proved to be quite efficient to detect an incipient contamination fault.

The WP quantifiers derived from probability distributions are either normalised over the total signal energy or the node energy. The normalisation is important for the successful diagnostic application. In practice, it is very difficult to compare different bearing conditions using the quantifiers without normalisation, unless the amplifier is carefully tuned to achieve equal amplitudes for different signals. Without normalisation the signal with higher amplitude will produce the wavelet coefficients with higher values. This will generate unpredictable effects on the quantifiers. For example, the signal from a localised defect without normalisation will generate much higher amplitude coefficients in band B than those shown in Figure 7-3, where the class belonging to the localised defect moves upwards on the abscissa. If these coefficients were directly used for diagnosis, the classification would often fail. The same situation will also happen for the other quantifiers.

From the viewpoint of implementation, it is expected that the orthogonal wavelet basis would have higher vanishing moments (order) N . This is because the higher the vanishing moments, the better representation the signal will have and signals would be approximated efficiently by fewer non-zero wavelet coefficients. See Section A.7 for details. However for the Daubechies orthogonal wavelets, a trade off has to be made between filter support size and vanishing moment. The filter support size has to be at least $2N-1$ for moment N . The implementation of longer support size will increase the cost and slow down the computation. This study shows the classification error increases when wavelet order N increases for some good quantifiers (F1 and F2) with satisfactory performance overall. An optimal implementation therefore relies on a balance between the increasing classification error probability and the cost of increasing support size.

In practice, diagnostic schemes for BFDD with minimal length of signal segment and dimension(s) are often preferable because increasing length or dimension is also costly for implementation. Unfortunately, analysis shows that the total classification error probability increases significantly as the length of signal segment decreases. It is also shown that a two-dimensional diagnostic scheme

using a pair of quantifiers can help to reduce the classification error probability, but the reductions are only significant when lower wavelet orders are used. Therefore, an optimal implementation also needs to consider how to balance between the increasing classification error probability and the cost of increasing length of signal segment or dimension(s).

7.9 Summary

This chapter reported the study of WP quantifiers as a new tool for BFDD. To the best of our knowledge, this work is the first study on the application of WP quantifiers in BFDD. Particularly, this work provides an alternative method for detecting and diagnosing incipient contamination faults. The WP quantifiers construct a quantitative time-frequency analysis: Relative Energy (RE) measures the normalised energy of the WP node; Total Entropy (TE) measures how the normalised energies of the WP nodes are distributed in the frequency domain; Node Entropy (NE) describes the uncertainty (i.e. the degree of disorder) of the normalised coefficients of the WP node.

By applying the WP quantifiers on AE signals from faulty bearings, this study shows that both localised defects and advanced incipient contamination faults can be successfully detected and diagnosed if the appropriate quantifier is chosen. The Bayesian method is also introduced to quantitatively analyse and evaluate the performance of the WP quantifiers. It is shown how the optimal quantifier for an application can be found by comparing the total classification error probabilities of quantifiers. In this case study on an iGX dry vacuum pump, the RE of band B [20 kHz, 30 kHz] is the optimal quantifier for BFDD. The performance of the WP quantifier also depends on the wavelet order and the length of signal segment. Reducing the Daubechies wavelet order or the length of signal segment will deteriorate the performance of the quantifier. It is also shown that a two-dimensional scheme can also help to improve the diagnostic performance and that the improvements are more significant when lower wavelet orders are used.

Chapter 8

Discussion and Future Work

8.1 Discussion

The rolling element bearing is one of the most common mechanical components in critical industrial rotating machinery. Since the failure of bearings will cause the machine to malfunction and may quickly lead to catastrophic failure of the machinery, it is very important to detect any bearing deterioration at an early stage. Following the previous research works in University of Leicester based on vibration measurements, this Ph.D. study adopts Acoustic Emission (AE) as a measurement method to provide good solutions for detecting both the localised defect and the incipient contamination fault of bearing. The contamination fault at an incipient stage is difficult to detect by using traditional vibration measurements. The bearing faults to be studied are both typical examples carefully selected by BOC Edwards experienced engineers and introduced in test facilities by Barden Bearing specialists. Therefore, they replicate the bearing damage and wear of the dry vacuum pump in natural semiconductor operating conditions.

The difficulties in processing, interpreting and classifying the information from AE signals limit the success of AE technique. An effective signal processing method for a BFDD system depends on the nature of acquired signal. The AE signals, with broadband and high frequency nature, are different from the vibration signals acquired from traditional accelerometers, which necessitate the development of new signal processing method. Signal processing is important that it bridges the stages of sensing and fault diagnosis. Applying the appropriate signal processing methods is essential to the success of a BFDD system.

Based on the cyclostationary model of impact signals and probability law governing the intervals, chapter 5 proposed a novel BFDD method called LocMax-Interval method to detect a localised bearing defect. The new method includes a new algorithm named LocMax algorithm for finding out the timings of impact signals and the interval distribution generated for localised defect detection. By examining whether the interval distribution is regular, a localised defect can be detected without *a priori* knowledge on Characteristic Defect Frequencies (CDFs). Since the involvement of noise will limit the effective range of the parameter values used in the LocMax algorithm, the DWT denoising method was then studied to improve the effective range.

Further improvement of LocMax-Interval method is achieved by pre-processing the AE signals with the UDWT decomposition and denoising. Chapter 6 showed the UDWT denoising, which combined Wavelet Maxima Chain search and thresholding, not only outperformed the DWT denoising on improving the effective parameter range, but also improved the diagnostic capability of the LocMax-Interval method. The BFDD problem (including localised defect, fault free and incipient contamination fault) can be solved using the recommended UDWT-Minimax denoising as a pre-processing approach for LocMax-Interval method.

The applications of WP quantifiers and Bayesian method for BFDD were studied in chapter 7. This is the first study on the application of WP quantifiers in BFDD. The WP quantifiers construct a quantitative time-frequency analysis for BFDD: Relative Energy (RE) measures the normalised energy of the WP node; Total Entropy (TE) measures how the normalised energies of the WP nodes are distributed in the frequency domain; Node Entropy (NE) describes the uncertainty (i.e. the degree of disorder) of the normalised coefficients of the WP node. The result shows that the BFDD problem can be solved if the appropriate WP quantifier is applied. Based on the quantitative performance evaluated by the Bayesian method, the study also shows that the optimal quantifier for an application can be found out by comparing the total classification error

probabilities of quantifiers. This quantitative study method has further applications in BFDD: Quantitative study can also be performed to investigate the relationships between the performance of the quantifiers and some factors. These factors include the wavelet order, length of signal segment and dimensionality of diagnostic scheme, which are all important in implementation.

8.2 Future Work

Since the broad existence of cyclostationary processes in physical and man-made processes, such as ultrasonic imaging of materials and biological tissues, medicine (EEG, ECG), radar, communication systems and cyclic vibration signals produced by Internal Combustion (IC) engines (Gardner *et al.* 2006), the proposed LocMax-Interval method and UDWT denoising are promising to find applications for cyclostationary signal detection in these areas.

The traditional detection techniques can detect localised defects efficiently and provide information on the locations of defects. However, the techniques fail to detect and diagnose some advanced bearing faults such as incipient contamination because these types of fault may not produce an increase in modulated characteristic frequency at an early stage. Time-frequency analysis methods such as DWPT give us deep insight into these challenging problems by using the WP quantifiers. The author believes the proposed method and quantifiers can be used to solve even more challenging problems in the field of condition monitoring. Besides the EEG and condition monitoring, the study methods and WP quantifiers introduced in Chapter 7 are also very promising in many other applications to extract features from signals.

It was pointed out that the Rényi entropy concept could be used to derive a generalised statistical moment, of which the classical statistical moment kurtosis is a special case. When relaxing the mean value property from an arithmetic mean to an exponential mean, the Rényi entropy becomes the well known Shannon entropy (Tao *et al.* 2007). The concept of Rényi entropy can be applied to derive some new mathematical tools, for example, the Rényi entropy-based TE and NE

quantifiers. The derivation and explanation of these new WP quantifiers with their applications will be an interesting direction for future study.

Chapter 7 introduced the Bayesian classifier to quantitatively evaluate the performance of the WP quantifiers. There λ_{ki} is assumed to be 1 for any $k \neq i$, i.e. any misclassification had an equal penalty. However, this tends to be a naive assumption. In the real world, some of the misclassification may imply more serious consequences. As a result, they should be penalised more heavily rather than equally. The actual penalty evaluation is closely linked to the industrial requirement. The unequal penalty terms will change the decision boundary of the Bayesian classifier and this change may cause different “best” quantifiers to be found.

Along with the reduction of AE transducer price, these new methods are expected to have very promising applications. To implement these methods will be the next critical step for real-time applications.

Chapter 9

Conclusions

This thesis summarised the investigations and research carried out by the author in Bearing Fault Detection and Diagnosis (BFDD). The main conclusions are:

- Acoustic Emission (AE) measurement is suitable to detect both localised defects and incipient faults of bearing.
- Some new BFDD systems using AE signals are developed and their effectiveness of the proposed BFDD systems is demonstrated on the iGX dry vacuum pump.
- A new signal processing method called LocMax-Interval method is proposed. It is feasible to detect localised defects when some knowledge is unavailable.
- The Un-decimated Discrete Wavelet Transform denoising can be used as a pre-processing approach to improve the detection capability and the effective parameter range of the LocMax-Interval method.
- The Wavelet Packet quantifiers, including Relative Energy, Total Entropy and Node Entropy are introduced for BFDD in the first time.
- The Bayesian method can be used to analyse and evaluate the performance of different quantifiers. The best quantifier is found with the minimal Bayesian classification error probabilities.
- The performance of the WP quantifier also depends on the wavelet order and the length of signal segment. Reducing the wavelet order or the length of signal segment will deteriorate the performance of the quantifier.
- A two-dimensional scheme can improve the diagnostic performance.

Bibliography

Altmann, J., Mathew, J. (2001) "Multiple Band-pass Autoregressive Demodulation for Rolling Element Bearing Fault Diagnosis," *Mechanical Systems and Signal Processing*, vol. 15(5), pp. 963-977.

Antoni, J. (2006) "The Spectral Kurtosis: A Useful Tool for Characterising Non-stationary Signals," *Mechanical Systems and Signal Processing*, vol. 20, pp. 282-307.

Antoni, J. (2007) "Cyclic Spectral Analysis of Rolling Element Bearing Signals: Facts and Frictions," *Journal of Sound and Vibration*, vol. 304, pp. 497-529.

Antoni, J., Randall, R.B. (2002) "Differential Diagnosis of Gear and Bearing Faults," *Journal of Vibration and Acoustics*, vol. 124, pp. 165-171.

Antoni, J., Randall, R.B. (2003) "A Stochastic Model for Simulation and Diagnostics of Rolling Element Bearings with Localized Faults," *Journal of Vibration and Acoustics*, vol. 125, pp. 282-289.

Antoni, J., Randall, R.B. (2004a) "Unsupervised Noise Cancellation for Vibration Signals: Part I - Evaluation of Adaptive Algorithms," *Mechanical Systems and Signal Processing*, vol. 18, pp. 89-101.

Antoni, J., Randall, R.B. (2004b) "Unsupervised Noise Cancellation for Vibration Signals: Part II - A Novel Frequency-domain Algorithm," *Mechanical Systems and Signal Processing*, vol. 18, pp. 103-117.

Antoni, J., Randall, R.B. (2006) "The Spectral Kurtosis: Application to the Vibratory Surveillance and Diagnostics of Rotating Machines," *Mechanical Systems and Signal Processing*, vol. 20, pp. 308-331.

Blanco, S., Figliola, A., Quiroga, R.Q., Rosso, O.A., Serrano, E. (1998) "Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function," *Physical Review E*, vol. 57, no. 1.

Braun, S. et al. (1986) *Mechanical Signature Analysis: theory and application*, Academic Press

Bibliography

Inc. (London) Ltd.

Choudhury, A., Tandon, N. (2000) "Application of Acoustic Emission Technique for the Detection of Defects in Rolling Element Bearing," *Tribology International* vol. 33, pp. 39-45.

Chui, C.K. (1992) *An Introduction to Wavelets*, Academic Press.

Daubechies, Ingrid (1992) *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics.

Donoho, D.L., Johnstone, I.M. (1994) "Ideal Spatial Adaptation via Wavelet Shrinkage," *Biometrika*, vol. 81(3), pp. 425-455.

Donoho, D.L., Johnstone, I.M. (1995) "Adapting to Unknown Smoothness via Wavelet Shrinkage," *J. Am. Stat. Assoc.*, vol. 90, pp. 1200-1224.

Ebersbach, S., Peng, Z. (2008) "Expert System Development for Vibration Analysis in Machine Condition Monitoring," *Expert Systems with Applications*, vol. 34, pp. 291-299.

FAG Bearing Corp., Rolling Bearing Damage: Recognition of Damage and Bearing Inspection. Publ. No. WL 82 102/2.

Feng, Y.H., Thanagasundram, S., Schlindwein, F.S. (2006) "Discrete Wavelet-based Thresholding study on Acoustic Emission Signals to Detect Bearing Defect on A Rotating Machine," *Proceeding of ICSV13 Conference*, Vienna.

Feng, Y.H., Schlindwein, F.S. (2007) "Improving Detectability on Localized Bearing Defect using Acoustic Emission Signal," *Proceeding of WCEAM-CM Conference*, ISBN: 978-1-901892-22-2, Harrogate, United Kingdom.

Gardner, W. A. *et al.* (2006) "Cyclostationarity: Half a Century of Research," *Signal Processing*, vol. 86.

Gardner, W.A, Spooner, C.M. (1992) "Signal Interception: Performance Advantages of Cyclic-Feature Detectors," *IEEE Transactions on Communications*, vol. 40, No. 1.

Gestri, G., Piram, P. (1975) "On the Autocorrelation of One Class of Non-stationary Random Point Process," *Biol. Cybernetics*, vol. 17, pp. 199-205.

