

**Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester**

by

Suguna Thanagasundram

**BEng (Electrical Engineering), National University of Singapore
MSc in Information & Communication Engineering, University
Leicester (2003)**

**Department of Engineering
University of Leicester**

May 2007

UMI Number: U232147

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U232147

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

FAULT DETECTION USING AUTOREGRESSIVE MODELLING TECHNIQUES

by

Suguna Thanagasundram

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.

Suguna Thanagasundram

May 2007

Abstract

The use of spectral analysis for fault detection and diagnostics in real-time has been conservative due to concerns over large processing requirements, especially when large sample sizes and high sampling frequencies are used. In this work, it is shown how such concerns can be allayed, to a large extent, by Autoregressive (AR) modelling, as the AR method has enhanced resolution capabilities compared to the Fast Fourier Transform (FFT) technique even when small sample sizes are used and requires a sampling rate just slightly above the Nyquist rate to give good parameter estimates. The use of a parametric method of AR modelling for fault diagnosis and prognosis is a relatively new concept in the field of condition monitoring.

In this thesis, a new methodology is proposed that combines AR modelling techniques and pole-related spectral decomposition for the detection of incipient single-point bearing defects for a vibration-based condition monitoring system. Vibration signals obtained from the ball bearings of a dry vacuum pump operating in normal and faulty conditions are used as the test signals and are modelled as time-variant AR series.

The positions of the poles, which are the roots of the AR coefficient polynomial, vary for every frame of vibration data. It is a known fact that as defects such as spalls and cracks start to appear on the ball bearings, the amplitude of the vibrations of characteristic defect frequencies increases. This is seen as the poles moving closer to the unit circle as the severity of the defect increases. Simple statistical indicators such as the power and frequency of each bearing defect spectral component can be extracted from the residual and position of the AR poles. These indicators can be effectively used for fault classification to distinguish between the no-fault and defective cases as the difference between them is significant.

(300 words)

Acknowledgements

One afternoon, while I was nearly finishing my thesis for my MSc studies in Information and Communications Engineering at University of Leicester, I arranged a meeting to meet up Dr. Fernando Soares Schlindwein who informed me that the department has won some funding from the Engineering and Physical Sciences Research Council (EPSRC) to conduct research on some novel fault prediction methods for a dry vacuum pump and enthusiastically encouraged me to apply for the scholarship. I spent the next two weeks intrigued by the thoughts of working on controller mechanisms to prevent early pump failure and pondered whether I can use my electrical background effectively to tackle the mechanical problem. Little did I realize, that my life will be intertwined with working with the dry vacuum pump for the next three years and it will take me lots of help from lots of people and a new experience into the academic world and a new education to understand and tackle the problem.

I would like to take this opportunity to thank my project supervisor, Dr. Fernando Soares Schlindwein, for his guidance, patience and knowledgeable insight that he has provided me during the course of the work. From that initial meeting right up to the reading of this thesis, he has spent a lot of time through weekly project meetings, progress reports and prompt email replies to all the constant queries of mine. Fernando's involvement with the project - as an advisor, teacher, mentor and friend - has been energetic, total and indispensable. I cannot thank him enough.

I am also grateful to all my lecturers in University of Leicester, especially my co-supervisor, Professor Sarah Spurgeon and my internal examiner, Dr. Michael Pont, for their supervision, guidance and moral support throughout my coursework.

I wish to acknowledge the assistance of electrical laboratory technicians, Rashmikant Patel, Dipak Raval and Bilal Haveliwala for their help and assistance in providing equipment and invaluable technical advice on issues related to this project. I would also like to express appreciation and thanks to other members of the Fault Prediction group, Christos Kitsos, Yanhui Feng and especially John Twiddle, for their guidance and assistance in this project and technical discussions on condition monitoring. Thanks also to many of my friends and colleagues who made it all

worthwhile- Devaraj Ayavoo, Nitin Patel, Andrew Norman, Teera Phatrapornnant, Dr. Michael Short, Mouazz Nahas and Susan Kurian.

The research described in this dissertation was supported financially by an EPSRC studentship (GR/S42866/01). I would like to express my gratitude to EPSRC and the Department of Engineering for allocating me the award. I would like to thank BOC Edwards for the supply of the iGX100 dry vacuum pump which was used as the main test-bed of experimentation. Special mention should go to Jeremy Watson, Larry Marini and Nick Tyler from BOC Edwards for their collaboration on providing us the test equipment and making arrangements for fitting, refitting and transport of pump for fault data acquisition and testing at University of Leicester. I would also like to thank Nick Dowding from Barden Bearings for providing us the whole set of faulty ceramic ball bearings on which our experimental investigations were based on.

Finally, I am greatly indebted to my parents for their moral and financial support they have provided me throughout my years of development and for always believing in me. Last but not least, this thesis would have been impossible without my supportive and caring fiancé, C.K.SenthilKumar. Thank you!

Table of Contents

1. INTRODUCTION	1
1.1 OVERVIEW	1
1.2 ORIGINAL CONTRIBUTIONS	8
1.3 THESIS STRUCTURE	10
2. BACKGROUND	13
2.1 INTRODUCTION	13
2.2 THE DRY VACUUM PUMP INTRODUCED	14
2.2.1 The Advantages of Going Dry	16
2.2.2 Condition Monitoring of Dry Vacuum Pumps.....	16
2.2.3 Ball Bearing Defect Frequencies	18
2.2.4 Defining the Ball Bearing Defect Frequencies	18
2.2.5 Conclusion: The Dry Vacuum Pump Explained and the Ball Bearing Defect Frequencies Calculated.....	22
2.3 THE AR MODEL	23
2.3.1 Reviewing the Basic Principles of AR Modelling	23
2.3.2 Autoregressive Model	23
2.3.3 Autoregressive Moving Average Model.....	26
2.3.4 Advantages of Using the AR Model	27
2.3.5 Disadvantages of Using the AR Model.....	29
2.3.6 Algorithms for Autoregressive Modelling.....	30
2.3.6.1 Yule-Walker.....	30
2.3.6.1.1 Yule-Walker Equations	31
2.3.6.1.2 Levinson-Durbin Recursion.....	32
2.3.6.2 Other Estimation Methods	33
2.3.6.2.1 Burg	33
2.3.6.2.2 Covariance and Modified Covariance	33
2.3.6.3 Reason for Choosing the Yule-Walker Method.....	33
2.3.7 Review of AR Modelling in Condition Monitoring Studies	34
2.3.8 Conclusions: The AR Model Explained	37
2.4 OVERVIEW OF CURRENT VIBRATION BASED DIAGNOSTIC TECHNIQUES.....	38
2.4.1 Time Based Techniques	39
2.4.2 Frequency Domain Techniques	41
2.4.2.1 AR Modelling As a Spectral Tool.....	43
2.4.2.2 AR Versus FFT	44
2.4.3 Time Frequency Analysis	48

2.4.4	Conclusion: Frequency Based Parametric Method of AR Modelling is an Appropriate Method.....	49
2.5	SUMMARY.....	50
3.	PLATFORM OF EXPERIMENTATION	51
3.1	INTRODUCTION	51
3.2	MOUNTING THE ACCELEROMETERS ON THE DRY VACUUM PUMP	51
3.3	ELLIPTIC LOW-PASS ANTI-ALIASING FILTER	52
3.4	AN EVALUATION OF AN ALTERNATIVE IMEMS VIBRATION SENSOR.....	54
3.4.1	Instrumentation	54
3.4.2	ADXL Vibration Signals	57
3.4.3	Resonance	61
3.4.4	Conclusion : ADXL105- A Low Cost Sensor Solution.....	66
3.5	FINDING THE RUNNING SPEED OF THE PUMP	66
3.6	SUMMARY	69
4.	SELECTION OF AR MODEL ORDER.....	70
4.1	INTRODUCTION	70
4.2	AR ORDER SELECTION CRITERIA.....	71
4.2.1	Motivation.....	73
4.2.2	Method of Finding the Optimal Order	74
4.2.3	Results and Discussion	75
4.2.3.1	Finding the Optimal Order.....	75
4.2.3.2	Verification of the Optimal Order.....	78
4.2.4	Conclusion: Optimum AR Model Order Found for Vibration Signals.....	82
4.3	PROPOSITION OF A NEW AND EASY APPROACH TO FIND THE OPTIMAL MODEL ORDER.....	82
4.3.1	Motivation.....	83
4.3.2	Problem Formulation	83
4.3.3	Methodology	84
4.3.4	Results and Discussion	84
4.3.4.1	ADXL Signal at Increasing Rotating Speeds.....	84
4.3.4.2	ADXL Signal Spectrum – Effect of Increasing Frame Length.....	90
4.3.4.3	ACF and PACF Plots for ADXL Signal at Increasing Rotating Speeds.....	93
4.3.4.4	Determination of Order from ACF and PACF plots	94
4.3.4.5	Brüel& Kjær (B&K) Vibration Signals at 100 Hz and 60 Hz	95
4.3.5	Conclusion: A Rule of Thumb for Finding the Optimal Model Order	98
4.4	SUMMARY	99

5. SELECTION OF OPTIMUM AR SAMPLE SIZE	100
5.1 INTRODUCTION	100
5.2 INVESTIGATION ON SELECTION OF OPTIMUM AR SAMPLE SIZE	100
5.2.1 Motivation.....	101
5.2.2 Using the Prediction Error as an Indicator.....	101
5.2.3 Methodology	103
5.2.4 Results.....	103
5.2.4.1 Increasing the Pump Speed	103
5.2.4.2 Increasing the Frame Size	106
5.2.5 Conclusion: Optimum Sample Size is Speed Dependent.....	109
5.3 INVESTIGATION ON IMPACT OF SAMPLING FREQUENCY	110
5.3.1 A Theoretical Explanation	110
5.3.2 Motivation.....	112
5.3.3 Methodology	112
5.3.4 Results.....	113
5.3.5 Conclusion: Length of Record Traversed Back is More Important than Sampling Rate for Model Order Selection in AR Modelling.....	117
5.4 SUMMARY	118
6. RESONANCE DEMODULATION	119
6.1 INTRODUCTION	119
6.2 RESONANCE DEMODULATION EXPLAINED.....	119
6.3 DEMODULATION USING HILBERT TRANSFORM	122
6.4 FINDING THE RESONANCE BANDWIDTH.....	123
6.5 THE NEED FOR RESONANCE DEMODULATION.....	126
6.6 FINDING THE OPTIMUM ORDER FOR THE DEMODULATED SIGNALS	129
6.7 SUMMARY	130
7. DEFINING A FAULT DETECTION TOOL	131
7.1 INTRODUCTION	131
7.2 MOTIVATION.....	131
7.3 PRELIMINARIES: LINKING AR SPECTRA WITH THE AR POLES IN THE Z-DOMAIN.....	132
7.4 LINKING THE AR POLES AND THE CHARACTERISTIC BEARING DEFECT FREQUENCIES.....	137
7.5 AN OVERVIEW OF THE AR POLE TRAJECTORY FAULT DETECTION TOOL	140
7.6 A DIAGRAMMATIC ILLUSTRATION OF THE FAULT DETECTION TOOL.....	142
7.7 SUMMARY	145

8. AR POLE BASED MONITORING.....	146
8.1 INTRODUCTION	146
8.2 TRACKING THE CRITICAL POLE MOVEMENTS.....	146
8.3 TESTING THE EFFECT OF FRAME SIZE	152
8.4 DETERMINING THE THRESHOLD USING ROC ANALYSIS	155
8.5 FRAME SIZE VERSUS INCREASE IN ACCURACY OF CLASSIFICATION.....	157
8.6 CONDITION NUMBER OF CRITICAL POLES.....	160
8.7 AR POLE BASED TRAJECTORY TECHNIQUE VERSUS THE FFT-BASED TECHNIQUE	163
8.8 REAL-TIME COST ESTIMATION OF BOTH THE AR-BASED AND FFT-BASED TECHNIQUES	168
8.9 SUMMARY	172
9. CONCLUSIONS AND FUTURE WORK.....	175
9.1 CONCLUSIONS.....	175
9.2 FUTURE WORK.....	178
REFERENCES	182
A APPENDIX	194
THE WORKING PRINCIPLES OF THE ROOTS AND CLAW DRY VACUUM PUMP	194
FIGURE SHOWING INSTRUMENTATION TO CAPTURE VARIOUS SIGNALS FROM THE IGX100 DRY VACUUM PUMP AND CORRESPONDING INSTRUMENTATION LIST.....	196
A TYPICAL CERAMIC BARDEN BALL BEARING MOUNTED ON THE HV AND LV ENDS OF THE DRY VACUUM PUMP.....	198
PARTS OF A BEARING	199
A BEARING WITH AN INNER RACE CRACK	200
B APPENDIX	201
THE EXPERIMENTAL PLATFORM.....	201
8 POLE ELLIPTIC LOW-PASS FILTER.....	204
FREQUENCY RESPONSE OF LOW-PASS ELLIPTIC FILTER.....	205
CONDITION MONITORING VI.....	206
CIRCUIT FOR THE I MEMS ADXL105 ACCELEROMETER.....	209
FIGURE SHOWING THE ADXL105 PROTOTYPE BUILT	210
BOTH ACCELEROMETERS MOUNTED ON PUMP	211

List of Related Publications & Reports

The following publications have been produced with the author as the main writer during the period of research. The more important results are discussed in the thesis and the author has referred to the publications in the text of the thesis.

Suguna Thanagasundram, Y. Feng, I.S.M. Abou Rayan and F.S. Schlindwein (2005) "*A case study of Auto-Regressive Modelling and Order Selection for a Dry Vacuum Pump*", ICSV12-Twelfth International Congress on Sound and Vibration, Lisbon, Portugal, 11-14 July 2005.

Suguna Thanagasundram, K.R. Gurung, Y. Feng and F.S. Schlindwein (2006) "*AR Pole Trajectory in Condition Monitoring Studies*", ICSV13-13th International Congress on Sound and Vibration, Vienna, Austria, 2-6 July 2006, ISBN: 3-9501554-5-7.

Suguna Thanagasundram and F.S. Schlindwein (2006a) "*Autoregressive based diagnostics scheme for detection of bearing faults*", ISMA2006-International Conference on Noise and Vibration Engineering, Leuven, Belgium, Pages 3531-3546, 18-20 Sept 2006, ISBN: 90-73802-83-0.

Suguna Thanagasundram and F.S. Schlindwein (2006b) "*Autoregressive Order Selection for Rotating Machinery*", The International Journal of Acoustics and Vibration, Volume 11, Number 3, Pages 144-154, ISSN 1027-5851.

Suguna Thanagasundram and F.S. Schlindwein (2006c) "*Comparison of integrated micro-electrical-mechanical system and piezoelectric accelerometers for machine condition monitoring*", Proceedings of the Institution of Mechanical Engineering-Part C, Journal of Mechanical Engineering Science, Volume 220, Number 8, Pages 1135-1146, ISSN: 0954-4062.

Suguna Thanagasundram and F.S. Schlindwein (2006d) "*A Fault Detection Tool Using Analysis from an Autoregressive Model Pole Trajectory*", Journal of Sound and Vibration, Accepted for publication with minor revisions, ISSN: 0022-460X.

Suguna Thanagasundram and F.S. Schlindwein (2006e) "*Finding the required sample size for vibrational analysis using autoregressive modelling techniques*", Submitted to ASME Journal of Vibration and Acoustics, In Review, ISSN 1048-9002.

Suguna Thanagasundram and F.S. Schlindwein (2007). "*A Fault Detection Tool Using AR Pole Trajectory* ", WCEAM CM 2007- Second World Congress on Engineering Asset Management (EAM) and The Fourth International Conference on Condition Monitoring, Harrogate, UK, 11-14 June 2007, ISBN: 978-1-901892-22-2.

Suguna Thanagasundram, Y. Feng and F.S.Schlindwein (2005) "*A Labview based system for condition monitoring of a dry vacuum pump using AR modelling techniques*", Poster presentation at Festival of Postgraduate Research, University of Leicester, 9th June 2005.

Suguna Thanagasundram, K.R. Gurung, Y. Feng and F.S. Schlindwein (2006) "*Exploring pole positions for fault diagnostics from dry vacuum pumps*", Poster presentation at Festival of Postgraduate Research, University of Leicester, 13th June 2006.

List of Figures

2-1	BALL BEARING PARAMETERS	19
2-2	DETECTING BEARING INNER RACE DEFECT FREQUENCY (MCINERNEY AND DAI 2003).	21
2-3	DIAGRAM SHOWING AN AUTOREGRESSIVE PROCESS.	24
3-1	MEASUREMENT OF NOISE LEVEL IN BOTH ACCELEROMETERS.	59
3-2	NOISE FLOOR MEASUREMENTS OF ACCELEROMETERS.	60
3-3	RESONANCE MEASUREMENTS IN TIME DOMAIN.	62
3-4	RESONANCE MEASUREMENTS IN FREQUENCY DOMAIN.	63
3-5	DIFFERENCE BETWEEN THE FREQUENCY SPECTRA OF ADXL AND B&K PIEZOELECTRIC ACCELEROMETER IN LINEAR SCALE.	64
3-6	SET SPEED VERSUS ACTUAL SPEED OF PUMP.	67
3-7	PUMP SPEED SET AT 110 Hz BUT UNDER INCREASING LOADS.	68
4-1	ORDER SELECTION CRITERIA GIVEN AS A FUNCTION OF MODEL ORDER USING THE YULE-WALKER ESTIMATION METHOD.	77
4-2	ANALYSIS OF THE ORDER OF AR MODEL FOR THE VARIOUS ORDER ESTIMATION CRITERIA.	77
4-3	HISTOGRAMS OF "BEST" ORDER.	79
4-4	VERIFICATION OF ORDER PREDICTION CRITERIA USING A "TRUE" AR PROCESS.	80
4-5	EFFECT OF INCREASING SAMPLE SIZE FOR MDL CRITERION.	81
4-6	BEHAVIOUR OF ORDER SELECTION CRITERIA FOR ADXL105 VIBRATION SIGNALS (100 Hz).	86
4-7	BEHAVIOUR OF ORDER SELECTION CRITERIA FOR ADXL105 VIBRATION SIGNALS (80 Hz).	87
4-8	BEHAVIOUR OF ORDER SELECTION CRITERIA FOR ADXL105 VIBRATION SIGNALS (60 Hz).	88
4-9	SPECTRA OF ADXL105 VIBRATION SIGNAL.	91
4-10	DETERMINATION OF ORDER FROM ACF AND PACF PLOTS	94
4-11	VIBRATION SIGNAL FROM BRÜEL& KJÆR (B&K) AT 100 Hz.	96
4-12	VIBRATION SIGNAL FROM BRÜEL& KJÆR (B&K) AT 60 Hz.	97
5-1	EFFECT OF SAMPLE SIZE ON THE VARIANCE OF THE PREDICTION ERROR (PE).	104
5-2	ACF AS A FUNCTION OF SAMPLES.	105
5-3	INCREASING THE FRAME SIZE FROM 250 TO 1000 SAMPLES (IN TWO DIMENSIONAL FORMAT).	107
5-4	AR SPECTRA FOR INCREASING FRAME SIZES FROM 0 TO 2000 SAMPLES (IN THREE DIMENSIONAL FORMAT).	107
5-5	SAME AS FIGURE 5-4, BUT WITH AR MODEL ORDER OF 60.	108
5-6	EFFECT OF INCREASING SAMPLING FREQUENCY ON OPTIMAL AR MODEL ORDER.	111

5-7	PREDICTION ERROR VERSUS MODEL ORDER FOR THREE DIFFERENT PUMP SPEEDS (20 SAMPLES PER REVOLUTION).	113
5-8	PREDICTION ERROR VERSUS MODEL ORDER FOR THREE DIFFERENT PUMP SPEEDS (30 SAMPLES PER REVOLUTION).	114
5-9	PUMP VIBRATION SIGNAL ACQUIRED FROM PUMP RUNNING AT 100 HZ SAMPLED AT INCREASING SAMPLING FREQUENCIES FROM 2 KHZ TO 8 KHZ. PREDICTION ERROR VERSUS MODEL ORDER PLOTTED FOR INCREASING MODEL ORDERS.	116
6-1	STEPS INVOLVED IN THE DEMODULATED RESONANCE TECHNIQUE AND THE CORRESPONDING WAVEFORMS PRODUCED.	121
6-2	BROADBAND SPECTRUM (0 TO 10 KHZ) FOR PUMP WITH NON-DEFECTIVE BEARING.	125
6-3	BROADBAND SPECTRUM (0 TO 10 KHZ) FOR PUMP WITH THE DEFECTIVE BEARING (INNER RACE FAULT).	125
6-4	FFT AND AR SPECTRA FOR ADXL105 VIBRATION SIGNAL.	126
6-5	SAME VIBRATION SIGNAL USED AS FOR FIGURE ABOVE.	126
6-6	OBTAINING THE STATISTICAL PARAMETERS	127
6-7	FINDING THE OPTIMUM ORDER FOR THE DEMODULATED DEFECTIVE VIBRATION SIGNAL.	129
7-1	MONITORING THE PEAKS OBTAINED FROM THE AR MODEL.	134
7-2	POLE ZERO REPRESENTATION.	135
7-3	DEFINING THE PARAMETERS OF AN AR POLE.	136
7-4	THEORETICAL LOCATION OF CHARACTERISTIC DEFECT FREQUENCIES IN THE Z PLANE.	138
7-5	TYPICAL MOVEMENT OF POLE AS DEFECT BECOMES MORE SEVERE (FROM 1 TO 3)	139
7-6	LOCUS OF A PARTICULAR POLE VERSUS EVOLUTION OF TIME.	139
7-7	BLOCK DIAGRAM OF THE PROCEDURE FOR BEARING FAULT DETECTION USING AR POLE TRACKING	144
8-1	DISTANCE OF BPFİ POLE FROM ORIGIN.	149
8-2	POWER OF BPFİ POLES.	150
8-3	ANGLE BPFİ POLE TRAVERSES.	151
8-4	CUMULATIVE AREA TRAVERSED BY THE BPFİ POLES.	152
8-5	DISTRIBUTION OF POLES FROM A PUMP RUNNING IN NORMAL CONDITIONS (RIGHT) AND FOR A CASE WITH A BEARING WHICH HAS AN INNER RACE FAULT (LEFT).	154
8-6	DISTRIBUTION OF NORMAL AND FAULTY POLES POWER PARAMETER.	156
8-7	DISTRIBUTION OF NORMAL AND FAULTY POLES USING DISTANCE FROM THE CENTRE OF THE UNIT CIRCLE AS THE DECISION PARAMETER (2000 SAMPLES).	156
8-8	DISTRIBUTION OF NORMAL AND FAULTY POLES USING DISTANCE FROM THE CENTRE OF THE UNIT CIRCLE AS THE DECISION PARAMETER (5000 SAMPLES).	156
8-9	ROC CURVES FOR INCREASING FRAME SIZES FOR THE DISTANCE OF THE POLES FROM THE CENTRE OF THE UNIT CIRCLE AS A DECISION PARAMETER.	158

Fault Detection Using Autoregressive Modelling Techniques

8-10	SENSITIVITY (A) AND SPECIFICITY (B) CURVES FOR A FRAME SIZE OF 4000 SAMPLES.	159
8-11	DISTRIBUTION OF NORMAL AND FAULTY PEAKS FOR THE FFT-BASED TECHNIQUE PLOTTED FOR FRAME SIZE OF 2000 SAMPLES.	164
8-12	DISTRIBUTION OF NORMAL AND FAULTY PEAKS FOR THE FFT-BASED TECHNIQUE PLOTTED FOR FRAME SIZE OF 5000 SAMPLES.	164
8-13	DISTRIBUTION OF NORMAL AND FAULTY PEAKS FOR THE FFT-BASED TECHNIQUE PLOTTED FOR FRAME SIZE OF 10000 SAMPLES.	164
8-14	ROC CURVES FOR INCREASING FRAME SIZES FOR FFT-BASED METHOD USING THE MAGNITUDE OF THE MAXIMUM PEAK AS A DECISION PARAMETER.	165
8-15	FAULTY PEAK DETECTED AT 555 HZ FOR BOTH AR AND FFT-BASED TECHNIQUES WHEN A FRAME SIZE OF 2000 SAMPLES IS USED.	167
8-16	FAULTY PEAK DETECTED AT 555 HZ FOR BOTH AR AND FFT-BASED TECHNIQUES WHEN A FRAME SIZE OF 5000 SAMPLES IS USED.	167
8-17	FAULTY PEAK DETECTED AT 555 HZ FOR BOTH AR AND FFT-BASED TECHNIQUES WHEN A FRAME SIZE OF 10000 SAMPLES IS USED.	167
8-18	VI SHOWING SEQUENCE STRUCTURE USED IN LABVIEW TO GET EXECUTION TIME OF CODE.	170
A-1	THE ROOT AND CLAW MECHANISM.	194
A-2	CROSS SECTION OF A THREE-STAGE PUMP, WITH ONE ROOTS AND TWO CLAW TYPE STAGES.	194
A-3	ROOT STAGE. REPRODUCED WITH PERMISSION FROM (EDWARDS 1993)	195
A-4	CLAW STAGE. REPRODUCED WITH PERMISSION FROM (EDWARDS 1993)	195
A-5	MONITORING THE VARIOUS PUMP PARAMETERS OF THE IGX100 DRY VACUUM PUMP IN OUR LABORATORY.	196
A-6	BARDEN BEARING CERAMIC BEARING AT HIGH VACUUM (HV) AND LOW VACUUM (LV) ENDS OF IGX100 DRY VACUUM PUMP.	198
A-7	BEARING COMPONENTS-INNER RACE , OUTER RACE AND THE BALL BEARINGS.	199
A-8	CRACK ON INNER RACE OF BEARING. CRACK WAS APPROXIMATELY 2 MM WIDE AND 2 MM DEEP.	200
B-1	SCHEMATIC OF THE COMPLETE DATA ACQUISITION SYSTEM.	202
B-2	SCHEMATIC SHOWING THE 8 POLE ELLIPTIC LOW-PASS FILTER WITH A CUTOFF FREQUENCY OF 10 KHZ.	204
B-3	FREQUENCY RESPONSE OF 8TH ORDER ELLIPTIC FILTER.	205
B-4	USING MATLAB SCRIPT NODES IN THE LABVIEW ENVIRONMENT	206
B-5	ADXL105 ACCELEROMETER CIRCUIT.	209
B-6	ADXL105 PCB WITH SURFACE MOUNT COMPONENTS	210
B-7	ADXL105 AND B&K ACCELEROMETERS MOUNTED ON HIGH VACUUM END OF DRY VACUUM PUMP.	211

List of Tables

2-1	COMPARING AR AND FFT SPECTRAL ESTIMATION METHODS.	47
3-1	PIEZOELECTRIC ACCELEROMETER PERFORMANCE COMPARED TO ADXL105 AS GIVEN IN DATASHEET SPECIFICATIONS.....	57
4-1	WORKING OUT THE THEORETICAL p_{min} VALUES FOR VARIOUS ROTATING SPEEDS AND SAMPLING RATES	84
5-1	STUDYING THE EFFECT OF SAMPLING FREQUENCY BY FIXING THE NUMBER OF SAMPLES PER REVOLUTION AND ADJUSTING THE SAMPLING FREQUENCY UNTIL THE SAME NUMBER OF POINTS ARE OBTAINED PER SHAFT REVOLUTION FOR THE THREE DIFFERENT PUMP SPEEDS.....	113
5-2	WORKING OUT THE REQUIRED NUMBER OF SAMPLES PER COMPLETE REVOLUTION OF SHAFT ROTATION AT EACH OF THE PUMP SPEED AND SAMPLING FREQUENCY FOR THE 100 HZ PUMP VIBRATION SIGNAL.....	115
8-1	CONDITION NUMBERS FOR INCREASING FRAME SIZES FOR NON-FAULTY AND FAULTY DATA OBTAINED USING THE 'COND' FUNCTION IN MATLAB.	162
8-2	BENCHMARKING THE AR AND FFT METHODS FOR 3 DIFFERENT FRAME SIZES.	170
A-1	INSTRUMENTATION LIST OF THE IGX100 DRY VACUUM PUMP.....	197

Abbreviations

A/D	Analog to Digital
AC	Alternating Current
ACF	Autocorrelation Function
ADC	Analog to Digital Converter
ADXL105	A type of surface micromachined iMEMs accelerometer of Analog devices
AI	Analog Input
AIC	Akaike Information Criterion
AO	Analog Output
AR	Autoregressive
ARMA	Autoregressive Moving Average
BK	Brüel & Kjær
BOC	British Oxygen Company
BPFI	Bearing Pass Frequency Inner Race
BPFO	Bearing Pass Frequency Outer Race
BSF	Ball Spin Frequency
CAT	Criterion AR Transfer
COTs	Commercial-Of-The-Shelf
FFT	Fast Fourier Transform
FIC	Finite Information Criterion
FPE	Final Prediction Error
FTF	Fundamental Train Frequency
HFRT	High Frequency Resonance Technique
HV	High Vacuum
IC	Integrated Chip
iGX	A type of pump in the BOC Edwards pump series
IIR	Infinite Impulse Response
iMEMs	integrated Micro Electrical Mechanical system
LPE	Linear Prediction Error
LV	Low Vacuum
LVM	LabView Measurement
MA	Moving Average

MDL	Minimum Description Length
MEM	Maximum Entropy Method
MPAR	Minimum Phase Autoregressive
MSE	Mean Square Error
NI	National Instruments
NMPAR	Non Minimum-Phase AR
PACF	Partial Autocorrelation Function
PCB	Printed Circuit Board
PE	Prediction Error
PSDE	Power Spectral Density Estimation
PSDs	Power Spectral Densities
PVD	Physical Vapor Deposition
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
RPM	Revolutions Per Minute
SAB	Single Amplifier Biquad
SNR	Signal to Noise Ratio
STFT	Short-Time Fourier Transform
UCA	Uncommitted Amplifier
VI	Virtual Instrument
WVD	Wigner-Ville Distribution
YWE	Yule-Walker Equations

1. INTRODUCTION

The research work described in this thesis was intended as a vehicle to develop further the theory associated with the traditional diagnostic techniques applied to rotating machinery such as time domain methods (Alfredson and Mathew 1985b), cepstral analysis (Zheng and Wang 2001), also known as cepstrum analysis, in which the spectrum of a logarithmic spectrum is calculated, envelope detection techniques (Randall, Antoni *et al.* 2000), wavelet transforms (He, Zhao *et al.* 1996) and vibration spectrum analysis (Mathew and Alfredson 1983; McInerny and Dai 2003) to predict failure in such machines. The main motivation for this work was the development of a simple and efficient fault detection tool based on vibration measurements for the dry vacuum pump.

In this chapter, an introduction to the work carried out in this research is given, covering an overview of the thesis itself and the goals and scope of the research are presented. The main contributions of the thesis are clearly listed. The key invention, which is a vibration based fault detection tool based on the tracking of AR pole movements in the complex z domain, is briefly described. The organization of this thesis is given at the end of the chapter. Many of the subjects mentioned in this chapter will be discussed in more detail in later chapters.

1.1. Overview

In many industries worldwide today, unplanned downtime as a result of machinery failure is a costly and time-consuming affair affecting the workers' safety, production efficiency and return on investments. There is a need to find machine faults at an early stage and one way of doing this is to use condition monitoring (Isermann 1995; Isermann and Balle 1997; Nandi and Toliyat 1999). Condition monitoring is the process of tracking the health of a machine by monitoring characteristic parameters or variables of the machine such as vibrations, acoustic emissions, pressure measurements, temperature and electrical properties (i.e., current and voltage measurements) to detect existing or developing faults in the components of the machine. A fault is defined as an unallowable deviation of at least one of the characteristic properties of the machine. Changes in the condition of the machine

manifest themselves as increased vibrations, higher temperatures, louder audible sounds, increased acoustic emissions transmitted in the structural components of the machine and higher currents drawn by the motor of the machine. Condition monitoring should be inexpensive, robust and simple to use. The main motivation of this work was to find a simple and easy condition monitoring tool to detect the development of faults in a dry vacuum pump. The particular machine that was used as the test bed in this project belongs to an important class of rotating machines not previously studied in this context. It is the BOC-Edwards iGX dry vacuum pump which is part of BOC's so-called 'dry' vacuum pump range.

There are many ways to classify machines but for the purpose of this work, it is convenient to consider two broad classes, respectively representing machines designed to mostly run with variable load and speed and machines designed to mostly run with fixed load and speed. A car engine is a good example of the former and a vacuum pump, of the latter. Hence a "steady state" machine is defined to be a machine in which all system values attain constant values at "steady state" operation (Esch 2002). It is a known fact that "steady state" machines have to be designed to withstand the effects of transients, such as starting loads, but it is not primarily these issues that concern us here. From the perspective of this research, the important point is that, for most of the operating time, the measurable variables will not be dynamically changing and, hence, any degradation in machine performance needs to be inferred from slow drifts in nominally constant values and/or in the statistics of "stationary" signals such as vibrations. Many rotating machines, such as pumps in vacuum systems, are, in contrast, expected to work in "steady state" most of the time. Predictable transient loads are still present in these situations but they are present for a much smaller fraction of the time. Unpredictable transients are often very brief. Both types of transient phenomena are promising sources of diagnostic information for fault detection in rotating machinery but the challenge identified as the main research problem in this thesis lies in looking for small and subtle changes in the "steady state" operation of the dry vacuum pump. The dry vacuum pump can be taken to be an example of a "steady state" machine as, during most of its operating time, it runs at a constant speed, producing signals of constant mean and variance, except at certain operating conditions where the speed fluctuates due to variations in the load and

during start up and shut down conditions, when it produces transient signals. The dry vacuum pump is hence defined as a type of “quasi-steady state” rotating machinery.

The main objective of the work is the complex challenge of fault detection in a “quasi-steady state” rotating machine. The dry vacuum pump can be made to operate at fixed rotating speeds and at constant operating conditions producing “quasi-steady state” signals. By analysing the various signals produced by the pump such as pressure, temperature, sound and inverter current (drawn from the 3 phase 2 pole AC asynchronous induction motor driving the pump) and particularly the vibration signals (Mathew 1989; Sabin 1995; Tang, Tan *et al.* 2001), faults can at first be detected and by studying the behaviour of the signals as faults develop (Mathew and Alfredson 1984; Mathew and Alfredson 1986; Mathew, Kuhnell *et al.* 1987b; Zhuge, Lu *et al.* 1990), faults can then be predicted at a later stage, if fault patterns can be established.

Fault detection and prognosis of equipment is an established technique in the prevention of costly machinery failure in many industries (Mathew, Kuhnell *et al.* 1987a; Yang, Mathew *et al.* 2005; Zhang, Mathew *et al.* 2005). The equipment monitored are different types of rotating machines and the reliability, availability and maintainability of these machines are critical to the overall performance and operation of the industries to maintain their competitiveness in a global marketplace. Some examples of these type of “quasi-steady state” rotating machines are piston pumps, rotary pumps, screw pumps, motors, engines, compressors and fans (Peck and Burrows 1994; Chen, Du *et al.* 1995). The various types of faults that can be monitored from these rotating machines are misalignment, mechanical looseness, gear faults, bearing faults, cavitation problems and lubrication deficiencies (Stammers 1989; He, Zhao *et al.* 1996). It is becoming more and more important to receive early warning of any problem before failure and outage occurs.

The dry vacuum pump is itself of high value (costing about £10,000) and, since it supports a vacuum system involved in silicon chip manufacture, the secondary effects of breakdown are often expensive. This type of pump is prevalently used in the wafer fabrication process in the semiconductor industry for applications such as lithography, Physical Vapor Deposition (PVD) process and silicon and metal etching and implanting sources (Bachmann and Kuhn 1990) and it has been estimated that

pump failure can contribute to significant loss of valuable products e.g. loss of wafer batches in excess of \$100,000 (Troup and Dennis 1991). Furthermore, loss of production time whilst the pump is repaired or replaced incurs additional costs to the operator. Therefore continuous condition monitoring of the pump is desirable to allow scheduling of maintenance or the replacement of a pump without affecting production. Dry vacuum pumps are not only economically important in their own right (British export value of several hundred million pounds) but they are also representative of a much wider class of similar machines which run mostly in “steady state” but have unpredictable failure modes. This machine is, therefore, considered to be a good exemplar process to work on and these requirements provide the motivation for the development of diagnostic techniques which can be used to predict impending failure of dry vacuum pumps and other similar machinery.

The condition monitoring and fault diagnosis scheme that will be developed aims to detect possible faults on different subsystems on the dry vacuum pump before these faults become serious. One subsystem under consideration is the set of bearings of the dry vacuum pump. Bearings are a very common component in industrial machinery and their importance in efficient and effective process operation has generated considerable interest in the field of condition monitoring.

Monitoring bearing vibration in a pump system is highly cost-effective in minimizing the pump downtime, both by providing advance warning for appropriate actions to be taken and by ensuring that the system does not deteriorate to a condition where emergency action is required. Bearing defects manifest themselves as either excessive wear or damage in the rolling ball elements as well as in the inner/outer races of the bearings (McInerny and Dai 2003). Fault identification of ball bearing-related phenomena using condition maintenance techniques has been the subject of extensive research for the last two decades (Sawalhi, Randall *et al.*; McFadden and Smith 1984a; McFadden and Smith 1984b; McFadden and Smith 1985; Tandon and Choudhury 1999; Choudhury and Tandon 2000; Ho and Randall 2000; McFadden and Toozhy 2000; Randall 2001; Randall, Antoni *et al.* 2001; Randall 2004; Antoni and Randall 2005; Harsha 2006a; Harsha 2006b; Tandon, Ramakrishna *et al.* 2007; Tandon, Yadava *et al.* 2007). One of the possible approaches to fault monitoring of the bearings is the processing of mechanical vibration signals obtained from the external

housings in which the bearings are mounted for extraction of diagnostic features. This technique is more commonly known as vibration signature analysis and there are many conventional procedures based on time harmonic and power spectrum analysis that have shown considerable success in detecting the presence of failures in the machines' components even at an incipient stage (Smith 1982; Alfredson and Mathew 1985b; Alfredson and Mathew 1985a; Hoffman and van der Merwe 2002; Du and Yang 2007; Sheen 2007; Zarei and Poshtan 2007).

This work has been limited to the study of bearing faults in the dry vacuum pump but can be easily extended to the study of other types of mechanical faults in the pump. Discussed in the thesis is the use of vibration measurements. The vibration signal was chosen as the principal signal for investigation as it was known as to be a good indicator of rolling element faults (Tandon and Choudhury 1997). The analysis, however, can be done with other types of "steady state" stationary signals such as acoustic emissions and sound.

The main investigation was validated using vibration signals obtained from the BOC Edwards iGX dry vacuum pump. This pump has a single row of deep groove ceramic bearings at both the low and high vacuum ends. If faults were to develop in the bearings, these can be identified from vibration signals as the signals obtained from the faulty conditions are different from those obtained in the non-faulty conditions.

Many methods currently used in the analysis of spectral properties of mechanical vibrations and acoustic emissions produced by rotating machines (Venkatesan GT, Danlu Zhang *et al.* 1999; Andrade, Esat *et al.* 2001) are soundly based in knowledge of the physical structure of the machine. However there is much about key stages in such methods that is arbitrary (or empirically derived), indicating scope for optimisation and the need for them to be used in combination with other methods. Amongst the different time and frequency domain based diagnostic techniques, vibration analysts often rely on the Power Spectral Densities (PSDs) of vibration data to monitor the health of moving parts of machinery. The PSD spectral components of the vibration signatures allow the identification of several types of faults. Common failures such as bearing faults and gear problems can be detected by

trending major frequency components and their amplitudes. The PSD can also be used to identify other types of vibration-related faults such as imbalance, misalignment, cavitation problems, shaft imbalance and lubricant deficiency (Mathew 1987) but our studies will concentrate mainly on identifying bearing faults.

Most of the PSD methods used are widely based on the FFT (Fast Fourier Transform) technique. However, the FFT-based spectrum analysis method does suffer from some shortcomings. One of the major setbacks of the spectral estimation technique is that a large number of frequency components have to be monitored due to the complexity of the system. A standard approach in evaluating the power at a particular frequency implies the computation of the whole spectrum first and then estimation of the power of a particular frequency of interest. Another concern of the FFT technique is that a large enough sample size has to be used for the spectral estimation for reasonable resolution capabilities as the resolution of the FFT is inversely proportional to the frame size utilized (S. Lawrence Marple Jr. 1987). This might not be appropriate in real-time applications. The major shortcomings of the FFT-based methods are discussed in detail in Section 2.4.2 of the thesis.

An interesting statistical signal processing alternative is to evaluate directly the frequencies of interest. In this case only those frequencies have to be estimated instead of the whole spectrum. This provides a reduction in computing time and effort, facilitating real-time estimation. In this study, characteristic bearing defect frequencies are extracted from the pole frequencies of a parametric time series AR model (Kay and Marple 1981; Broersen 2006). The AR model is used to decompose the signals into a set of poles which have a correspondence to the peaks of the signal's PSDs. Using the AR estimation technique, it is not necessary to obtain the whole spectrum. Instead, the evaluation of the pole frequencies of interest from the derived AR parameters would suffice, as AR modelling allows spectral decomposition. This often just involves the calculation of the AR coefficients and the variance of the input vibration signal (Mainardi, Bianchi *et al.* 1995). Small order AR models can efficiently estimate the pole frequencies which correspond to the poles of the bearing defect frequencies.