Harris, N. (2001) *Modern vacuum practice* (2nd ed.).

Harris, T.A. (1966) *Rolling Bearing Analysis*, John Wiley & Sons, Inc.

Ho, D., Randall, R.B. (2000) "Optimisation of Bearing Diagnostic Techniques using

Bibliography

- Simulated and Actual Bearing Fault Signals," *Mechanical Systems and Signal Processing*, vol. 14(5).
- Holroyd, T. (2001) "Condition Monitoring of Very Slowly Rotating Machinery using AE Techniques," *14th International congress on Condition monitoring and Diagnostic Engineering Management (COMADEM 2001)*, Manchester, UK.
- Holroyd, T. (2003) *Acoustic Emission & Ultrasonic Monitoring Handbook*. Coxmoor Publishing Company.
- Holroyd, T. and Randall, R.B. (1993) "Use of Acoustic Emission for Machine Condition Monitoring," *British Journal of Non-Destructive Testing*, vol. 35(2), 75-78.
- Hong, H., Liang, M. (2007) "K-Hybrid: A Kurtosis-Based Hybrid Thresholding Method for Mechanical Signal Denoising," *ASME Journal of Vibration and Acoustics*, vol. 129, pp. 458-470.
- Howard, I. (1994) "A Review of Rolling Element Bearing Vibration: Detection, Diagnosis and Prognosis," DSTO-RR-0013, Defence science and technology organisation Canberra (Australia).
- Isermann, R. (2005) "Model-based fault-detection and diagnosis-status and applications," *Annual Review in Control*, vol. 29, pp. 71-85.
- Jardine, A.K.J., Lin, D., Banjevic, D. (2006) "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483-1510.
- Kay, S.M., Stanley Lawrence Marple, JR. (1981) "Spectrum Analysis - A Modern Perspective," *Proceeding of IEEE*, vol. 69, No.11.
- Landini, L., Verrazzani, L. (1990) "Spectral Characterization of Tissues Microstructure by Ultrasounds: A Stochastic Approach," *IEEE Transactions on Ultrasonics. Ferroelectrics. And Frequency Control*, vol. 37, No. 5.
- Lei, Y., He, Z., Zi, Y., Hu, Q. (2007) "Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs," *Mechanical Systems and Signal Processing*, vol. 21, pp.2280-2294
- Li, C.J., Ma, J. (1997) "Wavelet Decomposition of Vibrations for Detection of Bearing Localized Defects," *NDT & E International*, vol. 30, No. 3, pp. 143-149.

Bibliography

- Lin, J., Qu, L.S. (2000) "Feature Extraction Based on Morlet Wavelet and Its Application for Mechanical Fault Diagnosis," *Journal of Sound and Vibration*, vol. 234(1), pp. 135-148.
- Lin, J. Zuo, M.J., Fyfe, K.R. (2004) "Mechanical Fault Detection Based on the Wavelet De-Noising Technique," *ASME Journal of Vibration and Acoustics*, vol. 126, pp. 9-16.
- Lou, X., Loparo, K.A. (2004) "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, pp.1077-1095.
- Mallat, S. (1989) "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.11, No.7.
- Mallat, S. (1999) *A Wavelet Tour of Signal processing (2nd Ed.)*, Academic Press.
- Mallat, S., Hwang, W.L. (1992) "Singularity Detection and Processing with Wavelets," *IEEE Trans. Inf. Theory*, vol. 38, No. 2.
- Maru, M.M. *et al.* (2007) "Study of Solid Contamination in Ball Bearings through Vibration and Wear Analyses," *Tribology International*, vol. 40, pp. 433-440.
- Mba, D., Raj B. K. N. Rao (2006) "Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines: Bearings, Pumps, Gearboxes, Engines and Rotating Structures," *The Shock and Vibration Digest*.
- McCormick, A. C. and A. K. Nandi (1996) "Comparison of artificial neural networks and other statistical methods for rotating machine condition classification", Proceedings of the 1997 IEE Colloquium on Modelling and Signal Processing for Fault Diagnosis, Sep 18 1996, London, UK, pp. 2-1, IEE, Stevenage, Engl.
- McFadden, P.D., Smith, J.D. (1984a) "Model for the Vibration Produced by A Single Point Defect in A Rolling Element Bearing," *Journal of Sound and Vibration*, vol. 96(1), pp. 69-82.
- McFadden, P.D., Smith, J.D. (1984b) "The Vibration Monitoring of Rolling Element Bearings by the High-frequency Resonance Technique: A Review," *Tribology International*, vol. 17(1), pp. 3-10.
- McFadden, P.D., Toozhy, M.M. (2000) "Application of Synchronous Averaging to Vibration Monitoring of Rolling Element Bearings," *Mechanical Systems and Signal Processing* vol. 14(6), 891-906.
- McInerny, S. A., Dai, Y. (2003). "Basic Vibration Signal Processing for Bearing Fault

Bibliography

Detection", *IEEE Transactions on Education*, 46(1), pp. 149-156.

Miettinen, J., Andersson, P. (2000) "Acoustic Emission of Rolling Bearing Lubricated with Contaminated Grease," *Tribology International*, vol. 33, pp. 777-787.

Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J. (2000) *Wavelet Toolbox- for use with MATLAB*, The MathWorks, Inc.

Nikolaou, N.G., Antoniadis, I.A. (2002) "Rolling Element Bearing Fault Diagnosis using Wavelet Packets," *NDT&E International*, vol. 35, pp. 197-205.

Paya, B.A., Esat, I.I., Badi, M.N.M. (1997) "Artificial Neural Network Based Fault Diagnostics of Rotating Machinery using Wavelet Transforms as A Preprocessor, " *Mechanical Systems and Signal Processing* , vol. 11(5), pp. 751-765.

Peng, Z.K., Chu, F.L. (2004) "Application of Wavelet Transform in Machine Condition Monitoring and Fault Diagnostics: A Review with Bibliography," *Mechanical Systems and Signal Processing*, vol. 18, pp. 199-221.

Percival, D.B., Walden, A.T. (2000) *Wavelet Methods for Time Series Analysis*, Cambridge University Press, England.

Pizurica, A. (2002) "Image Denoising using Wavelets and Spatial Context Modeling," Ph.D. Thesis, University of Gent, Belgium.

Qiu, H., Lee, J., Lin, J., Yu, G. (2006) "Wavelet Filter-based Weak Signature Detection Method and Its Application on Rolling Bearing Prognostics," *Journal of Sound and Vibration*, vol. 289, pp. 1066-1090.

Rafiee, J., Arvani, F., Harifi, A., Sadeghi, M.H., (2007) "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical Systems and Signal Processing*, vol.21, pp.1746-1754.

Randall, R.B., Antoni, J., Chobsaard, S. (2001) "The Relationship between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals," *Mechanical Systems and Signal Processing*, vol. 15(5), pp. 945-962.

Randall, R.B. (2004a) "State of the Art in Monitoring Rotating Machinery: Part 1," *Sound and Vibration*.

Randall, R.B. (2004b) "State of the Art in Monitoring Rotating Machinery: Part 2," *Sound*

Bibliography

and Vibration.

Rényi, A. (1961) "On measures of entropy and information," *Proceeding of 4th Berkeley Symposium on Mathematics of statistics and probability*, vol. 1, pp. 547-561.

Rioul, O., Vetterli, M. (1991) "Wavelets and Signal Processing", IEEE SP Magazine.

Rosso, A. Blanco, S. *et al.* (2001) "Wavelet entropy: A New Tool for Analysis of Short Duration Brain Electrical Signals," *Journal of Neuroscience Methods*, vol. 105, pp. 65-75.

Samanta, B., Al-Balushi, K.R. (2003), "Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings using Time-domain Features," *Mechanical Systems and Signal Processing*, vol. 17(2), pp. 317-328.

Sawalhi, N. (2007) "Diagnostics, Prognostics, and Fault Simulation for Rolling Element Bearings," Ph.D. Thesis, The University of New South Wales, Australia.

Saxena, A., Saad, A. (2007) "Evolving an Artificial Neural Network Classifier for Condition Monitoring of Rotating Mechanical Systems," *Applied Soft Computing*, vol. 7, pp.441-454.

Schindwein, F.S., Boardman, A., Vali, S. *et al.* (2004) "Noninvasive Determination of Fetal Heart Rate and Short Term Heart Rate Variability using Solely Doppler Ultrasound with Autocorrelation," *Proceedings of the 2nd International Conference on Medical Signal and Information Processing*, Malta.

Scruby, C.B. (1987) "An introduction to acoustic emission," *J. Phys. E: Sci. Instrum.* Vol. 20, pp. 946-953.

Shannon, C.E., Weaver, W. (1959) *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.

Shao, Y., Nezu, K. (2005) "Design of Mixture Denoising for Detecting Fault Bearing Signals," *Journal of Sound and Vibration*, vol. 282, pp. 899-917.

Shi, D.F., Wang, W.J., Qu, L.S. (2004) "Defect Detection for Bearings Using Envelope Spectra of Wavelet Transform," *Journal of Vibration and Acoustics*, vol. 126.

Strang, G., Nguyen, T. (1996) *Wavelets and Filter Banks*. Wellesley-Cambridge Press, MA.

Sugumaran, V., Ramachandran, K.I. (2007) "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing," *Mechanical Systems and Signal Processing*

Bibliography

Processing, vol. 21, pp.2237–2247.

Sun, Q., Tang, Y. (2002) "Singularity Analysis using Continuous Wavelet Transform for Bearing Fault Diagnosis," *Mechanical Systems and Signal Processing*, vol. 16(6), pp. 1025-1041.

Sweldens, W. (1996) "Wavelets: What next?," *Proceedings of the IEEE*, Vol. 84(4), pp. 680-685.

Tandon, N., Choudhury, A. (1999) "A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings," *Tribology International*, vol. 32, pp. 469-480.

Tao, B. *et al.* (2007) "Rényi Entropy-based Generalized Statistical Moments for Early Fatigue Defect Detection of Rolling-Element Bearing," *Proc. IMechE vol. 221 Part C: J. Mechanical Engineering Science*.

Theodoridis, S., Koutroumbas, K. (2003) *Pattern Recognition (2nd Ed.)*, USA: Academic Press.

Tse, P.T., Peng, Y.H., Yam, R. (2001) "Wavelet Analysis and Envelope Detection for Rolling Element Bearing Fault Diagnosis - Their Effectiveness and Flexibilities," *Journal of Vibration and Acoustics*, vol. 123.

Vetterli, M., Kovacevic, J. (1995) *Wavelets and Subband Coding*. USA: Prentice Hall, Inc..

Ville, F., Coulon, S., Lubrecht, A.A. (2006) "Influence of Solid Contaminations on the Fatigue Life of Lubricated Machine Elements," *Proc. IMechE vol. 220 Part J: Journal of Engineering Tribology*.

Wang, L., Hope, A.D. (2004) "Bearing Fault Diagnosis using Multi-layer Neural Networks," *Insight*, vol. 46, pp. 451-455.

Widodo, A., Yang, B. (2007) "Review: Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, pp. 2560-2574.

Yang, D.M., Stronach, A.F., MacConnell, P., Penman, J. (2002) "Third-order Spectral Techniques for the Diagnosis of Motor Bearing Condition using Artificial Neural Networks," *Mechanical Systems and Signal Processing*, vol. 16(2-3), pp.391-411.

Yang, Y., Yu, D., Cheng, J. (2006) "A Roller Bearing Fault Diagnosis Method based on

Bibliography

- EMD Energy Entropy and ANN, " *Journal of Sound and Vibration*, vol. 294, pp. 269-277.
- Zhang, S., Asakura, T., Xu, X.L., Xu, B.J. (2003) "Fault diagnosis system for rotary machine based on fuzzy neural networks," *JSME International Journal. Series C: Mechanical Systems, Machine Elements and Manufacturing*, vol. 46, pp. 1035-1041.
- Zhang,L., Jack,L.B., Nandi,A.K. (2005) "Fault Detection using Genetic Programming, " *Mechanical Systems and Signal Processing*, vol. 19, pp.271-289.
- Zio, E., Gola, G. (2009) " A Neuro-fuzzy Technique for Fault Diagnosis and Its Application to Rotating Machinery," *Reliability Engineering and System Safety*, vol. 94, pp.78-88.
- Zunino, L., Perez, D.G., Garavaglia, M. Rosso, O.A. (2007) "Wavelet Entropy of Stochastic Processes," *Physica A* 379 pp. 503-512.
- Zuo, M.J., Lin, J., Fan, X.F. (2005) "Feature Separation using ICA for a One-dimensional Time Series and Its Application in Fault Detection," *Journal of Sound and Vibration*, 287, 614-624.

Appendix A

Wavelet Methods

Since many wavelet techniques and terminology are frequently used in this thesis, the intention of this appendix is to provide a brief introduction and explanation to some major wavelet methods. Wavelet methods include a set of related methods based on the idea of a wavelet transform. From Continuous Wavelet Transform (CWT), it extends to other wavelet transform variations, such as the Un-decimated Discrete Wavelet Transform (UDWT), Discrete Wavelet Transform (DWT) and Discrete Wavelet Packet Transform (DWPT). For more comprehensive readings, refer to Daubechies 1992, Chui 1992, Vetterli and Kovacevic 1995, Strang and Nguyen 1996, Mallat 1999, Percival and Walden 2000.

This chapter starts by introducing the CWT in Section A.1 and explaining the relationship between different transform variations in Section A.2. Section A.3 introduces the Multiresolution Analysis (MRA) theory established by Mallat (1989), which provides a natural framework for understanding wavelet transforms. Sections A.4-A.6 summarise the mathematical descriptions and filtering conventions for different transform variations. See also Mallat 1999, Percival and Walden 2000 for more details. Orthogonal and biorthogonal wavelet bases have very different characteristics which lead to different applications. Section A.7 summarises these differences. Sections A.8-A.10 review some important wavelet-based techniques which will be further studied and exploited for BFDD in the following chapters. The chapter finishes with a summary in Section A.11.

A.1 Introduction to Wavelet Representation

The initial motivation of using wavelet transforms is to study signals at different scales with each scale corresponding to a specific frequency band. The multi-scale (i.e., multiresolution) representations provided by wavelet transforms are of great interest in science and engineering. They form a mechanism for decomposing a complicated signal into many frequency bands, each of which carries certain concise information. This rich and compact information eases the task for people to analyse and make decisions on the complicated signals. Wavelet theory was introduced and developed as a unifying framework in mid of 1980s (Rioul and Vetterli 1991). CWT provides good localisation in the time domain by matching the signal with a highly localised basic function-wavelet basis. CWT generates a continuous time-scale analysis which is particularly interesting for transient signal analysis.

Recall the standard Fourier transform of a signal $f(t)$

$$(Ff)(f) = \int_{-\infty}^{\infty} e^{-j\omega t} f(t) dt. \quad (\text{A-1})$$

This gives a global representation of the frequency content of signal $f(t)$, but it misses the information concerning time localisation. The timing of the occurrences of high frequency bursts cannot be revealed by the spectrum. The spectrum is the energy distribution associated with Fourier transform. However, these high frequency bursts often carry important information.

A typical example is the situation where high frequency burst signals are superimposed on a low frequency carrier. See Figure A-1 (left). The blue arrows point out the regular occurrences of burst signals. However, the occurrence of burst signals cannot be revealed by the conventional Fourier-based spectrum. See Figure A-1 (right). Therefore, the interesting information on the occurrence frequency of the burst signals needs to be discovered by other means.

Appendix A

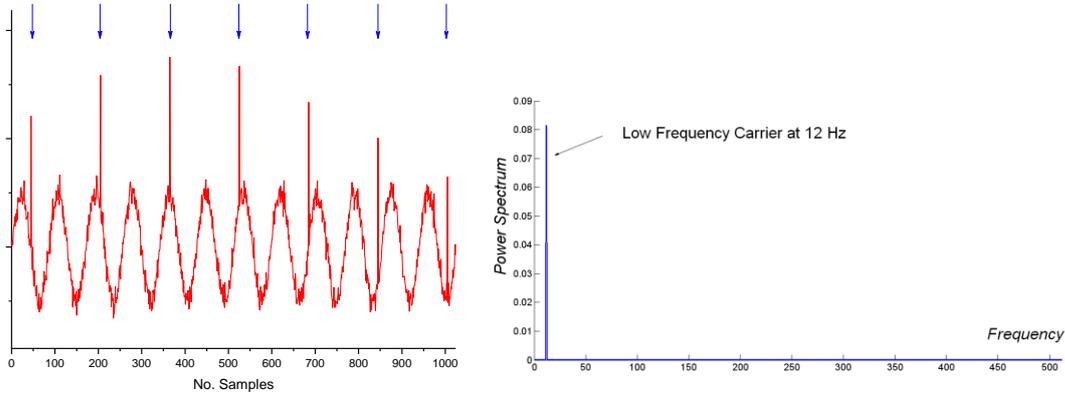


Figure A-1 Modulated burst signals on carrier (left), Y-axis is amplitude of arbitrary unit, and Fourier spectrum (right).

Daubechies (1992) explains the difference between Short Time Fourier Transform and CWT. The windowed Fourier transform provides a description for signal in the time-frequency domain. It windows a signal in time before applying the Fourier transform:

$$(F_g f)(f, u) = \int_{-\infty}^{\infty} f(t) \bar{g}_{\omega, u}(t) dt = \int_{-\infty}^{\infty} f(t) g(t-u) e^{-j\omega t} dt. \quad (A-2)$$

A popular choice for window function g is a Gaussian function. This transform is also called Short Time Fourier Transform (STFT) or Gabor Transform. STFT analyses signals using the window function $\bar{g}_{\omega, u}(t) = g(t-u) e^{-j\omega t}$ which is the same envelope function g , translated to different time location u and modulated with frequency component $e^{-j\omega t}$. The spectrogram is the energy distribution associated with STFT.

The wavelet transform provides a description in the time-scale domain. The CWT of a signal $f(t)$ is:

$$(W_{\psi} f)(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a, b}(t) dt = |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt, \quad (A-3)$$

$$(W_\psi f)(a, b) = \langle f, \psi_{a,b} \rangle. \tag{A-4}$$

The parameters a (scale) and b (time) allow us to obtain scale and time localisation. The wavelet basis ψ satisfies the following two basic properties (Percival and Walden 2000). Firstly, the integral of $\psi(t)$ is zero:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \tag{A-5}$$

Secondly, the square of $\psi(t)$ integrates to unity:

$$\int_{-\infty}^{\infty} \psi^2(t) dt = 1. \tag{A-6}$$

The difference between STFT and CWT lies in the elementary function used to analyse the signal. CWT analyses signals using the wavelet basis $\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right)$, which is the *same* function ψ translated by b , compressed or stretched by a . The wavelet basis is very compressed for high frequency, which allows CWT a better localisation at time than STFT. Figure A-2 shows the elementary functions for these two transforms.

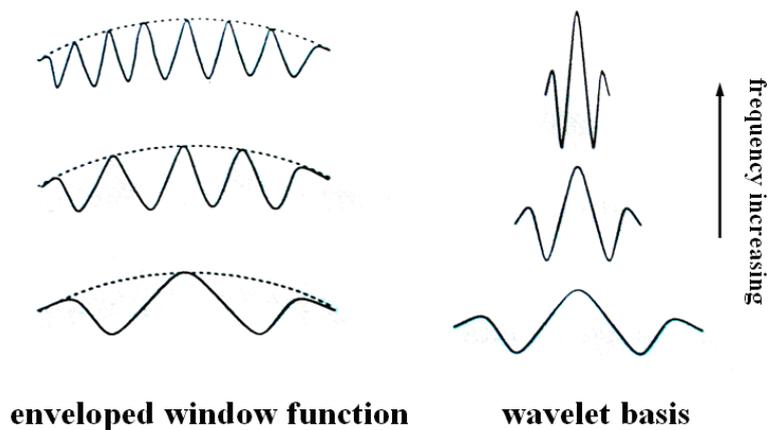


Figure A-2 Elementary functions for STFT (left) and CWT (right) to analyse the signal (Vetterli and Kovacevic 1995).

Appendix A

The two-dimensional CWT is obtained when the continuous integer parameter a is used. The scalogram is the energy distribution associated with the wavelet transform. CWT can effectively reveal the occurrences of burst signals. Figure A-3 shows that the burst signals are highlighted in the CWT scalogram. Scales a are continuous, changing in the scalogram of CWT from 1 to 32. The burst-generated coefficients converge to small scales which correspond to high frequencies.

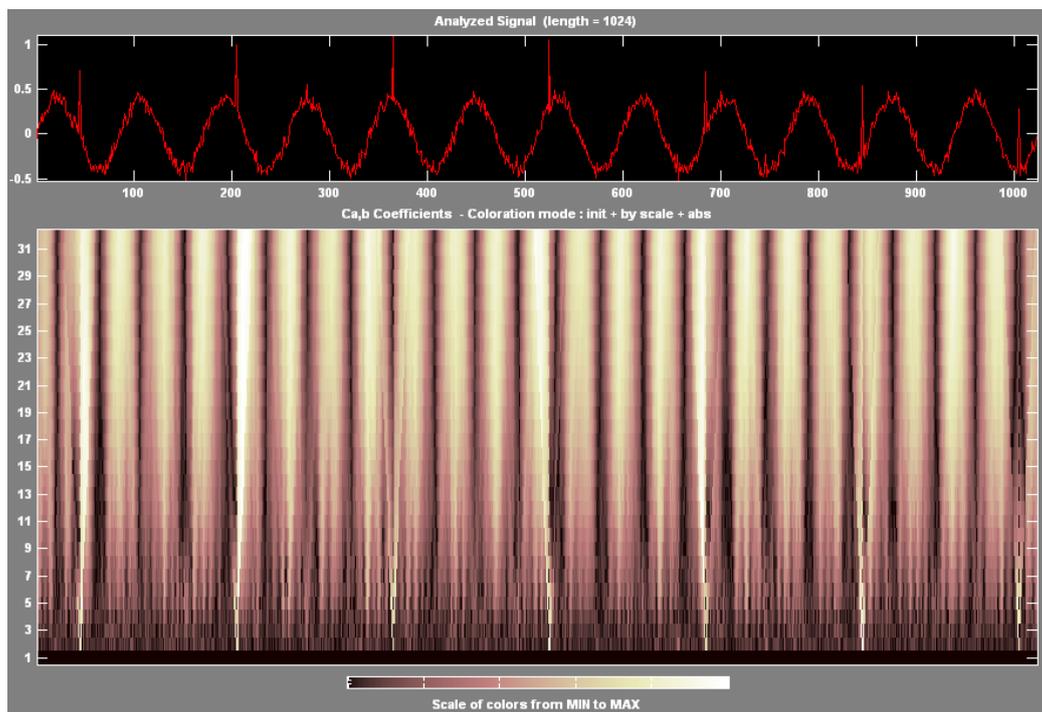


Figure A-3 High frequency bursts (top) are highlighted in scalogram (bottom). Scale parameters are 1: 32.

A.2 Beyond the CWT: DWT, UDWT and DWPT

The CWT is highly redundant and computationally intensive, which is inconvenient for many engineering applications. Simplicity can be achieved by subsampling the parameters a and b . The DWT and UDWT can be considered as two important variations of CWT from the view of subsampling, although their

computation methods are quite different.

The DWT can be considered as the subsampled version of CWT at dyadic scales (i.e., parameter a is selected as 2^j , $j=1, 2, 3, \dots$) and within a given dyadic scale, parameter b (time shift) is selected at the multiples of 2^j (Percival and Walden 2000). When the scale j increases, the number of the corresponding coefficients at scale j will drop down at the rate of 2^j . See Figure A-4 for the comparison between DWT and the other transforms.

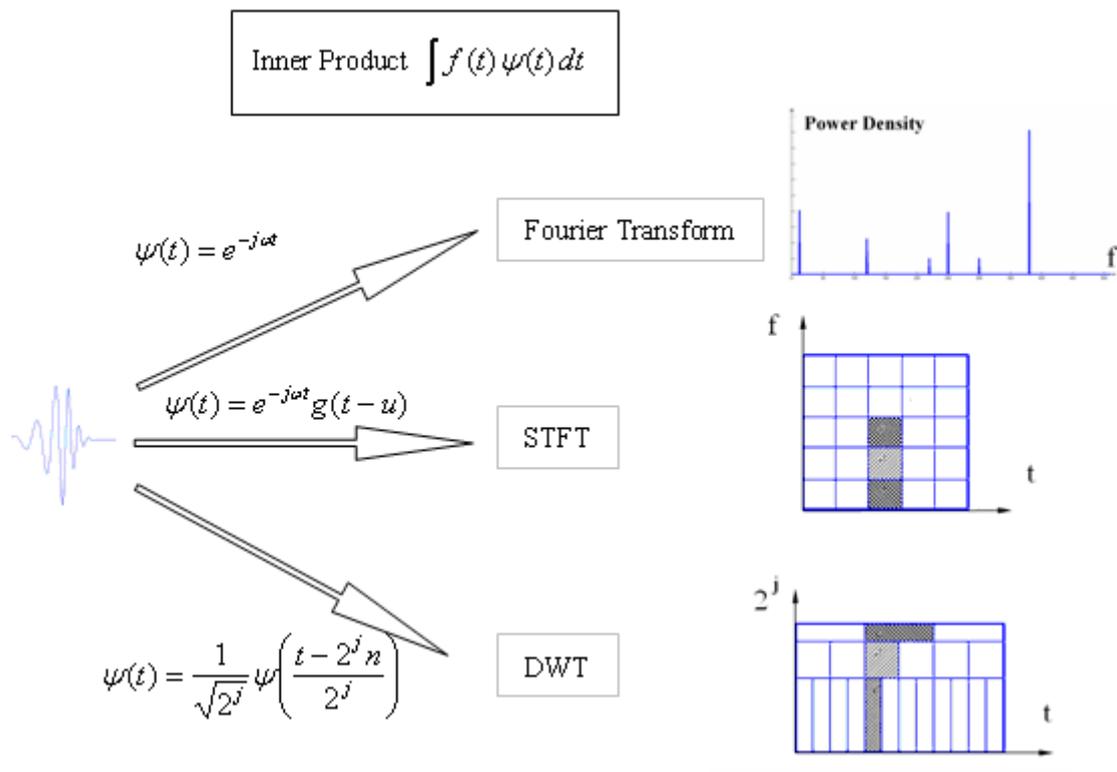


Figure A-4 Elementary functions and frequency (or scale) representations for different transforms.

The numerical computation method for DWT was developed by Mallat (1989) via a fast filter bank algorithm - the Fast Wavelet Transform (FWT) which is faster than the classical Fast Fourier Transform (FFT). Because the DWT can be

formulated in terms of filters, the notion of scale is also related to a certain frequency band. The equivalent filter yielding the wavelet coefficients for scale 2^j corresponds to a band-pass filter with a pass-band given by $[f_s/2^{j+1}, f_s/2^j]$ for $j \geq 1$, which is a dyadic decomposition of the Nyquist frequency (i.e., half of the sampling frequency).

The UDWT can be considered as the subsampled version of CWT at dyadic scales (i.e., parameter a is selected at 2^j , $j=1, 2, 3, \dots$). In contrast to DWT, UDWT keeps the same number of coefficients at all the scales as the original signal. See Figure A-5 for the sampling of time-scale plane for DWT and UDWT. The translation-invariant UDWT is more suitable for transient signal analysis which requires precise time localisation. The UDWT is computed via a filter bank algorithm called in French the *algorithme à trous*, (the filter is obtained by inserting $2^j - 1$ zeroes between successive coefficients, hence the French name, which means “holes algorithm” or “gaps algorithm”) which has the same computational complexity as the FFT (Mallat 1999).