The technique of AR estimation has been very widely used as a spectral analysis tool in geophysical (Kay 1988; Maiwald, Dalle Molle *et al.* 1993) and

biomedical applications (Schlindwein and Evans 1989; Keeton and Schlindwein 1998; Boardman, Schlindwein *et al.* 2002; Kelly, Burke *et al.* 2002), forecasting in economic studies and linear prediction models in speech coding (Oppenheim and Schaffer 1975; S. Lawrence Marple Jr. 1987; Marcek 2000; Prasolov 2004; Ishizuka, Kato *et al.* 2005; Batina, Jensen *et al.* 2006; Gaubitch, Ward *et al.* 2006) but its application in fault prediction studies is less common. Previous studies have reported the use of the AR technique as a model based linear one-step ahead predictor of the signal (McCormick, Nandi *et al.* 1998; Wang 2001c) in fault detection studies but pole-related spectral decomposition and tracking the trajectory of ‘critical poles’ for fault detection purposes is a new method. The AR technique also only requires a fraction of the samples that are required by the FFT method for the same resolution (Kay and Marple 1981). When compared to the traditional FFT method, the resolution of the AR technique is higher due to its implicit extrapolated autocorrelation sequence. This means that smaller sample sizes can be used for PSD estimation. The main limitation of the FFT method is that it does not work well for short data records and has a limited frequency resolution. The AR technique can overcome these limitations by allowing the use of smaller sampling rates and sample sizes while achieving a superior spectral resolution compared to the FFT methods. The choice of technique to be implemented has been driven by the aim of real-time implementation, as the ability to produce spectral estimates from short segments of data is important for the analysis of fault detection of vibration signals from the dry vacuum pump. The key novelty of the work was the development of a tool in software using AR modelling techniques that can detect the onset of incipient single-point bearing defects from the vibration data of the dry vacuum pump based on the AR pole positions in the z domain.

In summary, these are key issues this thesis will try to address. When there are many Commercial-Of-The-Shelf (COTS) products available which can be used as spectral analysis tools for frequency estimation of signals for condition monitoring purposes, why is it useful to employ an alternative spectral estimation tool using AR modelling methods? Is spectral estimation the only way fault detection can be done using AR modelling? What benefits are attained by the proposed method of tracking the AR pole trajectory movements in the z domain as opposed to existing techniques?

The current methods used in fault detection are first reviewed and then the basic principles of AR modelling are explored in order to use it effectively for fault detection. Construction of a fault detection tool that can be systematically used for classification and quantification of bearing faults based on vibration data from a type of “quasi-steady state” rotating machine (the dry vacuum pump) is considered. In the latter half of the thesis, the AR fault detection tool is applied to real data obtained from the iGX dry vacuum pump to show its effectiveness in detecting an important class of bearing fault, which is the inner race fault.

1.2. Original Contributions

Applying AR modelling techniques for fault detection studies in a “quasi-steady state” machine has resulted in the following original contributions:

- [1] This work is the first to introduce the concept of tracking of pole movements in the complex z domain approach for vibration based fault detection studies. One of the objectives of the research is to study the behaviour and stability of the poles as fault conditions develop in the pump. From the position of the poles inside the unit circle, classification and quantification of the main spectral peak of defect frequencies can be easily performed, leading to the possibility of having frame to frame monitoring of spectral parameters of interest. To test the efficacy of the scheme, the AR pole trajectory tool was applied to increasing frame sizes of vibration data captured from a dry vacuum pump in the laboratory and the performance of the classification scheme was tested with Receiver Operating Characteristic (ROC) analysis. The proposed method achieved high sensitivity and specificity rates. This computational method is very attractive for condition monitoring applications as it provides a more immediate comprehension of the spectral process characteristics when expressed in terms of poles and AR spectral components. The pole representation facilitates easier understanding of the spectral characteristics of the process because of the one-to-one correspondence between the poles and the AR spectral peaks. This main original work presented in the thesis has been published in (Thanagasundram, Gurung *et al.* 2006), (Thanagasundram and Schlindwein 2007) and (Thanagasundram and Schlindwein 2006d).

[2] As part of the research, an investigation was also conducted to compare the performance of two types of vibration sensors, a piezoelectric accelerometer and an integrated Micro Electrical Mechanical system (iMEMs) surface micromachined accelerometer, to explore alternative cost-effective solutions for acquiring reliable vibration data for diagnostics. Surface micromachined accelerometers are a new technology and their usage for vibrational analysis has been conservative due to concerns over increased noise levels and tolerance to high temperatures. It has been shown that such concerns can be allayed and that surface micromachined accelerometers can be an effective low cost high-quality alternative for machine condition monitoring. This second original contribution has been published in (Thanagasundram and Schlindwein 2006c).

[3] Some interesting investigations on finding out the optimum AR model order, correct AR sample size to be used and determination of the appropriate estimation method to be used in the context of the study of application of vibration data for fault detection in the dry vacuum pump were done. As a result of the research conducted, a new rule of thumb was also proposed for determining the minimum model order for Autoregressive-based spectrum analysis for rotating machinery. These original contributions have been published in (Thanagasundram, Feng *et al.* 2005), (Thanagasundram and Schlindwein 2006b) and (Thanagasundram and Schlindwein 2006e).

In summary, a fault diagnosis system in which AR modelling techniques are used as the principal tool for spectral estimation has been established. This way of fault detection and classification is new, as opposed to the more traditional way of conducting vibrational analysis, where FFT analysis is used for transformation of vibration data from time to frequency domain. The work on application of AR modelling techniques to fault detection has resulted in a novel concept of using the trajectory of the pole locations as a fault classification tool. This method has great potential in terms of its applicability to fault detection and fault prediction as the signal space is reduced tremendously and this aids fault classification. This new method has interesting potential applications in automatic system diagnosis, prognostics-related systems and condition monitoring and forms the main core of the work done, resulting in an original contribution.

1.3. Thesis Structure

The remaining part of this thesis is organized as follows:

Chapter 2 presents the major bodies of theories covered in this thesis, including the background research conducted to understand the research problem as a whole and its related literature. This chapter can be broken into three main sections.

In the first section, the working mechanics of the dry vacuum pump are introduced. The ball bearings of the dry vacuum pump upon which the methodology introduced in Chapter 3 was applied to obtain data for most of the analysis presented in this thesis are introduced and explained. The ball bearing defect frequencies are also defined.

The second section explains the principles of AR modelling. Understanding the basic fundamentals of AR modelling underpins the link between AR modelling and the proposed fault detection tool described in the later chapters. The theory of the various estimation methods such as the Yule-Walker and Burg and their variations are presented. Justification on the selection of the choice of estimation method of Yule Walker is given. The use of AR modelling techniques in condition monitoring studies is also reviewed.

The third section reviews the theory and practice of alternative vibration-based diagnostic techniques currently used in industry. The FFT method, which is a popular spectral analysis technique, is introduced. The boundaries and limitations of both AR and FFT methods are considered.

Chapter 3 describes the methodology adopted to solve the research problem presented in the thesis. The chapter gives an overview of the experimental setup, the data acquisition system and the sensors used in acquiring the vibration data from the dry vacuum pump. The performance of two types of vibration sensors, the traditional piezoelectric accelerometer and an integrated Micro Electrical Mechanical system (iMEMS) surface micromachined accelerometer were evaluated to explore alternative,

cost effective solutions for acquiring reliable vibration data for diagnostics. The work in this chapter has been published in (Thanagasundram and Schlindwein 2006c).

Chapter 4 introduces the notion of AR order selection criteria. This chapter can be broken down into two main sections.

The first section presents a thorough investigation conducted to determine the optimal AR model order for the vibration data obtained from the dry vacuum pump using various formal order selection criteria. Numerical examples are given and the need for the estimation of the optimum order is validated. The work in this chapter has been published in (Thanagasundram, Feng *et al.* 2005).

The second section is an extension of the work done in the first section and presents a practical, simple rule for determining the minimum model order for autoregressive-based spectrum analysis for rotating machinery based on key findings arising from the author's original work. The work in this section has been published in (Thanagasundram and Schlindwein 2006b).

Chapter 5 investigates two important facets in the implementation of the AR modelling. They are the choice of the sampling size and sampling frequency to be used in the implementation of AR modelling. This chapter can be broken down into two main sections.

The first section analyses the need for selection of an optimal sample size for AR modelling for our studies. Finding the optimum sample size required to encapsulate the vibration signal's behaviour for a given model order is important in the development of AR fault detection tool introduced in the later chapters and this aspect is investigated in this section. The work in this chapter has been published in (Thanagasundram and Schlindwein 2006e).

The second section conducts an investigation to find out what is the optimal sampling frequency to be used. First a qualitative explanation is given on what is the effect of increasing the sampling frequency on the selection of AR model order. Then

this explanation is justified with actual analysis conducted on data obtained from the dry vacuum pump at three different pump speeds..

Chapter 6 discusses resonance demodulation technique with particular respect to application of Hilbert transform. An explanation of why the application of the resonance demodulation enhances the chance of successful fault detection is given. The work in this chapter has been published in (Thanagasundram and Schlindwein 2006a).

Chapter 7 formalizes the notion of the AR fault detection tool and explains the preliminaries. The novel concept of tracking pole movements in the complex z domain for vibration based fault detection studies is clarified.

Chapter 8 presents results obtained with the AR pole-based fault detection tool described in Chapter 7. Numerical examples are given with real vibrational data from the dry vacuum pump. The performance of the tool for increasing frame sizes, validated using ROC analysis, is also presented. An investigation was also conducted to explore the stability of the poles for faulty and non-faulty data through the calculation of condition numbers. The work in this chapter has been published in (Thanagasundram, Gurung *et al.* 2006), (Thanagasundram and Schlindwein 2006d) and (Thanagasundram and Schlindwein 2007).

The performance of the proposed AR pole-based trajectory tool is compared with the FFT-based method for fault detection of data obtained from the dry vacuum pump. The computational costs of both methods are benchmarked and it is justified why the AR-based method is more favourable for real-time implementation.

The last chapter summarizes the work done and gives some concluding remarks.

2. BACKGROUND

2.1. Introduction

This chapter presents a brief introduction to the dry vacuum pump, which serves as the example of the “quasi-steady state” rotating machine in the practical component of the work. The working principles of the dry vacuum pump are described. The ‘Root and Claw’ mechanisms are explained in detail and the advantages of using dry vacuum pumps are presented. A detailed discussion of why condition monitoring of dry vacuum pumps is important in industry is also presented. The ball bearing defect frequencies which are relevant to the studied research problem are defined and calculated in Section 2.2.3.

This chapter also outlines the theory of AR modelling that is going to be implemented in the following chapters. The text is not intended to give an exhaustive description of the AR technique, which has now been documented in numerous books (Jenkins and Watts 1968; S. Lawrence Marple Jr. 1987; Kay 1988; Box, Jenkins *et al.* 1994). Instead, a brief review of the key formulae is given here, with a view to their use as part of a pre-processing fault detection tool for vibrational analysis of the dry vacuum pump.

In the Section 2.4, an overview is also given of the extensive research done in the area of alternative vibration-based diagnostic techniques. While these techniques are established and have been proved to work effectively in the area of fault detection, especially the detection of ball bearing faults, each method does have some drawbacks. None has targeted the problem from a real-time aspect as the diagnosis of the health of dry pumps in real-time is an important objective in the context of the problem studied. The approach used here has been to develop a real-time online fault condition monitoring tool for the dry vacuum pump right from the onset of the beginning of the research.

2.2. The Dry Vacuum Pump Introduced

A dry vacuum pump is a positive displacement mechanical rotary pump that can attain a vacuum without the use of fluid lubricants or sealant between rotors and stator in the pumping chamber (Duval 1989). Hence it is also known as the “oil-free” pump. Dry vacuum pumps can be based on different principles: the multistage Roots blowers, screw compressors, multistage dry pistons, turbomolecular drag pumps, the scroll type and the multistage Claws (Vogelsang, Verhulsdonk *et al.* 1999; Akutsu, Matsuoka *et al.* 2000; Lessard 2000). The BOC Edwards dry pumping technology (Edwards 1993) is based upon successive pairs of non-contacting pump rotors which transfer gases from the process to the atmosphere. The particular BOC Edwards iGX pump utilised in this project is a modular multistage pump that has one stage of Roots and four stages of Claws (refer to Figure A–1 and Figure A–2 in APPENDIX A). This pump is classified as a high vacuum pump as it can attain an ultimate vacuum level of up to 0.005 mbar (a pump is termed a high vacuum pump if it can attain base pressure lower than 0.1 mbar). Different models of the pump can have different pumping capacities and the model that we use has a peak pumping speed of $100 \text{ m}^3\text{h}^{-1}$. The typical construction of the pump consists of a pump housing having an inlet and an outlet for the passage of gas to be pumped, 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 and ball bearings at the inlet and outlet to support the rotor shafts.

In the iGX dry vacuum pump, there is one stage of a pair of two lobed rotors in the form of figure 8 and they are known as ‘Roots’ (Figure A–3 in APPENDIX A). The Roots are valveless positive displacement devices which interlock, synchronize and rotate in opposite directions. Very large displacement speeds can be obtained and the Roots have high throughput (Wycliffe 1987a). The Roots mechanism is very effective in the pressure region 10 to 10^{-3} mbar but only when delivering against low pressure differentials. Hence the Roots mechanism has a low compression ratio but a high throughput. The Root stage gives four compressions per revolution.

The iGX pump also has 4-stages of self-valving rotors (Figure A–4 in APPENDIX A) that are shaped like Claws at the outlet stage. Claws are true

compressors. The claw mechanism achieves high compression ratios but because of its self-valving action, flow is more restrictive and it has a lower throughput. Claws deliver efficiently against high pressure differentials. The four pairs of claw shaped rotors on the two shafts rotate in opposite directions to trap and compress the process gas that flows along the axis of the shafts between the claw pairs. The claw stage gives two compressions per revolution.

The iGX pump also features BOC Edwards patented feature (Edwards 1993), reverse claw arrangement (refer to Figure A–1 in APPENDIX A). This means that 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 first stage and creates a short direct gas path that allows direct transfer of gas from one stage to the next, through a common port in the interstage partition. The advantage of having this feature is that, in dusty applications, dust contained in the gas rushing through the pump is more likely to pass through the outlet rather than drop out and accumulate between the pump stages, thus preventing blockages (Troup and Turrell 1989).

By combining both the Roots and claw mechanisms, optimum performance has been achieved in the iGX dry vacuum pump. The specific design achieves high compression ratios by having several sets or stages of rotors in series. The reduced pumping speed performance of the claw design is improved by substituting a Roots stage at the pump inlet, the compression ratio of the Roots mechanism being more efficient in the lower pressure regions. This is the reason why the iGX pump has one stage of Roots at the inlet and four stages of Claws at the outlet.

The iGX dry vacuum pump normally runs hot since it does not have a cooling fluid to remove the heat generated by the compression of the gas. Repeated compressions that take place in the carry over volume can raise the temperature of the pump well above 100°C. The running temperature of the pump is regulated by an external cooling water jacket with an adjustable thermostatic valve which prevents the pump from overheating (refer to Figure A–2 in APPENDIX A).

2.2.1. The Advantages of Going Dry

The dry vacuum pump has been successfully used in semiconductor manufacturing industries and clean room environments (Troup and Turrell 1989; Konishi and Yamasawa 1999; Davis, Abreu *et al.* 2000) where corrosive gases are abundant, temperatures are high and effluents resulting from the processes can cause seizure of the pumps. Oil sealed rotary pumps have certain disadvantages in such conditions and could not operate continuously in such hostile and aggressive environments, even with expensive modifications such as use of filters, traps, forced lubrication and use of inert corrosive resistant materials for the construction of the pump housing (Maurice, Duval *et al.* 1979; Bez and Guarnaccia 1990). With the advent of oil-free pumping technology, dry vacuum pumps have become an effective and reliable alternative for such applications (Troup and Dennis 1991). Dry vacuum pumps require no internal lubrication and have non-contacting rotors to pump the process vapours; hence they have no rubbing components to wear. Since they do not use oil, there is no oil back migration and reduced atmospheric pollution (Wong, Laurenson *et al.* 1988). Hence dry vacuum pumps offer a great advantage in comparison to oil-sealed pumps, where contamination of the process and environmental constraints are problematical. The dry vacuum pumps can also easily handle the toxic vapours since corrosion only takes place in presence of moisture. Also the heat produced by the pump minimizes the deposition of condensable solids used in the low pressure chemical vapour deposition and metal etching processes, as it encourages the condensable products to remain in the vapour state (Edwards 1993).

2.2.2. Condition Monitoring of Dry Vacuum Pumps

The aggressive medium in which the pump operates means that, although the key factors which affect pump reliability are internal, instrumentation located within the flow paths of the pump and connecting pipe-work has a short life expectancy and low reliability (Zakrzewski, May *et al.* 1988). Very little work has been attempted in the vacuum community for a systematic and comprehensive approach in studying condition monitoring of dry vacuum pumps. The only published work relating to the

fault detection of dry vacuum pumps has been conducted by the following authors (Sabin 1995; Lim, Cheung *et al.* 2004), however the results presented in these studies are very preliminary.

Any proposed diagnostic system must be capable of accurately and reliably inferring faults from a set of instruments of a type which can be located externally. A pump with a proven, cost-effective self-diagnostic capability such as that described, will be desirable for potential customers and therefore of commercial advantage to BOC-Edwards. Some work has been attempted to evaluate the running conditions of the pump (Figure A-5 in APPENDIX A). Other signals have also been monitored, apart from the vibration signals, for the development of the condition monitoring system. The figure shows an illustration of how temperature sensors and the coolant water flow are monitored on the iGX100 pump. Also there are other devices used to measure exhaust gas pressure, vacuum pressure, air mass flow rates, gear speed and pump motor inverter current and voltage measurements. Table A-1 in APPENDIX A gives a detailed description of the transducers utilized and their respective mounting positions on the pump.

Whilst the design of dry vacuum pumps has remained relatively constant over the last ten years, more emphasis is being employed in the evolution of intelligent pumping systems with embedded sensors and self-diagnostic capabilities as dry vacuum pumps have become a critical part of the wafer manufacturing process (Hablanian 1981; Hablanian 1986; Duval 1987; Wycliffe 1987b; Hablanian 1988; World Pumps 2002). Pump failure can be a costly process. Also some customers are affected by the inconvenience of down time. Hence high reliability and availability have become primary customer requirements. Early detection of incipient faults is vital to avert costly failures. Today, in most wafer fabrication plants, vacuum pumps are fitted with integrated sensors and are controlled by microprocessors that monitor the status of the health of the dry vacuum pumps (World Pumps 2006). By trending and processing the signals by condition monitoring software, early symptoms of pump degradation and malfunction can be predicted. Smart vacuum pump condition monitoring algorithms can deliver advanced warnings of problems that can avoid costly equipment failures and unscheduled breakdowns (Berges 1987; World Pumps 1999; Myerson 2000; World Pumps 2005). Such software can also help in the pump

maintenance. It is also the current drive among dry vacuum users for the preference of online condition monitoring and remote real-time condition monitoring as this can save manpower and can establish connectivity by enabling many dry pumps to be monitored at the same time over the internet.

2.2.3. Ball Bearing Defect Frequencies

The purpose of this section is to define the characteristic ball bearing defect frequencies. The dry vacuum pump being like other rotating machinery has ball bearings at the inlet and outlet ends. By studying the frequency spectra of the vibration signature from the dry vacuum pump, bearing failures that can cause pump failure can be detected.

Spectral lines of the spectra produced by frequency-based techniques described above are correlated with the characteristic bearing frequencies if there are any defects present (Wang 2006). The degree of correlation is monitored and, if amplitude of the peak of a characteristic bearing frequency exceeds a threshold value defined by specific reference standard or baseline spectrum obtained from a pump running in normal no-fault conditions, a bearing defect is identified and diagnosed. The diagnosis can identify the exact location of the defect bearing as well as specify which component of the bearing is defective.

2.2.4. Defining the Ball Bearing Defect Frequencies

Rolling-element bearings generally consist of two concentric rings, namely an inner ring and an outer ring, between which a set of balls or rollers rotate in raceways and the cage (refer to Figure 2-1). The five bearing parameters that must be known to calculate the bearing defect frequencies are, B_D - the ball or roller diameter, P_D - the pitch diameter or cage diameter, N_B - the number of rolling elements, α - the contact angle and f - the shaft rotational frequency. The contact angle of the bearing is defined as the angle between a plane perpendicular to the bearing axis and a line joining the two points where the rolling element makes contact with the inner race and outer race.

Formulae have been developed to calculate bearing defect frequencies for every bearing geometry, inner raceway, outer raceway and rolling elements (Tandon and Choudhury 1997). These characteristic bearing defect frequencies, which are related to the raceways and the balls or rollers, can be calculated once the bearing dimensions and the rotational speed of the machine are known.

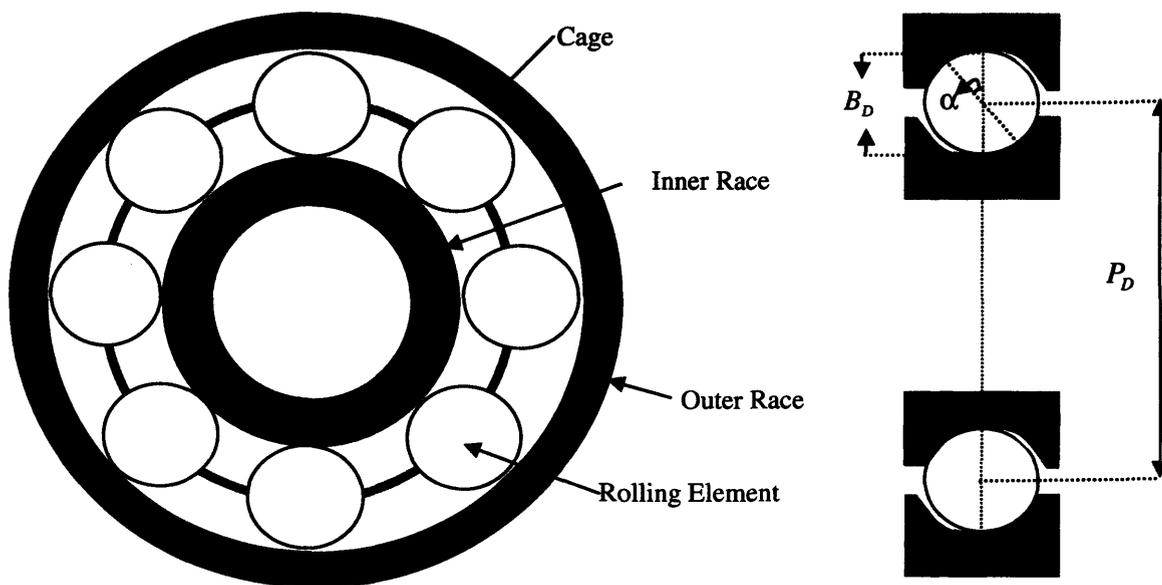


Figure 2-1 Ball Bearing Parameters

All ball bearings have four distinct components--the inner race, the outer race, cages or retainers and the balls or rollers. The inner race of the bearing is connected to the rotating shaft and the outer race is fixed to a stationary housing.

For a bearing with a stationary outer race and an inner rotating race, characteristic defect frequencies ([2.1] to [2.4]) can be obtained for flaws in the outer race, inner race, ball bearings or in the cage as follows, assuming that there is no slippage for the rolling elements.

$$\text{Bearing Pass Frequency Outer Race (BPFO)} = f \left(\frac{N_B}{2} \right) \left[1 - \frac{B_D}{P_D} \cos(\alpha) \right] \quad [2.1]$$

$$\text{Bearing Pass Frequency Inner Race (BPFI)} = f \left(\frac{N_B}{2} \right) \left[1 + \frac{B_D}{P_D} \cos(\alpha) \right] \quad [2.2]$$

$$\text{Ball Spin Frequency (BSF)} = \frac{f}{2} \left(\frac{P_D}{B_D} \right) \left[1 - \left(\frac{B_D}{P_D} \cos(\alpha) \right)^2 \right] \quad [2.3]$$

$$\text{Fundamental Train Frequency (FTF)} = \left(\frac{f}{2} \right) \left[1 - \frac{B_D}{P_D} \cos(\alpha) \right] \quad [2.4]$$

Defective bearing components generate a unique frequency response in relation to the dynamics of bearing motion and the mechanical vibrations produced are a function of the rotational speeds of each component. A single-point defect produces one of the four characteristic fault frequencies defined above, depending on which surface of the bearing contains the fault. Single-point defects on the bearings begin as localized defects on the raceways or ball elements and as the ball elements pass over the defective areas, small collisions occur, producing mechanical shockwaves in the form of damped sinusoidal impulses. These impacts then excite the natural frequencies of mechanical resonance in the machine. This process occurs every time a defect collides with another part of the bearing with a period determined by the location of the defect and the geometry of the bearing. The rate of occurrence is equal to one of the characteristic bearing fault frequencies. Theoretical models of single and multiple point defects of the vibration produced by a faulty bearing under constant and varying radial loads have been established by McFadden and Smith (McFadden and Smith 1984a; McFadden and Smith 1985). These models take into account the impulse series generated by a point defect in a bearing modelled from first principles as a function of the rotation and geometry of the bearing, the modulation of the periodic signal caused by non-uniform bearing load distribution, the transfer function of the vibration transmission from the rolling element bearing to the transducer as well as the exponential decay of vibration.

Vibrational analysis techniques can be used to monitor these frequencies in order to determine the condition of the bearing. Upon inception of a defect in the bearing component, some or all of the characteristic frequencies and their harmonics begin to emerge in the spectrum. Each defect present in the bearing produces vibration

either at a basic frequency or at some complex combination of several basic frequencies. More severe defects produce vibrations of greater amplitudes and result in harmonics.

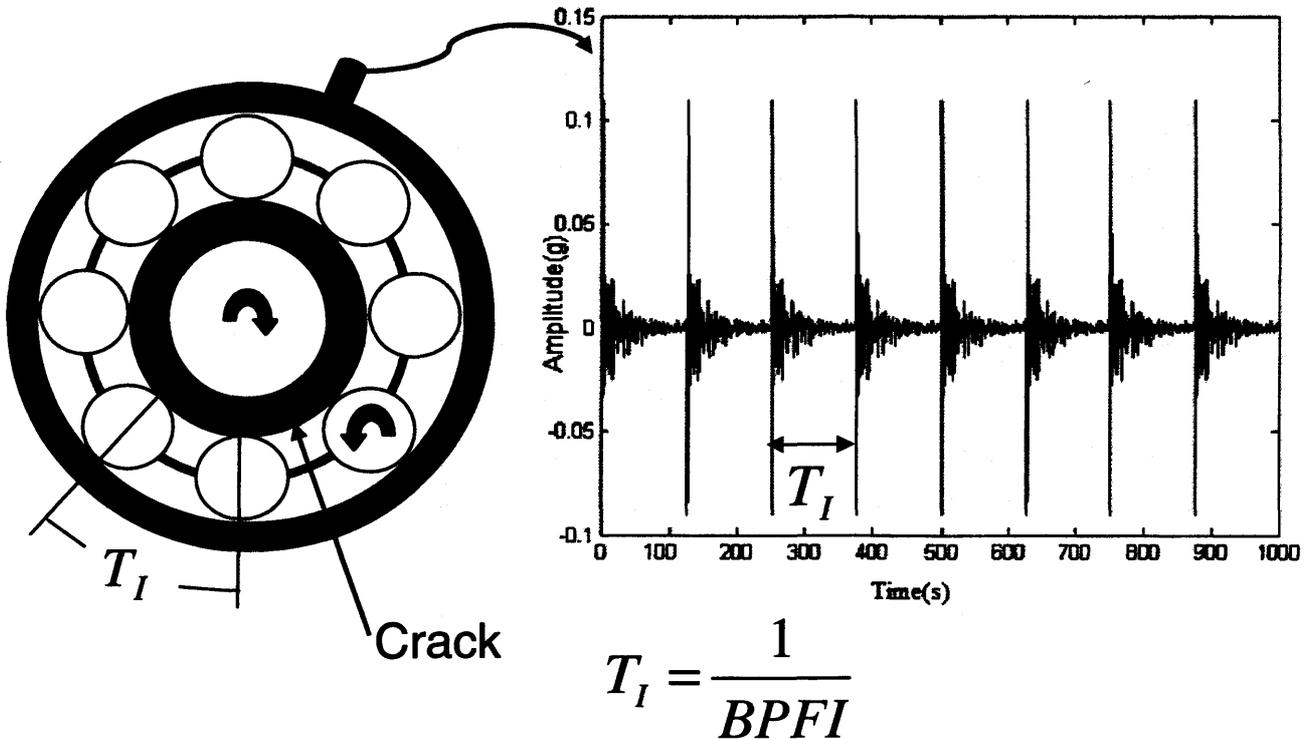


Figure 2–2 Detecting Bearing Inner Race Defect Frequency (McInerny and Dai 2003).

For instance, the time response signal captured by a vibration transducer mounted on the bearing housing caused by a ball bearing which has a localised single point defect on the inner race is shown in Figure 2–2. Periodic impulses are produced at a rate of the inverse of the impact frequency, and in this case, the inverse of inner race defect frequency (BPF I). The bearing rings at its natural resonating frequency and the response decays quickly because of damping.

The Barden bearing specifications for the BOC Edwards iGX dry vacuum pump that was used as the test bed in this experimentation are: number of balls = 9, pitch diameter = 46.2 mm, ball diameter = 9.5 mm and contact angle = 24.97 degrees. A picture of the whole bearing and the bearing components is shown in Figure A–6 and Figure A–7 respectively. The theoretical ball bearing defect frequencies BSF, BPFO, BPF I, FTF were calculated using equations ([2.1] to [2.4]) for a pump set running at 100 Hz and were found to be at 468 Hz, 366 Hz, 534 Hz and 41 Hz

respectively. The actual ball bearing defect frequencies BSF, BPFO, BPFI, FTF were obtained from vibrational data were estimated to be at 464 Hz, 363 Hz, 530 Hz and 40 Hz respectively. The slightly smaller frequencies obtained were due to the running speed of the pump being less than the actual set speed of the pump. The calculation of the actual running speed of the pump will be explained later in Section 3.5 in Chapter 3. Slippage in the bearings is another reason which has to be accounted for (slippage is typically around 2-5%). It should be noted that the bearing defect frequencies are directly proportional to the shaft rotating speed or, in other words, the rate at which the inner race is rotating. The outer race and inner race frequencies, BPFO and BPFI, vary linearly with the number of rolling elements. Each bearing defect frequency develops a unique characteristic vibration signature in the frequency spectrum. The vibration spectrum will have peaks at these defect frequencies and at their harmonics too. By identifying which defect frequencies are present in the frequency spectrum, we can in turn identify which bearing component is defective. A bearing with a seeded inner race crack was used to simulate the fault conditions in the study conducted. A picture of the defective bearing with the inner race 'crack' is shown in Figure A-8.

2.2.5. Conclusion: The Dry Vacuum Pump Explained and the Ball Bearing Defect Frequencies Calculated.

This section has reviewed the fundamental working principles of the dry vacuum pump and has explained why condition monitoring of dry vacuum pumps is important. The ball bearing defect frequencies have also been defined. How these ball bearing defect frequencies are produced due to the occurrence of defects which can be located at different parts of the bearing component has been explained. The fault detection scheme here is limited to work under the assumption that there is only one fault present, i.e. a single point defect. However, in reality, multiple faults can develop simultaneously. In such a case, the scheme can be modified to incorporate the diagnosis of multiple faults. Though the test fault conditions were artificially induced, the bearing faults were introduced in test facilities by Barden Bearing specialists to replicate bearing damage and wear in the dry vacuum pump in natural semiconductor operating conditions.

2.3. The AR Model

The AR model is a stochastic model that stems from linear prediction and allows for the demand of high-resolution spectral estimation. The AR model based approach is probably the most promising one for implementation into an automated diagnostic system due to its simplicity in formulation and in application. In addition, the parameter estimation algorithms for the AR model are relatively mature, always converge and are computationally efficient.

In this section, the theory of AR modelling is first examined and its merits are discussed for application to fault diagnosis. An overview is also given of other work on the application of AR modelling to condition monitoring studies.

2.3.1. Reviewing the Basic Principles of AR Modelling

The parametric models consist of a rational system function and where the innovation is a white noise process. There are three types of parametric methods and they are : the Autoregressive (AR) process model, the Moving Average (MA) process model and the Autoregressive Moving Average (ARMA) process model (Kay and Marple 1981).

2.3.2. Autoregressive Model

In an AR model (Makhoul 1975) the current value of a time series $x[n]$ at discrete time instant n is expressed as a linear combination of p previous values plus an error term [2.5]. $e[n]$ is the innovation white noise of the process with zero mean and variance σ^2 . p is the order of the model. a_k are known as autoregressive coefficients.

$$x[n] = -\sum_{k=1}^p a_k x[n-k] + e[n] \quad [2.5]$$

[2.5] is formally equivalent to the standard multiple linear regression model where $x[n]$ is the dependent variable and $x[n-1], \dots, x[n-k]$ are independent variables. The value of $x[n]$ is predicted from the past values and $e[n]$ represents the one-step prediction error.

Taking the z-transform of $x[n]$ and $e[n]$, yields $X[z]$ and $E[z]$ in the frequency domain. Equation [2.5] expressed as a linear filter in the z-transform domain is stated as [2.6].

$$X[z] = -X[z] \cdot \sum_{k=1}^p a_k z^{-k} + E[z] \quad [2.6]$$

The rational transfer function of the system, $H[z]$, can be defined by [2.8] relating the input to output where z is the forward shift operator.

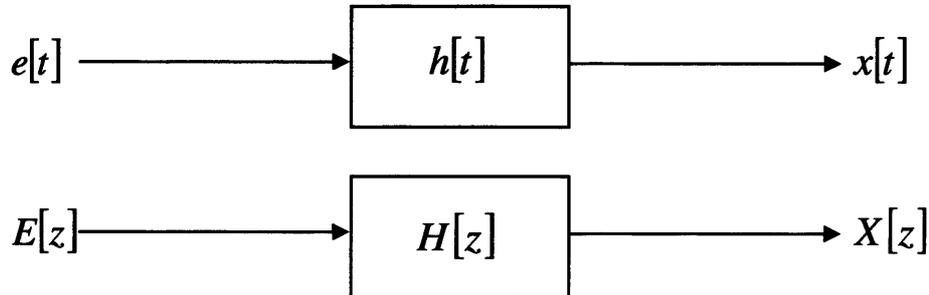


Figure 2–3 Diagram showing an autoregressive process.
With a white noise driving source input $e[t]$, transfer function $h[t]$ and output $x[t]$.
Representation in time domain (above) and frequency domain (below).

$$X[z] = H[z] \cdot E[z] \quad [2.7]$$

$$H[z] = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad [2.8]$$

This is called an Autoregressive (AR) or all-pole Infinite Impulse Response (IIR) model of order p and is usually denoted as an AR(p) model. The Power Spectral Density (PSD) of the output of the linear filter, $P[z]$, is related to the PSD of the input signal $P_e[z]$, which is noise. $P_e[z]$ is defined as

$$P[z] = H[z] \cdot H^* \left[\frac{1}{z} \right] \cdot P_e[z] \quad [2.9]$$

The PSD of white noise is a constant and is given by the variance, σ^2 , of the driving input signal, and Δt , the sampling interval, is the inverse of the sampling rate f_s .

$$P_e[z] = \sigma^2 \Delta t \quad [2.10]$$

$P[z]$ can be obtained by evaluating along the unit circle $z = \exp(j2\pi f \Delta t)$ for $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ once the a_k coefficients are known and is given by [2.11].

$$P[z] = \frac{\sigma^2 \Delta t}{\left| 1 + \sum_{k=1}^p a_k z^{-k} \right|_{z=e^{j2\pi f \Delta t}}^2} \quad [2.11]$$

2.3.3. Autoregressive Moving Average Model

The rational transfer function of the linear system in [2.8] can be more generally expressed as [2.12]

$$H[z] = \frac{B[z]}{A[z]} = \frac{1 + \sum_{k=1}^q b_k \cdot z^{-k}}{1 + \sum_{k=1}^p a_k \cdot z^{-k}} \quad [2.12]$$

The output sequence of the system then becomes the following difference equation.

$$x[n] = -\sum_{k=1}^p a_k x[n-k] + \sum_{m=0}^q b_m e[n-m] \quad [2.13]$$

The second part of this equation is a weighted average of the previous values of the input noise series and is termed a moving average (MA) or all zero model with order q and denoted as MA(q). The MA model is suitable for modelling data sequences with a spectrum of deep valleys rather than sharp peaks. [2.13] is a combination of the AR and MA models and is called an ARMA model or a pole-zero model or order p and q and denoted as ARMA(p,q). According to the Wold decomposition theorem (Box, Jenkins *et al.* 1994; Akaike 1969; Makhoul 1975; Kay and Marple 1981; S. Lawrence Marple Jr. 1987; Broersen 2000a), any ARMA or MA process can be represented by an AR model of infinite order. The most important implication of the Wold decomposition is that reasonable approximation can still be achieved by a higher AR model order of any MA or ARMA process, although an less efficient model may be chosen.

If there is significant background noise in the signal (S. Lawrence Marple Jr. 1987), the PSD is characterised by poles and zeros and the use of a low order all-pole AR model would yield poor results. In this case the use of an ARMA model would be able to achieve better results. For the bearing signals it is generally believed that the signal is much greater than the background noise (normally electrical noise for voltage measurements) and, with the use of an appropriate anti-aliasing filter (more details of

the anti-aliasing filter used in the study will be given in Section 3.2), the signal to noise ratio of the measured vibration signal can be enhanced. Hence the use of an AR model seems appropriate and sufficient in the investigation conducted.

2.3.4. Advantages of Using the AR Model

In this work the use of AR model has been selected for three primary reasons. The first fact is that an autoregressive model can be identified well for a system with sharp peaks. The AR modelling method is proven appropriate for estimation of power spectra with sharp peaks but not deep valleys as in the case of bearing faults. This is due to the all-pole nature of the AR model. Many practical vibration generating systems have very few deep nulls (McCormick, Nandi *et al.* 1998) so they are suited well for AR modelling. The spectrum of ball bearing vibration may be precisely classified into this category.

The second fact deals with the identification of the model. AR parameters and the autocorrelation sequence of the signal are related by a set of linear equations if Yule-Walker estimation is used. AR parameters may thus be estimated efficiently as solutions to linear equations. These are computationally straightforward to solve and always converge.

The third reason is the sensitivity of the AR parameters to variations in the conditions of the system (Junsheng, Dejie *et al.* 2006). Traditional vibration condition monitoring algorithms are based on the processing of vibration signals through various means of filtering and conditioning. Whilst these processes have evolved into well established procedures, their application still requires a high level of expertise in the selection of filter banks, spectral lines and other components to be removed (McCormick and Nandi 1996; McCormick and Nandi 1998). This has been a hindrance for automation of fault detection. The AR model can model the system dynamics and transients through variations in the autoregressive parameters of the AR model. The fault can be identified by the parameters (the AR coefficients and the variance of the prediction error) of the AR model after the AR model of the vibration signal is established without constructing the mathematical model of the system or

constructing the fault mechanism. One may not have *a priori* knowledge on the characteristics (e.g. tooth number, cross gear interactions) of the monitored system (Wang 2001b; Wang 2001a). Detection of fault conditions for which models are not available also becomes easier. Fault diagnosis can be achieved by having models for a number of known conditions and a time series of unknown condition can then be assigned the condition of the model which predicts the time series best.

2.3.5. Disadvantages of Using the AR Model

One of the major concerns when using AR modelling for fault detection studies is the selection of the model order. The model order needs to be estimated as part of AR PSD estimation. The trade-off between resolution and variance is determined by the order in the AR spectra. Use of too low a model order results in a highly smoothed spectrum, masking the peaks of frequencies of interest. Conversely, use of too high an order significantly increases the resolution and may introduce spurious detail into spectrum because of spectral splitting. Spectral line splitting is a phenomenon observed as a result of two or more closely spaced peaks occurring in the spectral estimate where only one peak should have been present. This behaviour was first documented by Fougere (Fougere, Zawalick *et al.* 1976), where it was noted that the extra poles were generated by additional AR parameters due to the use of a higher order than required and this gave rise to the spurious peaks. It was also noted that spectral line splitting tends to disappear as the length of the data record increases. Choosing a model with the smallest order and that best describes the true spectrum is an important principle in model selection and this is referred to as the principle of parsimony. The spirit of parsimony is against the selection of higher orders and states that the simpler the AR model is, the better it is. The true model order p is not known *a priori*. This is the main reason why many are deterred from using AR models as they have to determine the model order beforehand and have to spend time and effort in predicting the right order for the AR model.

2.3.6. Algorithms for Autoregressive Modelling

There are four methods for the estimation of the AR parameters: Yule-Walker or autocorrelation method, Burg or lattice filter method, Least squares forward method or covariance method and Least squares forward backward method, also known as modified covariance method (Kay and Marple 1981; Broersen 2000b).

2.3.6.1. Yule-Walker

The autocorrelation method or the Yule-Walker approach is described in detail in this section. In estimating the AR parameters, it is usually convenient to relate the a_k coefficients to the autocorrelation function $R_{xx}(n)$ of the time series $x[n]$. By multiplying the shifted conjugate of $x[n]$, i.e. $x[n] \times x^*[n-k]$ and taking the expectation on both sides of [2.5] (AR equation), the model parameters can be shown related to the Autocorrelation Function (ACF) [2.14].