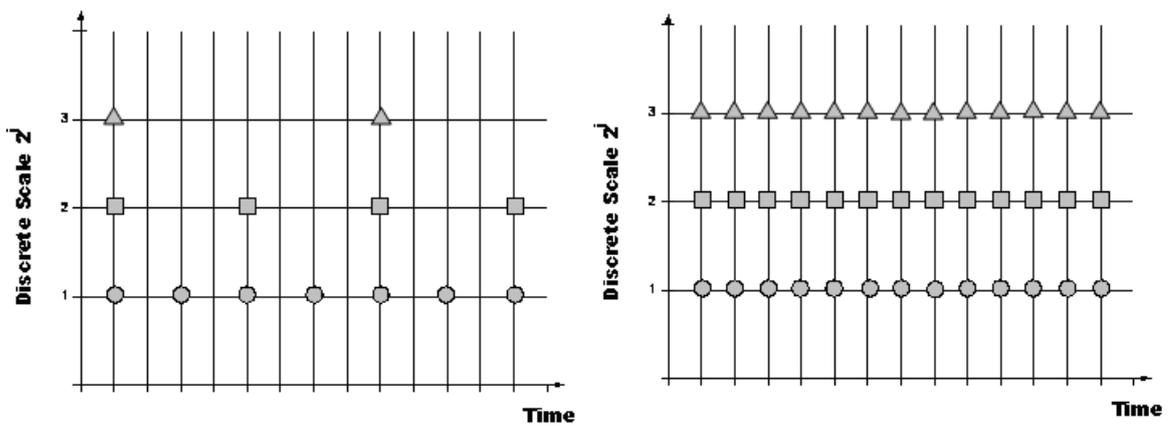


Figure A-5 Sampling of time-scale plane for DWT (left) and UDWT (right). Circle, square and triangle represent wavelet coefficients at dyadic scale 1, 2, 3, respectively.

The DWPT is a more flexible version of the DWT, which yields a time-frequency like decomposition using the similar computational method to the FWT (Percival

and Walden 2000). By adaptively decomposing the dyadic scales which correspond to octave frequency bands, smaller frequency separations called Wavelet Packet (WP) nodes are possible. This flexible decomposition allows better insight of the time-frequency structure of signal. Besides the fast computation and adaptive decomposition, DWPT also has some important advantages over STFT: The basis of DWPT can be flexibly chosen for practical implementation considerations; the contaminating noises that are concentrated in some frequency bands can also be easily eliminated. The relationship between different transform variations is illustrated in Figure A-6.

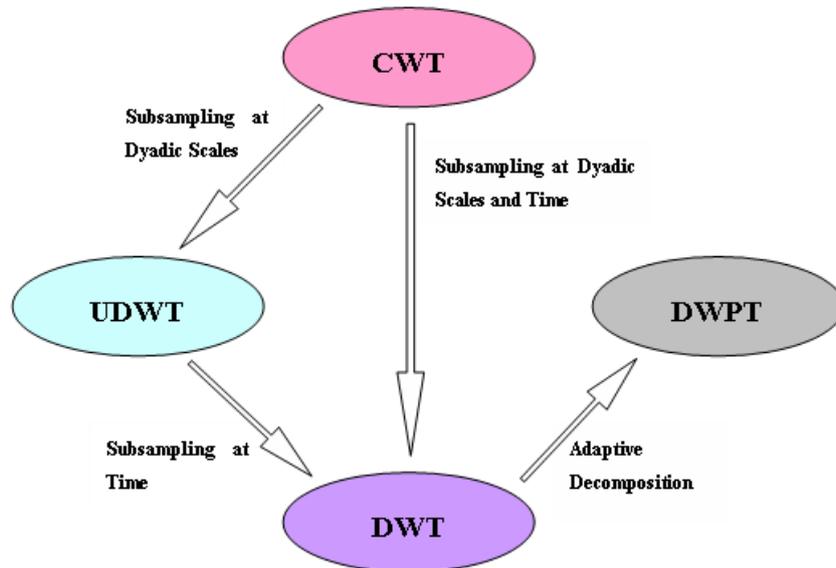


Figure A-6 Relationship between different transform variations.

A.3 Multiresolution Analysis

The Multiresolution Analysis (MRA) theory (Mallat 1989) provides a framework for the understanding of wavelet methods. The nested approximation spaces V_j are increasing $\dots V_2 \subset V_1 \subset V_0 \subset V_{-1} \dots$ and the detail space W_j is the difference between successive space V_{j+1} and V_j . The properties of subspaces V_j are summarised below (Strang and Nguyen 1996):

Appendix A

- a) $V_{j+1} \subset V_j$ (nested spaces) and $\bigcap V_j = \{0\}$ (emptiness) and $\overline{\bigcup V_j} = L^2(\mathbb{R})$ (completeness);
- b) Scale invariance: $f(t) \in V_j \Leftrightarrow f(t/2) \in V_{j+1}$;
- c) Translation invariance: $f(t) \in V_j \Leftrightarrow f(t - 2^j n) \in V_j$;
- d) Translation-invariant basis: V_0 has biorthogonal basis or orthogonal basis $\phi(t - n)$.

The detail spaces W_j associated with orthogonal basis have the properties:

$$V_{j+1} + W_{j+1} = V_j \quad \text{and} \quad V_j \cap W_j = \{0\} \quad \text{and} \quad \bigcap W_j = \{0\}.$$

The detail spaces W_j associated with biorthogonal basis have the following properties (Strang and Nguyen 1996):

- e) $V_{j+1} + W_{j+1} = V_j$ and $\tilde{V}_{j+1} + \tilde{W}_{j+1} = \tilde{V}_j$;
- f) $V_j \cap \tilde{W}_j = \{0\}$ and $\tilde{V}_j \cap W_j = \{0\}$.

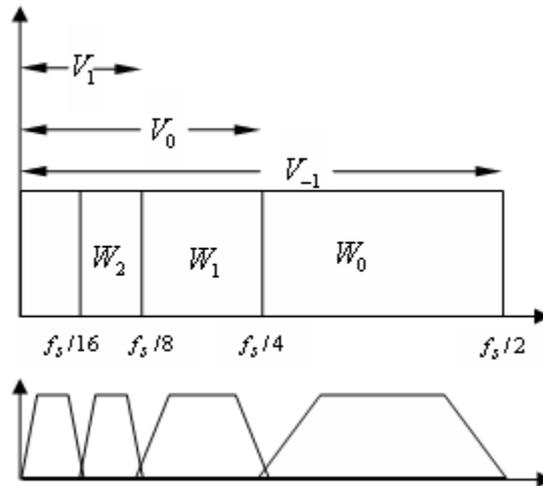


Figure A-7 Divisions of frequency domain for approximation and detail spaces.

The MRA decomposes successively each approximation V_j into a coarser approximation V_{j+1} plus the detail W_{j+1} . The detail W_{j+1} is the difference

between two approximations V_j and V_{j+1} . Figure A-7 shows the divisions of spectrum for approximation and detail spaces. Note that the MRA of signal is defined in the spaces of approximation and details, *not* for the scaling and wavelet coefficients. Therefore, the analysis which directly uses those coefficients can bring convenience in computation, e.g., the analysis based on UDWT, which will be introduced in Section A.5 and A.9.

A.4 Discrete Wavelet Transform

This Section summarise the mathematical descriptions and filtering conventions for DWT. Details refer to Mallat 1999, Percival and Walden 2000.

Let real signal \mathbf{X} be a N dimensional vector $\{X(t):t=1,\dots,N\}$, where the sample size N is taken to be an integer multiple of 2^{J_0} . The DWT of scale J of \mathbf{X} is given by $\mathbf{C}=\mathbf{H}\mathbf{X}$, where \mathbf{C} is a $J+1$ dimensional vector of DWT coefficients and \mathbf{H} is a $(J+1)\times N$ matrix defining the DWT. J is the scale to be decomposed, such that

$$\mathbf{C} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_J \\ \mathbf{A}_J \end{bmatrix} \quad \text{and} \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_J \\ \mathbf{G}_J \end{bmatrix},$$

where the wavelet coefficients $\mathbf{D}_j = \mathbf{H}_j\mathbf{X}$, and scaling coefficients $\mathbf{A}_j = \mathbf{G}_j\mathbf{X}$.

\mathbf{D}_j is a $N_j = N/2^j$ dimensional vector of wavelet coefficients; \mathbf{H}_j is a $N_j \times N$ dimensional matrix. \mathbf{A}_j is a N_j dimensional vector of scaling coefficients; \mathbf{G}_j is a $N_j \times N$ dimensional matrix.

Signal \mathbf{X} can be synthesized from the sum of details and approximation, which defines a MRA of \mathbf{X} such that

Appendix A

$$\mathbf{X} = \mathbf{H}^T \mathbf{C} = \sum_{j=1}^J \mathbf{H}_j^T \mathbf{D}_j + \mathbf{G}_J^T \mathbf{A}_J = \sum_{j=1}^J W_j + V_J, \quad (\text{A-7})$$

and the energy of the signal is kept by the wavelet and scaling coefficients:

$$\|\mathbf{X}\|^2 = \|\mathbf{C}\|^2 = \sum_{j=1}^J \|\mathbf{D}_j\|^2 + \|\mathbf{A}_J\|^2. \quad (\text{A-8})$$

The above expressions can be written in terms of inner product, such that (f and \mathbf{X} are interchangeable notations):

$$V_j = \sum \langle f, \phi_{j,n} \rangle \tilde{\phi}_{j,n} \quad \text{and} \quad W_j = \sum \langle f, \psi_{j,n} \rangle \tilde{\psi}_{j,n}, \quad (\text{A-9})$$

where $\tilde{\phi}$ and $\tilde{\psi}$ are the dual bases of scaling ϕ and wavelet ψ . For the orthogonal basis, the dual basis is itself, i.e., $\tilde{\phi}(t) = \phi(t)$ and $\tilde{\psi}(t) = \psi(t)$. The inner products are the scaling coefficients ($\mathbf{A}_j = \{a_j(n)\}$) and wavelet coefficients ($\mathbf{D}_j = \{d_j(n)\}$):

$$a_j(n) = \langle f, \phi_{j,n} \rangle = \frac{1}{\sqrt{2^j}} \int f(t) \phi\left(\frac{t-2^j n}{2^j}\right) dt, \quad (\text{A-10})$$

$$d_j(n) = (Wf)(t, n) = \langle f, \psi_{j,n} \rangle = \frac{1}{\sqrt{2^j}} \int f(t) \psi\left(\frac{t-2^j n}{2^j}\right) dt. \quad (\text{A-11})$$

In the numerical method for computation, the DWT wavelet and scaling coefficients are obtained using the perfect reconstruction filter banks (Strang and Nguyen 1996, pp.103). See Figure A-8 for the filter bank computation.

In the typical DWT decomposition, the scaling coefficients $a_j(n)$ and wavelet coefficients $d_j(n)$ are:

$$a_j(n) = \langle f, \phi_{j,n} \rangle \quad \text{and} \quad d_j(n) = \langle f, \psi_{j,n} \rangle, \quad (\text{A-12})$$

or, written in terms of recursive filtering convention:

$$a_{j+1}(n) = \sum g(t-2n)a_j(t), \quad (\text{A-13})$$

$$d_{j+1}(n) = \sum h(t-2n)a_j(t). \quad (\text{A-14})$$

The translation $2n$ notation specifies the downsampling by 2 at each level of the decomposition.

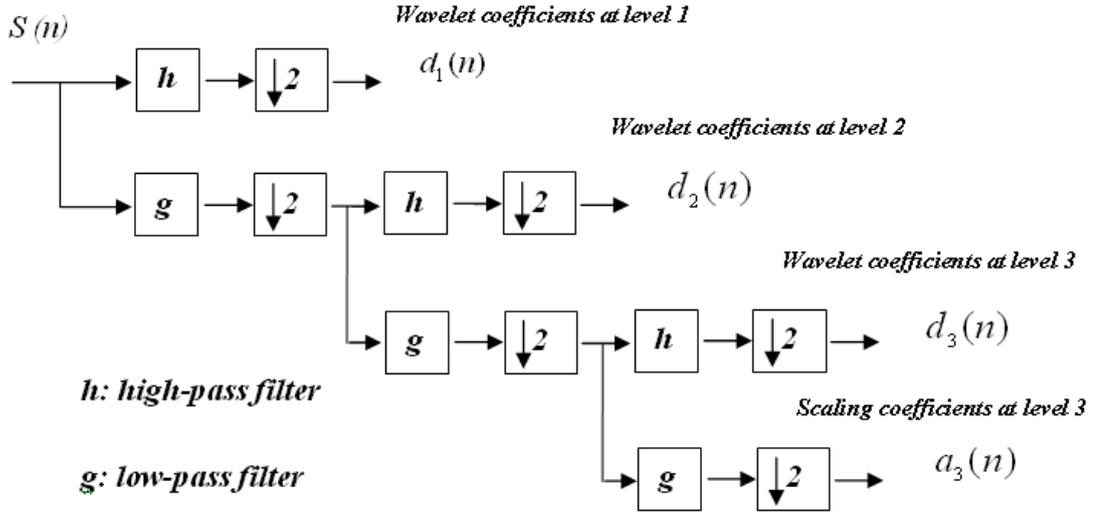


Figure A-8 Filter banks computation for DWT (3-level decomposition).

For simplicity, the coefficients can be viewed as the direct result of filtering signal \mathbf{X} :

$$a_j(n) = \sum_{l=1}^{L_j} g_{j,l} X_{2^j n-1-l \text{ modulo } N}, \quad t = 1, \dots, N_j; \quad (\text{A-15})$$

$$d_j(n) = \sum_{l=1}^{L_j} h_{j,l} X_{2^j n-1-l \text{ modulo } N}, \quad t = 1, \dots, N_j. \quad (\text{A-16})$$

$N_j = N / j$ is the length of coefficients vector $\mathbf{W}_{j,n}$ at scale j ; L_j is the filter width at j scale $L_j = (2^j - 1)(L - 1) + 1$; $h_{j,l}$ and $g_{j,l}$ are the equivalent filters for the computation of wavelet and scaling coefficients at scale j .

The scaling coefficients and wavelet coefficients are the subsampled sequences, which make the interpretation of the original signal very difficult, particularly for transient events. Therefore, the MRA using DWT needs to construct details and approximation spaces. However, the DWT is not translation invariant and the MRA will depend critically on where the signal is separated (Percival and Walden 2000, pp161). The starting points of each signal segment will greatly affect the analysis results. To obtain a translation-invariant wavelet representation, only the scales are subsampled at dyadic scales without subsampling the translation parameters. By this means, the translation-invariant Un-decimated Discrete Wavelet Transform is obtained, which will be introduced in the following Section.

A.5 Un-decimated Discrete Wavelet Transform

This Section summarise the mathematical descriptions and filtering conventions for UDWT. More details refer to Mallat 1999, Percival and Walden 2000.

Let the real signal \mathbf{X} be a N dimensional vector $\{X(t) : t = 1, \dots, N\}$, where the sample size N is any positive integer. For the scale J to be decomposed, the Un-decimated Discrete Wavelet Transform (UDWT) of \mathbf{X} is given by $\hat{\mathbf{C}} = \hat{\mathbf{H}}\mathbf{X}$, where $\hat{\mathbf{C}}$ is a $J+1$ dimensional vector of UDWT coefficients and $\hat{\mathbf{H}}$ is a $(J+1) \times N$ matrix defining the UDWT.

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{D}}_1 \\ \hat{\mathbf{D}}_2 \\ \vdots \\ \hat{\mathbf{D}}_J \\ \hat{\mathbf{A}}_J \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{H}} = \begin{bmatrix} \hat{\mathbf{H}}_1 \\ \hat{\mathbf{H}}_2 \\ \vdots \\ \hat{\mathbf{H}}_J \\ \hat{\mathbf{G}}_J \end{bmatrix},$$

The UDWT wavelet coefficients $\hat{\mathbf{D}}_j = \hat{\mathbf{H}}_j \mathbf{X}$, and scaling coefficients $\hat{\mathbf{A}}_j = \hat{\mathbf{G}}_j \mathbf{X}$. $\hat{\mathbf{D}}_j$ is a N dimensional vector of wavelet coefficients (with the *same* length as the signal \mathbf{X}); $\hat{\mathbf{H}}_j$ is a $N \times N$ dimensional matrix. $\hat{\mathbf{A}}_j$ is a N dimensional vector of scaling coefficients; $\hat{\mathbf{G}}_j$ is a $N \times N$ dimensional matrix.

The signal \mathbf{X} can be synthesized from the sum of details \hat{W}_j and approximations \hat{V} , which defines a MRA of \mathbf{X} . Such that

$$\mathbf{X} = \hat{\mathbf{H}}^T \hat{\mathbf{C}} = \sum_{j=1}^J \hat{\mathbf{H}}_j^T \hat{D}_j + \hat{\mathbf{G}}_J^T \hat{A}_J = \sum_{j=1}^J \hat{W}_j + \hat{V}_J, \quad (\text{A-17})$$

and the energy of the signal is also kept by the wavelet and scaling coefficients:

$$\|\mathbf{X}\|^2 = \|\hat{\mathbf{C}}\|^2 = \sum_{j=1}^J \|\hat{\mathbf{D}}_j\|^2 + \|\hat{\mathbf{A}}_J\|^2. \quad (\text{A-18})$$

The above expressions can be written in terms of inner product, such that (f and \mathbf{X} are interchangeable notations):

$$V_j = \sum \langle f, \phi_{j,n} \rangle \tilde{\phi}_{j,n} \quad \text{and} \quad W_j = \sum \langle f, \psi_{j,n} \rangle \tilde{\psi}_{j,n}, \quad (\text{A-19})$$

where $\tilde{\phi}$ and $\tilde{\psi}$ are the dual basis of scaling ϕ and wavelet ψ . For the orthogonal basis, the dual basis is itself, i.e., $\tilde{\phi}(t) = \phi(t)$ and $\tilde{\psi}(t) = \psi(t)$. The inner products are the scaling coefficients ($\mathbf{A}_j = \{a_j(n)\}$) and wavelet coefficients ($\mathbf{D}_j = \{d_j(n)\}$):

$$\hat{a}_j(n) = \langle f, \phi_{j,n} \rangle = \frac{1}{\sqrt{2^j}} \int f(t) \phi\left(\frac{t-n}{2^j}\right) dt, \quad (\text{A-20})$$

$$\hat{d}_j(n) = (\hat{W}f)(t, n) = \langle f, \psi_{j,n} \rangle = \frac{1}{\sqrt{2^j}} \int f(t) \psi\left(\frac{t-n}{2^j}\right) dt. \quad (\text{A-21})$$

Numerically, the UDWT wavelet and scaling coefficients are also computed using the filter banks (see Figure A-9). Notice that the dyadic subsamplings are removed for UDWT.

As the coefficients are computed using the filter bank algorithm they can be viewed as the direct result of filtering the signal \mathbf{X} :

$$\hat{a}_j(n) = \sum_{l=1}^{L_j} \hat{g}_{j,l} X_{t-l \text{ modulo } N}, \quad t = 1, \dots, N; \quad (\text{A-22})$$

$$\hat{d}_j(n) = \sum_{l=1}^{L_j} \hat{h}_{j,l} X_{t-l \text{ modulo } N}, \quad t = 1, \dots, N. \quad (\text{A-23})$$

N is the length of coefficients vector $\mathbf{W}_{j,n}$ at all the scales; L_j is the filter width at j scale $L_j = (2^j - 1)(L - 1) + 1$; $\hat{h}_{j,l}$ and $\hat{g}_{j,l}$ are respectively the equivalent high-pass and low-pass filter coefficients for the computation of UDWT wavelet and scaling coefficients at scale j .

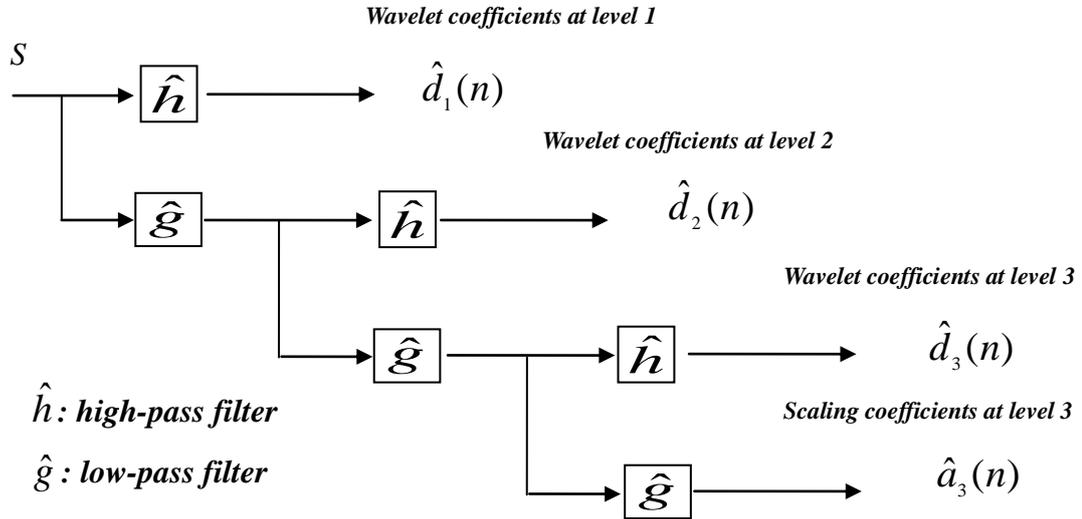


Figure A-9 Filter banks computation for UDWT (3-level decomposition).

A pattern search is particularly difficult if its representation depends on its location. To obtain the signal representations that are translation invariant is important in pattern recognition (Mallat 1999, pp146). The UDWT can provide this type of representation. When a pattern is translated, its UDWT representation will also be translated without modification. The UDWT has been discussed in the wavelet literature under different names, such as “shift translation invariant DWT”, “wavelet frames”, “translation invariant DWT”, “stationary DWT”, “time invariant DWT”, “non-decimated DWT” and “maximal overlap DWT”. The

UDWT is computed via a filter bank algorithm, the above mentioned '*algorithme à trous*'. This representation also allows direct analysis of the wavelet coefficients, which is impossible for DWT because the subsampling in time will greatly reduce the time resolution. The wavelet coefficients of UDWT are convenient for transient analysis. The precise timing locations can be found out by tracing the abrupt changes of wavelet coefficients. This singularity detection using UDWT is discussed in Section A.9.

A.6 Discrete Wavelet Packet Transform

This Section summarise the mathematical descriptions and filtering conventions for DWPT. More details refer to Mallat 1999, Percival and Walden 2000.

The DWPT is a more flexible version of the DWT, which yields a time-frequency like decomposition. By adaptively decomposing the dyadic scales, smaller frequency separations are possible. The WP nodes correspond to frequency bands. This flexible decomposition allows better insight of the time-frequency structure of signals. The DWPT can also be computed by the simple modification of the filter bank algorithm for the DWT. See Figure A-10. The DWPT generates a 'WP table' or 'WP tree'. The collection of nodes forming the indices of the WP table will be denoted by $T \equiv \{(j, k) : j = 1, \dots, J; k = 0, \dots, 2^j - 1\}$. The wavelet node (j, k) corresponds to the vector $\mathbf{P}_{j,k}$ of WP coefficients where $j = 1, \dots, J$ is the depth of the node, and $k = 0, \dots, 2^j - 1$ is the number of nodes that are on its left at the same depth.

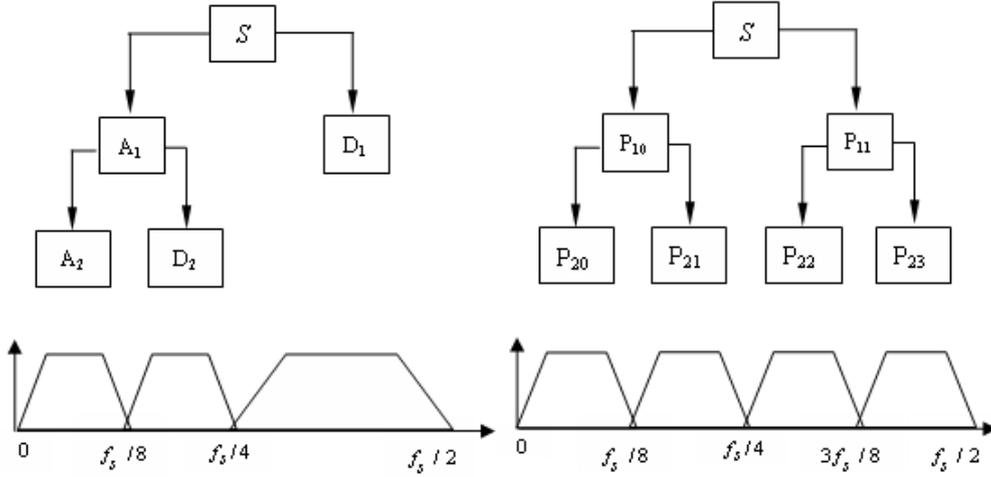


Figure A-10 Decomposition structures of DWT (left) and DWPT (right).

Let the real signal \mathbf{X} be an N dimensional vector $\{X(t) : t = 1, \dots, N\}$, where the sample size N is taken to be an integer multiple of 2^{J_0} . The DWPT of level J of \mathbf{X} yields an 2^j dimensional vector of DWPT coefficients, such that

$$\mathbf{C} = \begin{bmatrix} \mathbf{P}_{j,0} \\ \mathbf{P}_{j,1} \\ \vdots \\ \mathbf{P}_{j,2^j-2} \\ \mathbf{P}_{j,2^j-1} \end{bmatrix},$$

where the coefficients $\mathbf{P}_{j,k}$ are a $N_j = N/2^j$ dimensional vector associated with the frequency band $[f_s \cdot k/2^{j+1}, f_s \cdot (k+1)/2^{j+1}]$. $\|\mathbf{P}_{j,k}\|^2$, which can be interpreted as the energy in the frequency band, their sum preserves the signal energy:

$$\|\mathbf{P}_{j,k}\|^2 = E_{j,k} = \sum_{t=1}^{N_j} P_{j,k,t}, \quad (\text{A-24})$$

$$\|\mathbf{X}\|^2 = \sum_{k=0}^{2^j-1} \|P_{j,k}\|^2. \quad (\text{A-25})$$

The terminal WP nodes (\hat{j}, \hat{k}) preserve the energy of the signal \mathbf{X} , that is

$$\|X\|^2 = \sum_{\hat{j}=1}^J \sum_{\hat{k}=0}^{2^{\hat{j}}-1} \|P_{\hat{j},\hat{k}}\|^2. \quad (\text{A-26})$$

The WP coefficients can also be viewed as the direct result of filtering signal \mathbf{X} such that, the WP coefficients $\mathbf{P}_{j,k}$ of node (j, k) are:

$$\mathbf{P}_{j,k}(n) = \sum_{l=1}^{L_j} u_{j,k,l} X_{2^j(t+1)-1-l \text{ modulo } N}, \quad t = 1, \dots, N_j, \quad (\text{A-27})$$

where N_j is the length of coefficients vector $\mathbf{P}_{j,k}$ at level j and $N_j = N / j$, L_j is the filter width at j level $L_j = (2^j - 1)(L - 1) + 1$ and $u_{j,k,l}$ is the equivalent filter for WP node (j, k) . The wavelet quantifiers using DWPT will be introduced in Section A.10.

A.7 Wavelet Basis

Most applications of wavelet methods aim to efficiently represent signals with few non-zero wavelet coefficients. In other words, the goal is to produce as many as possible wavelet coefficients that are close to zero. The wavelet basis should be optimised to produce a maximum number of wavelet coefficients that are close to zero (Mallat 1999).