For a given the time series $x[n]=1,2,\dots,N$ with real values, the autocorrelation function at lag k is defined as:

$$R_{xx}(k) = \frac{1}{N} \sum_{n=1}^{N-k} (x(n) - \overline{x(n)}) (x(n+k) - \overline{x(n+k)}) \quad [2.14]$$

The normalized value of the autocorrelation is given by

$$ACF(k) = r_k = \frac{R_{xx}(k)}{R_{xx}(0)} \quad [2.15]$$

The ACF is a statistical measure of the dependence of the time series values at one time on the values at another time. The ACF of a discrete time series is simply the correlation of the process against a time-shifted version of itself. We have used the 'biased' estimator of the ACF. If the term N is replaced by $N-k$ in [2.14], then the ACF term is defined to be an 'unbiased' estimator. The biased estimator is generally

preferred as it tends to have a smaller mean square error and decays faster to zero with growing time lags than the unbiased estimator (Jenkins and Watts 1968).

2.3.6.1.1. Yule-Walker Equations

Assuming that the process under study is an AR process, the theoretical AR coefficients and the theoretical autocorrelation function satisfy a set of linear equations called Yule-Walker Equations (YWE). The Yule-Walker equations [2.16] allow computations of the $p+1$ model parameters from the $p+1$ autocorrelation coefficients by solving a set of $p+1$ linear equations. Earlier p was defined as the order of the AR model and σ^2 the variance of the noise component. The equations relating all these parameters are given in matrix form as:

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_p \\ r_1 & r_0 & \ddots & r_{p+1} \\ \vdots & \ddots & \ddots & \vdots \\ r_p & r_{p-1} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ \vdots \\ -a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad [2.16]$$

$$\sigma^2 = r_0 - \sum_{i=1}^p a_i r_{-i} \quad [2.17]$$

Based on these estimates, the AR parameters can be computed. In practice, when the precise form of the autocorrelation sequence is not known, it must be replaced by an estimate computed from the available time series $x[n]$. Since the estimation of the AR parameters only involves linear equations, there are well-established methods of estimating the AR coefficients and one of the computationally efficient methods is the Levinson-Durbin recursion.

The autocorrelation matrix is Hermitian $r_k^H = r_k$ and Toeplitz (identical element along any diagonal). Using the known or estimated (biased) autocorrelation function, the YWEs can be solved directly recursively using the Levinson-Durbin

algorithm. The Levinson-Durbin recursion is a computationally efficient method to estimate the AR coefficients since it exploits the fact that the autocorrelation matrix is both symmetric and is a Toeplitz matrix.

2.3.6.1.2. Levinson-Durbin Recursion

Partial Autocorrelation Function (PACF) can be seen as the normalized autocorrelation that remains at lag k after the effects of shorter lags ($1, 2, \dots, k-1$) have been regressively removed from the autocorrelation function at lag k (S. Lawrence Marple Jr. 1987). Lag k is equivalent to model order p in a linear regression model. The Yule-Walker equations [2.16] can be rewritten based on the partial autocorrelation coefficients for an AR (k) model and this is given by [2.18]

$$\begin{aligned} r_1 &= \phi_{k1} & + \phi_{k2}r_1 & + \dots & + \phi_{kk}r_{k-1} \\ r_2 &= \phi_{k1}r_1 & + \phi_{k2} & + \dots & + \phi_{kk}r_{k-2} \\ &\vdots & \vdots & + \dots & \vdots \\ r_k &= \phi_{k1}r_{k-1} & + \phi_{k2}r_{k-2} & + \dots & + \phi_{kk} \end{aligned} \quad [2.18]$$

where ϕ_{kk} is the partial autocorrelation function (derivation taken from (Box, Jenkins *et al.* 1994)). Levinson and Durbin have derived an iterative way of solving the Yule-Walker equations (Durbin 1959) and the values $\phi_{k1}, \phi_{k2}, \dots, \phi_{kk}$ can be solved recursively using the Levinson-Durbin algorithm.

$$PACF(k) = \phi_{kk} = \begin{cases} r_1 & k=1 \\ \frac{r_1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j} & k=2,3,\dots,K \end{cases} \quad [2.19]$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-1}$ for $j=1,2,\dots,k-1$. From the solution of YWEs, the estimation of the AR parameters a_k as well as power error σ^2 are obtained. In fact the

solution to the partial autocorrelation parameters looks familiar and it is the same solution as the solution of a_k autoregressive parameters.

2.3.6.2. Other Estimation Methods

2.3.6.2.1. Burg

The Burg method (de Waele and Broersen 2000) is a widely used AR estimation method and it is sometimes also known as the Maximum Entropy Method (MEM). The method calculates the AR coefficients directly from the data by estimating the reflection coefficients (negative PACFs) at successive orders in a least-square sense by employing the Levinson Durbin Recursion to derive the reflection coefficients (Broersen and Wensink 1993). The Burg method estimates the reflection coefficients by minimizing the sum of error powers from the forward and backward linear predictions using the Lagrangian multiplier method. The Burg method is more susceptible to spectral line splitting at higher orders than the other estimation methods.

2.3.6.2.2. Covariance and Modified Covariance

The covariance method for AR spectral estimation is based on minimizing the forward prediction error. The modified covariance method is based on minimizing the forward and backward prediction errors. The Burg and covariance methods produce comparable AR spectral estimates. The modified covariance method is best for sinusoidal components in data (S. Lawrence Marple Jr. 1987).

2.3.6.3. Reason for Choosing the Yule-Walker Method

The Yule-Walker, Burg, covariance and modified covariance parameter estimation methods are comparable in performance when large data records are used (Kay 1988). The performance of the Yule-Walker method deteriorates only when used with short data records (Kay 1988). Sufficiently large sample sizes which will capture the signal's characteristics have to be used if the Yule-Walker method is to guarantee accurate AR coefficients. This research issue of finding the optimal sample size is

addressed in Chapter 5 and the findings of this investigation are reported in Section 5.2. In this study, the Yule-Walker method with the Levinson-Durbin recursion has been chosen to find the AR coefficients (Kay and Marple 1981) as this estimation method is guaranteed to converge, resulting in a stable AR filter (roots are all within the unit circle). The Yule-Walker method produces a biased estimate of the residual variance however it has been shown the effect of bias on order selection is negligible (Broersen and Wensink 1993). The Yule-Walker estimation method was chosen mainly for its processing speed.

2.3.7. Review of AR Modelling in Condition Monitoring Studies

J.P. Dron (Dron, Rasolofondraibe *et al.* 1998; Dron, Rasolofondraibe *et al.* 2001; Dron, Bolaers *et al.* 2003; Dron, Bolaers *et al.* 2004) has studied the use of AR modelling for vibrational analysis of a forming press for a conditional maintenance program in 1998. It was noted that parametric methods are particularly worthwhile in the early detection of faults especially when two typical frequencies are close to one another and it was acknowledged that the model order selection is one of the major problems encountered when implementing parametric spectrum analysis methods.

Parametric modelling has been employed in fault diagnosis studies in low speed machinery by Mechefske (Mechefske 1993a; Mechefske 1993b; Mechefske and Mathew 1993). Mechefske noted that AR modelling is especially useful in low speed machinery as recording long periods of data in low speed machinery is impractical and AR method is beneficial in such cases as it can work with short data records and achieve the same resolution as the FFT method and at a fraction of the time taken. The effects of noise on the vibration signal has been studied by him in the paper (Mechefske 1993b). At low rotating speeds, the effects of background noise become dominant. Here Mechefske notes that AR modelling has performed consistently better than FFT-based techniques when signals are contaminated with various levels of noise as fault frequencies could be checked at a lower SNR than the FFT method. It is noted that, without being able to average the several frequency spectra, the FFT-based procedures had more difficulty in dealing with the noise than the AR-based methods. In another paper, Mechefske (Mechefske 1993a) investigates the optimum vibration signal length required by AR modelling needed to produce repeatable frequency

spectra and it is shown that the signal length is largely dependent on the speed of the equipment being monitored.

AR modelling, apart from being used as a spectral analysis tool, also has lots of potential as a model based automatic diagnostic system. This concept was researched by Baillie in 1996, where the concept of fault diagnosis using an observer bank of autoregressive time series models was investigated (Baillie and Mathew 1996). It was found that AR modelling requires much shorter lengths of data than traditional pattern classification tools such artificial neural networks and expert systems, which require large amounts of data training for successful fault prediction.

In 1996, Huang investigated a machine fault diagnostic system which was built by feeding AR parameters as inputs to a Fault Diagnostic Network based on fuzzy logic methodologies (Huang and Wang 1996). In 1998, McCormick goes one step further and analyses periodic time-varying autoregressive models for detection of bearing faults (McCormick, Nandi *et al.* 1998) to study the cyclo-stationarity nature of rotating machines. It was noted that using an AR-model based diagnostic system can provide a means to detecting machine faults even if data are only available from the machine in its normal condition. Using AR models as one-step ahead predictors, it is noted that it is possible to predict machine's behaviour even if it goes into a condition where no training data is available. McCormick (McCormick and Nandi 1996; McCormick and Nandi 1998; McCormick, Nandi *et al.* 1998) has noted that one of the advantages of using AR models is that analysis of the prediction error generated by using an autoregressive model of the machine vibrations as a one step ahead predictor may provide a set of features which could indicate the presence of faults even in cases where fault vibration data is not available.

In a paper by Mathew in 1996 (Baillie and Mathew 1996), a model-based condition monitoring system uses autoregressive techniques for detecting changes in a machine's health condition by monitoring the differences between the monitored and predicted signal as fault indicator and for classifying faults.

In a recent work in 2001, Wenyi Wang has effectively applied Minimum Phase Autoregressive (MPAR) approach for detection and diagnosis of gearbox faults for a

helicopter application at the Aeronautical and Maritime Research Laboratory (AMRL) (Wang and Wong 1999; Wang 2001c; Wang 2001a). A non-minimum phase AR model is identified for the gear signal by maximizing the kurtosis of the inverse filter error signal of the model. Sudden changes in the error signal are usually indications of the existence of localized gear faults. Under the assumption that gear mesh signals were derived from an AR system driven by Gaussian noise, an AR model was established for signals from the monitored gear under healthy conditions and then used as a Linear Prediction Error (LPE) filter. The future signals from the same gear, under healthy or fault conditions, were processed by the LPE filter. The output prediction error of this LPE filter should resemble random noise if the monitored gear remains in a healthy condition. However when a local fault (e.g. a tooth crack) is developed in the gear, the fault affected region would not be well predicted by the AR model that was established under healthy conditions. The LPE signal would reflect the changes caused by the fault. It was shown that the AR modelling approach outperforms the current gear fault diagnostic techniques, such as the residual signal method. Wenyi Wang notes that using this method there is no need to know the number of teeth on the monitored gear, the characteristics of the modulation waveforms or the number of gears on the same shaft. Moreover it is not important whether structural resonances or additional modulations are present in the signal whereas this information is essential to some of the current techniques.

Wenyi Wang noted that in terms of making an early detection of incipient local gear faults, the MPAR modelling method can still perform unsatisfactorily and proposed a more general linear prediction modelling method – a Non-Minimum Phase AR (NMPAR) method. Here it was assumed that the gear mesh signals were driven from a non-minimum phase AR system driven by non-Gaussian noise. The AR coefficients are estimated by kurtosis maximization using the inverse filter criteria. Kurtosis is a measure of Gaussianity of the residual signal and can be used as an index for fault detection. In conclusion, Wenyi Wang noted that the proposed NMPAR method by kurtosis maximization for gear fault detection has proved superior to current gear fault detection techniques from the following perspectives: 1) earlier and more convincing detection of gear tooth cracking 2) capability of detecting faults in complex gear boxes such as helicopter transmission gearboxes where cross-gear interactions are common. However the author noted the proposed technique requires

some knowledge of choice of some parameters in the cost function for maximization of the kurtosis function. Constraints had to be incorporated into the cost function to establish necessary and sufficient conditions for a blind signal recovery. It was noted that some work needs to be done for a more systematic approach based on a non-linear optimization algorithm suitable for gear fault detection.

In 2000, Fang Wen also applied autoregressive modelling techniques for fault detection with helicopter data (Fang, Willett *et al.* 2000). In this work, it was noted that “as regards signal processing, it appears that autoregressive (AR) coefficients from a simple linear model encapsulates a great deal of information in relatively few measurements; it has also been found that augmentation of these by harmonic and other parameters can improve classification greatly.” In this work, the results show promising fault detection accuracy, particularly when learning was based on autoregressive (AR) coefficient features. Near perfect fault recognition accuracy was reported with relatively small feature sets involving autoregressive coefficients. In this paper, Fang states that it is possible to use periodogram outputs explicitly as features for classification; however in general this implies a great many features and the usual curse of dimensionality may ensue. Since it is clear that spectral features do indeed yield much relevant information, a concise way of representing the spectrum is proposed here: the autoregressive (AR) coefficients. It is noted that AR coefficients are able to represent much global spectral information into a small dimension.

2.3.8. Conclusions: The AR Model Explained

The papers described above are pioneering in their recognition of the use of AR modelling for condition monitoring applications. However, none has reported the approach adopted here, that is to track the movement of the AR poles trajectory for fault classification. The development of the AR pole fault detection tool is discussed in detail in Chapters 7 and 8 of the thesis.

In this section, the basic principles of AR modelling have been reviewed. The use of AR modelling for PSD estimation brings about benefits in terms of improved resolution and smaller sampling rates compared to the more established method of FFT. This benefit of AR modelling will be explained more in detail in the next section

of this chapter. This aspect will be investigated further with supporting results in Chapters 4 and 5 of the thesis.

2.4. Overview of Current Vibration Based Diagnostic Techniques

In this section, various condition monitoring techniques which have shown effective performance in the detection of bearing faults such as the statistical methods, power spectral methods and time-frequency methods will be reviewed.

Every rotating machine exhibits a characteristic vibration signature that is unique to the machine. The dry vacuum pump being a “quasi-steady state” rotating machine also produces vibrations which can be used as an effective indicator for fault diagnosis. If the sources of vibration can be isolated, the exciting frequencies can be identified and, together with an understanding of the engineering and physics of the system producing it, its dynamic behaviour can be predicted at different operating conditions. The use of vibrational analysis for planned and predicted maintenance is an established technique (Mathew 1987). The monitoring of vibrations can give an indication of a machine’s condition while it is on-line and therefore can allow efficient scheduling of maintenance and early detection of potentially critical faults which can be costly to fix. Vibration diagnosis has been proven effective in the detection and diagnosis of bearing faults.

Vibrational analysis can be used to detect bearing failure, shaft/rotor imbalance, misalignment of couplings and bent shafts, lubrication problems and gear failure in rotating machines (Mathew and Alfredson 1983; Mathew and Alfredson 1984; Mathew and Alfredson 1986; Mathew and Alfredson 1987; Mathew, Kuhnell *et al.* 1987b; Mathew, Kuhnell *et al.* 1987a). Due to adverse operation conditions, improper installation or material fatigue, the bearings deteriorate over time. Bearing vibration can be used as a representative ‘health index’ for monitoring bearing working conditions. Bearing faults are the most common type of electric machine failures (Huang and Wang 1996). Early detection of these faults allows service to be performed during planned downtimes rather than during costly emergencies.

It is important to note that any rotating machine has many of the same vibration characteristics as any other rotating machine. This is true whether the machine is a 180 RPM water pump or a 10000 RPM gas turbine. That is why vibration analysis has been successful in condition monitoring, as the basic principles are the same.

2.4.1. Time Based Techniques

Rotating machinery produce signals that are randomly varying signals (Zhuge, Lu *et al.* 1990) and examples of these are displacement signals that can be measured with proximity probes, vibrations measured with accelerometers, sound measured with acoustic emission sensors and pressure signals measured with calibration gauges. One such way of achieving fault prediction and prevention is to use directly time domain analysis (Alfredson and Mathew 1985b; Dron, Bolaers *et al.* 2004) where statistical parameters such as the Root Mean Square (RMS), standard deviation, variance, crest factor, kurtosis, and skewness are used as trend parameters to detect the presence of incipient faults. The statistical parameters are then compared with baseline estimates obtained from machines in no-fault conditions and an increase in these statistical time domain figures may anticipate malfunction. Some of the time based statistical parameters which are mentioned above are defined as follows:

The RMS measures the overall vibration level of a rotating machine. The RMS of a discrete vibration signal is given by [2.20]. The RMS is calculated by the square root of the sum of the squares of the sample x_i divided by the sample size N , where x_i is from x_1 to x_N .

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i)^2} \quad [2.20]$$

The spread of the data about the mean value (i.e. the RMS) is known as standard deviation and is denoted by σ . The unbiased estimate of the standard deviation [2.21] is given by:

$$\text{Standard Deviation} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2} \quad [2.21]$$

where μ is the mean. Variance, which is the second central moment about the mean, is defined as σ^2 .

The skewness [2.22] which is the central third moment of the data is defined as:

$$\text{Skewness} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3 \quad [2.22]$$

The kurtosis [2.23] is the fourth central moment and is stated as:

$$\text{Kurtosis} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4 \quad [2.23]$$

The crest factor [2.24] is defined as:

$$\text{Crest Factor} = \frac{\text{Max}(x_i) - \text{Min}(x_i)}{\sigma} \quad [2.24]$$

Both the crest factor and kurtosis are measures of spikiness in the time domain waveform. When the bearings are in a good condition, the distribution of the amplitudes of the time domain signals produced by them are Gaussian-like (Dron, Bolaers *et al.* 2003) and their kurtosis values are close to 3. When the bearings are damaged, the appearance of defects such as spalls and cracks disturb the signals. Wide band impulse-like signals are generated when the bearings pass over the defect points, thus modifying the distribution. Typically the kurtosis value of a defective bearing signal is greater than 3.

It is difficult to monitor the development of faults using these indices alone. Using these statistical features individually, there would be a possibility of not detecting a change of condition of the machine as these scalar indicators are highly

non-linear and are very much dependent on external influences such as load and speed of the machine (Alfredson and Mathew 1985b). The scalar indicators are sensitive to variations of the rotation speed of shafts, band-pass filtering, resonance of the sensor and background noise. Values of statistical parameters like the kurtosis comes down to the level of an undamaged bearing when the defect is well advanced. Ocak, in his paper (Ocak and Loparo 2005) has stated that kurtosis and the crest factor are effective only at the early stages of bearing damage and are not sensitive at detecting faults when the severity of the damage increases. Other researchers have reported similar behaviour too (Tandon and Nakra 1992; Martin and Honarvar 1995; Heng and Nor 1998). It is noted that the vibration signals become more random-like and the values for the crest factor and kurtosis obtained decreased when the damage level of the bearing increased. The vibrations obtained from the defective bearings were less impulsive and more white-like, mimicking a normal bearing. Thus it is observed that the statistical analysis approach based on kurtosis and crest factor lacks the ability to detect the bearing defects at the later stages of their development.

2.4.2. Frequency Domain Techniques

The vibration signature produced by the machine is normally complex, being a mixture of all the vibration components generated by the machine such as gears (Mathew and Alfredson 1987), bearings (Mathew and Alfredson 1984; Mathew and Alfredson 1986) and other moving parts. Important features of the vibration signals are often not obvious in the time domain signal and it becomes necessary to perform spectral analysis. Essentially some signal processing enhancement technique must be employed to extract useful diagnostic information from the measured vibration signal.

Another approach towards vibration diagnostics is by using frequency domain methods (Alfredson and Mathew 1985a). The conversion of data from time domain to frequency domain is achieved using spectral analysis (Besancenez, Dron *et al.* 2001; Broersen 2002) to help identify the frequencies of interest. For example, if the machine is rotating at 100 Hz, it should be possible to identify a 100 Hz peak in the vibration frequency spectrum. These frequencies are known as forcing frequencies.

Having identified the various forcing frequencies, the next step is to identify what part of the machine is generating what frequency (Mathew, Kuhnell *et al.* 1987b). Then the next step is to measure the amplitude of the individual forcing frequencies and assess whether or not they are at acceptable levels.

Spectral analysis of rotating machinery consists of estimating the frequency content of a collection of sampled signals. The frequency spectra of the signals are compared with baseline spectra obtained from machinery run in normal or no-fault conditions (Mathew 1989). The spectral information of the largest peaks as well as the amplitudes and their phases can be monitored. The detection of new spectral peaks or changes in the Root Mean Square (RMS) values of frequency peaks of interest can indicate development of failures. This information can be used for data trend analysis. One-time measurements may not be indicative of the machine's health and normally trending of a machine element over a period of time is recommended (Alfredson and Mathew 1985b; Alfredson and Mathew 1985a; Mathew 1987). From trending plots, gradual changes in the condition of the machine elements can be detected. By monitoring the trending levels of vibration for a given machine, we can determine whether it is healthy or faulty and also predict the time until failure.

One very common application where this technique is frequently used is in the detection of bearing faults (Alfredson and Mathew 1985a) as was explained in Section 2.2.3. Bearing defect frequencies such as inner race fault, outer race fault, rolling element faults and cage frequency can be easily computed from the geometrical dimensions of the ball bearings (Tandon and Choudhury 1997). Then the defective spectra are compared with normal spectra to identify bearing faults (Alfredson and Mathew 1985a). Therefore having accurate frequency estimates is a key step to identification of bearing faults.

The FFT technique is the favoured method for spectral analysis as it is an established method and there exist many COTS products which easily aid in the implementation of the spectral analysis tool for frequency estimation of signals as part of a larger condition monitoring programs for fault detection in machines.

There exist other classes of frequency estimation methods and one example of these is the class of parametric modelling methods (Kay and Marple 1981) as was explained in Section 2.3.1. The parametric approach is based on modelling the signal under analysis as a realization of a particular stochastic process and estimating the model parameters from its samples. This method is very commonly used in seismic analysis and stock market forecasting (S. Lawrence Marple Jr. 1987) and in the field of biomedical such the study of the Doppler ultrasound frequency estimation and fetal heart rate studies (Schlindwein 1988; Schlindwein and Evans 1989; Schlindwein and Evans 1990, Boardman *et al.* 2002). The interest in using parametric spectral analysis compared to the FFT based techniques for fault detection and condition monitoring of equipment for rotating machinery has remained low. The main reason for this is because the order of the parametric models has to be determined beforehand (Mechefske and Mathew 1993) and has to be done accurately to get good frequency estimates.

In the following sections, the FFT and AR techniques are compared in terms of their PSD performance. The use of the parametric method of AR modelling as a spectral analysis tool is discussed in detail. The basic tenets of a PSD estimator are also examined.

2.4.2.1. AR Modelling As a Spectral Tool

There are a number of practical considerations to take into account when choosing a spectral estimator (Broersen 2002). Some of these include its performance in terms of resolution, variance and potential for real time application. Usually the PSD is defined in a nonparametric way by Fourier transform of an infinite autocorrelation sequence. Another possibility is to estimate PSD is to represent the examined process by a parametric description and then estimate PSD in terms of the model parameters. Such a PSD has some better properties than classical spectral estimators. For example, better frequency resolution is achieved by using parametric methods such as AR modelling or ARMA modelling mentioned in Sections 2.3.2 and 2.3.3. Classical estimation methods use windowed set of data or autocorrelation sequence and make the unrealistic assumption that values outside the window are implicitly zero and all data are periodic. The choice of window type and size requires exact knowledge of the true spectrum but this is hardly the case in real applications.

This leads to distortions in the spectral estimate. It has been quoted by Broersen (Broersen 2002) that “despite all efforts to construct efficient spectral estimators, the best or optimal spectral window can only be determined if the true spectrum is known *a priori*.” In model based methods more realistic assumptions about the data outside the window are made. The data is not assumed to be a periodic process and data is not multiplied with windows before the spectral transformation.

2.4.2.2. AR Versus FFT

Over the last 40 years, a primary tool for spectral analysis has been the Fast Fourier Transform (FFT). It is not required to explain much about the FFT technique as it is very tried and tested tool in many areas of engineering. The main limitation of FFT-based methods in the field of machine condition monitoring is the frequency resolution when applied to short lengths of signal.

Strictly speaking, the traditional FFT-based methods make the assumption the process is periodic and stationary. In practice, processes are not periodic and exhibit non-stationarity. The performance of FFT tools degrades when applied to non-stationary signals. One way to deal with non-stationary behaviour is to use Short Time Fourier Transform (STFT) or Wavelet analysis (these techniques will be explained more in detail in Section 2.4.3). These techniques are more applicable to transient signals. AR models have been proven to exhibit superior performance for nonstationary signals (Zhuge, Lu *et al.* 1990) than classical non parametric methodologies. The AR model doesn't assume periodicity and models the signals with flattest possible spectrum. Also AR-based spectral analysis can produce better spectral estimates for short segments of data since it is better able to characterize the time-varying behaviour of frequency estimates.

The frequency resolution characterizes the distinguishable minimum frequency difference between two sinusoids. Hence, for a given sampling frequency, the more samples you have, the higher the frequency resolution for the FFT-based method is. However, there is always a limit for the number of samples to use for spectral analysis as a bigger sample size translates to bigger memory buffers to hold the data as well as

longer processing time for each frame, inherently affecting the performance if a real time spectral analysis tool is to be implemented and a higher probability that the signal, over the extended time frame would be non-stationary. Also, in some situations, only short lengths of signal are available.

The relationship between the number of samples and frequency resolution of the FFT can be quantified by the following equation [2.25]:

$$\Delta f = \frac{f_s}{N} \quad [2.25]$$

For the AR PSD estimator the resolution of processes consisting of sinusoids in white noise [2.26] is given as

$$\Delta f = \frac{1.03 f_s}{p[SNR(p+1)]^{0.31}} \quad [2.26]$$

As can be seen, the frequency resolution of the FFT technique is bounded by the number of data samples available for use. The FFT method fails to work well for short data records and will have limited frequency resolution in such cases. The frequency resolution of the AR based method is a function of the model order and also the SNR.

The main advantage of the FFT PSD comes from the fact that it is computationally efficient. The use of FFT algorithm of Cooley and Tukey for the Fourier transformation leads to a reduced computer effort and this advantage in the computational influence is the main reason for the popularity of the FFT PSD method in the recent years (Broersen 2002). The processing requirement of the FFT method is proportional to $N \log_2 N$ where N is the sample size. The processing requirement of the AR method mainly depends on two factors, firstly the estimation method implemented and secondly it is proportional to p^2 , the square of the order. The Yule-Walker-based AR estimation method is computationally less expensive than the covariance or Burg counterparts. However it should be noted that even then the Yule-

Walker algorithm is slower than the FFT method. Schlindwein (Schlindwein 1988; Schlindwein and Evans 1989; Schlindwein and Evans 1990) has shown that the AR technique is slower than the FFT technique by a factor of 2.58 for the same sample size (N=256) and processor (TMS320C25) using the Yule-Walker method.

The main advantages and disadvantages of both PSD methods are summarized in Table 2-1.

Table 2–1 Comparing AR and FFT spectral estimation methods.

	AR	FFT
Resolution	Better than the FFT method. Dependent on model order p used.	Depends on frame size used.
Sampling Rate	Just above the Nyquist rate (where Nyquist rate is defined as twice the maximum frequency of the signal being sampled) is sufficient.	Typically 6 or 7 times the maximum frequency of the signal being sampled is required for good spectral estimates.
Spectral Variance	Less than FFT method.	Averaging required to reduce variance and to get smoother spectrum.
Applicability to non-stationary signals	Performs better than the FFT method because shorter segments can be used.	Performs poorly since longer segments are required.
Spectral Leakage	Non-existent as the technique does not assume periodicity of signal.	Spectral leakage is a problem that can be reduced with the use of suitable windows, but at the cost of frequency resolution
Performance in noise	Spectral resolution decreases with noise, but AR method performs better than FFT.	Performs poorly if SNR of signal is low.
Computational effort	Can be computationally intensive. Processing effort required is proportional to p^2 .	Requires less processing power than AR method. Processing power required is proportional to $N \log N$ of frame size N .

In this section, the reasons for choosing the parametric AR technique over the FFT technique for modelling the quasi-stationary vibration signals were explained. The main advantage of the AR approach comes from the fact that it can work with smaller sample sizes for the same resolution compared to the FFT method. Hence essentially the AR technique only requires a fraction of the samples that are required by the FFT method for the same resolution and may cost less in computational terms as fewer samples are used. This has an advantage especially in real time applications.

2.4.3. Time Frequency Analysis

Classical spectral estimators cannot correctly describe many interesting condition monitoring conditions where transient phenomena in the mechanisms of control must be investigated and where the basic assumption of stationarity cannot be assumed. The AR model is built upon a stationary process and does not perform well for non-stationary processes (Zhuge, Lu *et al.* 1990; McCormick and Nandi 1998; McCormick, Nandi *et al.* 1998). In such situations, the use of time-frequency analysis is more appropriate. Over the years, great interest has been focused on the development of spectral techniques for studying evolving non-stationary phenomena, both in the parametric and non parametric fields.

A time frequency distribution describes the evolution of the signal's energy in both time and frequency domains. Such distributions are particularly useful for analyzing signals with time-varying frequency content. By estimating the energy distribution of the signal in the time-frequency plane, it would be expected that changes in the amplitude of the signal could be located in both the time and frequency domains. Among the variety of time-frequency analysis techniques, the wavelet analysis, the spectrogram (the squared modulus of Short-time Fourier Transform(STFT)) and Wigner-Ville Distribution (WVD) are commonly used for fault diagnosis.

The wavelet analysis (Baillie and Mathew 1996; Fang, Willett *et al.* 2000; Abbaszadeh, Rahimian *et al.* 2002) is regarded as a powerful tool for detection of

sudden changes in non-stationary signals. It has the property of time-scale localization, which is particularly useful in the analysis of vibration signals with short time transient phenomena. The wavelet transform uses a finite duration basis function called the basis wavelet. The continuous wavelet transform is aimed at achieving a constant relative resolution i.e., good time resolution at high frequencies and good frequency resolution at low frequencies. This enables the effective analysis of signals consisting of high frequency content of short duration and low frequency content of long duration.

The spectrogram applies the Fourier transform for a short-time analysis window with the assumption that the signal satisfies the requirement of stationarity within the window. By moving the analysis window along the signal, it is expected that the time variation of the signal spectrum will be revealed. The spectrogram has many desirable properties, including a well-developed general theory. However serious difficulties arise when the spectrogram is used to deal with fast varying signals, such as transients. If a very narrow analysis window is chosen to catch sudden changes in the signal, it is very difficult to resolve the closely spaced frequency components of the signal. Thus there is a trade-off between time and frequency resolution for the Fourier-based spectrogram (Watts 1968).

The Wigner-Ville Distribution (Wang and Wong 1999; Yang, Mathew *et al.* 2005) provides a high-resolution representation in time and frequency for a non-stationary signal and possesses many interesting and important characteristics. However unlike the spectrogram, its energy distribution contains negative components that make interpretation difficult, and often has severe cross-terms between components in different time-frequency regions, potentially leading to further misinterpretation and confusion.

2.4.4. Conclusion: Frequency Based Parametric Method of AR Modelling is an Appropriate Method

The vibration signals produced by the dry vacuum pump are assumed to be quasi stationary and hence the use of the frequency based parametric method of AR modelling is appropriate in the context of the problem studied.

There are also other fault detection and diagnosis methods suggested by researchers which have been successfully implemented, especially for the analysis of vibration signals, resulting from ball bearings of machines. These are cyclostationary analysis (Randall, Antoni *et al.* 2000), higher order statistics (McCormick and Nandi 1996; McCormick and Nandi 1998) and cepstral analysis (Zheng and Wang 2001) methods. However, many of them are theoretically-based and have failed to address the problem from a real-time implementation point of view. Many of these methods require some complex and sophisticated analysis, which renders their application cumbersome for automated fault detection systems. A compromise has to be achieved between the method's resolution and the feasibility and simplicity required for real-time implementation. Real-time implementations of the fault diagnostic system in industrial machines do not permit high computational costs. As a solution to this, the model-based approach of AR modelling is proposed. Building a fault detection tool based on AR techniques gives us the possibility of a real-time implementation and the ability to detect faults with an acceptable computational cost.

2.5. Summary

In summary, this chapter has introduced the dry vacuum pump, outlined the basic equations of AR modelling, looked at alternative methods for vibrational analysis and has given the reasons for choosing the frequency-based parametric method of AR Modelling for the study conducted. The next chapter will look at the data acquisition setup and instrumentation used in acquiring the vibration signals from the dry vacuum pump.

3. Platform of Experimentation

3.1. Introduction

In this investigation, suitable transducers had to be chosen for studying the vibration measurements from the dry vacuum pump. Also, suitable instrumentation had to be constructed for collection of vibration signals from the pump. The measurement of machine vibration can be made using a wide array of sensors but in our case the accelerometers were selected and the performance of two different types of accelerometers was evaluated: ADXL105 iMEMs surface micromachined accelerometer and Brüel & Kjær (BK) 4370V accelerometer. Reasons for selecting these accelerometers will be given in this chapter. This work has been published in (Thanagasundram and Schlindwein 2006c).

3.2. Mounting the Accelerometers on the Dry Vacuum Pump

Two different types of accelerometer were mounted radially on different measurement points on the dry vacuum pump to capture the vibration signals (please refer to Figure B-1 in APPENDIX B). One was using a surface micromachined accelerometer ADXL105 (Doscher 1997; ADXL105 1999; Doscher 1999b; Doscher 1999a). The other was using a Brüel and Kjær 4370V accelerometer. For an in-depth analysis of the use of these two accelerometers, please refer to Section 3.4. It was noted that the measurement position was critical in affecting the characteristics of the vibration signals. The placement of sensors is of critical importance to achieving high quality measurement for machine condition monitoring and fault diagnosis and effective sensor placement strategy had to be adopted (Huang and Wang 1996; Tang, Tan *et al.* 2001). At the incipient stage, a defect-related impact is generally weak in magnitude and of short duration due to the small size of the defect (McFadden and Smith 1984a). To achieve high signal to noise ratio, the sensors need to be placed as closely as possible to the bearings, where signals due to structural defects are generated. Machine borne vibrations and other background noise are often encountered in the manufacturing environment and hence the signals generated by the defective bearing may be overwhelmed by noise contamination in both time and frequency domains. Also, for a sensor placed far away from the defect, considerable

signal attenuation can occur along the propagation path leading to poor signal to noise ratio. For bearing condition monitoring, vibration sensors have traditionally been placed on the outside surface of the mechanical housing within which the bearing to be monitored is located. Since the bearing of interest is mechanically linked to other machine structures, the sensors will pick up structural borne vibrations and environmental noise in addition to defect impact induced vibrations. For the pump, data was acquired from the bearing housing in vertical, radial and axial directions. However most of the analysis was done with data obtained in the radial direction as the radial measurement point was a convenient mounting point and the data obtained were of good quality.

3.3. Elliptic Low-pass Anti-Aliasing Filter

The signals from the ADXL105 accelerometer were filtered with an 8th order elliptic low pass anti-aliasing filter that was custom-built in our laboratory. An anti-aliasing filter is a filter used, before sampling or digitizing the signal, to restrict the bandwidth of the signal to satisfy the Shannon-Nyquist sampling theorem (Shannon 1998; Wyner and Shamai 1998). The purpose of the low-pass anti-aliasing filter built was to remove frequencies above 10 *kHz* since it had a cutoff frequency of 10 *kHz*. The vibration signals were oversampled at a frequency of 40 *kHz*. The maximum frequency that can that can be reconstructed without distortion is 20 *kHz* according to the Nyquist frequency (where Nyquist frequency is defined as half the sampling frequency of the discrete digitized signal (Ifeachor and Jervis 2003)). Of the 20 *kHz* range, only 10 *kHz* (in the range of 0 to 10 *kHz*) was filtered and retained by the low-pass filter to accommodate the practical constraints of anti-aliasing filters since an anti-aliasing filter with perfect stop band rejection (also known as brickwall filter) is impossible to realise no matter how good the filter is, and in reality, every realizable filter will permit some aliasing to occur (Proakis and Manolakis 1995). In summary, the use of the filter was for the following two reasons: One was to prevent aliasing of the vibration signals before digitization took place in the ADC (Analog to Digital Converter) cards. The other was to increase the gain of the signal by 2 to increase its SNR. The specifications of the filter and filter design are shown in Figure B-2 in APPENDIX B. A high order of 8 was implemented as it was desired to have a sharp cut-off rate and a large attenuation in the stop band. The frequency response of the

filter is shown in Figure B-3 in APPENDIX B. As can be seen from the frequency response, the filter constructed achieved a steep attenuation rate at the cut-off frequency of 10 kHz with an attenuation of almost 70 dB in the stop band. It is known that elliptic filters have a non-linear phase and hence a not-flat group delay, but the phase characteristics of the signal were not important in this study – where the amplitude of the vibration signal is of concern. The filter was constructed using a positive Single Amplifier Biquad (SAB) configuration to minimize the number of operational amplifiers used in the circuit board.

The signals from the Brüel and Kjær 4370V accelerometer were conditioned using a Brüel and Kjær 2692 preamplifier, which transduced the signal from charge to voltage form, changed its gain and band-pass filtered it from 0.1 Hz to 10 kHz. The analogue to digital conversion of the signals was performed with a 16-bit NI 6034E ADC card with 16 channels of Analog Input (AI) and 2 channels of Analog Output (AO). The key features of the card are that it has a maximum sampling rate of 200 kSamples/sec and an onboard clock which can be used to generate accurate pulses to control the sampling rate. The ADC board also features an onboard buffer, which is used as a temporary storage for the samples while the data is transferred to the computer's memory for processing, thus preventing data overflow.

The vibration signals were digitised at a sampling rate of $f_s = 40 \text{ kHz}$. The acceleration signals were suitably amplified by the on-board gain amplifier on the ADC card before the A/D conversion took place to make full use of the available dynamic range. The signals were originally sampled at 40 kHz. As noted earlier, this gives us a frequency range of 0 to 20 kHz to work with but after filtering with the low-pass elliptic filter, only the useful part from 0 to 10 kHz was retained (there was a gear mesh frequency of the pump occurring around 6 kHz). Later the data was digitally low-pass filtered at 1 kHz and downsampled by a factor of 20 to give the frequency range of 0-1 kHz as the bearing fault frequencies lie in the range from 0-1 kHz. Varying lengths of the signals were obtained as per our requirements. Most of the analysis was carried out using the ADXL105 vibration signals. The reason for doing so was because this signal had a good anti-aliasing filter which removed most of the observation noise (unwanted signals above 10 kHz) before the digitization process took place. This provided the opportunity to study the signal without complications

associated with significant amount of noise corruption. In addition the ADXL105 vibration signals were better modelled by an AR process than the B&K vibration signals. This observation is presented with supporting results of autocorrelation plots in Section 4.3.4.5.

For an overview of the details of the robust and efficient data acquisition system (hardware and software) for signal acquisition and processing for the purpose of condition monitoring for the dry vacuum pump, please refer to the section on Condition Monitoring VI in APPENDIX B.

3.4. An Evaluation of an Alternative iMEMs Vibration Sensor

This section presents a brief discussion on the use of inexpensive Micromachined Integrated Micro Electrical Mechanical System (iMEMs) accelerometers such as ADXL105 as an alternative to piezoelectric accelerometers for obtaining reliable and predictable data for diagnostics. The work presented here has been published in (Thanagasundram and Schlindwein 2006c).

3.4.1. Instrumentation

It was desired to monitor simultaneously and continuously the health of the iGX100 dry vacuum pump at several measurement points. The use of piezoelectric accelerometers for instrumentation is not practical on the basis of cost, since each needs an expensive associated charge amplifier. Charge amplifiers are a required piece of auxiliary equipment for piezoelectric accelerometers which are necessary to transduce the vibration signal from capacitance to voltage form, change its gain and band-pass filter it.

Alternative solutions had to be investigated. The ADXL105 accelerometer is a second-generation surface micromachined device that carries a differential capacitive sensor and electronic signal conditioning circuitry on a single integrated chip. In a recent article (Doscher 1999b), it has been demonstrated that surface micromachined accelerometers have reduced the cost of real time monitoring diagnostics from \$1000 to \$100 US dollars per point. Moreover, it has been illustrated how effectively micromachined accelerometers can be used to monitor machine vibrations (Doscher

1997) and it was also suggested that surface micromachined capacitive ADXL105 accelerometers can rival the performance of more expensive sensors such as the conventional piezoelectric accelerometers (Doscher 1999a) as they have a more stable sensitivity as a function of frequency and temperature as compared to piezoelectric accelerometers. Consequently, it was decided that the performance of an iMEMS accelerometer ADXL105 from Analog Devices would be evaluated as an alternative vibration sensor.

The ADXL105 measures accelerations with a full scale range up to $\pm 5g$ (g being the acceleration of gravity), has a sensitivity of 250mV/g , has a on-board 'Uncommitted Amplifier' (UCA) which can be used to change its output scale factor with external resistors or to add a 1 or 2-pole active filter, has a $225\mu\text{g}/\sqrt{\text{Hz}}$ noise floor, a 0-10 kHz bandwidth of frequency response and an on-board temperature sensor which can be used for compensating against temperature effects for high accuracy applications. The zero- g voltage is 2.5V but it is ratiometric to power supply.

In this application, it was decided to keep the sensitivity at 250mV/g so that the full acceleration range may be used. With a 2.5 V zero- g voltage, the accelerometer can measure $2.5\text{ V} \pm 0.25 \times 5\text{ V} = 2.5\text{ V} \pm 1.25\text{ V}$ without saturating the output waveform with a supply voltage of 5 Volts and having a safety margin of $\pm 1.25\text{ Volts}$. Also a summing amplifier stage with a potentiometer was added using the UCA, so that the zero- g voltage can be set at precisely 2.5 V . To keep a stable supply voltage of 5 V , a regulator stage was also added. In the laboratory, the circuit was built for the ADXL105 as shown in Figure B-5 in APPENDIX B. Surface mount capacitors and resistors were used to keep the Printed Circuit Board (PCB) size small and for easier mounting purposes. The PCB with the soldered components was then fixed into a small Nylon plastic box and the space filled up with epoxy. A picture of the ADXL105 sensor is shown in Figure B-6 in APPENDIX B.