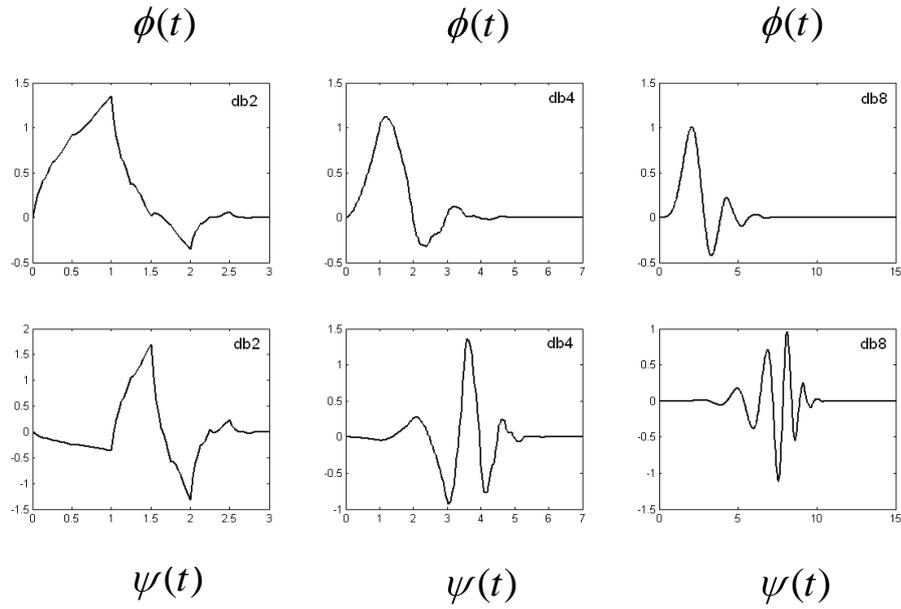


Figure A-11 Examples of Daubechies wavelets $db\ N$ (N is the number of vanishing moments).

Wavelet basis ψ has p vanishing moments if

$$\int t^k \psi(t) dt = 0 \quad (\text{A-28})$$

for $0 \leq k < p$. If the wavelet basis has enough vanishing moments p , then the wavelet coefficients are small at fine scales. However, the cost of vanishing moments is the support size. Compact support means the wavelet basis is optimal in the sense that it has a minimal support size for the given number of vanishing moments.

For the orthogonal basis, the support size has to be at least $2p - 1$ with vanishing moment p . Daubechies wavelets are compactly supported for a given number of vanishing moments. Three Daubechies wavelets are shown in Figure A-11. Chapter 7 shows a study on how Daubechies wavelet order influences the performance of wavelet quantifiers.

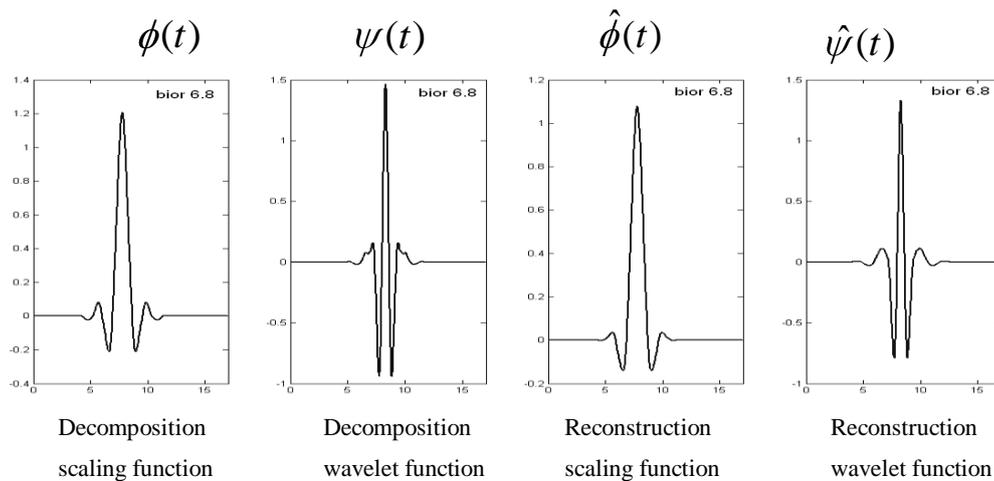


Figure A-12 Biorthogonal scaling and wavelet functions of bior 6.8.

Apart from the Haar (db1) wavelet, orthogonal wavelets with compact support cannot be either symmetric or antisymmetric. The wavelets are symmetric or antisymmetric if the coefficients of the filters used for wavelet implementation are symmetric or antisymmetric around the central coefficient. However, it is possible to have biorthogonal wavelets with compact support which are either symmetric or antisymmetric. For the biorthogonal wavelets, the compact support does not contradict the symmetry or antisymmetry. Symmetric or antisymmetric wavelets are implemented with perfect reconstruction filters having linear phase. Since the filter with linear phase has constant group delay, all frequency components have equal delay times. Linear phase is important in signal detection. The abrupt changes of wavelet coefficients at different scales can be traced to locate the singularities. Figure A-12 shows the biorthogonal wavelets bior 6.8. Chapter 5 shows a study using this biorthogonal wavelet for transient signal detection.

A.8 Wavelet Thresholding

Wavelet thresholding is a popular approach for signal denoising. The technique is performed on wavelet coefficients. See Figure A-13 for the procedure. The threshold is estimated from each scale and the coefficients at each scale are compared to the threshold. If the absolute value of coefficient is smaller than the

threshold, it is replaced by zero; otherwise, it is left unchanged or modified (Mallat 1999).

In hard thresholding, the coefficients greater than the threshold are left unchanged

$$D_h(w) = \begin{cases} w & \text{if } |w| \geq T \\ 0 & \text{if } |w| < T \end{cases} \quad (\text{A-29})$$

In soft thresholding, the coefficients greater than the threshold are reduced by an amount equal to the value of the threshold

$$D_s(w) = \begin{cases} \text{sgn}(w)(|w| - T) & \text{if } |w| \geq T \\ 0 & \text{if } |w| < T \end{cases} \quad (\text{A-30})$$

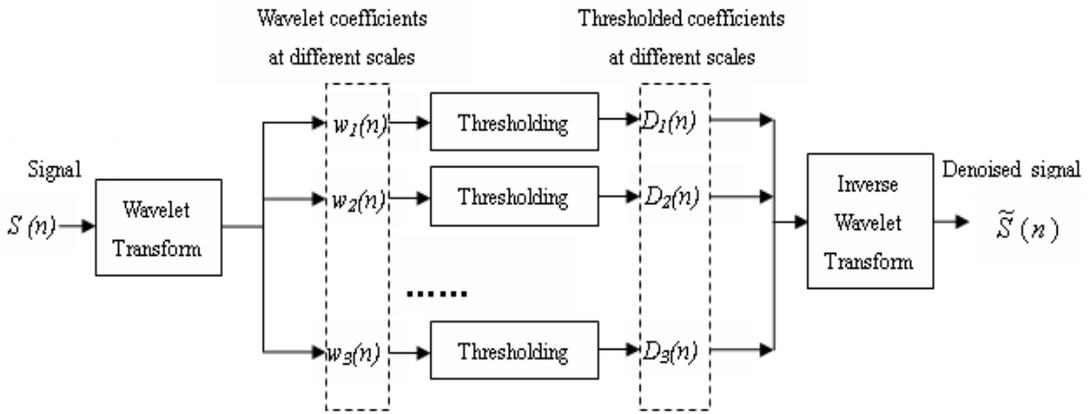


Figure A-13 Signal denoising via wavelet multi-scale thresholding and reconstruction.

Donoho and Johnstone (1994, 1995) systematically proposed the well known non-parametric thresholding estimators for wavelet denoising: SquareTwoLog, Stein Unbiased Risk Estimate (SURE) and Minimax. See also Misiti *et al.* 2000, Mallat 1999.

A signal \mathbf{S} of size L is contaminated by the addition of a noise \mathbf{W} . The measured signal \mathbf{X} is

$$\mathbf{X} = \mathbf{S} + \mathbf{W}, \quad (\text{A-31})$$

where \mathbf{W} is a Gaussian white noise of variance σ^2 . The signal \mathbf{S} is estimated

by the thresholded wavelet coefficients $D\langle \mathbf{X}, \psi \rangle$ projecting onto the dual basis space $\{\tilde{\psi}\}$

$$\hat{\mathbf{X}} = \sum D\langle \mathbf{X}, \psi \rangle \tilde{\psi}, \quad (\text{A-32})$$

where D is the thresholding function.

The SquareTwoLog method estimates the threshold as

$$T_{STL} = \sigma \sqrt{2 \log_e L}, \quad (\text{A-33})$$

where the noise variance σ is estimated at each scale as

$$\sigma = \frac{Med}{0.6745}, \quad (\text{A-34})$$

where Med is the median of wavelet coefficients.

Denote the wavelet coefficients w of signal \mathbf{X} at a scale by ascending order as w^{as} . The SURE method estimates the threshold T_{SURE} to be one of the wavelet coefficients in the sense it can minimise a quadratic loss function F

$$T_{SURE} = w^{as}(n) \Big|_{\min_n} (F(n)). \quad (\text{A-35})$$

The Minimax method minimises the maximum risk R

$$T_{\text{minimax}} = \inf_D \sup_S \{R(D, \mathbf{S})\}. \quad (\text{A-36})$$

where the risk $R(D, \mathbf{S}) = E\{\|D\mathbf{X} - \mathbf{S}\|^2\}$, and the threshold can be rescaled by the noise variance to $\sigma T_{\text{minimax}}$.

The above thresholding methods are convenient to perform because of their non-parametric nature. The sequence of thresholds estimated by the non-parametric thresholding estimators is SquareTwoLog, Minimax, SURE in descending order. This sequence is useful for designing a suitable thresholding strategy for signals with different degrees of noise contamination. For

observations with high SNR, strict thresholding (i.e. SquareTwoLog) can effectively eliminate noise and highlight the signal of interest. For observations with strong noise contamination, conservative (i.e. SURE) or moderate (i.e. Minimax) thresholding can help to improve the SNR. In Chapter 5, DWT thresholding is applied to improve the effective range of the parameter values used in the LocMax algorithm. Chapter 6 shows an UDWT denoising scheme, which combines the Wavelet Maxima Chain search and the thresholding principle, can achieve better performance than DWT thresholding.

A.9 Wavelet for Singularity Detection

Singularities and irregularities often carry the most important information of signals. In images, the locations of the object edges are provided by the discontinuities of the intensity. For some signals, such as the ECG signal and the AE signal used in this thesis, the most interesting information is carried by the transient characteristics such as impulses. Singularity detection is one of the most important applications of wavelet methods. This was reported in the pioneering work of Mallat and Hwang (1992) using UDWT. Singularities are detected by searching the abscissa where the wavelet modulus maxima converge at fine scales.

The term “modulus maximum” is used to describe the local maximum at point (u_0, a_0) such that the wavelet coefficient $|Wf(u, a_0)|$ is locally maximal at $u = u_0$. This local maximum is locally maximal at either the right or the left neighbourhood of u_0 . The modulus maximum that propagates from the finest scale to the coarse scales having the same signs corresponds to the signal transition. This reflects the important difference between signal transition and noise. Actual signal transition produces large coefficients across many scales. The signal transition shows persistence across scales, while noise attenuates very quickly as the scale increases. These associated modulus maxima are called Wavelet Maxima Chain (WMC) in this thesis.

In the work of Mallat and Hwang, the local regularity estimation is then applied as a criterion for selecting the significant coefficients. They estimate the local regularity, which is often measured by the Lipschitz exponent (also called *Hölder* exponent in the mathematical literature), from the decay of the modulus maxima across the scales. If the amplitude of the modulus maximum decreases strongly when the scale increases, it indicates the corresponding singularity has a negative Lipschitz exponent $\alpha < 0$. These maxima are mostly dominated by the noise and the whole WMC is removed. However, the modulus maximum produced by the signal transition would follow the decay at the rate 2^α and Lipschitz exponent $\alpha > 0$ when the scale decreases.

The above denoising algorithm shows the feasibility to discriminate a signal from noise by analysing the modulus maxima evolution across scales. Apart from the regularity estimation, approaches based on inter-scale ratios or inter-scale products of modulus maxima were also developed for the selection of significant coefficients (Pizurica 2002). Depending on the application, one could select the coefficient at a position k based on a significance measure m_k , which is computed from the observed wavelet coefficients. For example, one can define m_k as the local regularity estimation, coefficient magnitude, or the amount of inter-scale correlation at the position k . The singularity detection incorporated with the thresholding principle is used for denoising and shown in Chapter 6.

A.10 Wavelet or WP Quantifiers: Relative Energy and Entropy

Wavelet coefficients or WP coefficients are decomposed by the filter bank algorithm. These coefficients associated with specific frequency band preserve the energy of signal in the frequency band. Wavelet or WP quantifiers are then obtained by the expressions of these coefficients. The pioneer works on these quantifiers were reported by Rosso *et al.* (2001), Blanco *et al.* (1998) and Zunino *et al.* (2007) to quantitatively study electroencephalogram (EEG) signals.

Relative Energy (RE) is used for describing the energy distribution in the

frequency domain, normalised over the energy of the signal. To compute RE, the energies of coefficients associated with a specific frequency band are normalised over the energy of the whole signal. Because the coefficients can be viewed as the output of band-pass filtering, RE reflects the probability distributions of the frequency bands from the viewpoint of the frequency domain.