Piezoelectric accelerometers such as the Brüel and Kjær (B&K) 4370V accelerometer used as a comparison in this study have been around for many years. The desirable features of this type of accelerometer are its accuracy, low noise and operability at high temperatures. They are, however quite expensive and need extra signal conditioning circuitry such as preamplifiers. The specifications of the B&K

4370V accelerometer are compared to those of the Micromachined Integrated Micro Electrical Mechanical ADXL105 accelerometer in Table 3–1. Note that the working frequency range of the B&K piezoelectric accelerometer is from 0.1 to 4800 *Hz* and its resonant frequency lies outside this range at 16 *kHz*. So if there is a small excitation at 16 *kHz*, this will cause large amplified measurements at 16 *kHz* because of the resonance. For the ADXL105 accelerometer, the working frequency range is from 0.1 to 4800 *Hz* and the resonant frequency occurs at around 6.7 *kHz*.

Table 3–1 Piezoelectric Accelerometer Performance compared to ADXL105 as given in datasheet specifications

Accelerometer Property	Piezoelectric	ADXL105
Range	Up to 2000 g	±5 g
Sensitivity	100 pC/g	250 mV/g
Noise density	0.02 mg	0.225 mg
Temperature Range	-74 to 250°C	-40 to 85°C
Frequency Range	0.1 to 4800 Hz	0 to 10000 Hz (includes DC response)
Resonance	16 kHz	Around 7 kHz

3.4.2. ADXL Vibration Signals

The noise floor of the ADXL105 is proportional to square root of the measurement bandwidth required. As the measurement bandwidth increases, the noise floor increases and the Signal to Noise Ratio (SNR) of the measurement decreases. In this application, it was decided that the full frequency response up to 10 kHz would be monitored to include resonating frequencies for demodulated envelope spectra allowing detection of bearing defect frequencies. The minimum noise resolution for the ADXL105 was calculated from the following formula given in the datasheet (ADXL105 1999).

$$Noise(rms) = 225 \mu g (Hz)^{-1/2} \times (K \times B)^{1/2} \quad [3.1]$$

$$Noise(rms) = 225 \mu g (Hz)^{-1/2} \times (1 \times 10\,000)^{1/2} = 22.5$$

where K depends on the number of poles of the filter used with the uncommitted on-chip operational amplifier and B is the bandwidth of the measurement in Hz. Since there was no filter stage used in the circuit design, K is 1. Hence the minimum theoretical noise floor calculated for the ADXL105 accelerometer using Equation [3.1] is equivalent to 22.5 mg.

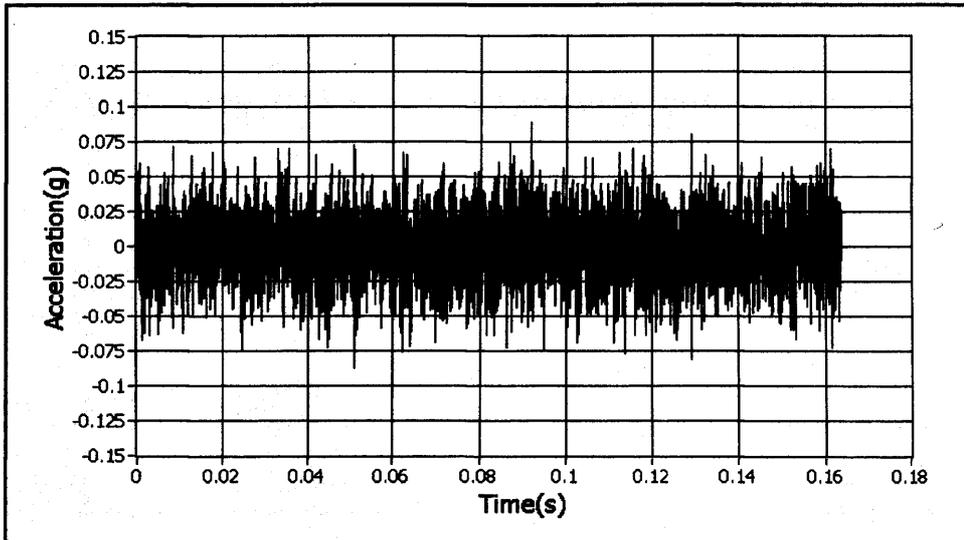
$$\text{Noise}(rms) = 20 \mu\text{g}(\text{Hz})^{-1/2} \times (B)^{1/2} \quad [3.2]$$

$$\text{Noise}(rms) = 20 \mu\text{g}(\text{Hz})^{-1/2} \times (10\,000)^{1/2} = 2$$

The equivalent theoretical noise floor can be calculated for the B&K piezoelectric Accelerometer using Equation [3.2] and is found to be 2 mg.

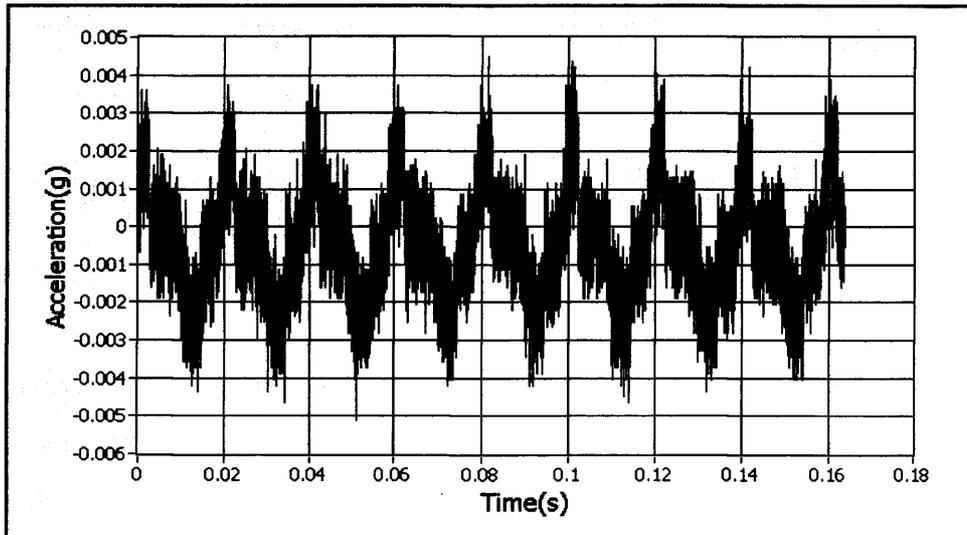
To compare the performance of the accelerometers, the noise levels in both accelerometers were quantified. The accelerometers were powered and mounted on the same measurement point on the pump (see Figure B-7 in APPENDIX B). The pump was not switched on. Spectral analysis of the acceleration signals gave estimations of the DC level (bin 0) and the AC level (the Root Mean Square (RMS) sum of the rest of the bins). These AC measurements are indicative of the noise present in both accelerometers. The experimental RMS value of ADXL105 noise measured in the laboratory (23 mg, as shown in Figure 3-1(a)) agreed very closely with the theoretical value of 22.5 mg calculated from equation [3.1]. In comparison, the RMS noise value of the Brüel and Kjær piezoelectric accelerometer (see Figure 3-1(b)) was measured to be 1.4 mg (this was slightly less than the value of 2 mg calculated using Equation [3.2]).

ADXL



(a) ADXL: RMS value of 23mg

B&K

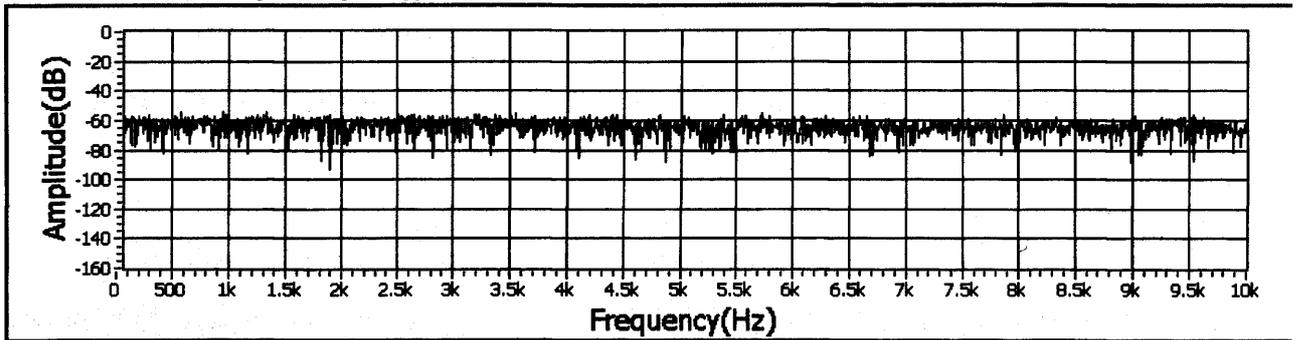


(b) B&K: RMS value of 1.4mg

Figure 3-1 Measurement of noise level in both accelerometers.

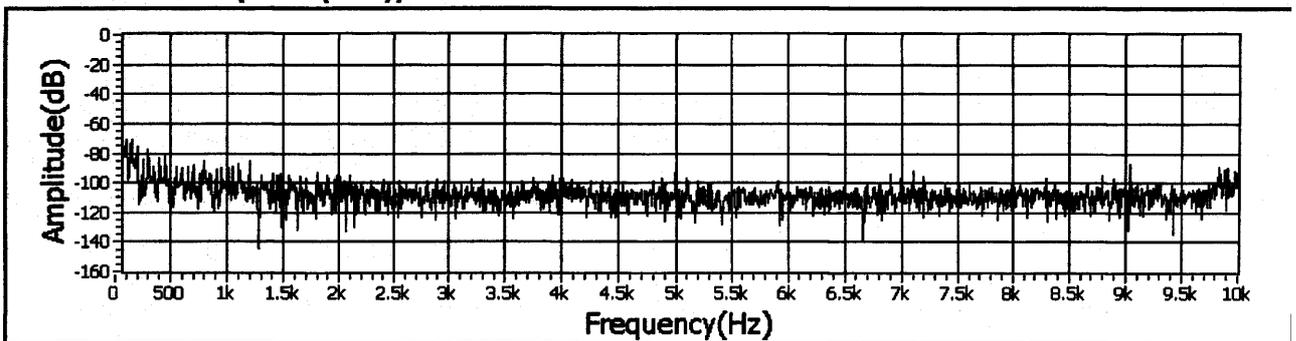
(a) ADXL105 and (b) Brüel& Kjær (B&K) 4370V accelerometer measured with 1.2% uncertainty in the testing environment (pump off, circuitry on). Waveforms shown are actual noise levels. RMS values found to give an indicative experimental noise level measurements. Notice the change of scale from (a) to (b).

ADXL Acceleration (FFT - (RMS))



(a) ADXL: -60dB noise floor

B&K Acceleration (FFT - (RMS))



(b) B&K: -100dB noise floor

Figure 3–2 Noise floor measurements of accelerometers.

(a)ADXL105 and (b) Brüel& Kjær (B&K) 4370V accelerometers in the frequency domain in the same testing environment (pump off, circuitry on).

Hence the measured noise level of the ADXL105 accelerometer was approximately 16 times greater than the noise level for the piezoelectric accelerometer. In the frequency spectrum, the ADXL105 noise floor was around -60dB and the Brüel and Kjær accelerometer noise floor was around -100 dB as shown in Figure 3–2.

The greater noise level of the ADXL105 limits the resolution of this accelerometer and thus its ability to detect small signal changes (Doscher 1997). This becomes important when measuring very low *g* amplitudes. This was not a major limitation for this study, as the pump is known to produce levels of vibration of up to ± 2 *g* (peak-to-peak) when running at 6000 RPM (revolutions per minute), or 100 *Hz*.

The accelerometer to be used for condition monitoring must be capable of accommodating peak-to-peak values of higher magnitude or at least two orders of the magnitude of vibration measured in the normal working conditions in order to be used

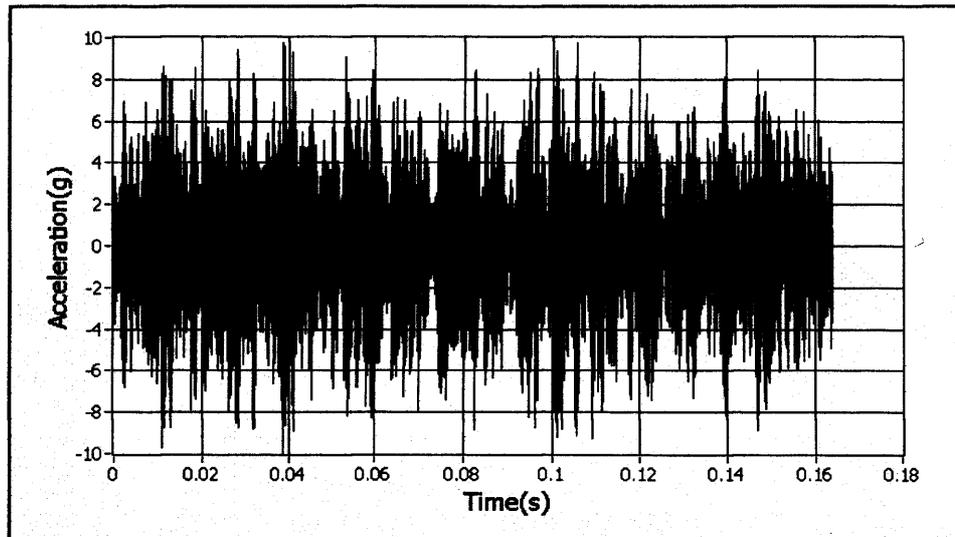
effectively in non-standard operations such as fault conditions. The higher noise level of the ADXL105 only becomes a major limitation when used in very low g level applications or when used to measure vibrations from very low speed rotating machinery (machines which have a rotating speed of much less than 100 Hz). Fault detection and diagnosis of low speed machinery (Barret 1993) is a subject of study itself where other types of transducers such as displacement transducers or proximity probes or special techniques such as pulse shock methods or high frequency resonance techniques (Prashad, Ghosh *et al.* 1985) have to be employed.

3.4.3. Resonance

It was also noted that the ADXL105 sensor exhibited resonance when the ADXL105 Cerpak package chip was not glued to the PCB. Resonance is due to the mechanical system made up of the mass of the package and the leads of the IC acting like "springs". Resonance is not desirable as it can saturate the output of the accelerometer. Much care had to be taken when soldering the ADXL105 chip and after soldering, epoxy glue was applied to stick the chip onto the PCB board to prevent the resonance. Figure 3–3 shows the time domain responses of the ADXL105 (a) when the chip was not glued and (b) when it was glued to the PCB.

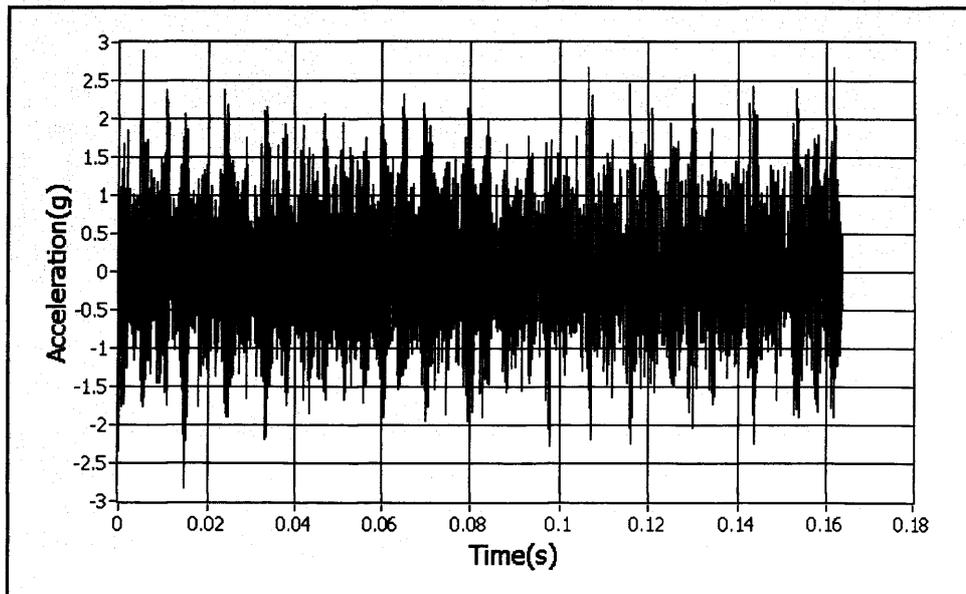
From the results, it can be seen that higher g levels are measured due to resonance when the ADXL105 chip is not glued to the PCB even though the pump was rotating at 110 Hz in both cases. The resonance can also be seen in the frequency domain spectrum in Figure 3–4(a). The vibration level of the ADXL105 accelerometer spectrum rises above the piezoelectric accelerometer spectrum around 6.7 kHz as measured in the experimental setup.

ADXL



(a)

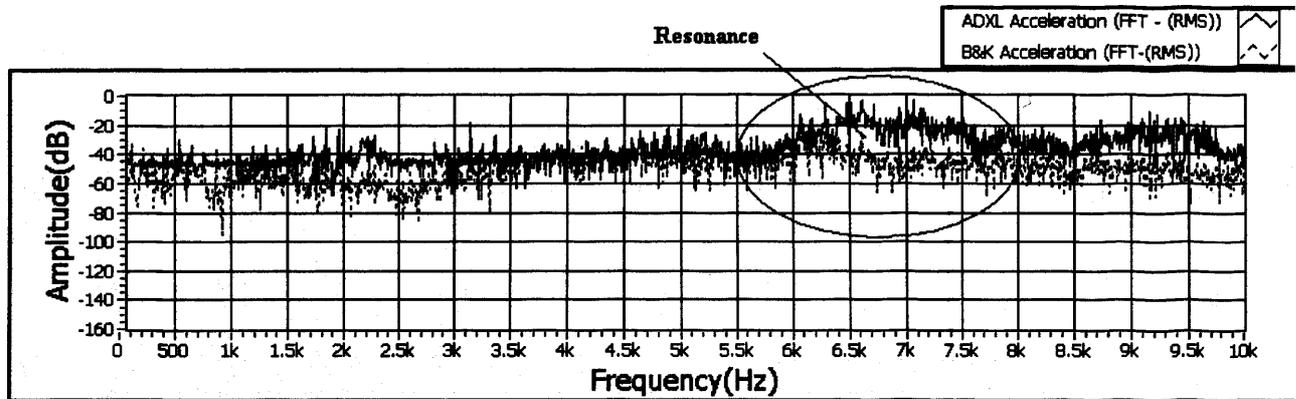
ADXL



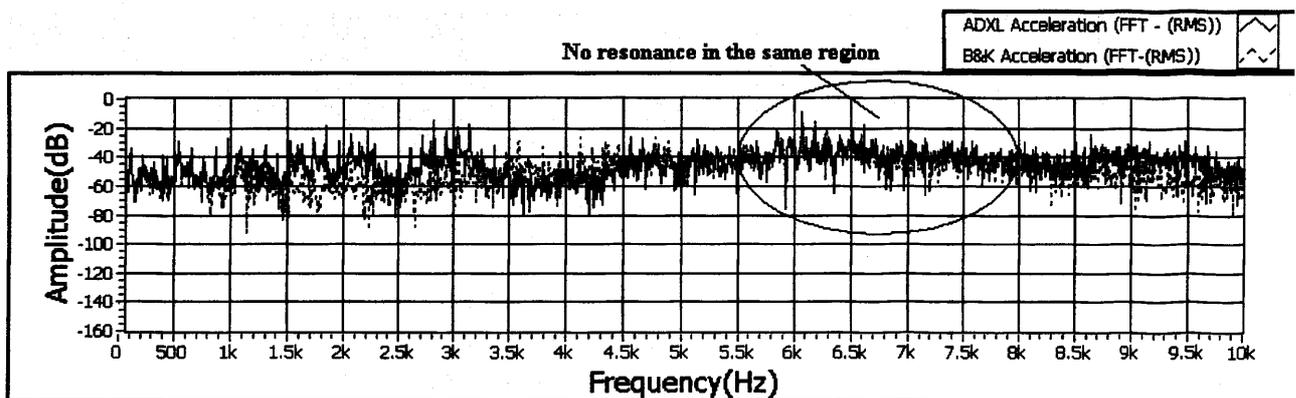
(b)

Figure 3-3 Resonance measurements in time domain.

(a) Resonance occurring when chip is not glued to PCB; and (b) when it is glued; Pump was rotating at 110 Hz in both cases. Notice the change of scale from (a) to (b).



(a)



(b)

Figure 3-4 Resonance measurements in frequency domain.

(a) Resonance occurring when chip is not glued to PCB (Rising of the ADXL105 vibration level near the resonating frequency 6.7 kHz and (b) when it is glued (the spectrum of the ADXL105 and that of the B&K piezoelectric accelerometer are very similar). The pump was rotating at 110 Hz in both cases.

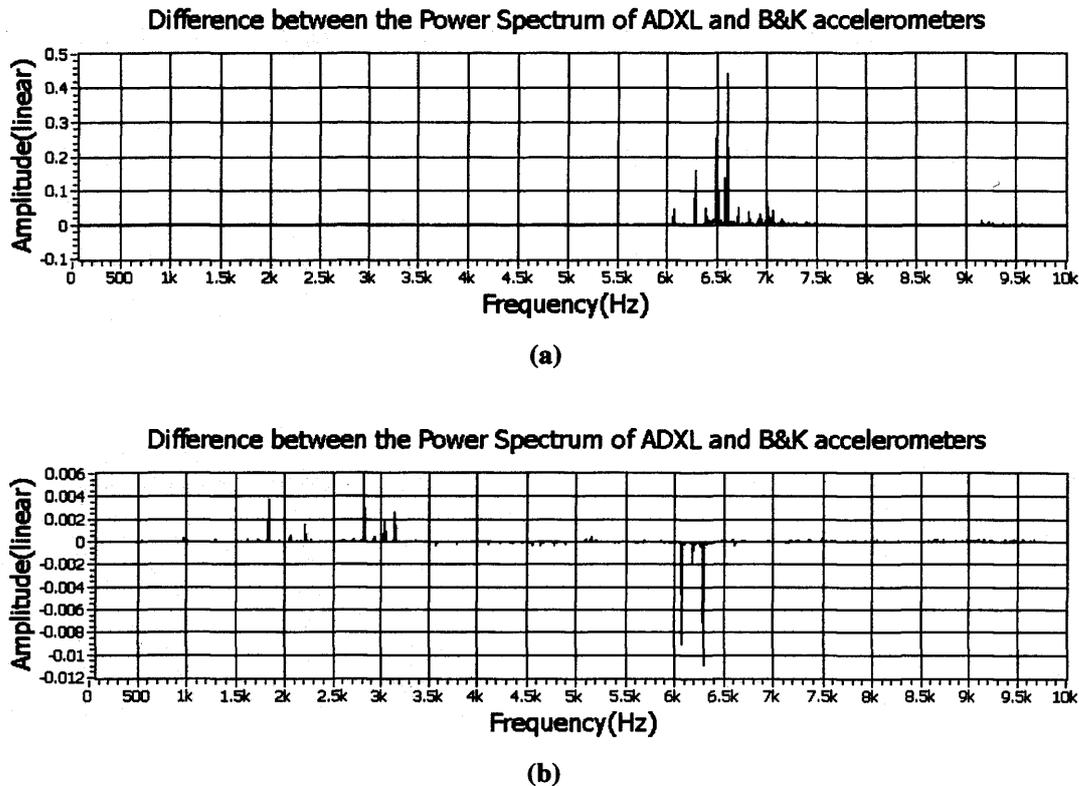


Figure 3–5 Difference between the frequency spectra of ADXL and B&K piezoelectric accelerometer in linear scale.

(a) Large differences in the amplitudes of the spectra are measured in the frequency range between 6 to 7.5 kHz when the ADXL105 chip is not glued to the PCB. (b) Only small differences in the amplitudes of the spectra are measured when the ADXL105 chip is glued to the PCB. The pump was rotating at 110 Hz in both cases.

In Figure 3–4(b), the spectra of both accelerometers are very similar when the ADXL105 chip was glued. The difference between the frequency spectra (obtained by subtracting the frequency spectrum of the ADXL accelerometer from the frequency spectrum of the B&K piezoelectric accelerometer) can be more clearly seen when the spectra are plotted in the linear scale. Figure 3–5(a) shows the difference in the spectra when the ADXL105 chip is not glued to the PCB and (Figure 3–5(b)) shows the difference in the spectra when the ADXL105 chip is glued to the PCB. When the ADXL105 chip is not glued to the PCB (Figure 3–5(a)), large differences as much as 0.5 g are measured in the frequency range between 6 to 7.5 kHz indicating resonance occurring around that region. When the ADXL105 chip is glued to the PCB, only small differences in the amplitudes are seen and these occurred around in the regions

between 1.5 to 3.5 *kHz* and 6 to 6.5 *kHz* but they are significantly much smaller and the largest difference is only measured to be around 0.011g (only about 1/5th of the magnitude of the difference measured when resonance had occurred).

3.4.4. Conclusion : ADXL105- A Low Cost Sensor Solution

The performance of the ADXL105 accelerometer has been verified and has been found to give the same quality of data as that of a piezoelectric accelerometer. The main advantages for the use of the ADXL105 accelerometer are its low cost, its ability to measure DC response, its better temperature stability and the presence of on chip signal conditioning circuitry. The main drawbacks would be that it cannot be used at high temperatures, it can resonate if proper mounting techniques are not adopted and it has a higher noise level than piezoelectric accelerometers. However, if proper measures are taken, all these limitations can be overcome and micromachined accelerometers can be used successfully for machine diagnostics, as validated by (Thanagasundram and Schlindwein 2006c). Since this is not a low g application, the higher noise level is not a great concern. The low cost of the ADXL105 sensors allows permanent sensor placement on multiple measurement points on the dry vacuum pump and makes it economically possible to extend on-line monitoring. This is an excellent solution for acquiring consistent, reliable and accurate data as many of the errors and inconsistencies of temporary mounting can be prevented. The ability to use more data also improves the success of automatic fault diagnostic techniques.

3.5. Finding the Running Speed of the Pump

The iGX dry vacuum pump is driven by a 3 phase, 2 pole AC asynchronous induction motor and there is also an inverter acting as variable frequency drive for controlling the pump's fundamental rotor rotational frequency. The relationship between the synchronous speed of the pump and the set reference speed is given by the following formula [3.3]:

$$\text{Synchronous Speed of motor (Hz)} = \frac{2 \times \text{Frequency}}{\text{Number of Poles}} \quad [3.3]$$

If, say, the reference frequency is set at 110 Hz and since the motor has 2 poles per phase, the synchronous actual running speed of the motor achieved should be 110 Hz. But in reality, the motor only achieves a speed slightly less than 110 Hz, for example 108 Hz. This is known as motor slip, is a function of the load, and is characteristic of

AC induction motors. These effects can be seen clearly in the vibration signature. This phenomenon was studied in the laboratory by setting the dry vacuum pump to increasing speeds from 50, 60, 70, 80, 90, 100 to 105 Hz and keeping the inlet pressure at 0 mbar whilst the outlet pressure was maintained constant at atmospheric pressure and the fundamental rotational speed of the shaft (which is equal to the actual synchronous speed achieved) was measured.

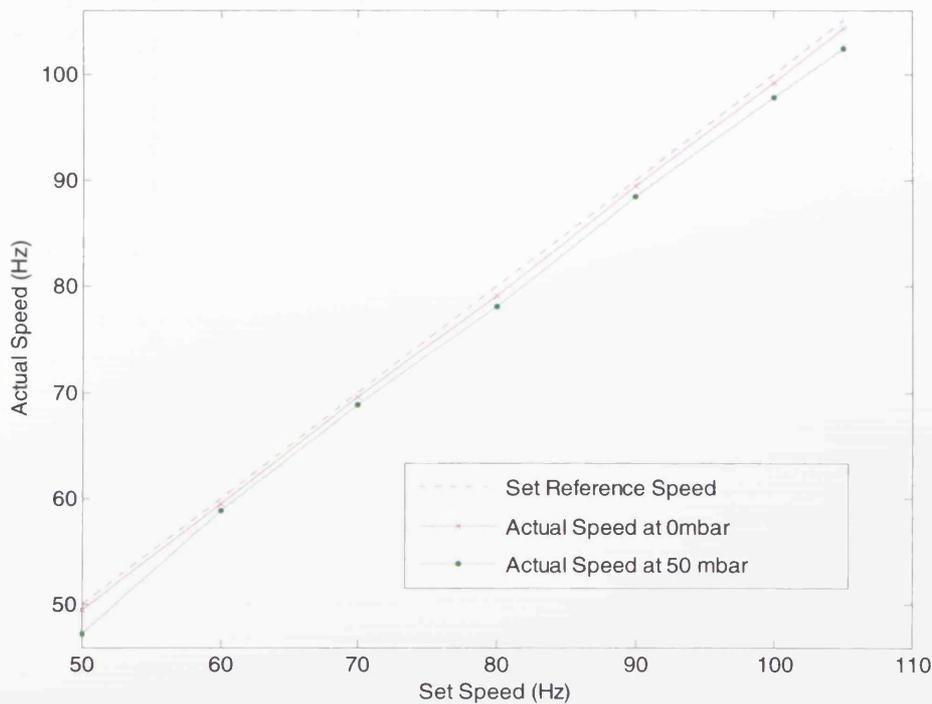


Figure 3-6 Set Speed versus Actual Speed of pump.

Actual Speed of Pump was always less than the Set Speed and the difference depends on the loading factor.

The inlet pressure set is taken to be equivalent to the loading factor of the pump. The fundamental rotational speed of the shaft (actual running speed) was noted by screening for the frequency of the first harmonic in the spectra of the vibration signature of the pump. The first harmonic is equal to the actual rotational speed achieved by the pump. Results are shown in Figure 3-6. It can be seen for that for any speed, the actual running speed was always less than the set speed. This is because of the motor slip. The vibration measurements for the pump were repeated under a higher

load condition (50 mbar). The slip increased with a bigger load as seen by the bigger difference between the set speed and actual speed achieved.

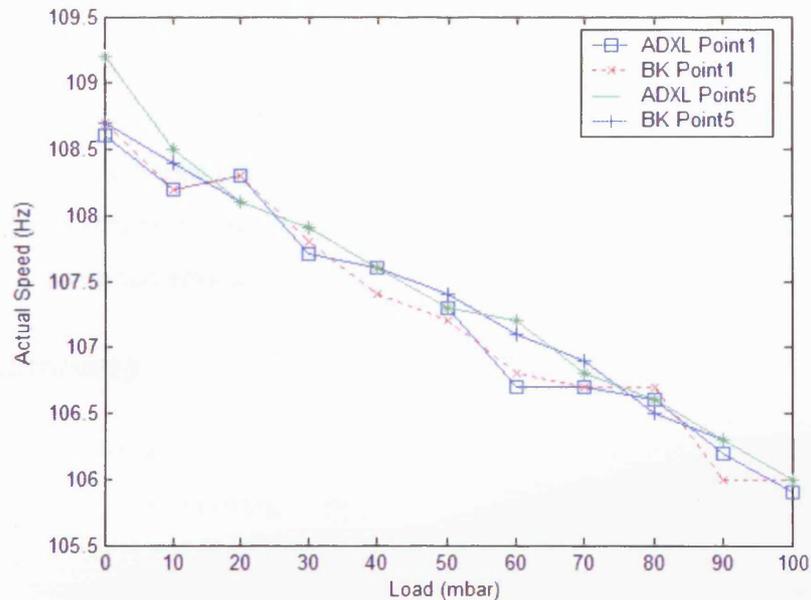


Figure 3–7 Pump speed set at 110 Hz but under increasing loads.

Inlet Pressure was increased from 0 to 100 mbar. Readings captured using two different accelerometers and at two different measurement points. It can be seen that as loading factor increased, the actual speed achieved by the pump also decreased because of increasing slip.

The experiment was repeated with the pump set at 110 Hz, inlet pressure was varied from 0 to 100 mbar and readings were taken from two different measurement points and using two different accelerometers. The accelerometers used are discussed more in detail in Section 3.4. Point 1 was at High Vacuum (HV) end and point 5 was the Low Vacuum (LV) end. Please refer to diagram Figure B–1 in APPENDIX B for the exact location of the measurement points. It was again observed that the higher the load applied to the pump, the more is the motor slip. Slip increased with a bigger load because of the higher mechanical damping experienced by the pump. The pump has to work harder when a higher load is set as it means that it pumps more molecules of air to achieve the required vacuum when the inlet pressure slip is higher. Slip causes the pump's total flow rate to be less than the theoretical flow rate, hence decreasing its efficiency and slowing down the pump.

Understanding these fundamental working principles of the dry vacuum pump is important as the in-depth knowledge gained, of the effects of different operating

conditions such as speed and load on the vibration of the pump, provided the foundation for the development of a sound fault detection tool right from the beginning of the project. It was understood that the first key step in the fault detection scheme was accurate determination of the running speed of the pump. The rotating speed of the pump's rotor shaft to which the bearing case was directly connected was often less than the set speed of the pump due to rotor slip as explained above. The effect of the actual speed was predominant in the detection of bearing faults as this was the speed that was used in the calculations of the bearing defect frequencies and had to be determined very accurately.

3.6. Summary

This chapter gave an overview of the data acquisition system for the dry vacuum pump condition monitoring system. The use of a surface micromachined accelerometer ADXL105 for spectral analysis of vibration data from a dry vacuum pump was also validated. The implementation details were explained and the schematics of the hardware circuits are given in APPENDIX B. It was also shown why it was necessary to accurately estimate the running speed of the pump for the calculation of the bearing defect frequencies in order to identify the presence of peaks associated to known faults in the spectra of faulty bearings.

4. SELECTION OF AR MODEL ORDER

4.1. Introduction

Model order selection is an important factor in parametric PSD estimation, affecting its performance as fault detection tool, and consequently, this aspect is investigated in this chapter.

In Section 2.3.7, a review was given of previous work applying parametric methods of AR modelling for fault detection and condition monitoring studies. However, none of the previously mentioned studies performed a detailed analysis comparing various different techniques for estimating the optimum AR model order. This is the main issue addressed in this chapter. The order selection criteria and some of the limitations in determining the optimum AR model order are analysed with supporting results.

Methods of AR order selection criteria such as AIC, FPE, MDL, CAT and FSIC are first introduced. The study consists of two main sections. In the first section, a study that was conducted to determine the optimum AR model order for vibration signals obtained from the dry vacuum pump is presented. It is desirable to minimise the computational complexity of the AR model by choosing the minimum value of the order that adequately represents the signal being modelled. This study consists of two parts. In the first part, the optimum order was determined for real data obtained from the pump. Then the optimum order determined was verified with simulated data representing “true” AR processes and of varying frame sizes to check whether the criteria indeed found the optimum AR model order. The contents of this chapter have been published in (Thanagasundram, Feng *et al.* 2005).

The second section presents a discussion on the formulation of a practical rule of thumb for the estimation of the minimum model order for Autoregressive (AR) based spectrum analysis for data from rotating machinery. The work in this section has been published in (Thanagasundram and Schlindwein 2006b).

4.2. AR Order Selection Criteria

Order selection criteria have been developed to indicate which model order to choose. Two of the most established criteria for order selection have been provided by Akaike. They are the Final Prediction Error (FPE) (Akaike 1969) and Akaike Information Criterion (AIC) (Akaike 1974). The FPE criterion [4.2] selects the order of the AR process so that the average error variance for a one-step prediction is minimised. In 1974, Akaike suggested an order selection criterion using maximum likelihood approach and this was the AIC criterion [4.1]. The AIC determines the model order by minimizing the information theoretic function. There have been reported to be many variants of the AIC method but the one used for this investigation is the original one with a penalty factor of 2. As N approaches infinity, the AIC and FPE criteria are asymptotically equivalent. In 1983, Rissanen developed the Minimum Description Length (MDL) [4.3] estimator (Rissanen 1984). It is called a minimally consistent criterion because $p \log_e(N)$ (the product of the AR model order and the log to the base of the exponential of the sample size) increases with N faster than with p . In 1974 another method was proposed by Parzen and is termed Criterion AR Transfer (CAT) Function (Parzen 1975) [4.4]. The 'optimum' order p is selected to be that in which the estimate of the difference between the mean square error of the true prediction filter and estimated filter is minimum. All these techniques for estimation of the AR order are termed as asymptotic information criteria.

Kay and Marple have cautioned that the results of spectra using these criteria have been mixed particularly when applied to actual data rather than simulated AR processes (S. Lawrence Marple Jr. 1987). The criteria presented can only be used as guidelines for initial order selection. They are known to work well with computer generated synthetic AR signals but may not work well with actual data, depending on how well such data can be modelled by an AR process.

All these criteria were designed to reduce the probability of under fit at the cost of over fit (Broersen 2000b). For all order estimation criteria listed in Equations [4.1] to [4.5], N is the sample size, k the trial model order and σ^2 the variance of the prediction error for the given model order. A cost function term in each of them,

known as the penalty term, penalises for the use of extra AR parameters above the ‘optimum’ order.

The performance of a more recent method of order selection criterion, Finite Information Criterion (FIC) (Broersen 1985; Broersen 1990; Broersen 1998; Broersen 2000b; Broersen 2000a; Broersen 2002; Broersen, De Waele *et al.* 2003) was also investigated. FIC has been claimed to perform better than the asymptotic criteria when the ratio p/N is large. Samples are called finite if $N < \infty$ and the ratio p/N is greater than a small value such as 0.1. This method has been stated to be more accurate than the asymptotic criteria for a finite number of samples as it has been shown by Broersen (Broersen and Wensink 1993) that the asymptotic criteria do not take into account the distinction that exists in practice between the different estimation methods. The FIC criterion calculates the variance coefficients depending on the method of estimation used hence this method is meant to be more robust than the other order selection methods. The FIC criterion is defined by [4.5] where v_i is the finite sample variance coefficient for the Yule-Walker method $v_i = (N - i)/\{N(N + 2)\}$.

$$AIC(k) = \ln(\sigma(k)^2) + (2k + 1)/N \quad [4.1]$$

$$FPE(k) = \frac{N + k + 1}{N - k - 1} \sigma(k)^2 \quad [4.2]$$

$$MDL(k) = \sigma(k)^2 \left(1 + \left(\frac{k+1}{N} \right) \ln(N) \right) \quad [4.3]$$

$$CAT(k) = \left[\frac{1}{N} \sum_{j=1}^k \frac{N-j}{N\sigma(j)^2} \right] - \frac{N-k}{N\sigma(k)^2} \quad [4.4]$$

$$FIC(k) = \ln(\sigma(k)^2) + 2 \sum_{i=1}^k v_i \quad [4.5]$$

The optimal model order is the order k that minimizes the above criterion Equations [4.1] to [4.5]. The variance of the prediction error is defined as

$$\sigma(k)^2 = \frac{1}{N-k} \sum_{n=k+1}^N \left\{ x(n) + \sum_{i=1}^k a_i x(n-i) \right\}^2 \quad [4.6]$$

All of the above-mentioned order selection criteria will be used to investigate the optimal model order selection for the vibration data obtained from the dry vacuum pump in the following sections.

4.2.1. Motivation

It has been pointed out earlier, in Section 2.3.5, that the disadvantage of the AR method is that the optimum model order is not known a priori and some experimentation with different orders is required before the right order can be selected for the given finite record of signal of length N samples (frame size). If the model order is too low, insufficient detail for confident fault diagnosis will be revealed in the spectra. If the model order is too high, then the spectral estimate is too peaky and spectral line splitting might occur (Fougere, Zawalick *et al.* 1976). Schlindwein and Evans have applied AIC, FPE and CAT model order criteria for spectral analysis of Doppler ultrasound signals (Schlindwein and Evans 1990). From this study, it has been concluded that, for spectral analysis of these signals, overestimating the model order is better than underestimating it. They also noted that that using a smaller frame size is more likely to produce an underestimation of the model order.

The task of finding the optimum order is not trivial. The main difficulty arising from using the order selection criteria is the need to apply the criteria to a large number of frames of data. The optimum order cannot be determined by applying it to a single frame of data because that frame might not be representative of the overall properties of the signal. Many frames of the signal have to be analysed before the optimum order could be determined accurately and conclusively. Some investigators compute probabilities while others use more graphical methods, such as histograms. The process becomes rigorous with the requirement of processing of orders for many frames of the signal.

Secondly, it is unwise to find the optimal order by the application of one order selection criterion alone. It has been reported by some researchers that certain criteria,

like FPE, tend to overestimate the model order (Dron, Rasolofondraibe *et al.* 2001). Criteria like the MDL (refer to Equation [4.3]) have the tendency to underestimate the model order. It is advisable to test the performance of a combination of criteria and, if all of them select the same order, then that order can be concluded to be the optimum order.

In this case, it was desired to keep the model order low as smaller orders translate to smaller processing requirements for spectral estimation. The computational time needed to calculate the autoregressive coefficients using the Levinson-Durbin recursive algorithm is proportional to p^2 , the square of the AR model order, therefore in real-time applications of the AR algorithm it is sensible to avoid high model orders in order to save unnecessary computation.

4.2.2. Method of Finding the Optimal Order

An assessment of the ability for estimating order of AR processes of frames of vibration signals of fixed sample sizes was investigated. The study consists of two main parts. In the first part, the optimal orders for describing 428 frames of 328 samples ($N=328$) of vibration signal using the five criteria were evaluated. The steps used in determining the optimum order for each of criteria are stated as follows. A maximum order, K , was chosen. In this case this maximum order was fixed to a value of 60 ($K=60$), as it was expected that the optimum order would not exceed this value for the chosen data set. Variance $\sigma(k)^2$ (refer to Equation [4.6]) of the prediction error was computed scanning orders from $1 \leq k \leq K$ using the Yule-Walker estimation method (Equation [2.16]). Then, for each specific criterion, the criterion values are obtained for the range $1 \leq k \leq K$ using Equations [4.1] to [4.5]. The optimum order is independently selected for each criterion as the minimum value for each method. The steps are repeated for all the frames in the signal as a Monte Carlo simulation.

The size of the data frames used for this analysis is $N=328$. The ADXL vibration signal captured from the High Vacuum end of the pump was used as the test signal.

4.2.3. Results and Discussion

4.2.3.1. Finding the Optimal Order

Figure 4–1 (a) shows the plot of how the variance of the prediction error for every model order p has been transformed into order criteria $AIC(k)$, $FPE(k)$, $MDL(k)$, $CAT(k)$ and $FIC(k)$ for orders $1 \leq k \leq K$. It can be seen that for low model orders, all the criteria have close values. A significant finding is that all the criteria have a local minimum occurring at an order around 25 for the given AR signal and sample size. Hence for this particular frame, the best order M is equivalent to 25.

For higher model orders, the criterion values do tend to deviate. However for some of the frames, FPE and CAT criteria chose higher “best” orders. This can be explained if one looks closely at Figure 4–1 (b). The penalty factors built in all the AR order selection criteria are monotonically increasing functions. In this particular frame, there was a kink occurring at order around 41. The penalty factors built in AIC, MDL and FIC criteria weren’t affected much by this kink and these three criteria selected order $p=25$ as the local minimum. FPE and CAT whose penalty seem to be increasing more slowly with k , select the next best local minimum as the “best” order. Since the optimum order was earlier defined to be minimum criterion value in the range $1 \leq k \leq K$, these two criteria indeed select the higher order as the “best” order though this was not supposed to be the case. At this point the influence the choice of the maximum order K has on the order selection should be questioned. If it had been selected somewhere around 40, this problem could have been avoided and all the criteria would have given similar results. It can be seen how sensitive asymptotic criteria like FPE and CAT can be to the maximum order selected. The performance of these two criteria has also degraded because a small sample size was used. In fact it is known that FIC was the only criterion designed to work well independently of the choice of K and with small number of observations.

Figure 4–2 shows order versus time by the five respective order selection criteria for the 428 frames of data tested. It can be confirmed that the FIC criterion is the best criterion for order selection and it was found to outperform the other criteria when the sample size was small (Broersen and Wensink 1993). The FIC criterion exhibits good performance when the sample size is finite. In this case, the sample size

is $N=328$. Hence $N < \infty$ and the ratio p/N (maximum order $K=60$ and $p/N = 60/328$) is greater than 0.1. The FIC criterion also selects the optimal model order independent of the maximum order K . The FIC criterion has performed considerably well and selected the “best” order of 25 for most of the frames. AIC and MDL have almost identical performance, except at some points where the selected order differs. The performance of FPE and CAT criteria was poor. There is always a split between the orders selected. It is either 25 or 45. This can be explained by the earlier discussion for Figure 4–1.

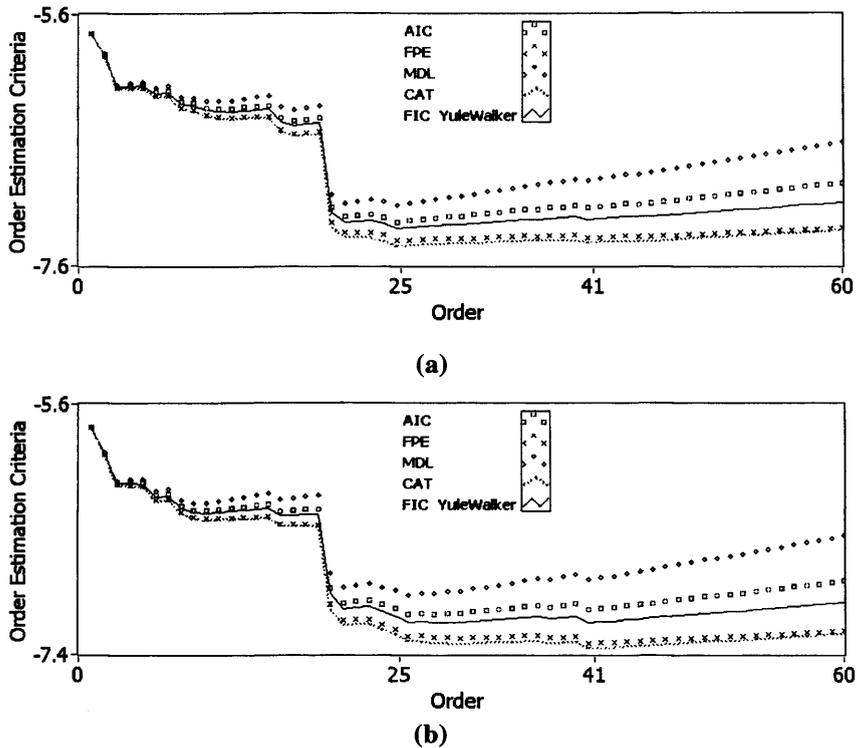


Figure 4-1 Order selection criteria given as a function of model order using the Yule-Walker estimation method.

A particular vibration dataset (sample size $N=328$) was used. (a) Here $K=60$ and the optimum order $M=25$ for all the order selection criteria. (b) Same simulations as used before but using another data frame (sample size also $N=328$). For this particular frame, the optimum order M selected for $AIC(k)=MDL(k)=FIC(k)=25$. But $FPE(k)$ and $CAT(k)$ chose a slightly higher order of 41.

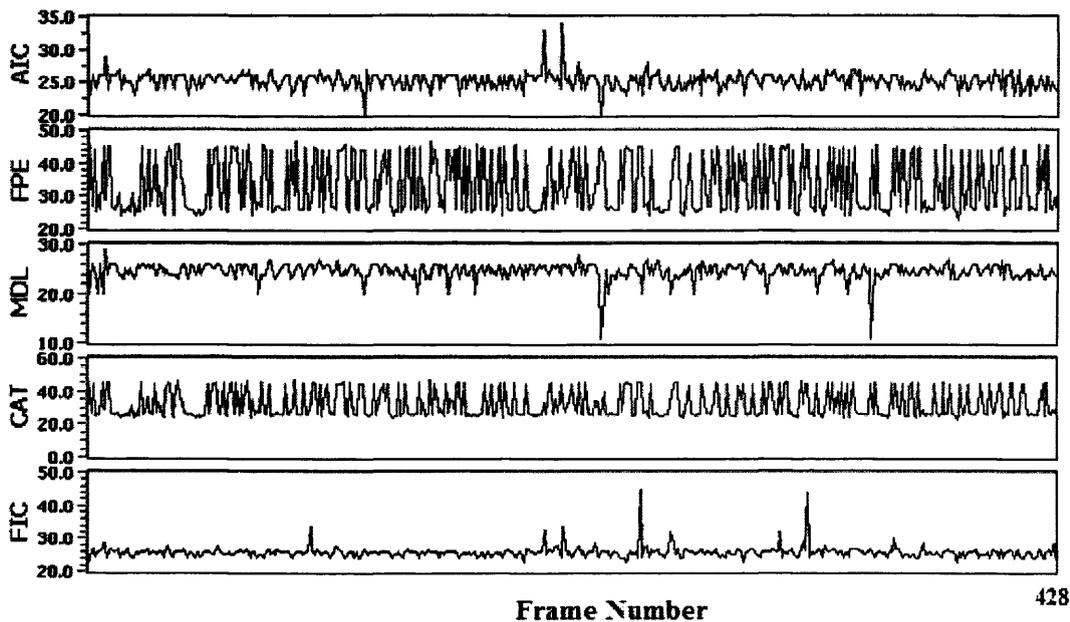


Figure 4-2 Analysis of the order of AR model for the various order estimation criteria. 428 frames of vibration signal each with a frame size of $N=328$ were used.

The distribution of model orders was then plotted as histograms of order occurrence versus model order. The order with the highest occurrence is identified as the “best” order, M for each order selection criteria. The occurrence of orders as histograms is given in Figure 4–3. The same trend of results can be observed here statistically. The distribution for each of the above criteria in selecting the optimum order is 42%, 27%, 28%, 30% and 46% in that order. The highest order selected does not exceed 45 for more than 98% of the frames. For the MDL criterion, the highest order selected is 23 but there is a clustering of orders around 23, 24 and 25. The splitting behaviour of the orders can also be observed for the FPE and CAT criteria.

4.2.3.2. Verification of the Optimal Order

From the previous investigation, for 428 frames of this particular vibration signal, the optimal order M is determined to be 25. The method used by investigators Schlindwein (Schlindwein and Evans 1989) and Boardman (Boardman, Schlindwein *et al.* 2002) is used to verify the ability of the criteria in identifying the optimal model order. In this method, a “true” AR process of known order (in this case, the optimal order determined earlier) is first generated by using coefficients extracted from the actual signal (this is done in order to preserve the signal characteristics) . Then all the five criteria are used to predict the (known) AR order. This is repeated for 1000 realizations of “true” AR processes each created with AR coefficients corresponding to a particular selected real signal. This allows an assessment of the abilities of the five criteria in estimating the correct order and the statistical performance of the different order selection criteria in identifying the optimum order.

In our case, from one typical frame the 25 a_k coefficients are obtained and used to generate 1000 sets of simulated series representing the AR process of various lengths $N=300, 500$ and 800 samples. The first 200 samples in each realization are discarded to minimise transient effects. Then the order selection criteria are again applied to the 1000 simulated AR processes to test their performance. The effect of increasing the sample size on the performance is also investigated.

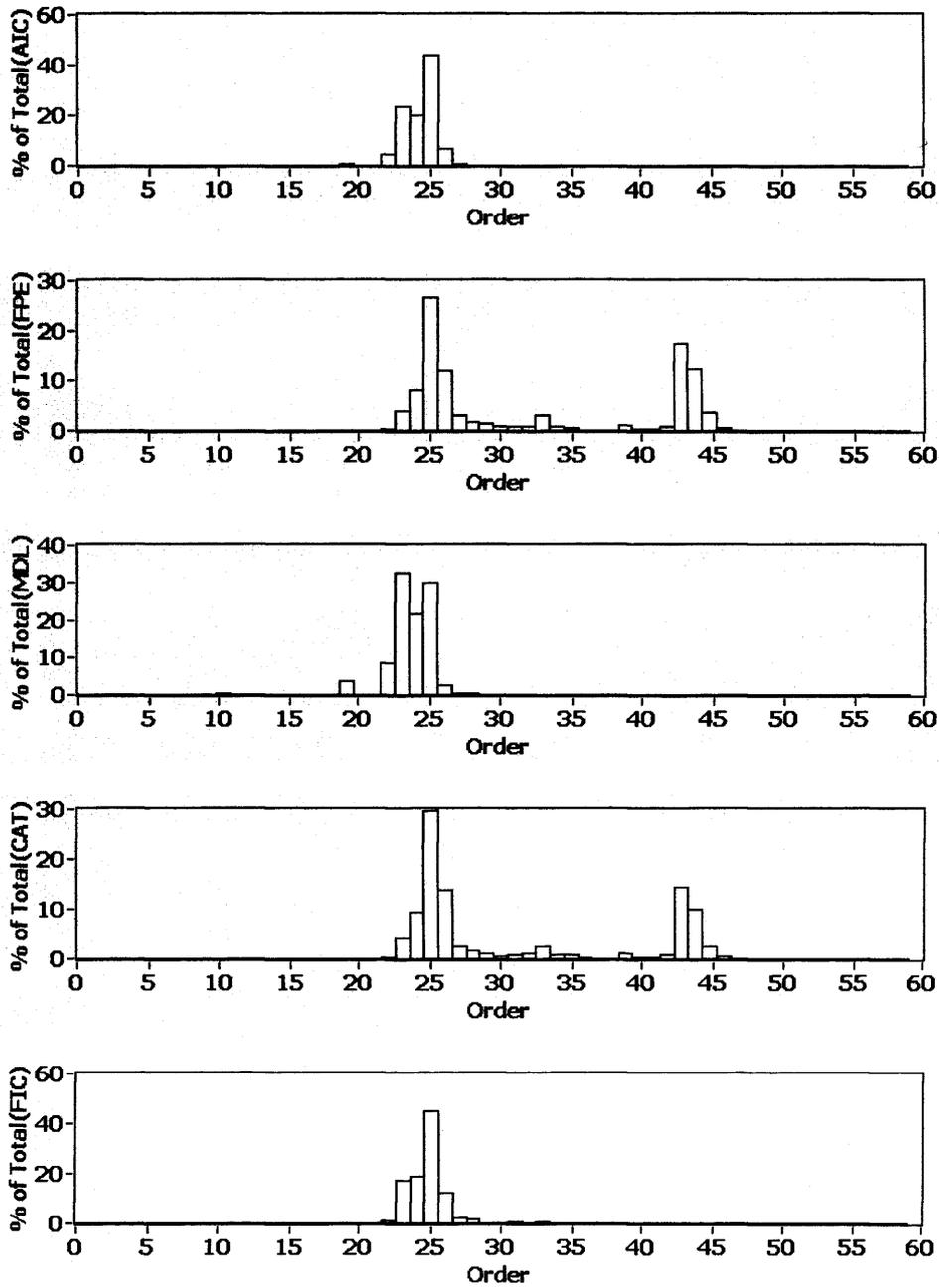


Figure 4-3 Histograms of "best" order.
428 frames of vibration signal of length N=328.

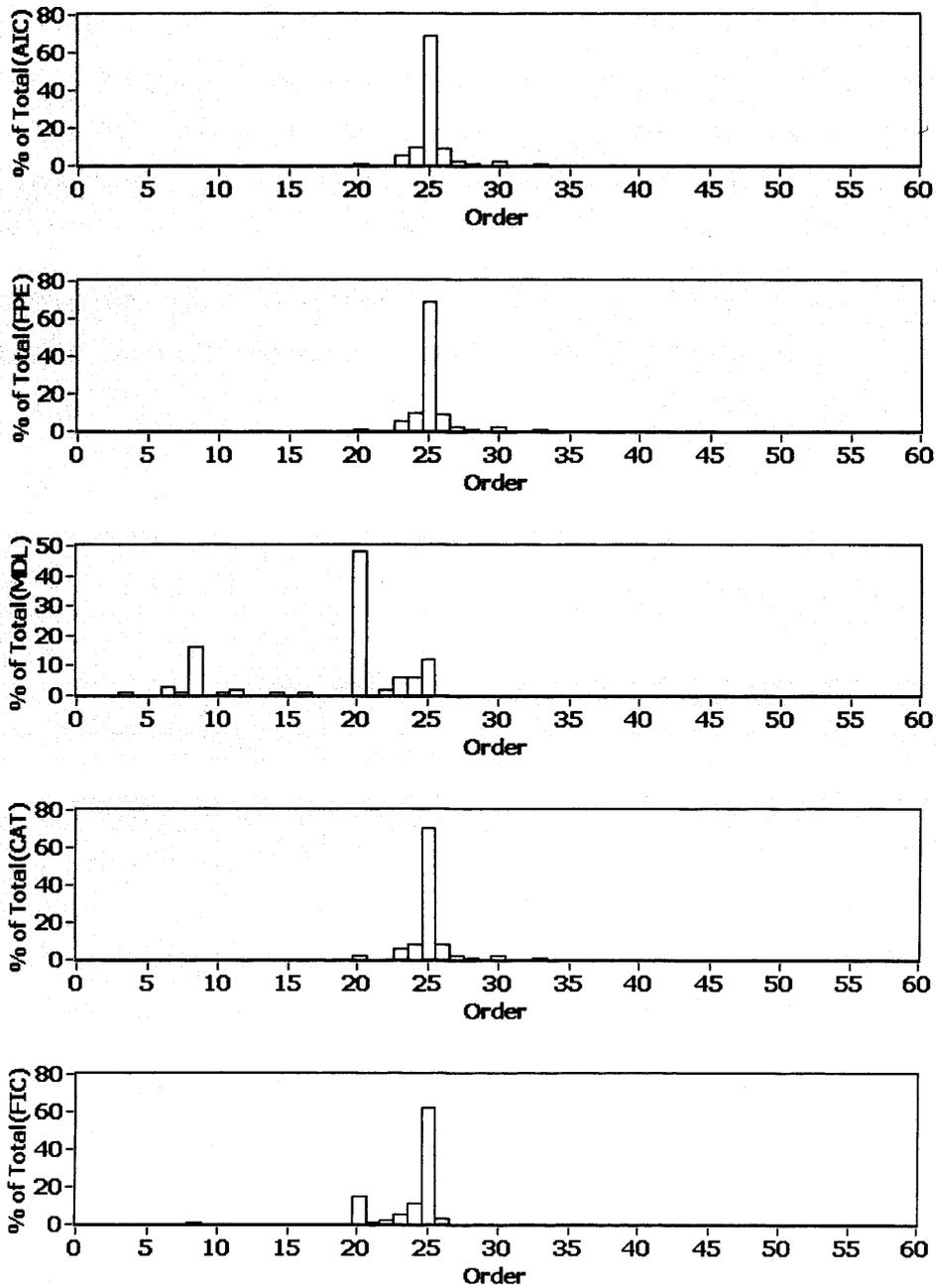


Figure 4-4 Verification of order prediction criteria using a “true” AR process.

M=25 and N=300.

In all cases the histograms have a peak corresponding to the “best” order $M=25$ except in the case of MDL, which estimated order $M=20$ (Figure 4–4). This is expected of MDL as it belongs to a class of criteria which tend to select under-fitted models. From the spread of orders, it is seen that overestimation of order is negligible in all cases. Underestimation of order occurs less frequently as sample size increases. This is good as over-fitting is less problematic than under-fitting. Performance is also improved for the MDL criterion, as expected, as the sample size increases (see Figure 4–5). The distribution of the highest order selected, in this case the optimal order of 25, increases from 70% to 90% as sample size increases from 500 to 800 samples.

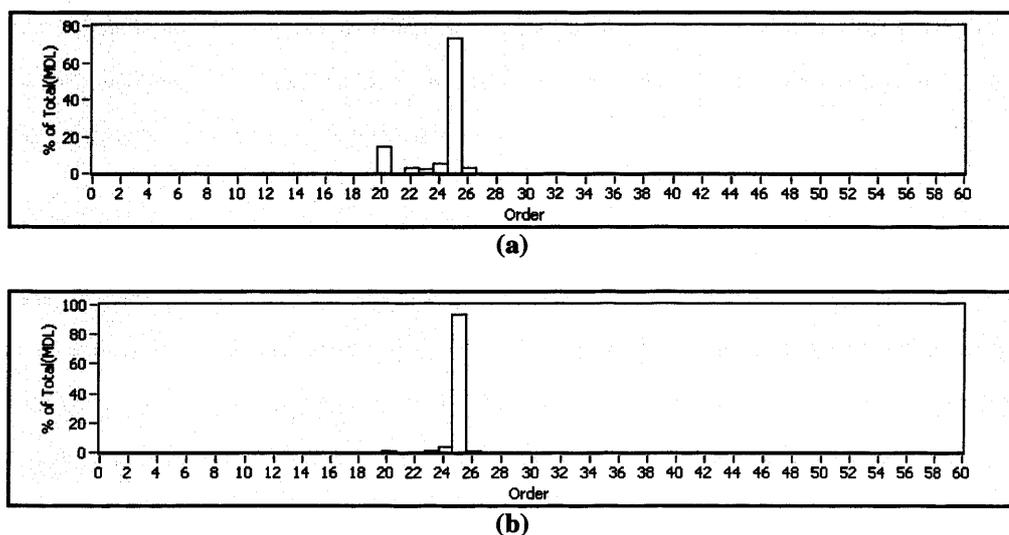


Figure 4–5 Effect of increasing sample size for MDL criterion.

(a) $N=500$ and (b) $N=800$.

4.2.4. Conclusion: Optimum AR Model Order Found for Vibration Signals

The optimal AR model order was estimated via five different order selection criteria for the vibration signal. Results show that the minimum order required can be as low as 25 and that for 90% of the frames the model order does not exceed 45 for a sample size of $N=328$. The findings show that all five criteria investigated perform well in determining the optimum order with little to choose between them unless when a small number of samples are available. Asymptotic criteria like AIC, FPE, MDL and CAT perform well only when the sample size is large. If the number of observations used is small, the ability of the order selection criteria in identifying the correct autoregressive order may be affected. In such cases, higher orders may be selected due to the occurrence of a second local minimum. The findings of this study can provide useful insight on the use of order selection criteria for finding the optimum order for any desired signal for fault diagnosis purposes.

It should be noted that in this study the optimum AR model order was determined for raw test vibration signals obtained from a pump with no faults. The test signal was also not demodulated. This was because the investigation was conducted at the beginning of the PhD studies and there was no faulty data available at the time when this study was conducted. However, this investigation was repeated once faulty data was available. It was found that the optimum order can be as low as 10 for the demodulated faulty vibration signatures and this optimal model order was the subsequent model order used for the implementation of the fault detection tool based on the AR pole trajectory.

4.3. Proposition of a New and Easy Approach to Find the Optimal Model Order

The focus of this section is to introduce an easy rule for estimation of the minimum model order for Autoregressive (AR) based spectrum analysis for data from rotating machinery. As a result of the work done in the previous Section 4.2, it was realised that one need not use traditional model order selection criteria such AIC, FPE, MDL, CAT and FIC to estimate the optimal order. Though these asymptotic criteria for

model order estimation are useful, a more straightforward method can be used, especially in the case of rotating machinery.

4.3.1. Motivation

From our earlier study on order selection criteria in Section 4.2, a clear repetitive trend was seen from the plots. Functions of all above mentioned order selection criteria had a steep decrease at a certain order. It was found that this minimum order p_{\min} was a function of the sampling rate, f_s , and the rotating speed of the machine $f_{machine}$. and this behaviour was found to be repeatable at different sampling rates and different rotating speeds.

4.3.2. Problem Formulation

The minimum order for rotating machinery, p_{\min} , was proposed and it is simply the ratio of the sampling rate, f_s to the rotating speed of the machine, $f_{machine}$, i.e.

$$p_{\min} = \frac{f_s}{f_{machine}}.$$

The explanation is simply that the minimum model order corresponds

to the number of samples collected over a full turn of the machine, that is, p_{\min} . The suggested AR model order p_{\min} is the number of sample points corresponding to one shaft revolution. This finding is not surprising as the AR model can be seen equivalent to the standard multiple linear regression model. A current sample of the signal is estimated as a linear summation of p previous samples where p is the model order. Hence the number of samples used for the regression must be equal to at least the number of samples in one complete revolution of the signal for the prediction error to be low. This is the minimal model order.

The Box-Jenkins methods of order determination using autocorrelation and partial autocorrelations plots are also used for justification of the selection of this minimal order. Most of the tests were carried using the vibration signals from an ADXL105 accelerometer (Thanagasundram and Schindwein 2006c). Comparisons were also made the B&K vibration signal. It will be shown that, although the B&K vibration signal has a higher Signal to Noise Ratio (SNR) than the ADXL105 vibration signal (as shown by noise floor measurement results in Section 3.4.2), the

B&K signal is less well modelled by an AR process, as it has an autocorrelation function which slowly decays with time. The argument is supported by experimental results and suggests that p_{\min} can be used as an initial optimal order for parametric modelling.

4.3.3. Methodology

The speed of the pump, f_{machine} , was varied at a fixed 0 mbar loading factor and test signals were collected with different sampling rates. The theoretical values for p_{\min} were worked out for the various speeds and sampling rates and are given in Table 4–1.

Table 4–1 Working out the theoretical p_{\min} values for various rotating speeds and sampling rates

f_{machine} (Hz)	f_s (Hz)	N_s (Number of Sample points per revolution)	$2N_s$	$3N_s$
100	2000	20	40	60
100	4000	40	80	120
100	5000	60	120	180
80	2000	25	50	75
80	4000	50	100	150
80	5000	75	150	225
60	2000	$33\frac{1}{3}$	$66\frac{2}{3}$	100
60	4000	$66\frac{2}{3}$	$133\frac{1}{3}$	200
60	5000	$83\frac{1}{3}$	$166\frac{2}{3}$	250

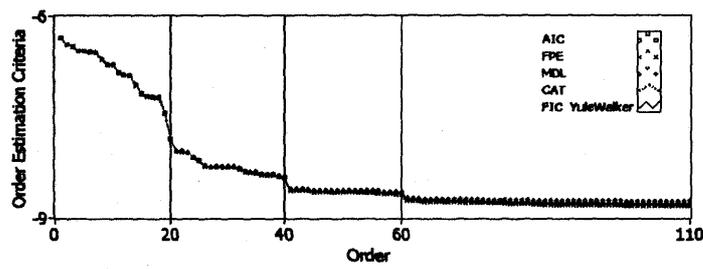
4.3.4. Results and Discussion

4.3.4.1. ADXL Signal at Increasing Rotating Speeds

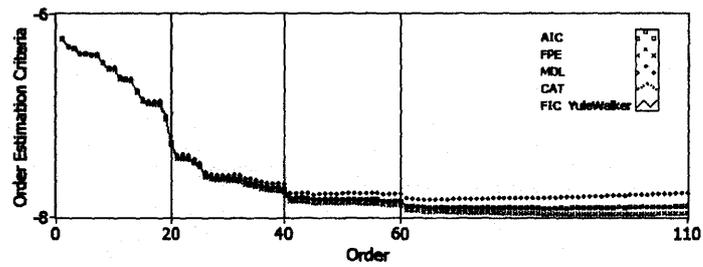
Order selection criterion values were calculated using equations [4.1] to [4.5] for the vibration signals obtained from the pump using the ADXL105 accelerometer. When data was acquired for the analysis to generate the diagrams in Figure 4–6, the speed of

the pump $f_{machine}$ was kept at 100 Hz. The sampling rate f_s was 2000 Hz for Figure 4–6 (a to c). p_{min} is 20 (refer to Table 4–1). Only the frame size N was varied. Looking at Figure 4–6(a), it can be seen that the behaviour of all order selection algorithms was very similar. Initially they showed a dramatic drop in their criterion values. Then this decrease becomes more gradual. The point where this occurs is the p_{min} order of 20. The criterion curves then flatten out and remain relatively constant until order 40 ($2p_{min}$) is reached, where the curves for the 5 order selection criteria show another sudden decrease. A third slight decrease in the criterion values is again observed at $3p_{min}$, at order 60. After order 60 there is no further ‘step’ decrease in prediction error.

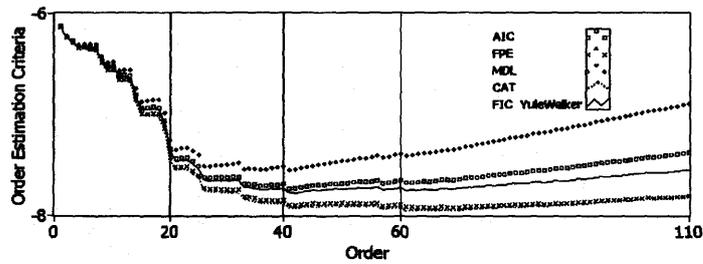
From the above we can conclude that there is a relationship between the prediction error (all the order selection criteria are functions of the prediction error) and the number of samples per revolution N_s (which is the same as p_{min}). It is expected of the prediction error of an AR model to decrease monotonically with the order, but looking at Figure 4–6(a), we can clearly see that the decrease in prediction error has occurred in step changes at multiples of p_{min} . For Figure 4–6(b and c), the criterion values do begin to increase. This effect is more clearly seen in Figure 4–6(c). This happened because the frame length was decreased from 4 s to 1 s and 0.25 s respectively. The number of samples N used for the order selection estimation has an effect on the behaviour of the criteria. If the ratio N/p is large, the penalty factor inbuilt in each of these order selection criteria has a greater effect. This explains the increase in the values observed in Figure 4–6(c) as a small frame size was used.



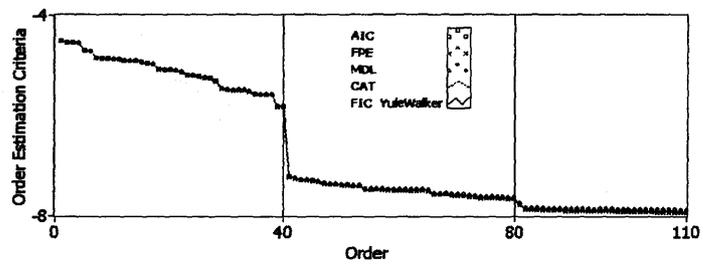
(a)



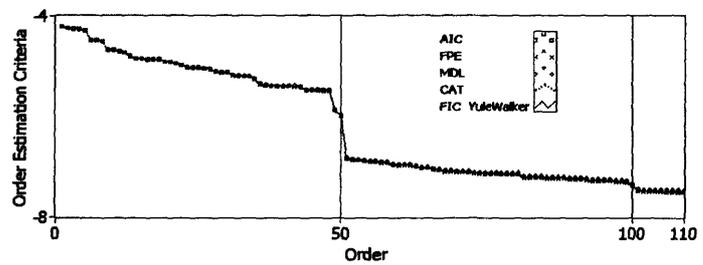
(b)



(c)



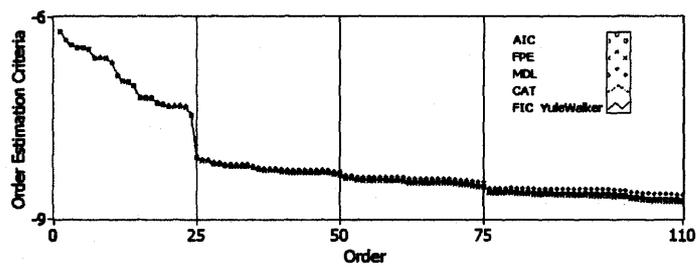
(d)



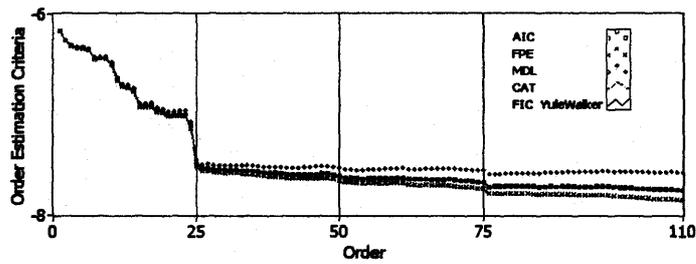
(e)

Figure 4–6 Behaviour of order selection criteria for ADXL105 vibration signals (100 Hz).

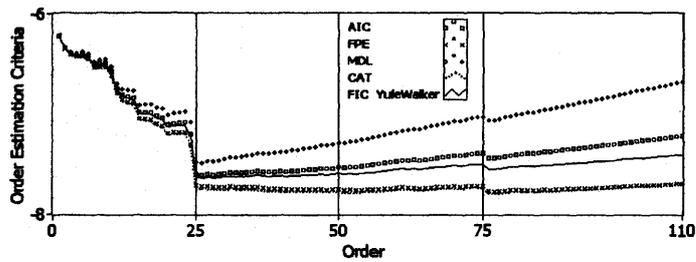
The speed of pump $f_{machine}$ was fixed at 100 Hz. Sampling rate f_s remained constant at 2000 Hz but the length of frames was set at 4 s, 1 s and 0.25 s respectively for figures (a) to (c) in that order. For (d) and (e) the length of frame was 4 s but sampling rate was 4000 Hz for (d) and 5000 Hz for (e).



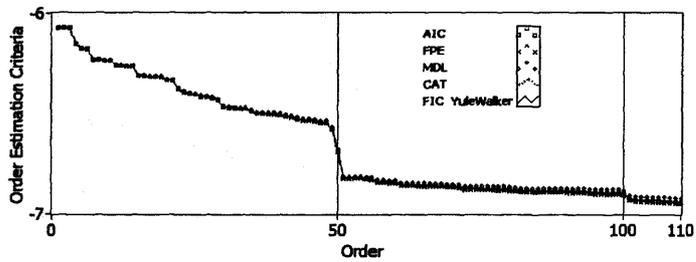
(a)



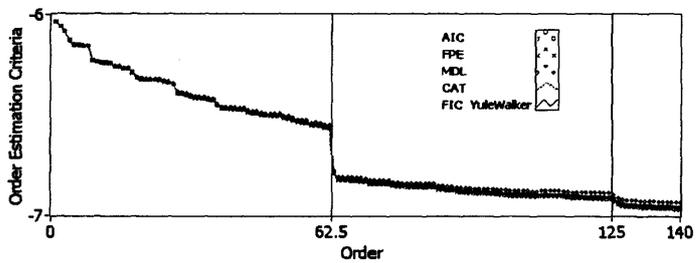
(b)



(c)



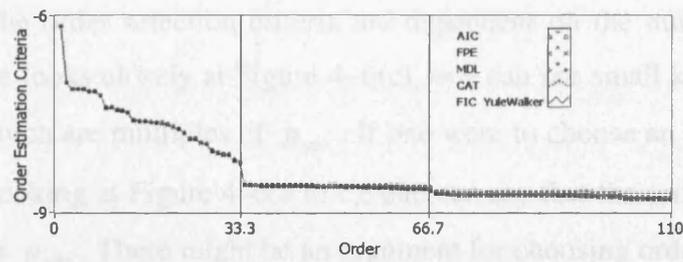
(d)



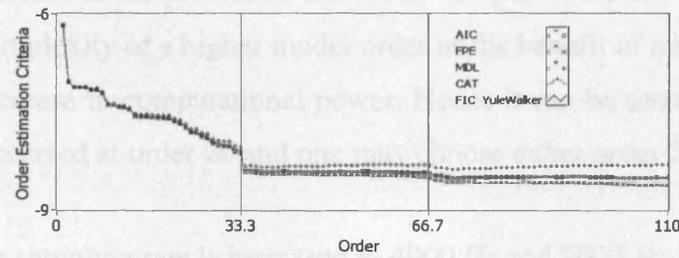
(e)

Figure 4-7 Behaviour of order selection criteria for ADXL105 vibration signals (80 Hz).

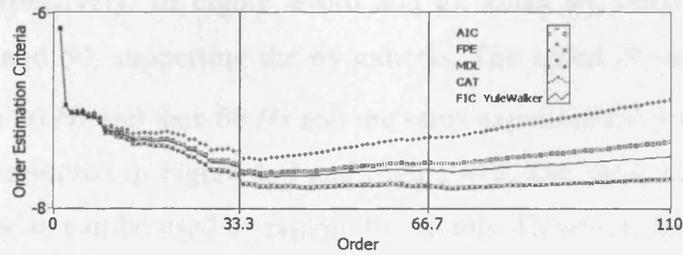
The speed of pump $f_{machine}$ was fixed at 80 Hz. The sampling rate f_s remained constant at 2000 Hz but the length of frames was set at 4 s, 1 s and 0.25 s respectively for figures (a) to (c), in that order. For (d) and (e) the length of frame was kept constant at 4 s but sampling rate was 4000 Hz for (d) and 5000 Hz for (e).



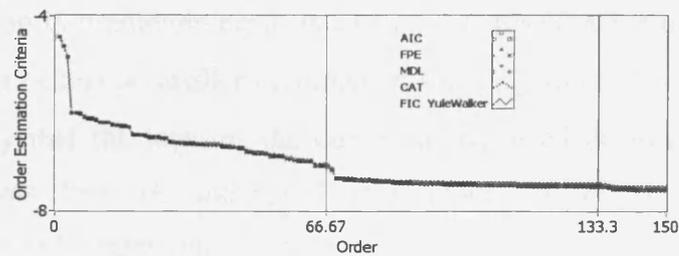
(a)



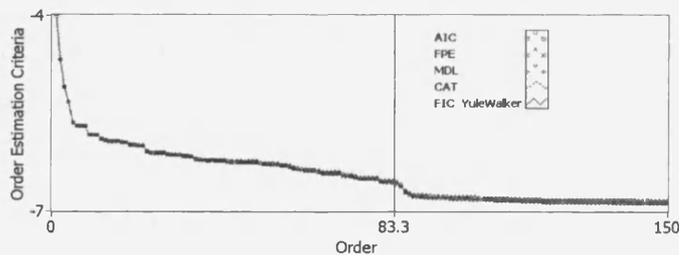
(b)



(c)



(d)



(e)

Figure 4–8 Behaviour of order selection criteria for ADXL105 vibration signals (60 Hz).

The speed of pump $f_{machine}$ was 60 Hz. Sampling rate f_s remained constant at 2000 Hz but the length of frames was set at 4 s, 1 s and 0.25 s respectively for figures (a) to (c), in that order. For (d) and (e) length of frame was kept constant at 4 s but sampling rate was 4000 Hz for (d) and 5000 Hz for (e).

Clearly the order selection criteria are dependent on the number of samples. Even then, if one looks closely at Figure 4–6(c), one can see small kinks occurring at 20, 40 and 60 which are multiples of p_{\min} . If one were to choose an optimal order for AR modelling, looking at Figure 4–6(a to c), one can say that the order cannot be less than 20, which is p_{\min} . There might be an argument for choosing order 40 as there is a further slight decrease in the prediction error. Choosing an order above 60 is not worth the additional complexity of a higher model order as the benefit of a better fit gained is not worth the increase in computational power. Hence it can be seen that the ‘knee of the curve’ has occurred at order 20 and one may choose either order 20 or order 40.

When the sampling rate is increased to 4000 Hz and 5000 Hz, the value of p_{\min} is 40 and 50 respectively. In Figure 4–6(d and e), kinks are observed occurring at multiples of 40 and 50, supporting the hypothesis. The speed of the machine f_{machine} was decreased to 80 Hz and then 60 Hz and the same experimental procedure repeated. The results are presented in Figure 4–7 and Figure 4–8. The same kind of analysis as that discussed above can be used to explain the results. However there was one effect to be noted. The size of the steps in the prediction error changed with the speed of the machine. The drop in prediction error was much less for 60 Hz than for 100 Hz. The knee of the curve occurs at smaller multiples of the p_{\min} value. Looking at Figure 4–8(a), we can say that the knee of the curve has occurred at around 33 (since the sampling rate was 2000 Hz and speed of machine was 60 Hz). If one had no knowledge of the order selection criteria, they could have easily estimated the optimal order from the p_{\min} formula we propose here because the order selection criteria had also predicted orders close to this value as verified by findings in Section 4.2.3.1. Using the p_{\min} formula, one can easily predict a ballpark figure of optimal order 20 for the 100 Hz vibration signal. The optimal order found by the order selection criteria for the 100 Hz vibration signal was 25 and this value is not far from the value predicted by the method proposed here. It should only be noted that if a higher speed of machine was used, then there is some advantage in using twice or thrice p_{\min} because of the characteristic behaviour that was observed.

4.3.4.2. ADXL Signal Spectrum – Effect of Increasing Frame Length

The section exploits the effect the frame length has on frequency resolution of the AR frequency estimates. Optimal model orders that were earlier determined using the order selection criteria are used and it is verified whether they are the right orders required to model this ADXL vibration signal's basic behaviour. For Figure 4–9, the speed of the pump $f_{machine}$ was fixed at 100 Hz and the sampling rate f_s was 2000 Hz. The figures on the left hand side ((a), (c), (e), (g) and (i)) were obtained for a frame length of 0.25 s while the figures on the right hand side ((b), (d), (f), (h) and (j)) were obtained for a frame length of 4 s.

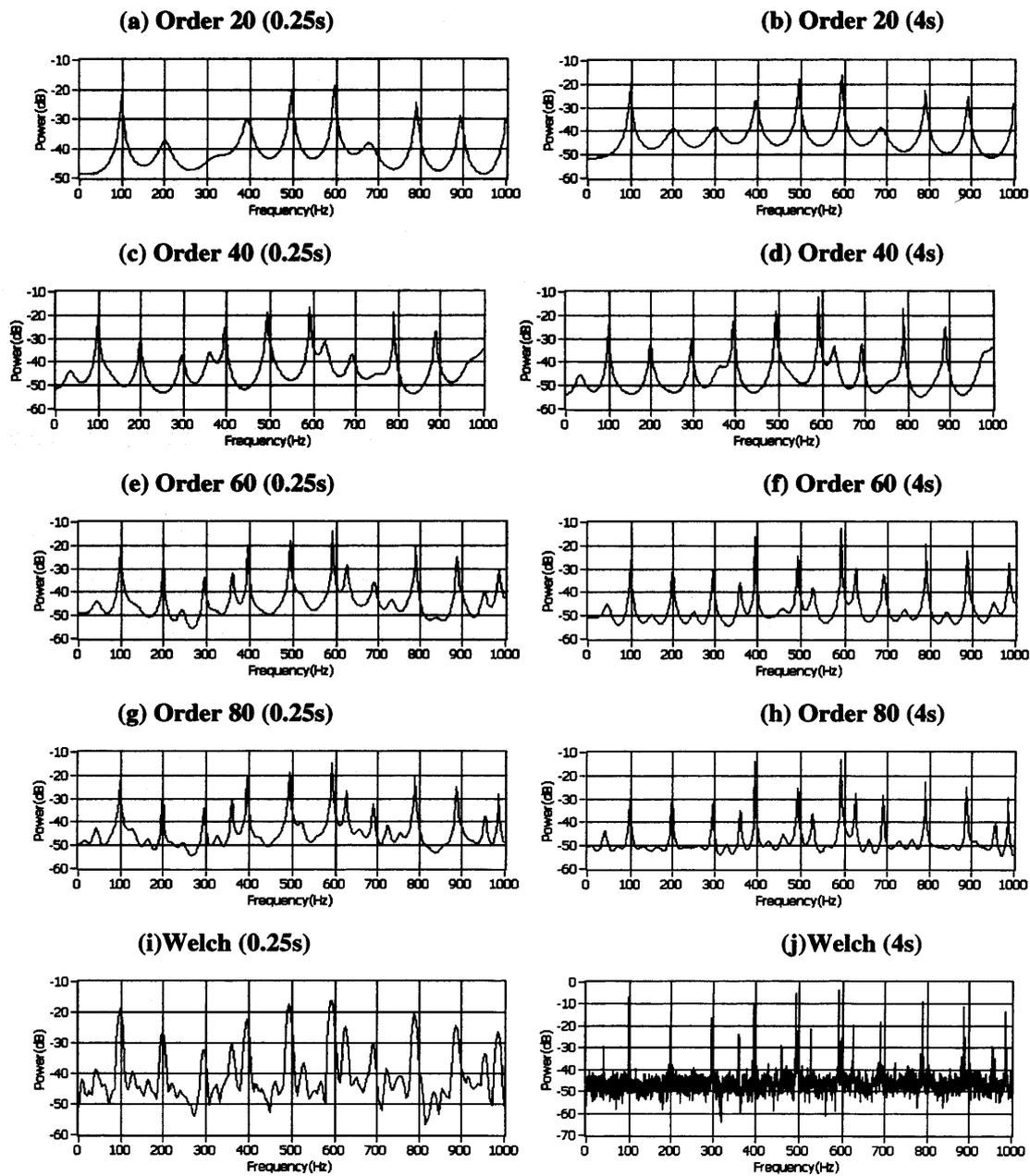


Figure 4-9 Spectra of ADXL105 vibration signal.