Entropy is used for measuring the uncertainty (i.e. the degree of disorder) of the signal. This concept plays an important role in many scientific and application fields, including statistical mechanics, information theory, communications, signal processing, data mining and machine learning. The Shannon entropy (Shannon and Weaver 1959) is the most popular quantifier of entropy using logarithm of base 2:

$$Q = -\sum_j p_j \cdot \log_2 p_j \quad (\text{A-37})$$

where $\{p_j\}$ is the probability distribution satisfying the condition

$$\sum_j p_j = 1. \quad (\text{A-38})$$

The value of the entropy reflects the degree of disorder. A bigger entropy value means greater disorder and vice versa. Depending on how $\{p_j\}$ is obtained, the entropy can measure the uncertainty of the signal in different ways. If the distribution $\{p_j\}$ is computed using the RE, the entropy will measure the degree of disorder for the frequency energy distribution. A signal with a narrow band spectrum can be viewed as a typical example of ordered frequency energy distribution with small entropy value. This kind of entropy is called Total Entropy (TE). If the distribution $\{p_j\}$ is computed using the normalised coefficient in a specific frequency band, the entropy will measure the degree of disorder of the coefficients in this frequency band. It shows how redundant the time series is. A signal with some periodic or self-similar patterns is an ordered process having a small entropy value. This second type of entropy is called Node Entropy (NE).

The wavelet or WP quantifiers construct a time-frequency analysis in a quantitative way. Chapter 7 shows the details on the application of these quantifiers for BFDD.

A.11 Summary

Wavelet methods include a set of processing and analysis methods based on the idea of a wavelet transform. This chapter briefly reviewed the principles of wavelet theory, the differences between the transform variations, the mathematical descriptions and filtering conventions for different transform variations.

The wavelet basis is very compressed for high frequency which allows good localisation at time. The DWT can be considered as the subsampled version of the CWT at dyadic scales and time. The UDWT can be considered as the subsampled version of the CWT at dyadic scales. Compared to DWT, the UDWT is translation invariant and is more suitable for the transient signal analysis which requires precise time localisation. This UDWT representation also allows direct analysis of the wavelet coefficients, which can bring convenience for transient analysis. By adaptively decomposing the dyadic scales of the DWT, smaller frequency separations are possible for the DWPT. This flexible decomposition allows better insight of the time-frequency structure of signal. The wavelet basis of the DWPT can be flexibly chosen for practical implementation considerations and the contaminating noises concentrating in some frequency bands can be easily eliminated. All these transform variations can be efficiently computed via the filter bank algorithm, which is one of the important characteristics of wavelet methods (Sweldens 1996).

Sections A.7-A.10 showed the theoretical basis for the understanding of the following chapters. Orthogonal and biorthogonal wavelet bases have very different characteristics, which lead to different applications. The Daubechies basis with compact support can efficiently represent signals with few non-zero wavelet coefficients, which can be put to good use in the applications concerning the statistics of wavelet coefficients. In singularity detection, biorthogonal bases

are often preferred because of their symmetry and consequent linear phase. Studies using these two types of wavelet bases are discussed in chapter 6 and chapter 5, respectively. Section A.8-A.10 reviewed the wavelet-based techniques which are studied for BFDD in the chapter 5-7. These include wavelet thresholding for signal denoising, singularity detection using UDWT, wavelet or WP quantifiers for quantitative time-frequency analysis.

Appendix B

A Localised Defect Detection using ACF

B.1 Introduction

The most commonly used signal processing technique for a vibration signal to detect the localised bearing defects is the envelope analysis or High Frequency Resonance Technique (HFRT). See Section 4.4. One of the difficulties with envelope analysis on a vibration signal is the determination of the best resonance frequency band to envelope at the pre-processing stage. One of the solutions for the problem is using an AE signal to detect the localised defects. The modulation of AE signatures at bearing Characteristic Defect Frequency (CDF) has been reported in many research works (Holroyd and Randall 1993, Holroyd 2001). The AE signals can be demodulated directly to reveal the CDF without the need to find out the resonance frequency band. Since AE signals have broad high-frequency range and very good time resolution, they can characterise the impact train (generated by a localised defect, which is transient in nature) well.

This appendix studies the detection of localised bearing defect using autocorrelation function (ACF). To facilitate the detection scheme, the Characteristic Defect Frequency (CDF) will be first expressed in the time domain as Characteristic Defect Interval (CDI). The relationship between CDF and CDI has been pointed out in Section 2.2, that is, CDI (expressed as the number of discrete sampling points) is the sampling frequency divided by CDF.

B.2 Autocorrelation Function

ACF is very powerful for finding out repeating pattern in a signal. It is a function of time lag m used to measure the associated degree of a signal itself at time lag

m :

$$R_{xx}(m) = \varepsilon\{x(n+m)x^*(n)\} = \lim_{M \rightarrow \infty} \frac{1}{2M+1} \sum_{n=-M}^{n=M} x(n+m)x^*(n). \quad (\text{A-1})$$

In practice, ACF has to be estimated from the available finite discrete signal $x(n)$. Assuming real signal $x(n)$ is the N data samples with indices from $n=0$ to $n=N-1$, the biased estimates of ACF is

$$R^B_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n+|m|)x(n). \quad (\text{A-2})$$

The unbiased estimates of ACF for lags indices $-(N-1) \leq m \leq N-1$ is

$$R^u_{xx}(m) = \frac{1}{N-|m|} \sum_{n=0}^{N-|m|-1} x(n+|m|)x(n), \quad (\text{A-3})$$

B.3 Detection Scheme

The following detection scheme for localised bearing aims to find out if there are repeating impulses in a signal and estimate the interval between these impulses. There are three steps for the detection scheme using ACF (Schlindwein *et al.* 2004): a signal is first transformed into envelope signal; a short time moving window is then used, as an approximation to a matched filter, to split the signal and their ACFs are calculated; finally, the positions of the local maximum coefficients of these ACF in a certain range are found out respectively.

The approaches to achieve envelope signals are different for vibration and AE signals. The envelope vibration signal needs to be achieved from raw signal via filtering around resonance frequency band and then demodulating. But it is difficult to determine the best frequency band of raw vibration signal to perform filtering. Refer to Section 3.3 for some details. In contrast, the envelope Acoustic Emission (AE) signal can be achieved easily via only demodulating raw signal.

B.4 Experimental Setup and Signal Processing

A five stage “Roots and Claw” dry vacuum pump with empty load was used as test bed. A defective bearing with inner race defect was mounted at the high vacuum side of pump. The speed of pump was set at 105 Hz (6300 rev min⁻¹) and the inlet pressure was set at 0 mbar. An AE transducer (PAC R3α) and an accelerometer were firmly held at the surface of the pump house near the high vacuum end in the radial direction. AE signals were amplified with gain of 1000 and filtered by a band-pass filter. Vibration signals were suitably pre-filtered with in-house built 8th order elliptic anti-aliasing filters with cut-off frequencies of 10 kHz.

Analogue signals were digitised by a 16-bit NI Analogue to Digital Converter (ADC) with the sampling rate 100 kHz and stored in the computer. The stored vibration signals were then band-pass filtered to isolate a resonant frequency range 8 kHz-10 kHz. The AE and vibration signals were demodulated and then downsampled to 25 kHz. For a sampling rate 25 kHz, the characteristic Ball Pass Frequency of Inner race (BPFI) is 550 Hz and the corresponding CDI is estimated as 45 (expressed as the number of discrete sampling points).

A short time moving window was then applied to split the signals into short segments with 50% overlap, and the ACFs of the segments were calculated via formula (A-3). Finally, the positions of the local maximum coefficient of these ACFs in a search range were found out. The search range here is set from 40 to 50.

B.5 Results and Summary

Figure B-1 and Figure B-2 below show the histograms of the found positions using the AE and vibration signals respectively. Figure B-1 is a typical Gaussian distribution that converges to its mean value 45 which matches with the theoretical value. In contrast, the distribution in Figure B-2 does not show this convergence.

Appendix B

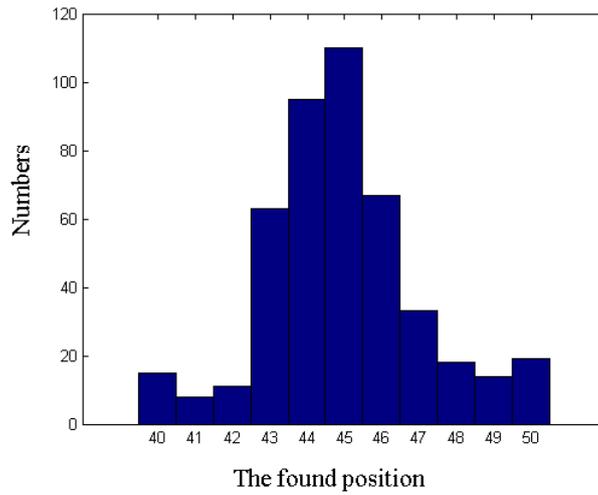


Figure B-1 The histogram of the positions found using the AE signals.

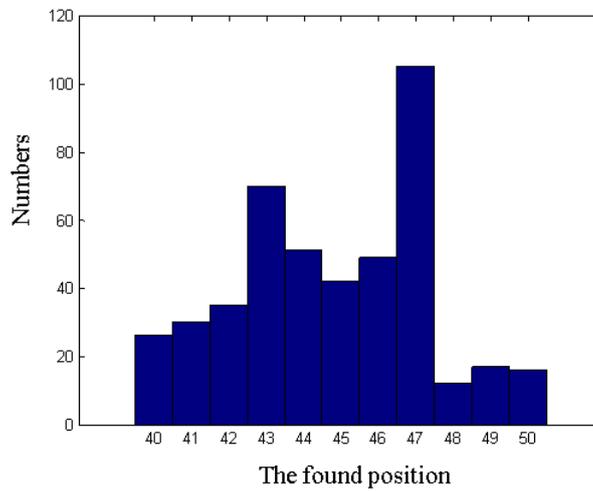


Figure B-2 The histogram of the positions found using the vibration signals.

This appendix studied a detection scheme for localised defect using autocorrelation function. It is shown that using AE signals can improve the detection performance.

Appendix C

MATLAB Codes

C.1 Cyclostationary Signals Simulation

```
%% Cyclostationary Signals Simulation (Cauchy distribution)
Par.N=1000000;
Par.CDI=143;
Par.rnd=floor(Par.N/Par.CDI);
cauchyn=zeros(1,Par.rnd);

!! Here generate a set of Cauchy distribution data which size is Par.rnd;
%% Store the integer part of generated data in cauchyn();
cauchyn=floor(abs(cauchyn));
%% limit data size for simulation;
for i=1:Par.rnd
    if cauchyn(1,i)>400
        cauchyn(1,i)=cauchyn(1,i-1);
    end
end

%% Set the deviation of noise;
Par.dev=0.5;

%% Here are options for different types of additive noise;
% Sim(1,:) =rand(1,Par.N);           %% Uniform noise;
% Sim(2,:) =abs(Par.dev*randn(1,Par.N)+2); %% Gaussian noise all greater than zeros;
% Sim(3,:) =abs(Par.dev*randn(1,Par.N)); %% white Gaussian noise;
% Sim(4,:) =zeros(1,Par.N);         %% without noise;

!! Choose one of the options;
Sim(5,:)=Sim(2,:);
%% Periodic impulse with Cauchy distribution;
```

Appendix C

```
ip=1; Par.p=2; Par.S=4*Par.dev;
while Par.p<Par.N
    Sim(5,Par.p)=Sim(5,Par.p)+Par.S;
    Par.p=Par.p+cauchyn(1,ip);
    ip=ip+1;
end

%% Noise estimation
thr.noise=wnoisest(Sim(5,1:1000));
%% Threshold estimation using SquareTwoLog, Minimax or SURE estimator;
thr.sqt =thselect(Sim(5,1:1000),'sqtwoLog');
thr.min =thselect(Sim(5,1:1000),'minimaxi');
thr.heur =thselect(Sim(5,1:1000),'heursure');
thr.rig =thselect(Sim(5,1:1000),'rigrsure');

!! Choose one of the threshold estimates for hard thresholding processing;
thr.se = thr. rig;
%% Signal denoising;
for i=1:Par.N
    if Sim(5,i)>thr.se
        Sim(6,i)=Sim(5,i);
    else
        Sim(6,i)=0;
    end
end
end
```

C.2 LocMax Algorithm and Interval Calculation

```

%% LocMax Detection algorithm
!! Here load in signal r(1,:) for detection;
ind.r=find(r(1,:)~=0);
l.i=length(ind.r);

l.do=180;      %% The condition of distance to terminate a local search;
j=1;
t=ind.r(1,1);  %% Initialise the position index vector;

%% The following program for timing detection of signal;
for i=2:l.i
    if ind.r(1,i)-t>l.do
        ind.do(1,j)=t;
        j=j+1;
        t=ind.r(1,i);
    elseif abs(r(1,ind.r(1,i)))>abs(r(1,t))
        t=ind.r(1,i);
    end
end
ind.do(1,j)=t;

%% Find out the intervals;
l.b=length(ind.do(1,:));
ki=1;
for i=1:l.b-1
    p(ki)=ind.do(1,i+1)-ind.do(1,i);
    ki=ki+1;
end

%% Plot the intervals histogram;
figure;hist(p',1:2:500);

```

C.3 UDWT Denoising Algorithm

```

!! Load in signal S.D for denoising;
Ld=2048;                %% Set the length of signal segment;
c=floor(length(S.D)/Ld); %% Calculate the number of signal segments;
ki=1;
kj=1;
km=1;
l.do=90;

%% calculate c realization;
for ii=1:c
    A(ii,:)=S.D(1+Ld*(ii-1):Ld*ii);
    clear ind;

    %% Stationary Wavelet Transform using biorthogonal wavelet basis
    swc=swt(A(ii,:),2,'bior6.8');
    %% Estimate threshold using SURE estimator;
    thr.se(ii) =thselect(swc(1,:),'rigrsure');

    %% consider the phase problem
    rswc(1,:)=rshift(swc(1,:));
    rswc(2,:)=rshift(swc(2,:));
    lswc(1,:)=lshift(swc(1,:));
    lswc(2,:)=lshift(swc(2,:));

    %% Here ignore the border coefficients to avoid border effect;
    swc(:,1:2)=0;
    lswc(:,1:2)=0;
    rswc(:,1:2)=0;
    swc(:,Ld-1:Ld)=0;
    lswc(:,Ld-1:Ld)=0;
    rswc(:,Ld-1:Ld)=0;

    %% Positive part; Only keep the Maximum Chain;
    pokeep(1:5,:)=zeros(5,Ld);