Speed of pump $f_{machine}$ was 100 Hz. Sampling rate f_s was 2000 Hz. Frame length was 0.25 s for those figures on left hand side and 4 s for figures on right hand side. Figures 5(a) to (h) show AR frequency estimates of order 20, 40, 60 and 80 respectively. 5(i) and (j) shows FFT frequency estimate obtained using the Welch method.

As the order was increased from 20 to 60, the resolution of the spectra improves. As order is increased from 60 to 80, there is not much improvement in terms of resolution, but only a slight increase in the PSD variance. In fact Figure 4–9 (e and g) and correspondingly Figure 4–9 (f and h) look very similar. Hence order 60 is sufficient for this machine speed and sampling rate. The Welch method, an averaged modified FFT periodogram, was used to obtain Figure 4–9(i) and Figure 4–9(j). A Hamming window and 50% overlap were used to obtain these frequency estimates. The frame size used was $N=500$ and $N=8000$ samples for Figure 4–9(i) and Figure 4–9 (j) respectively. The sizes of the sections were 256 and 4096 accordingly. There is some significant difference between the spectra in Figure 4–9(i) and Figure 4–9(j). The reason for this is that a much larger frame size N was used in Figure 4–9(j) than in Figure 4–9(i). Hence, the strong dependence of the FFT-based methods on frame size is clearly observed. It is well known that FFT-based techniques require large frame sizes to provide accurate frequency estimates. AR-based techniques can work with smaller frame sizes and hence lead to an improvement in the time resolution.

4.3.4.3.ACF and PACF Plots for ADXL Signal at Increasing Rotating Speeds

Statisticians use ACF and PACF plots to predict the model order of AR models. The ACF and PACF were defined in Section 2.3.6. Box and Jenkins have suggested that the examination of the behaviour of the Auto-Correlation function (ACF) and Partial Auto-Correlation function (PACF) of a time series can give information for the identification of the right type of model for its analysis and also aid in the selection of the right model order (Box, Jenkins *et al.* 1994). The plot of the ACF is an indication of the randomness in the data. The periodicity of a signal can also be seen in its ACF plot. If a signal contains a periodic component with period P , a peak in the curve occurs at integral multiples of P . The ACF for an AR(k) model has form of exponential decay or a damped sinusoid or a mixture of both. The partial autocorrelation function helps to determine the order of the AR process. If a time series is a AR(k) process, then the PACF plot for the signal converges to zero for orders greater than k .

For a time series of N observations, Bartlett's approximation sets the 95% confidence region at $\pm \frac{2}{\sqrt{N}}$. These approximate confidence bounds provide limits to help judge the statistical significance of the AR parameters calculated. If a parameter is outside the confidence interval limit, then it can be concluded the optimum order of the AR process has not been reached yet. Once they are within the limits, the residuals of the model are white and the optimum order has been reached.

4.3.4.4. Determination of Order from ACF and PACF plots

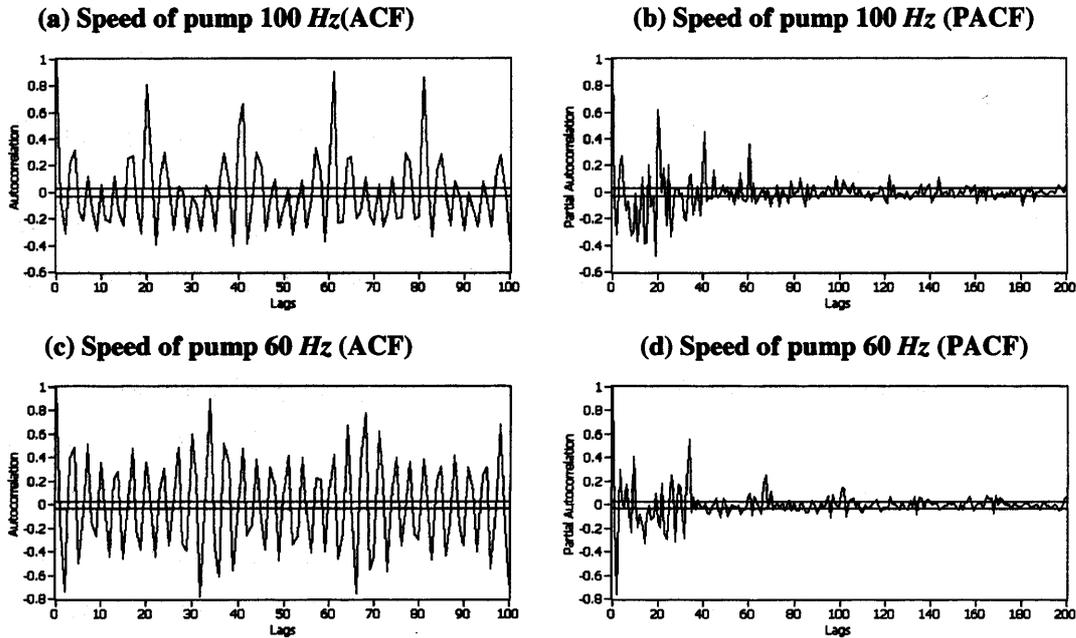


Figure 4–10 Determination of order from ACF and PACF plots

Speed of pump $f_{machine}$ was set at values as those indicated in the figures. ADXL105 vibration signal acquired at a sampling rate f_s of 2000 Hz. Length of frames was 2 s. Horizontal lines denote 95% confidence interval level.

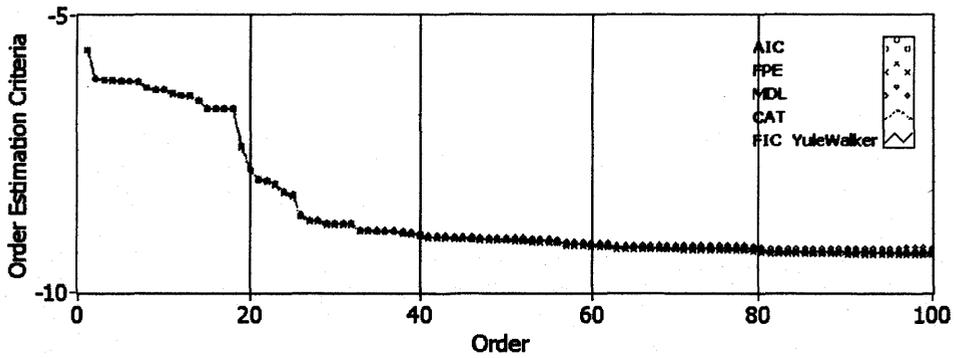
The vibration signals from this ADXL105 transducer are representative of a true AR process, as can be seen from their autocorrelation and partial autocorrelation plots. Figure 4–10(a and c) show the ACF plots of the vibration signals from the ADXL105 when the speed of the pump $f_{machine}$ was set to 100 Hz and 60 Hz respectively. The corresponding PACF plots are shown in Figure 4–10(b and d). The periodicity of the pump can be clearly seen in the ACF plots. The repetitive peaks occur at 20 and 33.3 accordingly. Looking at Figure 4–10(a and c), it can be seen that ACF has a periodic behaviour with the periods occurring at multiples of p_{min} . The envelope between the repetitive peaks is a decreasing exponential. It is to be noted that as the speed of the machine is increased, the process resembles a true AR process better. This can be seen in the well defined ACF plot in Figure 4–10(a). The order of the process is the point where the PACF plot cuts off. A 95% confidence limit has been used to judge the point where the function has died off. Referring to Figure 4–10(b), spikes can be seen occurring at 20, 40 and 60. At orders (or lags) above 60, the PACF becomes white noise-like. The same behaviour is observed in Figure 4–10(d). Here the spikes occur at multiples of

33.3. The amplitudes of the spikes decrease and there is a more marked decrease as the speed of the machine increases. An optimal order for Figure 4–10(d) would be 33 and this is the p_{\min} value for the vibration signal at that sampling rate and machine speed (refer to Table 4–1). This is in line with the behaviour observed with the order selection criteria plots.

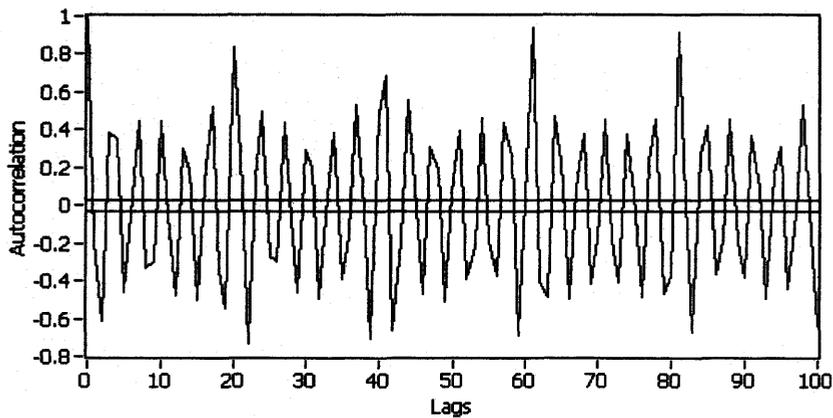
4.3.4.5. Brüel & Kjær (B&K) Vibration Signals at 100 Hz and 60 Hz

Experiments were repeated, in the same test conditions, using vibration signals obtained from the pump with a Brüel and Kjær (B&K) piezoelectric accelerometer. The specifications of the accelerometer were given in Section 3.4.1. This accelerometer has a frequency response from 0 to 4800 Hz and a smaller noise density specification of 0.02 mg/Hz (where g is the acceleration of gravity) compared to the ADXL105 accelerometer. It was described in Section 3.4.2 that the SNR of the vibration signal obtained using the Brüel and Kjær accelerometer is higher than that of the ADXL vibration signal. It was investigated whether the proposed technique would also work with its signal.

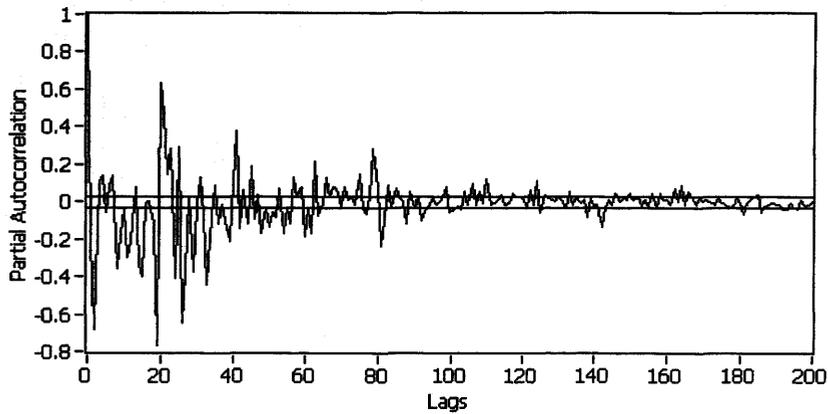
Results are shown in Figure 4–11 and Figure 4–12. The same behaviour is observed. The result that the optimal order would occur at p_{\min} also works for this signal. However, the statistical properties of these signals, mainly the autocorrelation plots, were different from the ADXL105 vibration signals. The decrease in prediction error at multiples of p_{\min} is less marked, as the SNR of this signal is higher. The Brüel and Kjær vibration signal has a slowly decaying ACF function (Figure 4–11(b)) which is indicative of a signal that is less well modelled by an AR process. The PACF plot of Brüel and Kjær vibration signal at 100 Hz cuts off below the 95% confidence limit only at high model order above 80 ($4 p_{\min}$). For the 60 Hz signal, the PACF plot never cuts off below the 95% confidence limit but the biggest decrease in value occurs around 33 (approximately around the p_{\min} value).



(a)



(b)

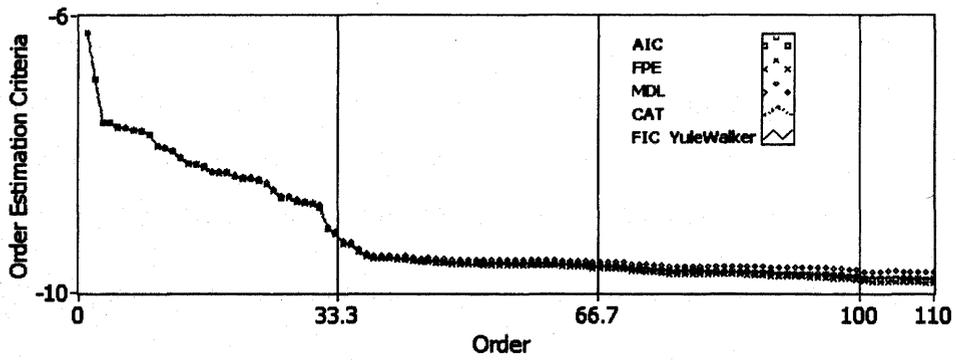


(c)

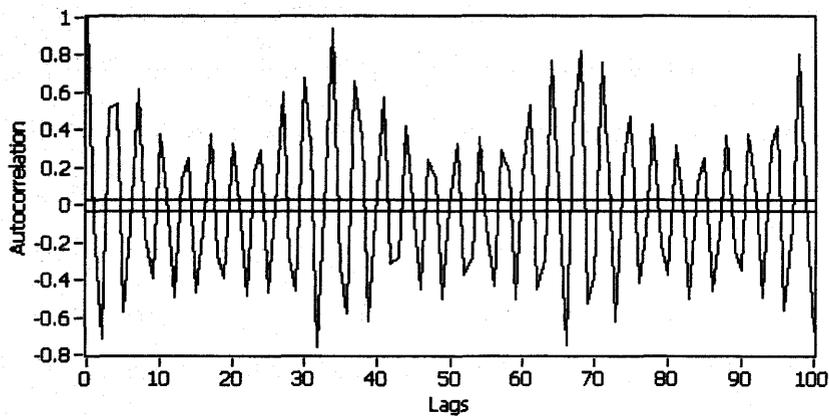
Figure 4–11 Vibration signal from Brüel& Kjær (B&K) at 100 Hz.

Speed of pump $f_{machine}$ was 100 Hz. Sampling rate f_s was 2000 Hz. Frame length was 2 s.

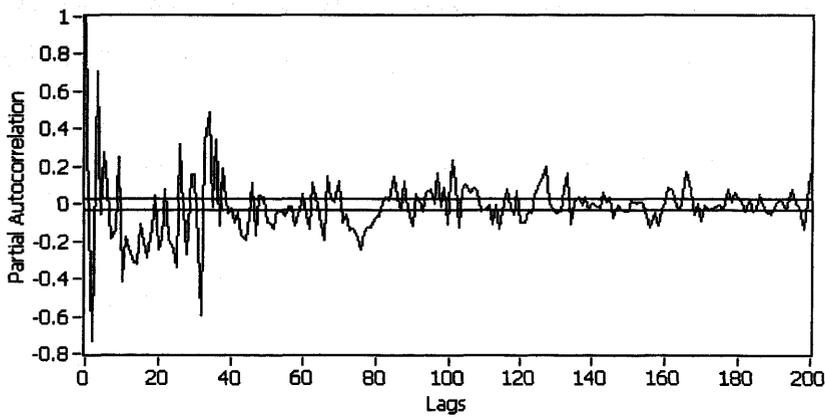
(a) Order selection criteria (b) ACF plot and (c) PACF plot.



(a)



(b)



(c)

Figure 4-12 Vibration signal from Brüel& Kjær (B&K) at 60 Hz.

Speed of pump $f_{machine}$ was 60 Hz. Sampling rate f_s was 2000 Hz. Frame length was 2 s.

(a) Order selection criteria (b) ACF plot and (c) PACF plot.

4.3.5. Conclusion: A Rule of Thumb for Finding the Optimal Model Order

This section has presented a numerical testing of a simple way of finding the optimum order for an AR model for the analysis of data from rotating machinery. The main motivation for this work was to devise a simple way to overcome the problem of AR order selection, especially for rotating machines, and thus propose the formulation of p_{\min} .

The suggested AR model order $p_{\min} = \frac{f_s}{f_{\text{machine}}}$ is the number of sample points in one complete revolution of the shaft at the various sampling frequencies and rotating speeds of the machine. Experimental results, using vibration data taken from a dry vacuum pump at different sampling rates and rotating speeds, show that at p_{\min} there is a marked reduction in the prediction error. For low speed rotating machinery, the optimal order is p_{\min} . As the speed of the rotating machine increases, there is some advantage in using twice or thrice p_{\min} to produce more accurate frequency estimates. Results were also justified using ACF and PACF plots and also for a different transducer (B&K accelerometer). The application of the p_{\min} method can ease the estimation of the minimum optimum AR model order required for rotating machinery. It is hoped this may lead to increased interest for use of AR modelling for spectral analysis of signals from rotating machinery for condition monitoring schemes

Results show that the method is an excellent indicator of what the initial order should be. From the results presented a few concluding remarks can be made:

1. AR modelling can be used for spectral analysis in condition monitoring mainly because of its ability to work with smaller frame sizes and yet achieve a resolution improvements (both in time and frequency) compared to FFT techniques.
2. AR modelling can be effectively used for fault diagnosis as the optimal order can be found easily with the formula p_{\min} we propose here.
3. This simple rule of thumb does not replace traditional order selection criteria but can be used as a ballpark minimum figure for the optimal order.
4. It was noticed that as the speed of the machine increases, it might be advantageous to use twice or thrice the p_{\min} order.
5. This formula also works with signals with different SNR.

6. In this work, the AR model was constructed for a test environment with a fixed loading factor. If there are changes in the loading factor, it is anticipated that the performance of the proposed technique still remains the same. The calculation of p_{\min} still applies as increasing the loading factor of the pump only increases the level of the strength of the vibration felt by the pump. The number of basic harmonics of the fundamental shaft frequency in the AR spectra remains the same, hence the minimum order is also expected to be the same.

4.4. Summary

This chapter has introduced the AR order selection criteria and has presented a detailed study on finding the optimal order for the vibration signals obtained from the dry vacuum pump using the order selection criteria. This chapter has also outlined an easy rule of thumb for finding the optimum order for AR modelling for signals attained from rotating machinery. The proposed method does not replace existing AR order selection criteria but is supposed to complement them. On these foundations, the next chapter proceeds to investigate the effects of the optimum sample size and sampling frequency for AR modelling.

5. SELECTION OF OPTIMUM AR SAMPLE SIZE

5.1. Introduction

Chapter 4 had discussed and addressed the problem of AR model order selection for rotating machinery. A rule of thumb using the p_{\min} formula was proposed in Section 4.3 for a 'quick and easy way' to establish the minimum AR model order needed in the context of rotating machinery when run at different running speeds.

This chapter aims to investigate the impact of the optimum sample size and optimum sampling frequency in applying AR modelling as a fault detection tool. This chapter is organized into two main sections. In the first section, an experimental study is conducted to determine the optimum sample size for AR modelling in order for the AR technique to be used effectively as part of a fault detection toolbox. The technique of the parametric method of AR model identification method is sensitive to the sampling frequency. In the second section of this chapter, a qualitative discussion is given to explain the impact of the sampling frequency on the estimation of the AR model order as well as a quantitative study with some findings is reported to support the arguments presented.

5.2. Investigation on Selection of Optimum AR Sample Size

In Section 2.4.2.2, from the literature review, it was acknowledged that the parametric method of AR modelling requires a smaller sample size than FFT (Fast Fourier Transform) technique for Power Spectral Density (PSD) estimation (Schlindwein and Evans 1989; Mechefske 1993a). The question of how small a sample size can be, and yet encapsulate the signal's behaviour, is unknown. The objective of the study reported in this section is to find the minimum sample size required for vibration data acquired at various speeds of the pump.

5.2.1. Motivation

The usage of the right amount of data is vital for accurate prognosis of malfunctions. The cost of collecting large amounts of data can be prohibitive in condition monitoring applications. The length of data needed clearly affects the precocity of the fault detector (the final objective of this wider research project is to have a real-time system with a deadline such that appropriate remedial action can be taken in time to save the Integrated Chip (IC) batch in the event of an imminent vacuum pump fault). It is important to know what is the minimum amount of data needed to produce detailed and repeatable spectra for fault detection. Another significant drawback of using too large a sample size is that it is difficult to ensure stationarity if too large a sample size is used (Pardey, Roberts *et al.* 1996; Poulimenos, Spridonakos *et al.* 2006). The vibration signal under study is considered to be quasi-steady stationary. To ensure stationarity the signal has to be divided into sufficiently short segments. If large sample sizes are used, the monitored segment departs from stationarity and the performance of the AR technique degrades. Thus the optimum sample size required to encapsulate a vibration signal's behaviour for a given model order has to be found.

5.2.2. Using the Prediction Error as an Indicator

The measured difference between the actual AR process and the signal formed by the estimated samples is termed the experimental Prediction Error (PE) and was defined by Equation [4.6] in Section 4.2. The experimental PE is a measure of the AR model's goodness of fit to new data generated by the same time series as data used for estimation of the model. The experimental PE always decreases for increasing model orders and can be used as guide in determining the optimal model order.

The theoretical Prediction Error (PE) for unbiased AR models is defined using Equation [5.1] and this is stated to be equivalent to its asymptotical expectation (Broersen 1998).

$$\textit{Theoretical Prediction Error (PE)} = \sigma^2 \left(1 + \frac{p}{N} \right) \quad [5.1]$$

where σ^2 is the variance of the signal of the input process and in an AR model, this driving signal is assumed to behave like white noise. p is the AR model order and N is the sample size.

The theoretical PE is influenced by σ^2 . σ^2 is theoretically a constant that is dependent on the energy in the driving signal and gives no information on the quality of the models. The theoretical PE is linearly dependent on p , the AR model order. Both the estimated PE and the theoretical PE are inversely dependent on the sample size, N . Both the theoretical and experimental PE are plotted and investigated in the next section because, while the experimental PE is subject to modelling errors such the dependence on the estimation method used (such as Yule-Walker or Burg), the theoretical PE clearly shows the effect on the PE due to sample size alone.

It is generally recommended that the ratio of the model order to the sample size p/N should be kept to less than 0.1 as using a ratio bigger than this affects the empirical statistics of the AR parameters and the model fit degrades depending on the estimation method used (Broersen 2000a). Hence the sample size cannot be too small for AR spectral estimation. However usage of too large a sample size is also generally not recommended as this leads to a wastage in computational effort and does not improve the quality of the model.

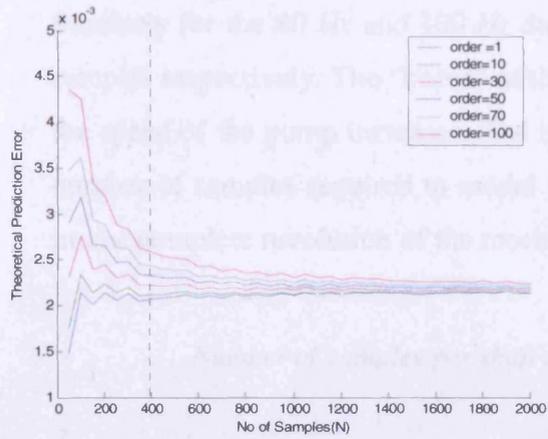
5.2.3. Methodology

To capture the vibration signals, the rotational speed of the pump was set at 60 *Hz*, 80 *Hz* and 100 *Hz* respectively at ultimate pressure of 0 mbar and the vibration data were acquired as detailed in Section 3.2.

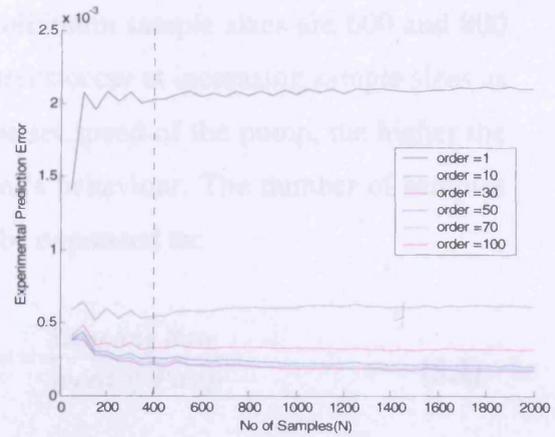
5.2.4. Results

5.2.4.1. Increasing the Pump Speed

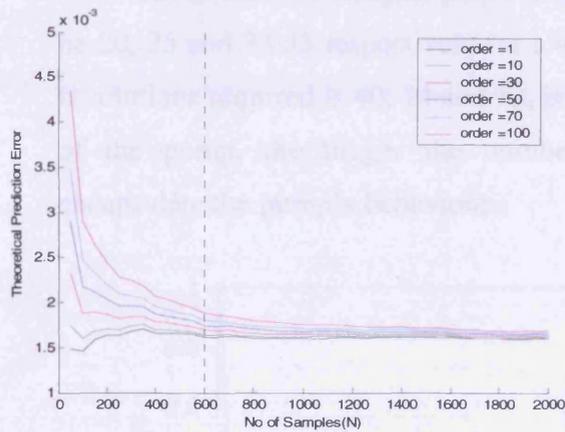
The effect of increasing the pump speed on the prediction error is investigated. The theoretical and experimental PEs are obtained for vibration data acquired from the pump at 60 *Hz*, 80 *Hz* and 100 *Hz* (Figure 5–1). The theoretical PEs decrease asymptotically with increasing samples sizes and reach constant values which are variances of the innovations used and in this case, the variances of the vibration signals used at the different speeds. The points where the theoretical PEs reach this constant value is deemed to be the required minimum sample size for the AR process. The experimental PEs decrease with increasing sample sizes, except at very low model orders of 1 and 10 where the experimental PE first seems to increase at small sizes and then decrease eventually. It was stated before that the optimal order was found to be 25 or more from the investigations reported in Chapter 4. The experimental PE is highly influenced by the model order used. Clearly both the theoretical and experimental PEs are functions of sample size N . From observing Figure 5–1, it can be seen that the theoretical PE increases from low to high orders while the reverse is true for the experimental PE. Looking at Figure 5–1 (a) and (b) and neglecting the trends seen for the low orders of 1 and 10 in the experimental PE, initially the PE decreases dramatically as N increases from 0 to 200 samples. The decrease is less drastic from 200 to 400 samples. For values of N greater than 400 samples, the decrease in PE as a function of N is very slow. Similar trends are observed at 80 *Hz* and 100 *Hz*. For the 60 *Hz* vibration data, both the theoretical and experimental PEs have the “knees” of the curves occurring at 400 samples and this is the minimum sample size required.



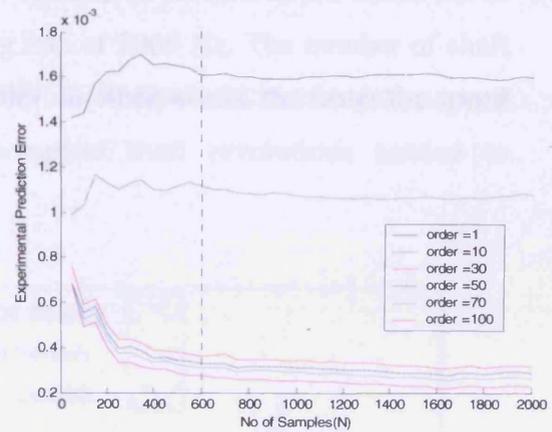
(a) 60 Hz



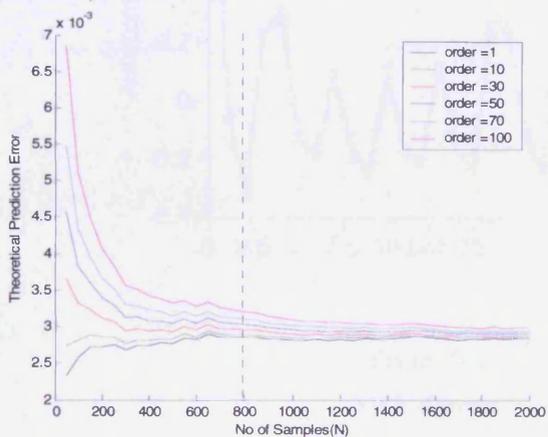
(b) 60 Hz



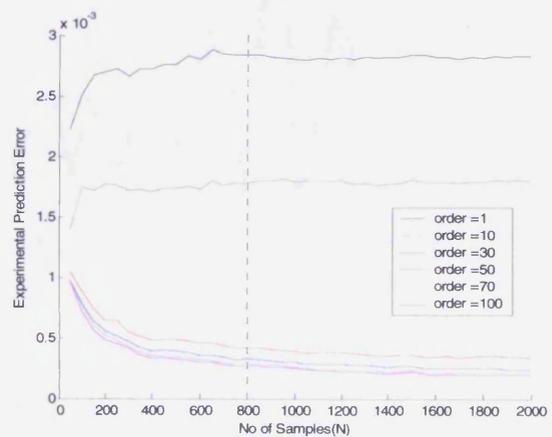
(c) 80 Hz



(d) 80 Hz



(e) 100 Hz



(f) 100 Hz

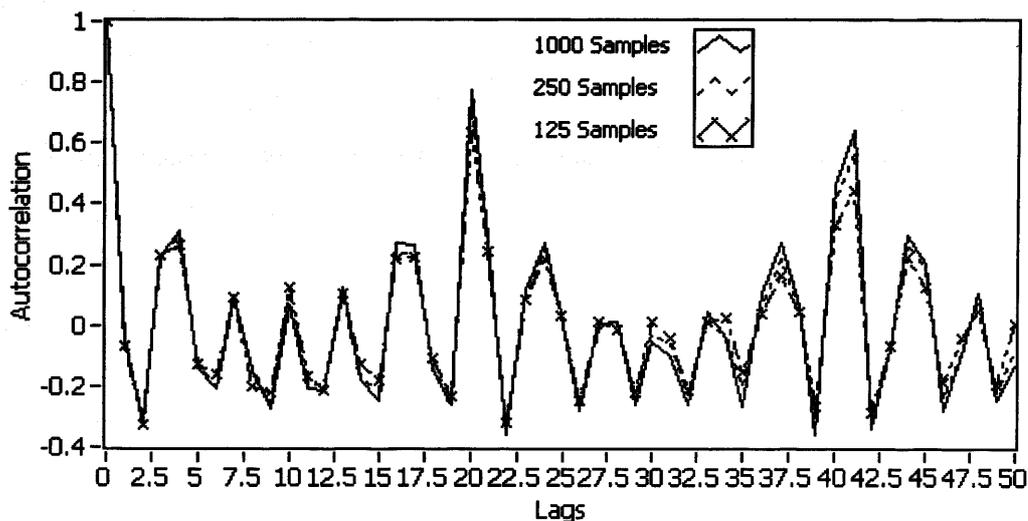
Figure 5–1 Effect of sample size on the variance of the Prediction Error (PE).

Plots of theoretical and experimental PEs are calculated using Equations [4.6] and [5.1]. Plots (a), (c) and (e) correspond to theoretical PEs and (b), (d) and (f) correspond to actual PEs. Plots (a) and (b) are for 60 Hz vibrational data, (c) and (d) for 80 Hz and (e) and (f) for 100 Hz respectively.

Similarly for the 80 Hz and 100 Hz data, the minimum sample sizes are 600 and 800 samples respectively. The “knees” of the PE curves occur at increasing sample sizes as the speed of the pump increases. The higher the set speed of the pump, the higher the number of samples required to model the signal’s behaviour. The number of samples in one complete revolution of the machine can be expressed as:

$$\text{Number of samples per shaft revolution} = \frac{\text{Sampling Rate}}{\text{Speed of Pump}} \quad [5.2]$$

The number of samples per revolution for 100 Hz, 80 Hz and 60 Hz works out to be 20, 25 and 33.33 respectively for a sampling rate of 2000 Hz. The number of shaft revolutions required is 40, 24 and 12, in that order. In other words, the faster the speed of the pump, the bigger the number of complete shaft revolutions needed to encapsulate the pump’s behaviour.



**Figure 5–2 ACF as a function of samples.
MSE decreases as the sample size increases.**

The experimental PE decreases with N because the Yule Walker estimation method was used to derive the AR coefficients. The Yule-Walker AR method uses the Autocorrelation Function (ACF) of the data to estimate the AR coefficients. A larger sample size will yield a more accurate estimate of the ACF. The Yule-Walker technique is based on the solution of a set of linear equations using the ACF of the samples. The Mean Square Error (MSE) of the ACF decreases as the number of data

points used increases. Fluctuations on the nonzero autocorrelation lags get smaller as the sample size increases (Figure 5–2). The sample variance for nonzero lags in the ACF drops off with \sqrt{N} . This translates to a decrease in the variance of the prediction error and a better fit of the AR model.

5.2.4.2. Increasing the Frame Size

In this section the effect of increasing the frame size N on the AR spectra is explored and the results are presented in Figure 5–3. Though from the earlier studies it was known that a minimum order of 25 would suffice, a slightly higher model order of 30 was used to generate the plots (in practice, it is better to overestimate the model order than underestimate it). Spectra were plotted for increasing frame sizes and the point where the spectra stabilised was deemed to be the point where the optimum sample size required had been reached. As N becomes larger, the peaks in the PSD become sharper and the valleys deeper as the AR modelling method is appropriate for estimation of power spectra with sharp peaks but not deep valleys (Kay and Marple 1981) as in the case of bearing faults. This is due to the all pole nature of the AR model. This characteristic is interesting for fault diagnosis of machines as it is the peaks of the characteristic defect frequencies which are important and are monitored for detection of malfunctions. The resolution of the AR spectra becomes better as N increases.

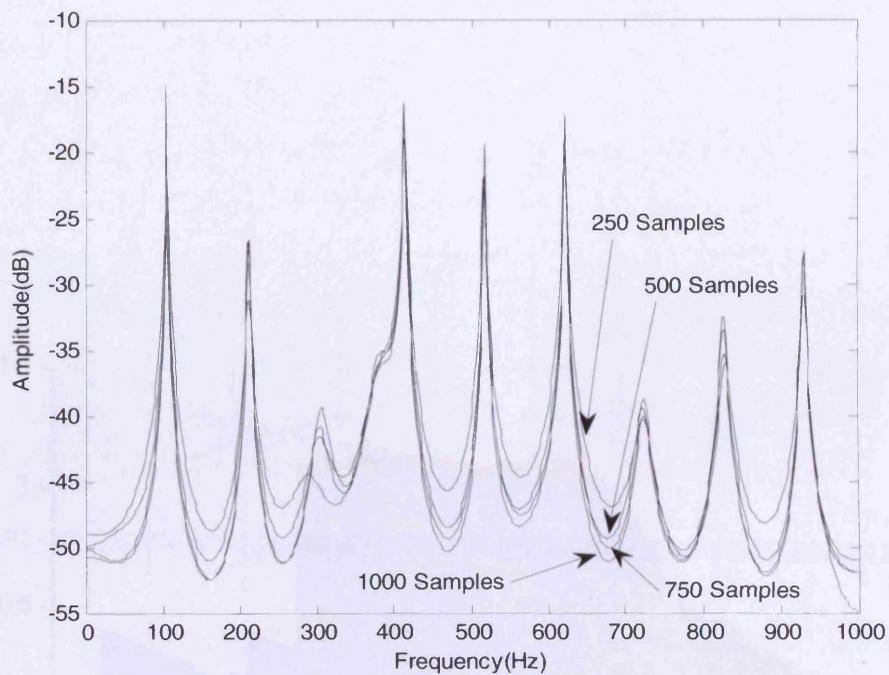


Figure 5-3 Increasing the frame size from 250 to 1000 samples (in two dimensional format). Plot generated by actual ADXL105 vibration data obtained from the pump running at 100 Hz.

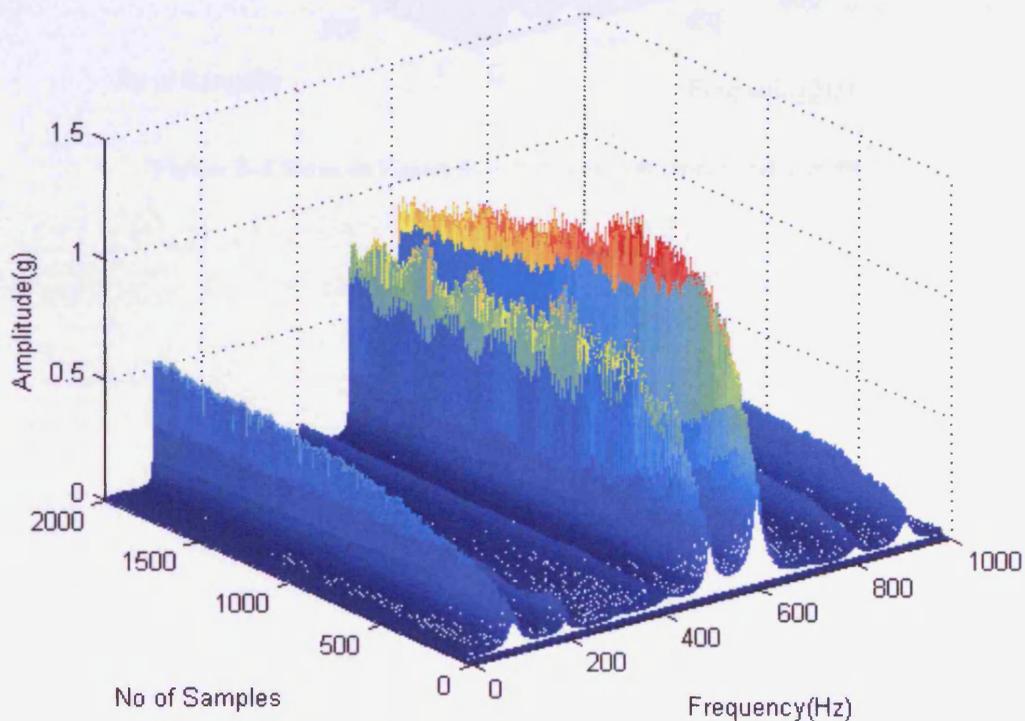


Figure 5-4 AR spectra for increasing frame sizes from 0 to 2000 samples (in three dimensional format).

Model order of 30 was used. ADXL105 vibration data obtained from the pump running at 100 Hz was used for plot.

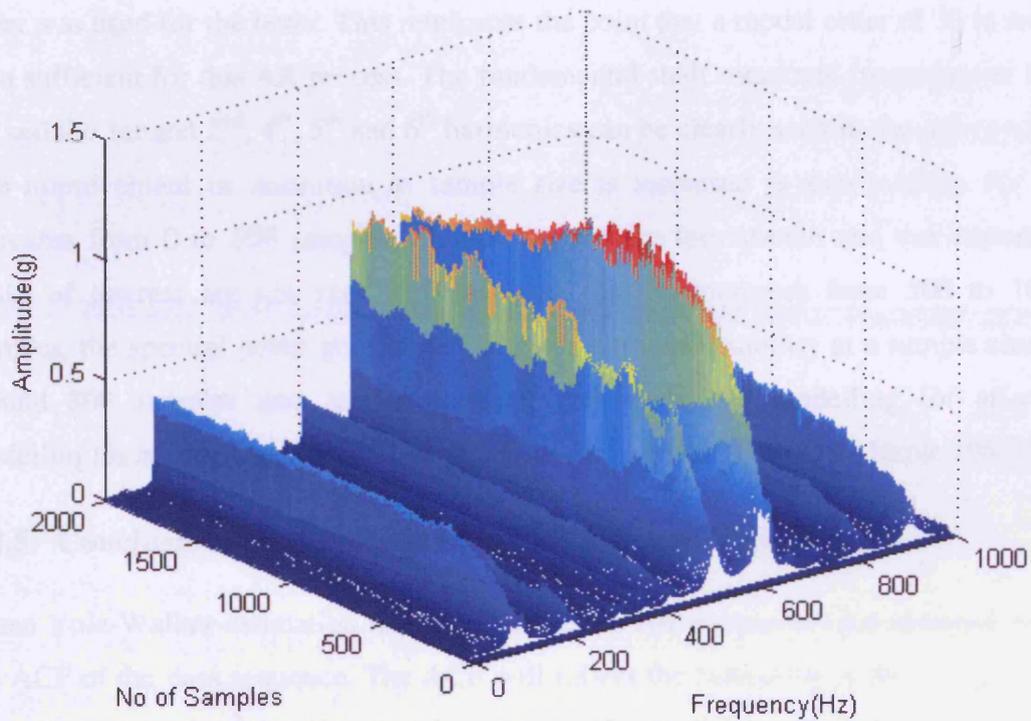


Figure 5-5 Same as Figure 5-4, but with AR model order of 60.

For Figure 5–4 and Figure 5–5, the pump was set to run at 100 *Hz* and vibration data acquired from the ADXL105 accelerometer. The only difference between the figures is that a model order of 30 was used for Figure 5–4 whilst an order of 60 was used for Figure 5–5. Both figures look very similar even though a higher order was used for the latter. This reinforces the point that a model order of 30 is more than sufficient for this AR process. The fundamental shaft rotational frequency of 100 *Hz* and the 1st and 2nd, 4th, 5th and 6th harmonics can be clearly seen in the AR spectra. The improvement in resolution as sample size is increased is also evident. As *N* increases from 0 to 500 samples, the AR spectra are too smooth and the important peaks of interest are not very well resolved. As *N* increases from 500 to 1000 samples, the spectral peaks get sharper. The spectra reach stability at a sample size of around 800 samples and remain constant thereafter. AR modelling (or all-pole modelling) is appropriate for such spectra with sharp peaks (Kay and Marple 1981).

5.2.5. Conclusion: Optimum Sample Size is Speed Dependent

When Yule-Walker estimation method is used, the AR parameters are obtained from the ACF of the data sequence. The ACF will reflect the behaviour of the true process better as the length of the data record increases. The prediction error of the AR model decreases with an increasing sample size. From our earlier studies it was determined that the resolution of an AR model mainly depends on the model order. Usage of a sample size bigger than the optimum sample size can only lead to a slight improvement in terms of resolution. The prediction error of the models can serve as a useful indicator to help determine the minimum sample size required for rotating machinery such as pumps. AR spectra of increasing frame sizes were also plotted and the sample sizes where the spectra stabilised agreed with values obtained from the prediction error plots. Results show that the optimum sample size is speed dependent and for 100 *Hz* it corresponds to 40 full shaft rotations of the pump. As the speed of the pump was increased, more samples were required and this corresponds to more full revolutions of the shaft of the pump. This behaviour can be generally attributed to the statistical properties of the signal being observed. It is hoped that findings reported in this study give some insight on finding the minimum sample size required for AR modelling in rotating machinery such as pumps, motors and gears. Though the study

was carried out on data obtained from a pump running in a perfectly healthy state, the same analysis methodology can be extended to data obtained from a pump with defects.

5.3. Investigation on Impact of Sampling Frequency

In AR model estimation, a suitable sampling frequency has to be selected for confident AR model identification. In this section, the choice of the use of a sampling frequency of 2 kHz reported in the investigations conducted in Chapters 4, 6, 7 and 8 is justified by showing that the effect of using a sampling frequency higher than the Nyquist rate is only to increase AR model order. A qualitative explanation is first given to illustrate what is the effect of increasing the sampling frequency on the estimated AR model order.

5.3.1. A Theoretical Explanation

The AR model order is a function of the required length of time (or the optimum record length) and the sampling rate utilised in the AR model identification. As stated in Equation [2.5] in Chapter 2, an AR model assumes a linear model structure and estimates a future sample of a signal as a linear summation of p previous points plus a noise term (which can be seen as the prediction error in the estimation) where p is the AR model order. The length of time the AR model has to trace in time to predict this future sample is a characteristic property of the signal to be examined and depends entirely on the statistical properties of the signal (see Figure 5–6). The minimum length of time one has to trace back to capture the signal's behavior is denoted by T and this is denoted as the optimal record length. In this example, if the signal is sampled at a sampling frequency of f_s , this means that one has to step back in time by 5 points to estimate the future sample of the signal. In this case, the model order p is then said to be optimal and is equal to 5. If the sampling frequency is doubled to $2f_s$, this means that one has to step back in time by 10 samples to traverse the same length of time. Now the model order p has doubled to become 10. So the effect of doubling the sampling frequency is to double the optimal model order $p_{optimum}$. If the sampling frequency is tripled to $3f_s$, this means that one has to step back in time by 15 points now to cover the same length of time required.

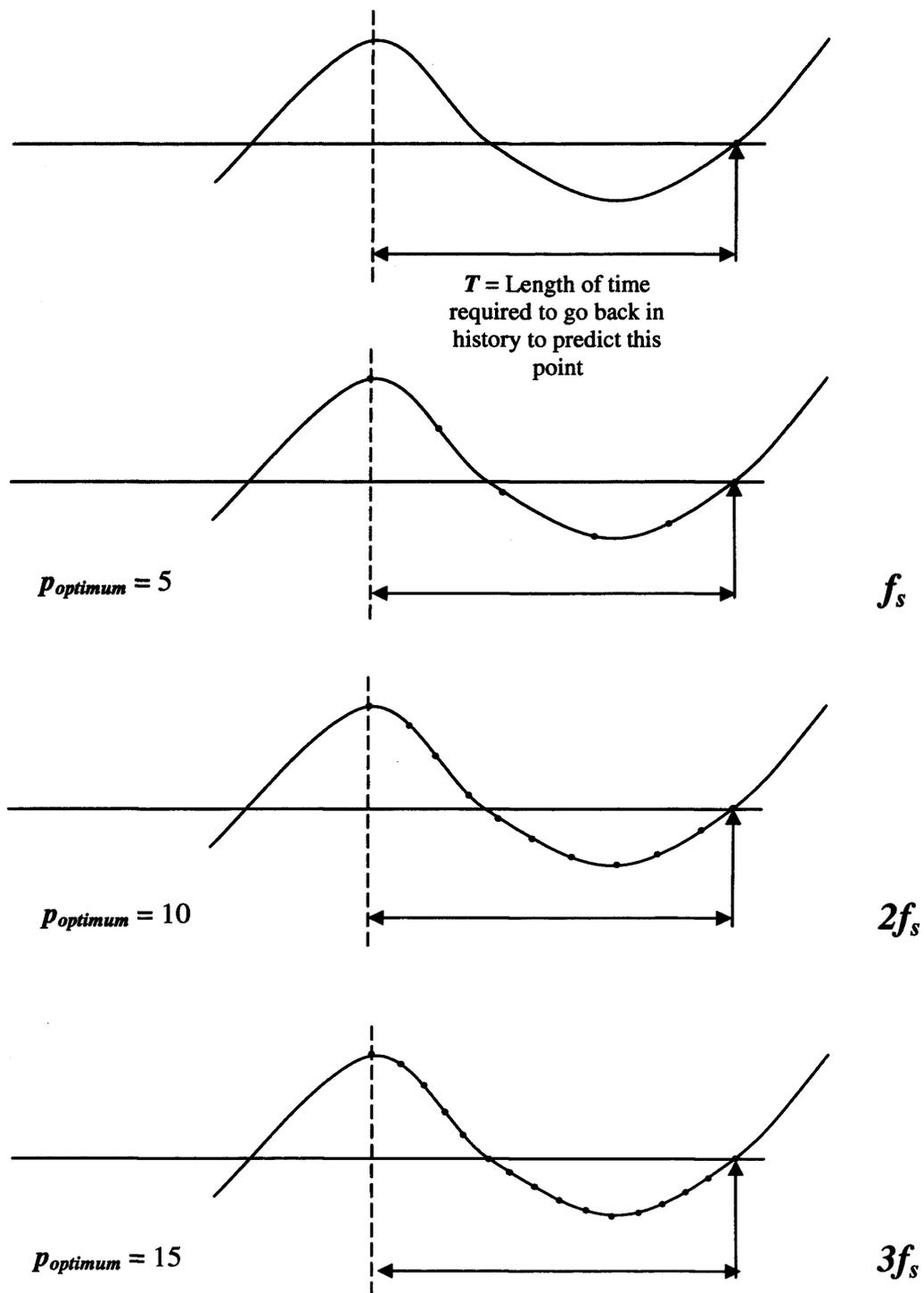


Figure 5–6 Effect of increasing sampling frequency on optimal AR model order.
 The only effect of increasing the sampling rate is to increase optimum AR model order. The length of time required to go back in history to capture the signal’s behaviour still remains the same even though the sampling frequency is increased.

The optimal model order $p_{optimum}$ becomes 15. By doing this again and again, an infinitely fine sampling of an infinitesimal time interval is obtained and this translates to an infinitely large sampling frequency and infinitely large model order. It can be seen the AR model identification method is sensitive to the sampling frequency and as a result, the only effect of using a sampling frequency more than the required Nyquist rate is to increase the AR model order.

5.3.2. Motivation

In the application studied in Section 5.2, the number of shaft revolutions is a non-dimensional quantity. Model order p is a non-dimensional quantity. Sampling frequency is not but it can be non-dimensionalized with respect to the rotating shaft frequency so that interesting comparisons can be made across models. Instead of merely increasing the sampling frequency and observing its implication on the AR model identification process, a better way of studying the effect of sampling rate has been adopted here and has been verified with actual data obtained from the dry vacuum pump as reported in the data collection section of Chapter 3.

5.3.3. Methodology

The sampling rate was adjusted until the same number of sample points per shaft revolution (at each of the 3 speeds studied) were obtained (refer to Table 5-1) and the prediction errors of the 3 different AR models, one for each pump speed, were plotted for increasing model orders. It was noted what was the minimum model order required for each of the signals to obtain acceptable performance.

Table 5-1 Studying the effect of sampling frequency by fixing the number of samples per revolution and adjusting the sampling frequency until the same number of points are obtained per shaft revolution for the three different pump speeds.

Pump Speed(Hz)	No of Samples per rev	Sampling Freq	Length of time for one complete revolution (ms)
60	20	1200	16.7
80	20	1600	12.5
100	20	2000	10
60	30	1800	16.7
80	30	2400	12.5
100	30	3000	10

5.3.4. Results

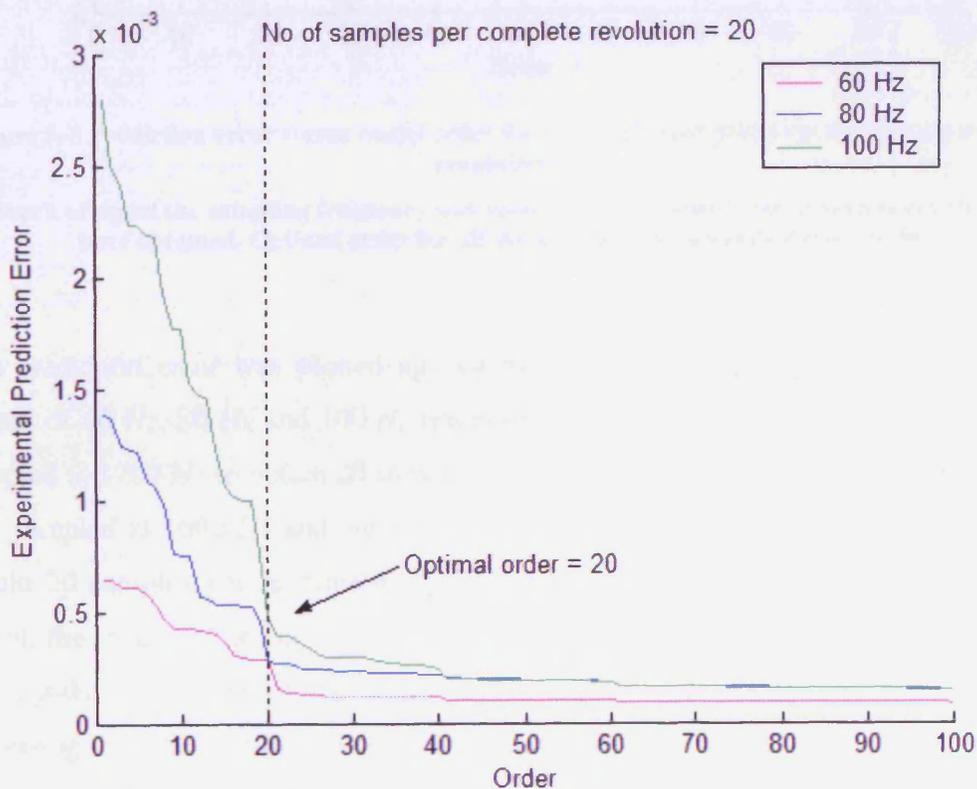


Figure 5-7 Prediction error versus model order for three different pump speeds (20 samples per revolution).

For each signal the sampling frequency was varied until 20 samples per revolution of the shaft were obtained. Optimal order for all three signals occurs at model order 20.

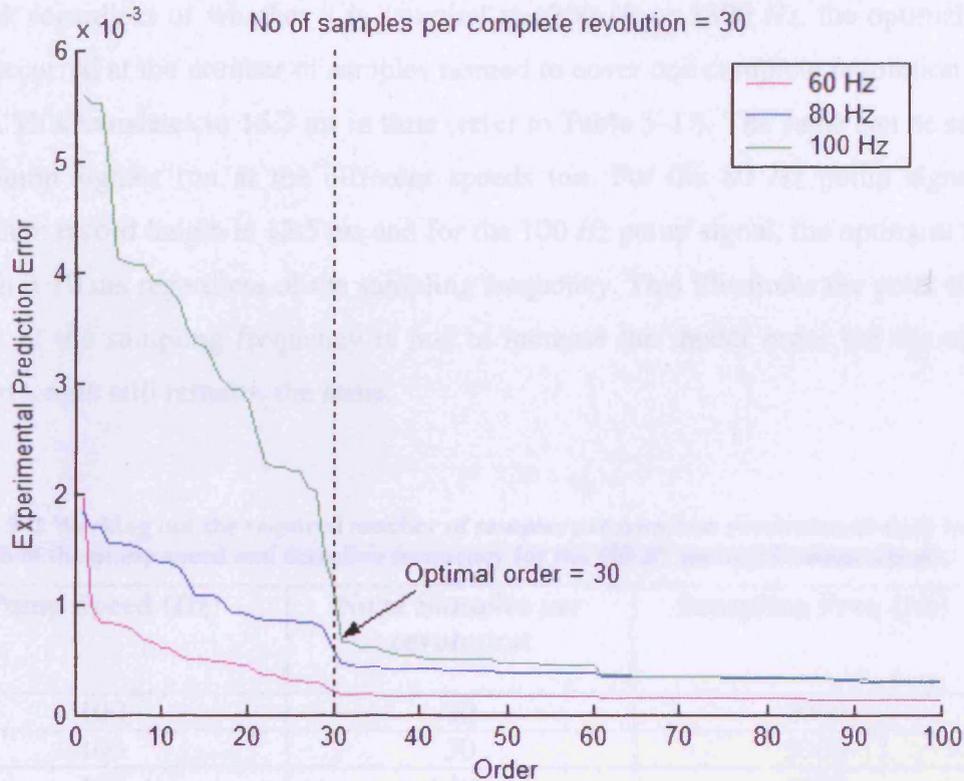


Figure 5-8 Prediction error versus model order for three different pump speeds (30 samples per revolution).

For each of signal the sampling frequency was varied until 30 samples per revolution of the shaft were obtained. Optimal order for all three signals occurs at model order 30.

The prediction error was plotted against the model order for three different pump speeds of 60 Hz, 80 Hz and 100 Hz respectively (Figure 5-7). The 60 Hz signal was sampled at 1200 Hz to obtain 20 samples per revolution of the shaft. The 80 Hz signal was sampled at 1600 Hz and the 100 Hz signal was sampled at 2000 Hz similarly to obtain 20 samples per revolution of the shaft (refer to Table 5-1). For the 60 Hz signal, the 'knee of the curve' occurs around model order 20. For the 80 Hz and 100 Hz signals, the 'knees of the curve' also occur around model order 20. It is not surprising that the optimum model order equals to the number of samples per shaft revolution. The same kind of behaviour was observed when the prediction error was plotted against the model order for three different pump speeds, but now for 30 samples per revolution of the shaft (refer to Figure 5-8). The 60 Hz signal was sampled at 1800 Hz to obtain 30 samples per revolution of the shaft. The 80 Hz signal was sampled at 2400 Hz and the 100 Hz signal was sampled at 3000 Hz similarly to obtain 30 samples per revolution of the shaft (refer to Table 5-1). For the 60 Hz pump

signal, regardless of whether it is sampled at 1200 Hz or 1800 Hz, the optimal order had occurred at the number of samples needed to cover one complete revolution of the shaft. This translates to 16.7 ms in time (refer to Table 5–1). The same can be said for the pump signals run at the different speeds too. For the 80 Hz pump signal, the optimum record length is 12.5 ms and for the 100 Hz pump signal, the optimum record length is 10 ms regardless of the sampling frequency. This illustrates the point that the effect of the sampling frequency is just to increase the model order but the optimal record length still remains the same.

Table 5–2 Working out the required number of samples per complete revolution of shaft rotation at each of the pump speed and sampling frequency for the 100 Hz pump vibration signal.

Pump Speed (Hz)	No of Samples per revolution	Sampling Freq (Hz)
100	20	2000
100	30	3000
100	40	4000
100	50	5000
100	60	6000
100	70	7000
100	80	8000

The prediction error is again plotted for the 100 Hz pump signal for increasing model orders. The sampling frequency is varied from 2 kHz to 8 kHz corresponding to a number of samples of 20 to 80 per complete revolution (refer to Table 5–2). For all the signals sampled at the various sampling frequencies, initially when the optimum AR model order is large relative to the number of data samples observed, the prediction error is large. The biggest decrease in prediction error occurs at the optimum model order. For the signal sampled at 2 kHz, the ‘knee of the curve’ occurs around model order 20. For the signal sampled at 3 kHz, the ‘knee of the curve’ occurs around model order 30. Hence at each of the sampling frequency, the optimum model order occurs at the number of samples per complete shaft revolution. For the 100 Hz pump signal sampled at the sampling frequencies of 4 kHz, 5 kHz, 6 kHz, 7 kHz and 8 kHz, the optimum model orders occur at 40, 50, 60, 70 and 80 respectively. The rationale behind this kind of behaviour can be explained by the same reasoning given for Figure 5–7 and Figure 5–8 . It is also to be noted that the higher the sampling

frequency, the bigger is the prediction error. The prediction error was the biggest for the signal which was sampled at the highest sampling frequency of 8 kHz. Hence a bigger number of samples per complete revolutions, in this case 80, was required before the prediction error stabilized and the biggest decrease in the prediction error occurs. By increasing the sampling frequency, it was found necessary to increase the model order for obtaining an accurate AR model. This reinforces the fact that it is better to sample the signal at a smaller sampling frequency as this gives smaller prediction errors and also lower model orders.

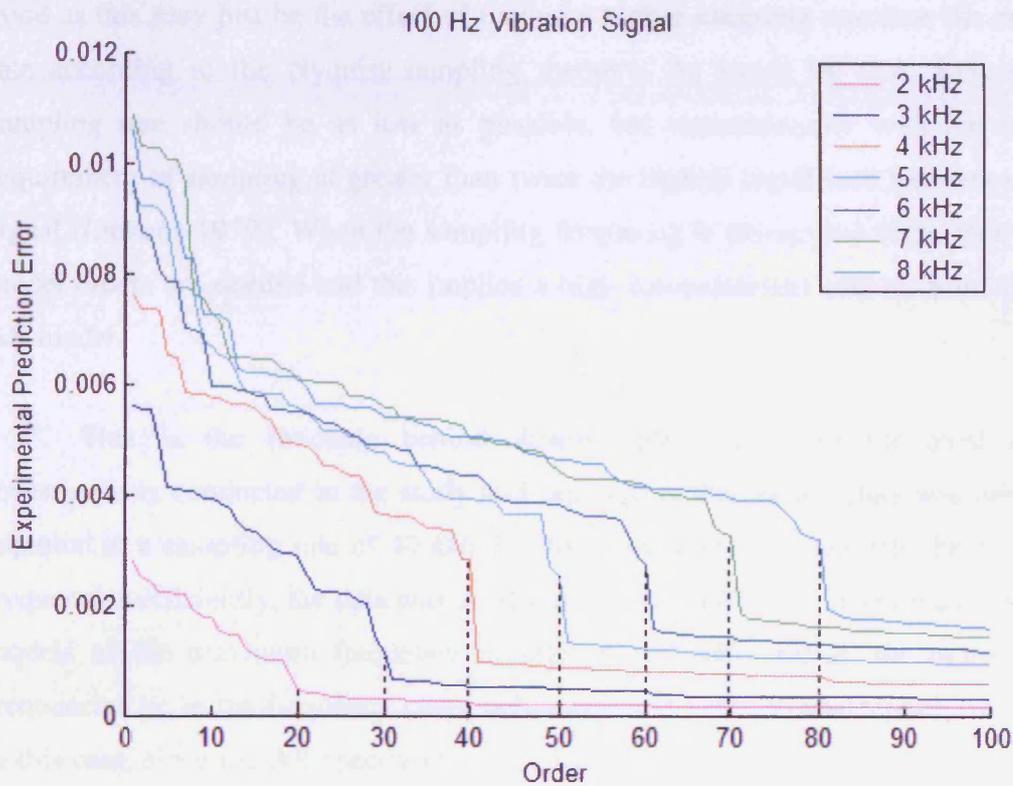


Figure 5–9 Pump vibration signal acquired from pump running at 100 Hz sampled at increasing sampling frequencies from 2 kHz to 8 kHz. Prediction error versus model order plotted for increasing model orders.

5.3.5. Conclusion: Length of Record Traversed Back is More Important than Sampling Rate for Model Order Selection in AR Modelling

This investigation has reinforced the fact that to determine the model order, the length in time we have to trace back is more important than the sampling frequency. The effect of increasing the sampling frequency is merely to increase the model order required. The order of AR models influences the estimation accuracy (Smail, Thomas *et al.* 1999) of model identification but having a large model order is not necessarily good as this may just be the effect of having a higher sampling rate than the required rate according to the Nyquist sampling theorem. As stated by D.A. Linkens, the sampling rate should be as low as possible, but commensurate with the Nyquist requirement of sampling at greater than twice the highest significant frequency in the signal (Linkens 1979). When the sampling frequency is chosen too high, then higher model orders are needed and this implies a high computational cost in estimating the AR model.

This is the rationale behind downsampling the data for most of the investigations conducted in the study and reported in this thesis. Data was originally captured at a sampling rate of 40 kHz but to examine the characteristic bearing fault frequencies efficiently, the data was downsampled to 2 kHz prior to estimating the AR models as the maximum frequency of interest is 1 kHz and all the bearing fault frequencies lie in the frequency range between 0 to 1 kHz. Downsampling was useful in this case, since the AR spectra fit over a smaller segment of the frequency domain, from 0 to 1 kHz. Downsampling had also extended the memory span the AR models corresponded in time for given fixed model orders and allowed the investigations to be carried out with relatively small AR model orders.

5.4. Summary

In this chapter, some interesting findings on determining the optimum sample size for AR modelling have been presented. The impact of increasing the sampling frequency on AR model order selection with particular respect to rotating machinery has been considered. It has been concluded that the optimum sample is speed dependent. Also it has been verified that a good optimal sampling frequency would be between two to three times the maximum frequency (respecting the Shannon's Nyquist sampling theorem) in the signal being observed. In the next chapter an important concept of pre-processing the vibration signals in the form of resonance demodulation, which is necessary to enhance the chances of successful fault detection, is introduced and investigated.

6. RESONANCE DEMODULATION

6.1. Introduction

In the proposed AR fault detection tool to be introduced in Chapters 7 and 8, a critical stage before spectral estimation of the vibration signals using AR modelling method was the pre-processing of signals using the demodulation resonance technique. The spectra of the raw damaged bearings show no remarkable features at the bearing defect frequencies and are dominated by the fundamental rotating speed of the machine. A method of conditioning of the signal before spectral estimation is necessary and this technique is explored in this chapter. This method is demodulated resonance analysis, also known as HFRT (High Frequency Resonance Technique) or envelope spectral analysis. Some of the results presented in this chapter have been published in (Thanagasundram and Schlindwein 2006a).

6.2. Resonance Demodulation Explained

The resonance demodulation technique (Harting 1978; Prashad, Ghosh *et al.* 1985) has been extensively used in the diagnosis of rolling bearings and the approach focuses on the analysis of the structural resonance excited by the fault-induced impacts. Vibration signals from defective bearings are weak impulses and are buried in the machinery induced vibrations and noise. Often environmental conditions such as the instantaneous speed variations, as well as the presence of multiple fault conditions, obscure the defective vibration signals that are required for reliable diagnostics. Application of spectral estimation techniques alone would not help detect the bearing defect frequencies. The demodulated resonance analysis technique has been shown to overcome some of the limitations of normal spectral analysis for bearing defect detection.

Harting proposed the demodulated resonance method (Harting 1978) and this technique relies on the fact that the defect generated vibration information is carried by high frequency resonances of the bearing elements or pump housing and we can make use of the high magnification present in the excitations to successfully extract the faults. The impulses of bearing defects are of extremely short durations compared

to the interval between impulses. The energy in the defective bearings' impacts is distributed across a wide frequency range and is easily masked by noise and frequency components generated by the machine. However the structural resonance in the bearing housing and surrounding structures acts as an amplifier to low-energy bearings impacts when the bearing housing is excited by the impacts.

There are many ways to employ the resonance demodulation technique. One way is to band-pass filter the signal to isolate one resonant frequency so as to exclude the vibration generated by other parts of the machine. The band-passed filtered signal is centred at the carrier frequency and has a bandwidth corresponding to the maximal modulating frequency. The carrier frequency is, in this case, the resonant frequency and the modulating frequency is the fundamental rotating shaft frequency. An envelope detector then demodulates the filtered signal and the resulting signal is studied using spectral analysis. If there is a defect in the bearing, this is denoted by the appearance of a frequency peak associated with that defect.

Another way to do resonance demodulation is to band-pass filter the signal around the resonant peak. The band-passed signal is then squared and low-pass filtered. The resulting signal is known as the squared envelope and describes the power of the envelope signal (Howard 1994). A third way of extracting the demodulated signal has been used, which involves applying the Hilbert Transform for the rectification. It has been showed by Randall that the squared envelopes of the analytic spectra of damaged bearings are highly correlated to the spectral components at the bearing defect frequencies and produce even better results as the application of the Hilbert Transform produces one sided spectra, thus avoiding the interference of the sum frequencies which result from two-sided spectra (Randall, Antoni *et al.* 2000). The key steps in demodulating the signal are shown in Figure 6–1.

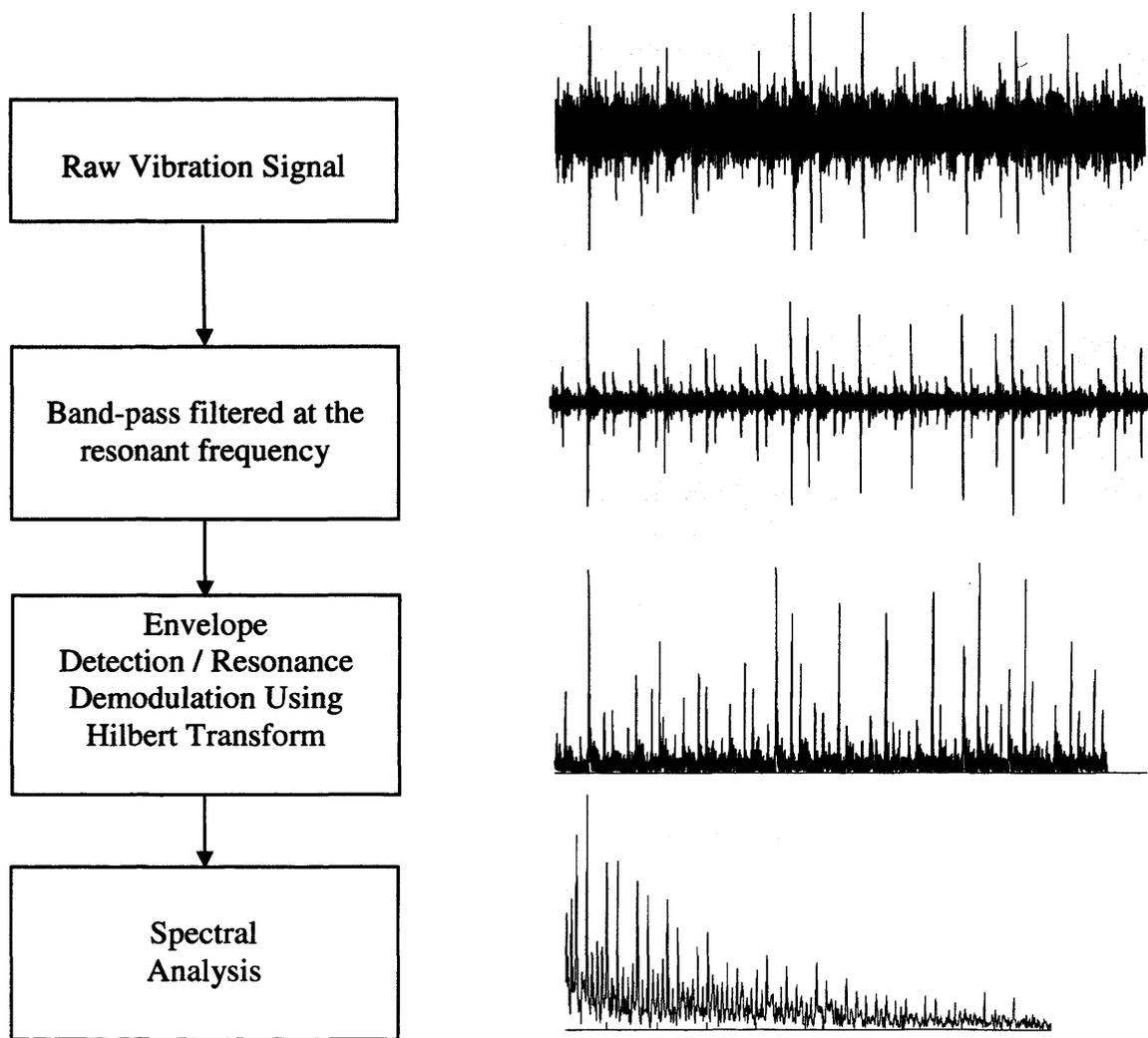


Figure 6-1 Steps involved in the demodulated resonance technique and the corresponding waveforms produced.

6.3. Demodulation using Hilbert Transform

The Hilbert transform can be used to demodulate the complex faulty bearing signal for isolation of the low frequency components from the high frequency signal to obtain the required and useful part for diagnosing the faulty bearings (Thrane 1984). The Hilbert transform is first applied to create an artificial complex signal from the time domain signal.

If $x(t)$ represents the time signal,

$$\bar{x}(t) = H[x(t)] \quad [6.1]$$

where H is the Hilbert transform. The analytic signal is obtained by

$$\hat{x}(t) = x(t) + j \cdot \bar{x}(t) \quad [6.2]$$

where j is $\sqrt{-1}$. This is the analytic signal whose real part is the original signal and whose imaginary part is the Hilbert Transform of the real part. The mathematical operation of the Hilbert Transform is defined as follows

$$\hat{x}(t) = x(t) * \frac{1}{\pi t} \quad [6.3]$$

The Hilbert transform can be thought of as an interesting kind of filter, in which the amplitudes of the spectral components are left unchanged, but their phases are shifted by a phase shift of 90° , positively or negatively according to the sign of the frequency term.

$$H(f) = -j \operatorname{sign}(f) = \begin{cases} -j & f > 0 \\ j & f < 0 \end{cases} \quad [6.4]$$

The positive frequency components are shifted by $-j$ (a phase shift of -90°) and negative frequency components are shifted by $+j$ (a phase shift of $+90^\circ$).

The complex envelope is obtained by

$$|\hat{x}(t)| = \sqrt{x(t)^2 + \bar{x}(t)^2} \quad [6.5]$$

The envelope of the demodulated signal is defined by the modulus of the analytic signal. It is always a positive function. The advantage of obtaining the analytic signal is that by transforming the time domain signal to the analytic signal, the negative frequencies are removed and only the positive components (with double amplitude) are retained facilitating further analysis. Also by applying the Hilbert transform, the positive band-pass centre frequency is translated to the origin to produce a baseband signal. Hence one can sample the resulting complex envelope signal with a smaller sampling rate.

6.4. Finding the Resonance Bandwidth

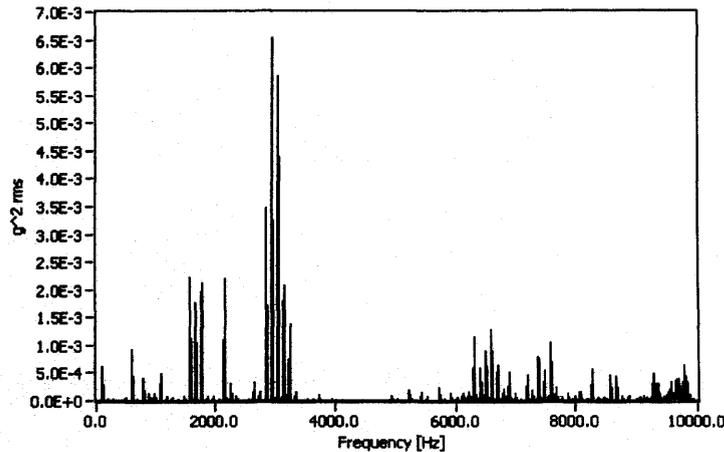
The tricky part of the envelope demodulation technique is that the most suitable bandwidth must be identified before demodulation takes place. The bandwidth of the filter should be chosen such that it covers the whole range of the resonance. The main idea is to identify the band-pass range so as to eliminate high amplitude signals not associated with the bearing faults, yet encompass a frequency range containing the bearing fault spectral components that are not negligibly small. Various high and low pass settings have to be experimented and normally their choice is related to the resonances of the system.

The best frequency band to envelope (the resonance bandwidth) has to be identified prior to the demodulation stage. For the faulty bearing with the inner race

crack, the resonance bandwidth was found to be between 6-8 kHz. Figure 6-2 shows the raw broadband spectrum (0 to 10 kHz) of the non-defective bearing run at 100 Hz. It can be seen that there is some energy in the frequency bands between 2 to 4 kHz and 6 to 8 kHz. Figure 6-3 shows the same raw broadband spectrum but for the defective bearing with an inner race fault run at 100 Hz. The excitation due to the defective bearing impacts dominates the spectrum. The elevated bearing energies between 6 to 8 kHz are indicative of a propagating bearing fault. Resonances excited for the defective bearing lie between 6 to 8 kHz. The resonances between 2 to 4 kHz fail to be excited. The amount of energy in the defective spectra is much more than for the non-defective spectra in the same processing band. So it is the frequency range between 6 to 8 kHz that needs to be band-pass filtered to isolate the structural resonance induced in the system by the defective bearings. Filtering the signal around this frequency extracts that part of the signal which is mainly composed of the resonant ringing pulses due to the bearing defect. The signal to noise ratio of the defective bearing signals in this region is high due to the resonances of the structure. The frequency range excited by the bearing defect remained constant for all pump speeds. This frequency range is digitally filtered using an elliptic Infinite Impulse Response (IIR) band-pass filter of order 10 between 6 to 8 kHz. Then the Hilbert Transform is applied to obtain the analytic signal as part of resonance demodulation prior to spectral estimation.

Figure 6-4 shows the spectra of the bearing with seeded inner race defect obtained using both the AR and FFT techniques for a pump run at 105 Hz. The vibration signals were not pre-processed utilizing resonance demodulation method before being frequency analysed. It can be seen that the spectra are dominated by the fundamental rotating frequency (100 Hz- actual speed of the pump less than the set speed due to slip) and its harmonics. The bearing defect frequency does not appear in the frequency spectrum because of its low amplitude with respect to the higher energy harmonics of the rotating speed. Figure 6-5 shows the spectra of the same vibration signal, but resonance demodulated now, before spectral estimation was performed. A cleaner spectrum is obtained, free from the fundamental rotating frequency and its harmonics and other unwanted components. The inner race defect frequency becomes visible. By band-pass filtering in the resonance bandwidth (6 to 8 kHz) and applying the Hilbert transform to demodulate the signal, the technique of resonance

demodulation has amplified the amplitude at the characteristic bearing defect frequency and has made the bearing defect more easily identifiable.



**Figure 6-2 Broadband spectrum (0 to 10 kHz) for pump with non-defective bearing.
Pump running at 100 Hz.**

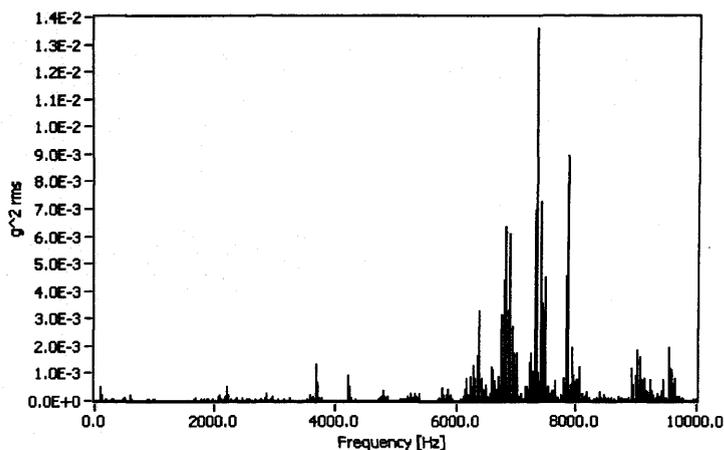


Figure 6-3 Broadband spectrum (0 to 10 kHz) for pump with the defective bearing (inner race fault).

Pump running at 100 Hz. Note resonance occurring in the 6-8 kHz region.

It is now the simple case of monitoring the level of the characteristic bearing defect frequency in the demodulated spectrum to measure the severity of the defect. Shiroshi and collaborators have shown that the size of the demodulation peak in the demodulated spectrum is linearly related to the severity of the defect when using vibration measurements (Shiroishi, Li *et al.* 1997). The normalized ratio of the

demodulation peak to the ‘carpet’ level (noise floor in the demodulation spectra) provides a quantitative measure of the bearing defect condition.

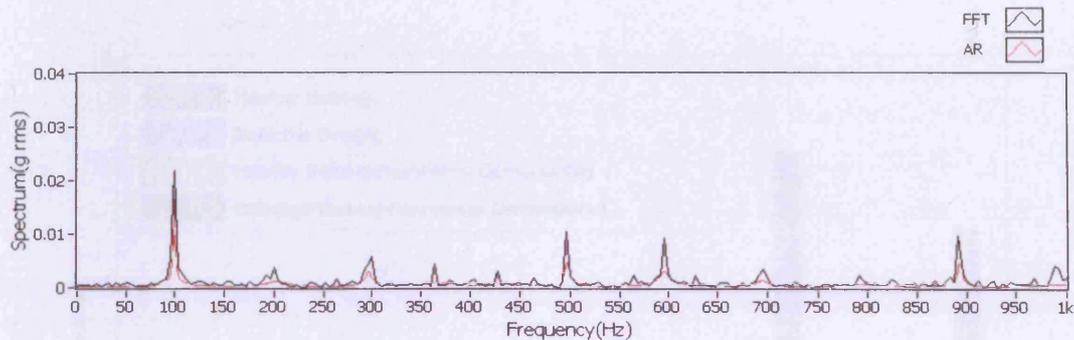


Figure 6–4 FFT and AR Spectra for ADXL105 vibration signal. Obtained from a bearing with a seeded inner race fault prior to resonance demodulation. Speed of pump was set to 105 Hz. Sampling rate f_s was 2000 Hz. Note absence of bearing defect frequency as spectra was dominated by fundamental rotating frequency and its harmonics.

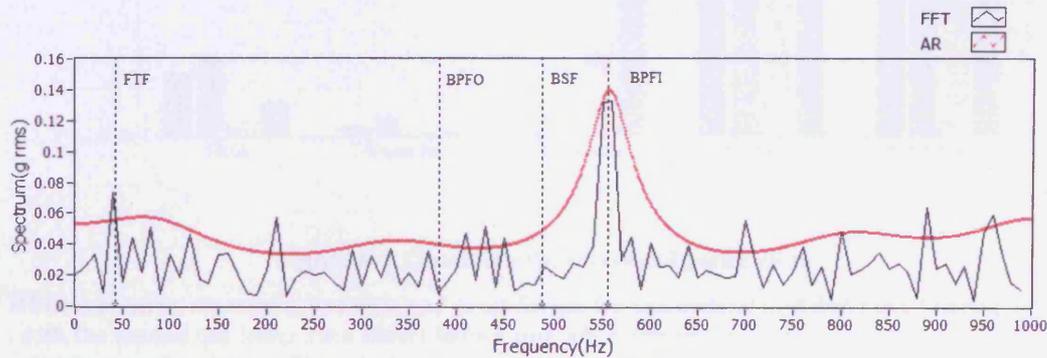


Figure 6–5 Same vibration signal used as for figure above. Time domain vibration signal resonance demodulated prior to spectral estimation. The Ball Pass Frequency of Inner Race defect is clearly evident at 555 Hz in both FFT and AR-based spectra.

6.5. The Need for Resonance Demodulation

The resonance demodulation technique is assessed with regards to its effectiveness in the detection of the bearing condition using the time statistical parameters introduced in Section 2.4.1. Using equations [2.21] to [2.24], the statistical parameters of RMS, variance, skewness, kurtosis and crest factor are obtained for the normal and defective

bearing signal with the implanted inner race defect before and after resonance demodulation.