```

Appendix C

```
for i=1:Ld
    for j=1:2
        if swc(j,i)>thr.se(ii) & swc(j,i)>lswc(j,i) & swc(j,i)>rswc(j,i);
            pokeep(j,i)=swc(j,i);
        else
            pokeep(j,i)=0;
        end
    end
end

for i=3:(Ld-2)
    if pokeep(1,i)~=0 & pokeep(2,i)~=0
        pokeep(3,i)=1;
        pokeep(4,i)=pokeep(1,i);
        pokeep(5,i)=pokeep(2,i);
    elseif pokeep(1,i)~=0 & pokeep(2,i-1)~=0
        pokeep(3,i)=2;
        pokeep(4,i)=pokeep(1,i);
        pokeep(5,i)=pokeep(2,i-1);
    elseif pokeep(1,i)~=0 & pokeep(2,i+1)~=0
        pokeep(3,i)=3;
        pokeep(4,i)=pokeep(1,i);
        pokeep(5,i)=pokeep(2,i+1);
    end
end

%% Negative part; Only keep the Maximum Chain;
nokeep(1:5,:)=zeros(5,Ld);

for i=1:Ld
    for j=1:2
        if swc(j,i)<thr.se(ii) & swc(j,i)<lswc(j,i) & swc(j,i)<rswc(j,i);
            nokeep(j,i)=swc(j,i);
        else
            nokeep(j,i)=0;
        end
    end
end
```

Appendix C

```
for i=3:(Ld-2)
    if nokeep(1,i)~=0 & nokeep(2,i)~=0
        nokeep(3,i)=1;
        nokeep(4,i)=nokeep(1,i);
        nokeep(5,i)=nokeep(2,i);
    elseif nokeep(1,i)~=0 & nokeep(2,i-1)~=0
        nokeep(3,i)=2;
        nokeep(4,i)=nokeep(1,i);
        nokeep(5,i)=nokeep(2,i-1);
    elseif nokeep(1,i)~=0 & nokeep(2,i+1)~=0
        nokeep(3,i)=3;
        nokeep(4,i)=nokeep(1,i);
        nokeep(5,i)=nokeep(2,i+1);
    end
end

%% Save the denoising results;
keep(1:3,:)=zeros(3,Ld);
keep(1:3,:)=nokeep(3:5,:)+pokeep(3:5,:);
end
```

C.4 Wavelet Packet Quantifiers

```

%% Algorithms for calculating Wavelet Packet Quantifiers;
!! Load in signals S.L, S.D, S.F for three different bearing conditions (localised, contaminated, fault free);
%% Sampling rate is 80 kHz;
!! Require installing MATLAB wavelet toolbox;

L=length(S.L);
Le=2048;          %% Set the length of signal segment;
W=floor(L/Le);   %% Calculate the number of signal segments;

%% Calculate the quantifiers of all realization; Decomposition using Haar wavelet;
for i=1:W

%% Calculate quantifiers for localised defect bearing;
wpt.L = wptdec(S.L(1+Le*(i-1):Le*i),2,'haar');
[cfs0,cfs1,cfs2,cfs3]=read(wpt.L,'cfs',3,'cfs',4,'cfs',6,'cfs',5);
cfs0=zeros(size(cfs0));
wpt.L=write(wpt.L,'cfs',3,cfs0);  %% Set 0-10kHz frequency part as zero;
cfs1=cfs1.^2/sum(cfs1.^2);
cfs2=cfs2.^2/sum(cfs2.^2);
cfs3=cfs3.^2/sum(cfs3.^2);

%% Calculate Node Entropy quantifier;
Entropy.T.L(i,:)=[-sum(cfs1.*log2(cfs1)) -sum(cfs2.*log2(cfs2)) -sum(cfs3.*log2(cfs3))];

%% Calculate Relative Energy quantifier;
E(i,:) = wenergy(wpt.L)/100;
%% Reorder energy vector into frequency order;
Energy.L(i,:)=[E(i,2) E(i,4) E(i,3)];

%% Calculate Total Entropy quantifier;
Entropy.W.L(i,:)=
[-Energy.L(i,1)*log2(Energy.L(i,1)) -Energy.L(i,2)*log2(Energy.L(i,2)) -Energy.L(i,3)*log2(Energy.L(i,3))];
TotEntropy.W.L(i)=sum(Entropy.W.L(i,:));
clear E; clear cfs1; clear cfs2; clear cfs3; clear cfs0;

```

Appendix C

```

%% Calculate quantifiers for contaminated bearing using the same algorithm;
wpt.D = wptdec(S.D(1+Le*(i-1):Le*i),2,'haar');
[cfs0,cfs1,cfs2,cfs3]=read(wpt.D,'cfs',3,'cfs',4,'cfs',6,'cfs',5);
cfs0=zeros(size(cfs0));
wpt.D=write(wpt.D,'cfs',3,cfs0);
cfs1=cfs1.^2/sum(cfs1.^2);
cfs2=cfs2.^2/sum(cfs2.^2);
cfs3=cfs3.^2/sum(cfs3.^2);
Entropy.T.D(i,:)=[-sum(cfs1.*log2(cfs1)) -sum(cfs2.*log2(cfs2)) -sum(cfs3.*log2(cfs3))];
E(i,:) = wenergy(wpt.D)/100;
Energy.D(i,:)=[E(i,2) E(i,4) E(i,3)];
Entropy.W.D(i,:)=[-Energy.D(i,1)*log2(Energy.D(i,1))
-Energy.D(i,2)*log2(Energy.D(i,2)) -Energy.D(i,3)*log2(Energy.D(i,3))];
TotEntropy.W.D(i)=sum(Entropy.W.D(i,:));
clear E; clear cfs1; clear cfs2; clear cfs3; clear cfs0;

%% Calculate quantifiers for fault free bearing using the same algorithm;
wpt.F = wptdec(S.F(1+Le*(i-1):Le*i),2,'haar');
[cfs0,cfs1,cfs2,cfs3]=read(wpt.F,'cfs',3,'cfs',4,'cfs',6,'cfs',5);
cfs0=zeros(size(cfs0));
wpt.F=write(wpt.F,'cfs',3,cfs0);
cfs1=cfs1.^2/sum(cfs1.^2);
cfs2=cfs2.^2/sum(cfs2.^2);
cfs3=cfs3.^2/sum(cfs3.^2);
Entropy.T.F(i,:)=
[-sum(cfs1.*log2(cfs1)) -sum(cfs2.*log2(cfs2)) -sum(cfs3.*log2(cfs3))];
E(i,:) = wenergy(wpt.F)/100;
Energy.F(i,:)=[E(i,2) E(i,4) E(i,3)];
Entropy.W.F(i,:)=
[-Energy.F(i,1)*log2(Energy.F(i,1)) -Energy.F(i,2)*log2(Energy.F(i,2)) -Energy.F(i,3)*log2(Energy.F(i,3))];
TotEntropy.W.F(i)=sum(Entropy.W.F(i,:));
clear E; clear cfs1; clear cfs2; clear cfs3; clear cfs0;

end

```

C.5 One Dimensional Bayesian Classification

```

!! Load in features;
!! Toolbox named STPRTool by Franc needs to be installed;
%% Access to website http://cmp.felk.cvut.cz/cmp/software/stprtool/;
%% trn is the 1-dimensional feature matrix;

trn.F(1,:)=Energy.F(:,1);
trn.D(1,:)=Energy.D(:,1);
trn.L(1,:)=Energy.L(:,1);

trn.F(2,:)=Energy.F(:,2);
trn.D(2,:)=Energy.D(:,2);
trn.L(2,:)=Energy.L(:,2);

trn.F(3,:)=Energy.F(:,3);
trn.D(3,:)=Energy.D(:,3);
trn.L(3,:)=Energy.L(:,3);

trn.F(4,:)=TotEntropy.W.F(:);
trn.D(4,:)=TotEntropy.W.D(:);
trn.L(4,:)=TotEntropy.W.L(:);

trn.F(5,:)=Entropy.T.F(:,1);
trn.D(5,:)=Entropy.T.D(:,1);
trn.L(5,:)=Entropy.T.L(:,1);

trn.F(6,:)=Entropy.T.F(:,2);
trn.D(6,:)=Entropy.T.D(:,2);
trn.L(6,:)=Entropy.T.L(:,2);

trn.F(7,:)=Entropy.T.F(:,3);
trn.D(7,:)=Entropy.T.D(:,3);
trn.L(7,:)=Entropy.T.L(:,3);

%% Number of the classes; %% Label the classes;
W=length(trn.F(1,:));
le.F=ones(1,W);
le.D=2*le.F;
le.L=3*le.F;

```

Appendix C

```
%% Set the percentage of training data;
ind=floor(0.3*W);
comp=3;

%% Calculate all the classification errors;
for i=1:7

    %% Using Gaussian Mixture Models to train the model;
    bayes_model.Pclass{1}=
    emgmm(trn.F(i,1:ind),struct('ncomp',comp,'tmax',2000,'eps_logL',0.1,'eps_Alphas',0.1));
    bayes_model.Pclass{2}=
    emgmm(trn.D(i,1:ind),struct('ncomp',comp,'tmax',2000,'eps_logL',0.1,'eps_Alphas',0.1));
    bayes_model.Pclass{3}=
    emgmm(trn.L(i,1:ind),struct('ncomp',comp,'tmax',2000,'eps_logL',0.1,'eps_Alphas',0.1));

    %% The priori probability is 1/3 for each class;
    bayes_model.Prior=[1/3 1/3 1/3];
    %% Using Bayesian classification;
    ypred.F=bayescls(trn.F(i,:),bayes_model);
    ypred.D=bayescls(trn.D(i,:),bayes_model);
    ypred.L=bayescls(trn.L(i,:),bayes_model);

    %% Calculate the classification errors;
    err.F(i)=cerror(ypred.F,le.F);
    err.D(i)=cerror(ypred.D,le.D);
    err.L(i)=cerror(ypred.L,le.L);
    err.s(i)=err.F(i)+err.D(i)+err.L(i);

    %% Remember here to clear all temporal vectors for next iteration;
    clear bayes_model;clear ypred;

end

error(1,:)=err.F;
error(2,:)=err.D;
error(3,:)=err.L;
error(4,:)=err.s;
error=roundn(error,-3);
```

C.6 Two Dimensional Bayesian Classification

!! Toolbox named STPRTool by Franc needs to be installed;

```
%% Load in Relative Energy;
```

```
trn.F(1,:)=Energy.F(:,1);
```

```
trn.D(1,:)=Energy.D(:,1);
```

```
trn.L(1,:)=Energy.L(:,1);
```

```
trn.F(2,:)=Energy.F(:,2);
```

```
trn.D(2,:)=Energy.D(:,2);
```

```
trn.L(2,:)=Energy.L(:,2);
```

```
trn.F(3,:)=Energy.F(:,3);
```

```
trn.D(3,:)=Energy.D(:,3);
```

```
trn.L(3,:)=Energy.L(:,3);
```

```
%% Load in Total Entropy;
```

```
trn.F(4,:)=TotEntropy.W.F(:);
```

```
trn.D(4,:)=TotEntropy.W.D(:);
```

```
trn.L(4,:)=TotEntropy.W.L(:);
```

```
%% Load in Node Entropy;
```

```
trn.F(5,:)=Entropy.T.F(:,1);
```

```
trn.D(5,:)=Entropy.T.D(:,1);
```

```
trn.L(5,:)=Entropy.T.L(:,1);
```

```
trn.F(6,:)=Entropy.T.F(:,2);
```

```
trn.D(6,:)=Entropy.T.D(:,2);
```

```
trn.L(6,:)=Entropy.T.L(:,2);
```

```
trn.F(7,:)=Entropy.T.F(:,3);
```

```
trn.D(7,:)=Entropy.T.D(:,3);
```

```
trn.L(7,:)=Entropy.T.L(:,3);
```

```
%% Number of the classes;
```

```
W=length(trn.F(1,:));
```

```
%% Label the classes;
```

```
le.F=ones(1,W);
```

Appendix C

```
le.D=2*le.F;
le.L=3*le.F;

%% Set the percentage of training data;
ind=floor(0.3*W);
comp=3;

%% trn1 is the 2-dimensional feature matrix;
N=4;
M=2;
trn1.F(1,:)=trn.F(N,:);
trn1.D(1,:)=trn.D(N,:);
trn1.L(1,:)=trn.L(N,:);

trn1.F(2,:)=trn.F(M,:);
trn1.D(2,:)=trn.D(M,:);
trn1.L(2,:)=trn.L(M,:);

%% Using Gaussian Mixture Models to train the model;
bayes_model.Pclass{1}=emgmm(trn1.F(:,1:ind),struct('ncomp',comp));
bayes_model.Pclass{2}=emgmm(trn1.D(:,1:ind),struct('ncomp',comp));
bayes_model.Pclass{3}=emgmm(trn1.L(:,1:ind),struct('ncomp',comp));

%% The priori probability is 1/3 for each class;
bayes_model.Prior=[1/3 1/3 1/3];

%% Using Bayesian classification;
ypred.F=bayescls(trn1.F,bayes_model);
ypred.D=bayescls(trn1.D,bayes_model);
ypred.L=bayescls(trn1.L,bayes_model);

%% Calculate the classification errors;
err.F=cerror(ypred.F,le.F);
err.D=cerror(ypred.D,le.D);
err.L=cerror(ypred.L,le.L);
err.s=err.F+err.D+err.L;

%% Visualization 2-d classification map;
figure;hold on;
```

Appendix C

```
ppatterns(trn1.F,'k. ');  
ppatterns(trn1.D,'b. ');  
ppatterns(trn1.L,'r. ');  
bayes_model.fun='bayescls';  
pboundary(bayes_model); %% Plot the boundary;
```