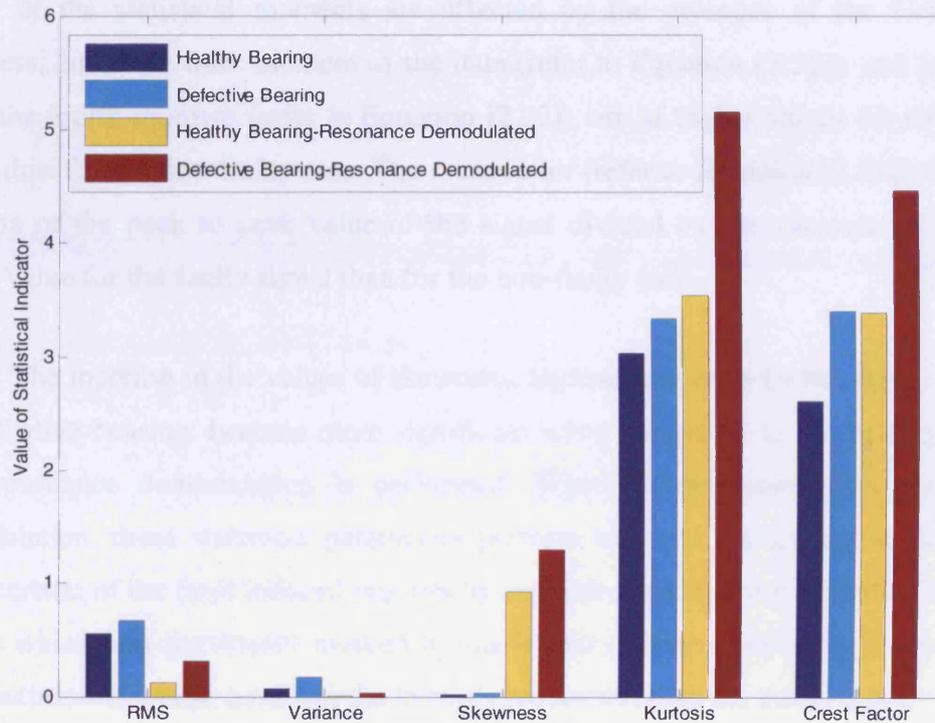


Figure 6–6 Obtaining the statistical parameters

RMS, variance, skewness, kurtosis and crest factor for the normal and defective bearing signal with the seeded the inner race defect before and after resonance demodulation. The vibration signals were band-pass filtered (6 to 8 kHz) and demodulated using the Hilbert transform for resonance demodulation.

Results are shown in Figure 6–6. For the RMS and variance figures, the statistical values obtained after resonance demodulation were of much lower values compared to those prior to demodulation. The process of band-pass filtering and demodulation using the Hilbert transform has removed a major portion of the vibration signals and subsequently the energy content of the demodulated signals was much less compared to the raw signals. The RMS and variance parameters are not useful in the detection of the bearing fault.

The vibration response of bearings, subjected to defects, is characterized by the presence of sharp peaks. The bearing vibration signal becomes more impulse-like with the presence of the localised bearing fault in the form of inner race crack. The skewness, kurtosis and crest factor parameters capture this behaviour better and give higher values since they respond more to the spikiness of the vibration signal. The values of the statistical moments are affected by the presence of the fault. The skewness, being the third moment of the data (refer to Equation [2.22]), and kurtosis, being the fourth moment (refer to Equation [2.23]), are of higher values for the faulty signal than for the non-faulty case. The crest factor (refer to Equation [2.24]) which is function of the peak to peak value of the signal divided by the variance, is also of higher value for the faulty signal than for the non-faulty case.

The increase in the values of skewness, kurtosis and crest factor, especially for the defective bearing, became more significant when compared to the healthy signal after resonance demodulation is performed. When pre-processed with resonance demodulation, these statistical parameters perform better in encapsulating the spiky characteristic of the fault induced impacts by extracting the useful part of the vibration signals which was previously masked by the higher energy modulating components. The maximum increase occurs in the kurtosis parameter, thus the kurtosis parameter is an effective defect indicator. The kurtosis value of the healthy raw signals was close to 3. The kurtosis value of the defective raw signals was slightly higher than 3 indicating a propagating fault (the difference between them being approximately equal to 0.2). The kurtosis value of the healthy demodulated signals was close to 3.5. The kurtosis value of the defective demodulated signals was close to a value of 5 (the difference between them being approximately equal to 1.5). Hence the difference in the kurtosis values between the healthy and defective signals had become even greater when the vibration signals were resonance demodulated.

The use of these statistical indicators can be used to identify that there is a bearing fault, however the time domain analysis method is not very helpful in identifying which bearing component is defective. The aim of this investigation is to show that the application of the resonance demodulation has extracted the more useful part of the vibration signals, that is, the impulse-like bearing signals due the impacts between the faulty component and the other parts of the bearing components. This

investigation has illustrated that prior processing in the form of resonance demodulation is therefore necessary before spectral estimation to enhance the chances of successful fault identification and detection.

6.6. Finding the Optimum Order for the Demodulated Signals

In Chapter 4, it was concluded that the optimum model order for AR modelling of the raw vibration signals was $p = 25$. When the vibration signals are demodulated, it was found that even lower model orders could be used. Figure 6–7 shows a plot of the order selection criteria versus model order for the 5 order selection criteria (refer to Equations [4.1] to [4.5]). The knee of the curves occurs at around model order 4. A model order of 10 was chosen (higher than what is required) to represent the demodulated normal and faulty bearing signals as the vibration signal was quasi-steady stationary and the model order chosen has to represent all the varying time segments. The characteristics of one vibration time segment may differ from another. Hence a higher model order was chosen to be on the safer side. This model order was found to be more than sufficient to represent the signals' characteristics in both faulty and non-faulty conditions.

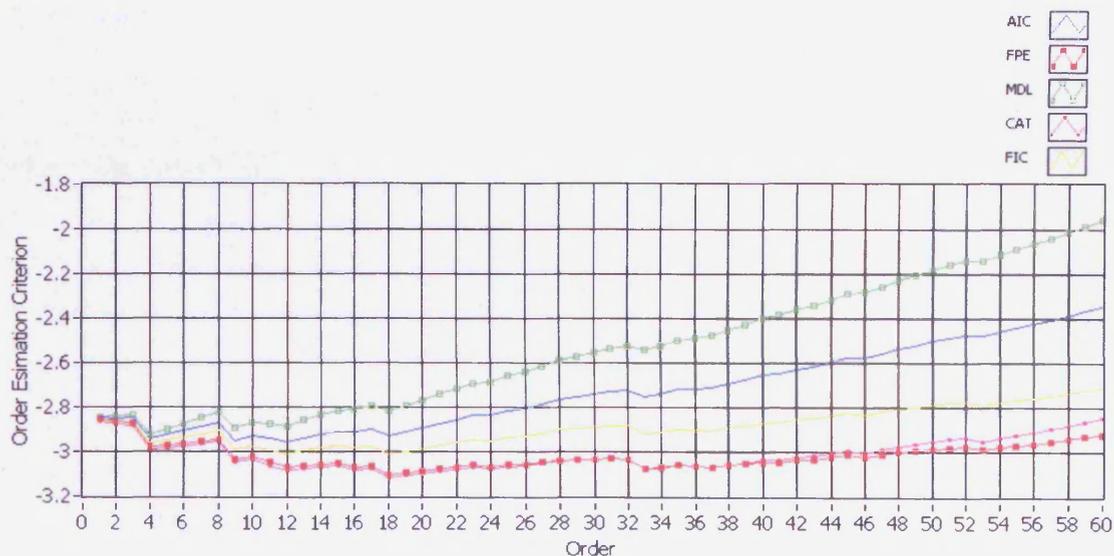


Figure 6–7 Finding the optimum order for the demodulated defective vibration signal.

6.7. Summary

This chapter discussed the usage of the resonance demodulation technique. The resonance excited by the vibration signal was identified and was band-pass filtered around the resonant peak (6 to 8 *kHz*) to extract the fault-induced bearing impacts. Hilbert transform was then applied to produce the analytic signal and the magnitude of the analytic signal is the demodulated signal. It has been shown that the application of resonance demodulation improves the sensitivity of the fault detection. The optimum model order was also found for the demodulated signals and justification was given of why a model order of 10 was chosen.

The next two chapters focus on the development of the fault detection tool using the AR pole trajectory method.

7. DEFINING A FAULT DETECTION TOOL

7.1. Introduction

So far, the investigations of the AR model have resulted in finding out the optimum AR model order for vibration signals obtained from the dry vacuum pump and establishing a new rule of thumb for determining the minimum model order for Autoregressive (AR) based spectrum analysis for rotating machinery (in Chapter 4) and determining the optimum AR sample size to be used (in Chapter 5). The application of the resonance demodulation technique to enhance the success of bearing fault identification has been investigated in Chapter 6. In this chapter, a novel fault detection tool is defined based on the groundwork laid by the previous chapters. The key step on the fault diagnosis of the ball bearings is the extraction of fault features and identification of the condition from the ball bearing vibration signals. The AR fault detection tool proposed here achieves this through pole-related spectral decomposition that allows the identification of the fault features. The tool is tested with real vibration data captured from the pump and results illustrating its performance are presented in the following chapter.

7.2. Motivation

In the extraction and investigation of the time variant spectral parameters, the pole representation is a very interesting approach, as it offers an immediate and easy comprehension of the associated behaviour of the vibration signals. The pole representation offers a qualitative insight into the quasi-steady stationary AR process and the spectral process characteristics can be more easily understood when expressed in terms of poles and AR spectral components. There is a one-to-one correspondence between the AR poles and the spectral peaks. By studying the movement of poles in the complex z -plane, the development and progression of bearing faults can be monitored. The tracking of pole locations offers a unique way of detecting bearing faults in rotating machinery and the proposed method has interesting potential applications in condition monitoring, diagnostic and prognostic-related systems, as

control engineers can relate easily to the idea of mapping poles in the complex domain.

The AR spectra can be decomposed into a set of relating poles as the transfer function of the AR model can be rewritten as an all-pole model. The AR poles are obtained by finding the roots of the AR coefficient polynomial in the transfer function. This set of AR poles has a one-to-one correspondence to the spectral peaks of the signals PSDs. The characteristic bearing defect frequencies, which were defined in Section 2.2.3, are directly extracted from the AR pole locations. The following section explains the relationship between the AR pole frequencies and the characteristic bearing frequencies and a mathematical derivation of how the pole frequencies can be obtained from the parametric time series AR model is given.

7.3. Preliminaries: Linking AR Spectra with the AR poles in the z-domain

The transfer function $H[z]$ of the AR polynomial was defined in equation [2.8]. By finding the roots of the AR coefficient polynomial in the denominator of $H[z]$, the poles, p_k , are obtained and the transfer function can be rewritten as [7.1]. Hence the model transfer function $H[z]$ can be expressed either in the form of AR coefficients a_k or poles p_k . These two forms of expressions contain essentially the same information on the studied AR process and define the process together with the value of the variance term σ^2 . An AR model's transfer function contains poles in the denominator plus only trivial zeroes in the numerator at $z=0$, so it is referred to as an "all-pole" model. Since the coefficients of $H[z]$ are real, the roots must be real or complex conjugate pairs. The number of poles in the z plane equals p , the AR model order.

$$H[z] = \frac{1}{(z + p_0)(z + p_1)\dots\dots(z + p_p)} = \frac{1}{z^{p+1} \cdot \prod_{k=0}^p (1 + p_k z^{-1})} \quad [7.1]$$

Why is the pole representation of the AR model [7.1] better than the AR coefficient representation [2.8] for fault detection? The AR filter coefficients a_k (defined in Section 2.3.2) are not stable and are highly dependent on the filter order and do not directly reflect the signal properties. The poles p_k , however, are more closely related to the spectral form and contain important information on the system condition that can be interpreted directly from the pole diagram. This is the reason why the pole representation of the AR transfer function is the preferred one. For a stable AR filter, the distance of all the AR poles from the origin, $|p_k|$, must be less than 1 and the nearer the pole is to the unit circle, the higher the corresponding peak in the AR spectrum. The magnitude of the AR spectral peaks is directly proportional to the estimated white noise variance σ^2 . Hence for two poles (each obtained separately from two different signals, with one signal having a higher energy than the other), even if both have the same distance from the centre of the unit circle, the pole derived from the signal with the higher energy, will translate to a bigger peak in the spectral domain because of the bigger variance of the white noise.

Each pair of complex conjugate poles in [7.1] has a one to one relation with a peak in the AR spectrum. A p^{th} order AR model with p poles will have a maximum of m peak frequencies in the AR spectrum where $m = p/2$ when p is even and $m = (p+1)/2$ when p is odd (the other pole being real). Not all poles give rise to sharp peaks in the AR spectrum. Only the poles which are close to the unit circle give rise to sharp peaks in the AR frequency spectrum (see Figure 7-1). The other poles are distributed around the unit circle to create an equiripple 'flat' PSD estimation. The symmetry with respect to the real axis is related to the fact that the signals have real values and this advantage can be conveniently exploited by disregarding the poles in the negative imaginary plane, reducing the redundancy of poles in the proposed pole tracking method.

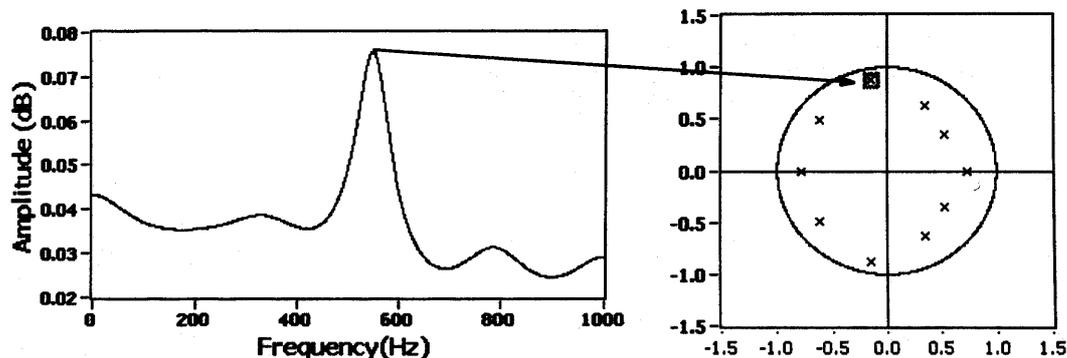
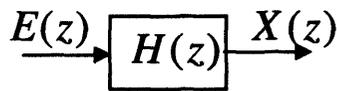


Figure 7-1 Monitoring the peaks obtained from the AR model.

In this case, AR model order was 8 and sampling frequency was 2000 Hz. Notice that it is only the pole nearest to the unit circle that gives rise to a sharp peak in the AR spectrum.

The process of transforming the AR time process to the AR pole positions is described pictorially in Figure 7-2. Figure 7-2(a) displays $x[t]$, the raw time domain signal to be processed. The transfer function described in Equation [2.8] is shown in Figure 7-2(b) first in the AR coefficients a_k representation and then being transformed to p_k poles values as defined by equation [7.1] (in this case, the AR identification process indicated has 10 coefficients). Figure 7-2(c) shows the PSD of the $x[t]$ signal, the AR spectra, $P(z)$, (defined by Equation [2.11]). Figure 7-2(d) shows the corresponding pole locations in the unit circle in the z-transform domain. Note that the spectral peaks in Figure 7-2(c) correspond to poles closer to the circle with the most dominant peak caused by the pole closest to the unit circle.

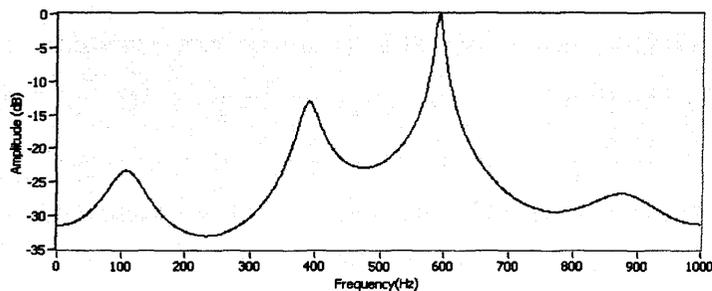
(a) Signal $x[t]$



$a_k =$	0.0964		$0.8706 + 0.3317i$
	0.7012		$0.8706 - 0.3317i$
	-0.1625		$0.3010 + 0.8722i$
	-0.1512		$0.3010 - 0.8722i$
	-0.0987		$-0.2844 + 0.9444i$
	-0.2099		$-0.2844 - 0.9444i$
	-0.0033		$-0.7550 + 0.3356i$
	0.4681		$-0.7550 - 0.3356i$
	0.1972		$-0.1241 + 0.5164i$
	0.2177		$-0.1241 - 0.5164i$

(b)

(c)



(d)

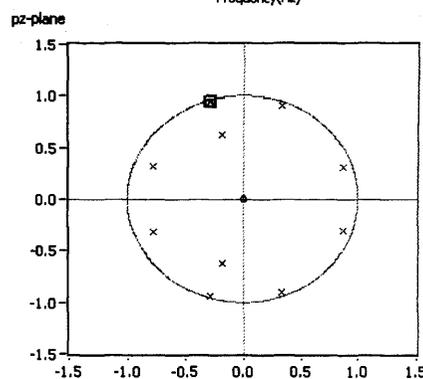


Figure 7-2 Pole Zero Representation.

(a) Signal $x(t)$ is modelled by an AR model, (b) Its transfer function $H(z)$ can be described in terms of coefficients a_k or poles p_k (c) the corresponding spectrum of the AR process is plotted (d) the pole diagram. The AR peaks are closely linked with the position of poles. The most dominant peak in the AR spectra is caused by the critical pole shown boxed in the pole zero diagram which is closest to the unit circle.

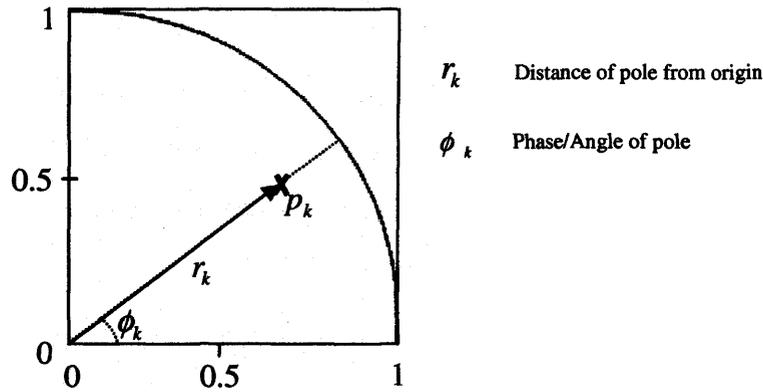


Figure 7-3 Defining the parameters of an AR pole.

Each pole p_k has a phase ϕ_k and a magnitude r_k which are given by the angular position and the distance from the origin (see Figure 7-3). By knowing the pole position inside the unit circle, the frequency f_k of each peak can be obtained from the phase ϕ_k of the pole if the sampling frequency f_s is known (equation [7.2]). Since the AR time series is quasi-stationary, the positions of the poles vary slightly with every frame of data. The trajectory mapped out by the poles can be quantified by finding the area traversed by the poles over a period of time. This aspect can be used as a parameter to indicate the condition of the bearing. The area mapped out by a characteristic bearing frequency pole will be different depending on whether they were obtained from no-fault or faulty conditions. The area mapped out by the migratory poles is given by equation [7.3].

$$f_k = 2 \cdot \pi \cdot \phi_k = \tan^{-1}(\text{Im}(z_k)/\text{Re}(z_k)) \times f_s / 2\pi \quad [7.2]$$

$$\text{Area} = \frac{1}{2} \cdot (\Delta r_k)^2 \cdot \Delta \phi_k \quad [7.3]$$

The power associated with each spectral pole p_k is estimated from the residues of the complex poles as proposed in (Johnsen and Andersen 1978). The transfer function in equation [7.1] can be rewritten as a partial fraction expansion. The residue (equation [7.4]) is simply the coefficient of the one-pole term $1/(1 - p_k z^{-1})$ in the

partial fraction expansion of $H[z]$ at $z = p_k$. Each residue r_k provides an estimate of the integrated power in the neighbourhood of the spectral frequency f_k associated with pole p_k . The spectral power P_k of the pole p_k is obtained by multiplying the real part of the residue term with the variance of the driving AR time series σ^2 and the scale factor n [7.5]. $n = 2$ for complex conjugate poles and $n = 1$ for real poles at either 0 Hz or at the Nyquist frequency.

$$r_k = z^{-1} \cdot (z - p_k) \cdot H(z) \Big|_{z=p_k} \quad [7.4]$$

$$P_k = \sigma^2 \cdot n \cdot \text{Re}[r_k] \Big|_{z=p_k} \quad [7.5]$$

7.4. Linking the AR Poles and the Characteristic Bearing Defect Frequencies

The preceding section described how the AR coefficients can be transformed to AR pole locations. The basic equations for finding the angle, area and power of the AR pole positions were also defined. In this section, the relationship between the AR pole locations and the characteristic bearing defect frequencies are formalized. The characteristic bearing defect frequencies for the ball bearings were defined in Section 2.2.3. The characteristic bearing defect frequencies can also be transformed to AR pole locations in the z domain. As an example, consider the case of a defective bearing with an inner race crack set to rotate at 105 Hz. The pole locations can be worked out for other defects and other rotating speeds but this simple example will allow us to elucidate the essence of our arguments without getting bogged down by details.

The theoretical ball bearing defect frequencies BSF, BPFO, BPF1 and FTF are estimated to be around 492 Hz, 384 Hz, 561 Hz and 43 Hz respectively when the pump's running speed is set to 105 Hz. The theoretical AR pole phase angles ϕ_k can be worked out if the sampling rate is known and in this case, the sampling rate f_s used was 2000 Hz. The angle of the theoretical pole locations of the characteristic

bearing defect frequencies are then worked out in degrees to be as shown in Figure 7-4 using equation [7.6].

$$\Delta \text{ angle (in degrees)} = \frac{\text{Characteristic Bearing Defect Frequency}}{f_s/2} \times 180^\circ \quad [7.6]$$

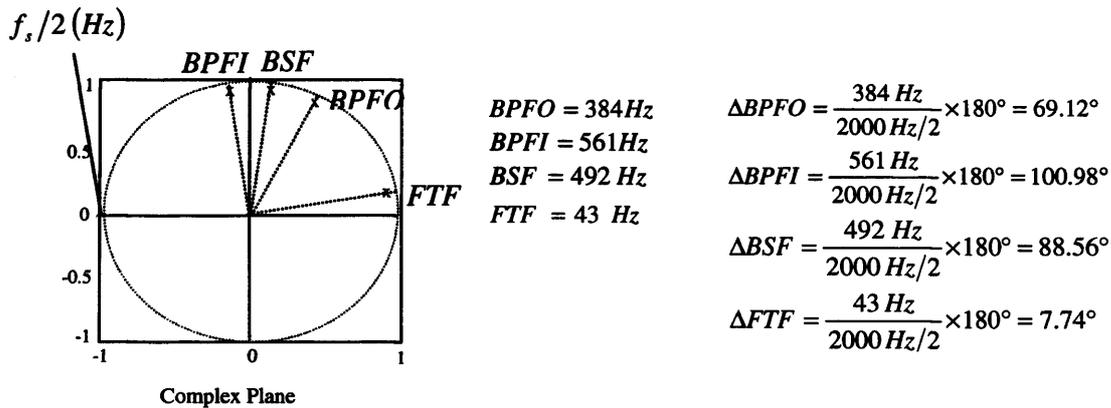


Figure 7-4 Theoretical location of characteristic defect frequencies in the z plane.

These can be worked out with standard reference formulas if the bearing dimensions are known.

Pump was set to 105 Hz and sampling frequency used was 2000 Hz.

The angles of the AR pole locations are related to the characteristic bearing defect frequencies. For a fixed rotating speed, the AR pole angles are fixed, as they are determined by the geometrical shape of the ball bearings used, but the distances of the pole locations from the unit circle are determined by the levels of vibration at that particular frequency. It is known that, as defects appear on the ball bearings and their severity increases with time, the amplitudes of the vibrations of characteristic bearing defect frequencies also increase (see Figure 7-5). This can be seen as the poles moving closer to the unit circle as the severity of the defect increases. The appropriate alarm level for the vibration signal can be determined from standards such as the ISO 10816 (ISO 1995) and ISO 7919 (ISO 1996). These can be translated to relative allowable amplitudes and alarm levels and hence corresponding pole displacements for the characteristic ball bearing defect frequencies.

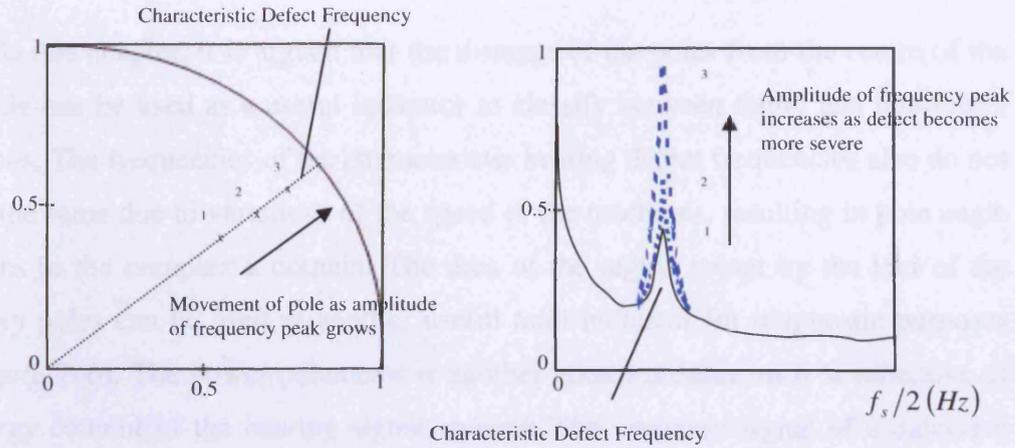


Figure 7-5 Typical movement of pole as defect becomes more severe (from 1 to 3)

The nearer the pole is to the unit circle, the bigger is the amplitude seen in the frequency spectrum. The amplitude of vibration of characteristic defect frequency increases as the pole moves closer to the unit circle.

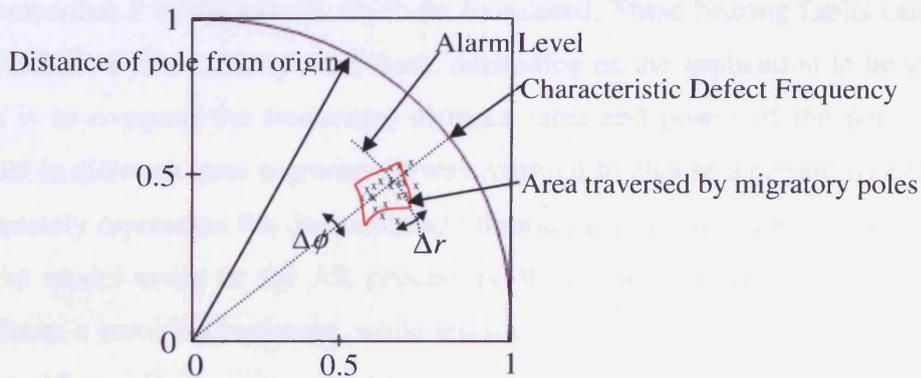


Figure 7-6 Locus of a particular pole versus evolution of time.

Note that pole is migratory and the area pole traverses can be calculated. Alarm levels for the characteristic defect frequencies can be determined from standards such as the ISO 10816 and, if pole crosses the alarm level, we can conclude that the defect is causing the characteristic frequency and hence the pole displacement is severe.

In this chapter, it is argued that the distance of the poles from the centre of the unit circle can be used as a useful indicator to classify between faulty and non-faulty conditions. The frequencies of the characteristic bearing defect frequencies also do not remain the same due to variations of the speed of the machines, resulting in pole angle variations in the complex z domain. The area of the region swept by the loci of the migratory poles can be used as another useful fault indicator for diagnostic purposes (see Figure 7–6). The power parameter is another good indicator, as it is reflective of the energy content of the bearing signal studied. The vibration signal of a defective bearing is normally much higher than the vibration signal obtained from good bearings, as the energy of certain frequency bands is excited due to the presence of faults.

7.5. An Overview of the AR Pole Trajectory Fault Detection Tool

The proposed fault detection tool consists of two phases: training and detection. It is assumed that S bearing faults are to be monitored. These bearing faults can be BPFO, BPFI, BSF, FTF or multiples of these, depending on the application to be studied. The idea is to compare the frequency, distance, area and power of the poles of the AR model in different time segments. It was required to choose a certain model order that adequately represents the demodulated vibration data for all time segments. Selection of the model order in the AR process is of critical importance. Too low an order produces a smoothed estimate, while too high an order may cause spurious peaks and spectral line splitting. This aspect has been clearly clarified in Chapter 4. In Chapter 6, it was justified why an AR model order of 10 was chosen for the demodulated vibration signals. A 10th order AR model was used to study the behaviour of the demodulated ADXL105 vibration signals mounted on the HV end for a pump operating in normal conditions and also for a pump fitted with a bearing with defects.

Training Phase:

1. Each training set of vibration data was formed constituting an observation of the data when there are no faults in the bearings. First, the AR parameters are found for the bearings in a healthy condition and subsequently, σ^2 , the variance of the input signal in the no-fault condition, is also derived.

2. Changes in the condition of the ball bearings and the non-stationarity of the vibration signals cause deviations in the AR coefficients, which, in turn, result in deviations in the AR pole positions. The positions of the AR poles vary from frame to frame. The AR poles, which are the roots of the AR coefficient polynomial, are obtained for each frame of data. There are p poles when a model order of p is used.
3. Out of these, the critical poles for each characteristic bearing defect frequency are found. There is a one-to-one mapping between the characteristic bearing defect frequencies and the angular frequencies of the critical poles (equation [7.2]). The poles of interest are the ones which have their angles closest to the angles corresponding to the characteristic bearing defect frequencies. These are termed the '**critical poles**' and they are monitored.
4. Next, parameters λ are derived from the critical poles in the normal condition. The parameter λ can be the area of the region swept by the migratory loci of the critical poles, their distances from the centre of the unit circle, or the power of the critical poles. The parameters λ reflect the characteristics of the time invariant spectral parameters and comprise important information of the system condition. The impulses produced by ball bearings with defects will be modelled by the changing system dynamics and this in turn will be reflected in the parameters λ . The parameters λ are the conditional indicators that discriminate between the damaged and non-damaged bearings.

Detection phase:

Boundary conditions can be established using the known angles of characteristic defect frequencies and acceptable levels obtained from standards, and this can help the optimum threshold for the classification boundary to be determined. Once the boundary conditions for the healthy signal are extracted and normalised, the procedure is repeated for the detection phase with vibration signals from a test bearing. If the test bearing is a bearing with a defect, the parameter λ of the critical poles will not normally be within the acceptable levels of the established boundary conditions (refer to Figure 7-6) or be above a certain predetermined threshold. A reasonable threshold level can be chosen based on the probabilities calculated from the

training data. In this case, the threshold value was determined using Receiver Operating Characteristics (ROC) analysis.

One important and useful measure of the performance of any diagnostic test is ROC curve analysis. A threshold is determined for separation of data into two classes based on some decision parameter and, depending on whether the subject falls below or above the cut-off level, the subject is termed 'positive' or 'negative'. In reality, there will be some overlap between the two classes of data and some subjects will be misclassified as 'false positives' and 'false negatives'. The sensitivity and specificity of the data are defined as

$$\text{Sensitivity} = \frac{\text{Number of true positive decisions}}{\text{Number of actual positive decisions}} \quad [7.7]$$

$$\text{Specificity} = \frac{\text{Number of true negative decisions}}{\text{Number of actual negative decisions}} \quad [7.8]$$

ROC curve analysis can be established by taking the parameters λ as the fault indicators, and then working out the condition and fault patterns of the ball bearings. Depending on what defect the bearing has, such as whether it is a defect on the inner race or outer race, the defect can be identified. The above explained diagnosis approach is under the assumption that only one fault is present. However, in reality, multiple faults can develop. In such a case, the scheme can be easily modified to incorporate the diagnosis of multiple faults.

7.6. A Diagrammatic Illustration of the Fault Detection Tool

Tracking the parameters λ is performed by processing each raw time domain sample through the stages shown in the diagram of Figure 7-7. This diagram summarises all the key steps in processing the data for obtaining the fault indicators from the critical poles. In summary, this diagram is a brief amalgamation of the all the topics addressed in this thesis. The vibration signals are captured from the bearings to be diagnosed, filtered with an anti-aliasing 10 kHz low-pass filter, and sampled at 40 kHz. The

filtered vibration signals were then amplitude demodulated by band-pass filtering around one of the resonant peaks to remove the structural resonance and to strengthen the weak bearing signals. The Hilbert transform is then applied to the band-passed signal to obtain the squared envelope and the resulting signal is the demodulated vibration signal. The resonance demodulation technique improves the signal-to-noise ratio of the vibration signatures for a more effective detection of bearing defects. The data is then downsampled to 2 kHz, as it was known that the frequencies of interest lie in the range from 0-1 kHz.

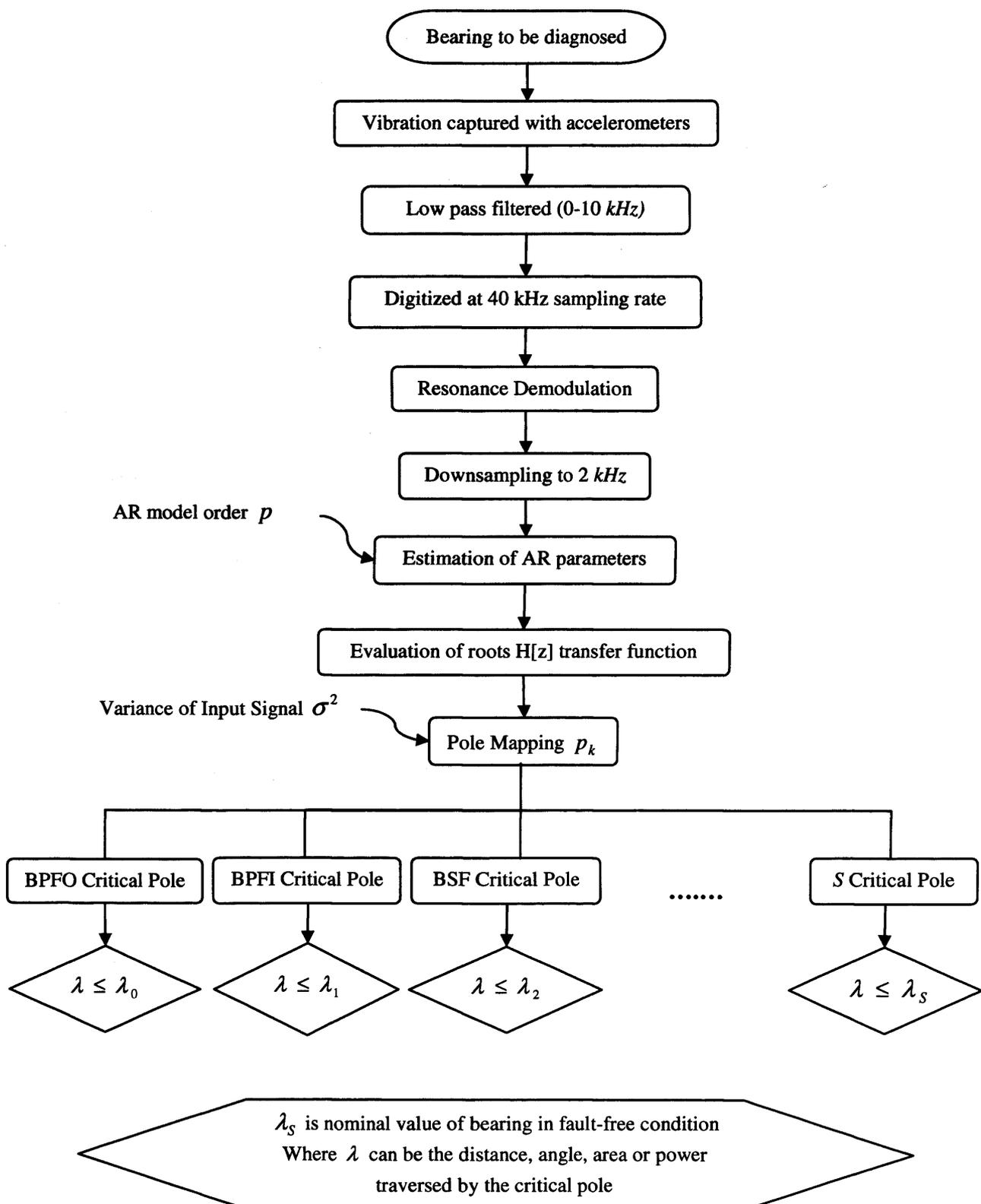


Figure 7-7 Block diagram of the procedure for bearing fault detection using AR Pole Tracking

The AR coefficients are obtained with fixed model order of 10 and subsequently, the AR pole positions are determined for each frame of data, as described by the previous section. The critical poles corresponding to the bearing defect frequencies are identified. From the positions of the AR critical poles, parameters λ are derived. The parameters λ are compared with those obtained from non-faulty conditions and the bearing is diagnosed to be whether in faulty condition or not.

7.7. Summary

A formal framework for automated assessment of bearing faults using the AR pole positions has been established in this chapter and defined as a fault detection tool. Metrics in the form of area, angle, power and distance of the AR pole positions (λ parameters) were defined for fault detection purposes. By quantifying the values of the parameters λ , bearing faults are to be identified. The characteristic bearing frequencies were modelled and mapped using the AR method in the z domain. The critical poles are those closest, in terms of angular position, to the characteristic bearing frequencies poles' positions. It is proposed that the parameters λ are then compared with baseline figures obtained from healthy bearings to detect the presence of faults. A higher power and a bigger distance from the origin of the AR poles will be indicative behaviours of faulty bearings.

The proposed method has useful applications in fault classification as faults can be detected by studying the movement of poles in the complex plane without having to compute the whole PSD spectra (normally done using the FFT method), which can be computationally intensive. The bearing defect frequencies can be decomposed to spectral pole positions and only the critical pole positions need to be tracked.

The novel fault detection tool is tested with actual data obtained from the pump. The effectiveness of the tool and results are presented in the next chapter.

8. AR POLE BASED MONITORING

8.1. Introduction

In the previous chapter, a fault classification scheme based upon the position of the coordinates of the Autoregressive (AR) poles was formally established. The AR coefficients for each time sample are obtained from the estimated AR model parameters. The poles, which are the roots of the AR coefficient polynomial, are then derived. The critical poles are identified and further decomposed into parameters λ . The parameters λ are extracted as features relating to the characteristic bearing defect frequencies. In this chapter, experimental tests are conducted to identify the differences in the behaviour of the parameters λ for vibration signals collected from two similar high speed dry vacuum pumps, one with a healthy set of bearings and another with a ball bearing with an inner race defect. The effectiveness of the AR pole trajectory fault detection tool is verified and characterised using ROC analysis. The stability of the AR models is also investigated by calculating the condition numbers in the faulty and non-faulty experimental conditions.

In Section 8.7, to demonstrate the competence of the proposed tool, the classification performance of the AR pole trajectory tool is compared with the FFT-based method. In the final section, both the mentioned techniques are benchmarked to give an estimate of their computational cost.

8.2. Tracking the Critical Pole Movements

A 10th order AR model was estimated for the ADXL105 vibration signals mounted on the HV end for a pump operating in normal conditions and also for a pump fitted with a bearing with an inner race fault. The BPF1 pole was identified as the critical pole. The loci of the BPF1 critical poles on the z plane were tracked in different time segments, considering the mean and standard deviations of the parameters λ , and seeking differences between results for damaged and undamaged bearings. The BPF1 pole was chosen as the dominant pole as the fault condition studied was the case of a pump fitted with a bearing with an inner race fault. If other bearing defects are to be monitored, the corresponding critical pole positions relating to the characteristic defect frequencies of significance can also be tracked. Frame sizes of 4000 samples were

used and 100 frames of data each of length 2 s were processed to monitor the movement of the BPF I pole corresponding to the inner race defect frequency. A 10th order AR model was used to estimate the parameters λ for successive time segments. The distance of the BPF I pole from the origin, the power of the BPF I pole, the angle mapped by the BPF I pole and the cumulative area traversed by the BPF I pole were plotted for the vibration data for both normal and faulty conditions.

Refer to Figure 8–1. The mean \pm standard deviation of the distance of the BPF I poles from the origin in the non-faulty case was 0.8005 ± 0.0580 . The mean \pm standard deviation distance of BPF I poles from the origin in the faulty case was 0.9303 ± 0.0148 . It can be seen that the mean distance of the BPF I poles was much larger for the faulty condition than in the normal case. This will be seen as bigger amplitude of spectral peaks associated with critical poles (amplitude of spectral peaks is proportional to the inverse of the distance of pole to the unit circle) when the AR spectrum is plotted for the faulty condition.

Refer to Figure 8–2. The mean \pm standard deviation of the power of the BPF I poles in the non-faulty case was $9.570 \times 10^{-4} \pm 2.831 \times 10^{-4}$ and in the faulty case 0.0369 ± 0.0086 respectively. Clearly the power of the faulty poles is much larger than the non-faulty case. The power and distance of the poles from the origin are related and are dependent on each other. Though the distance and power of the BPF I critical poles can be used as effective indicators for fault classification, the power is a better classifier than the distance since it incorporates the variance term of the signal (which is equal to the energy of the signal, and energy of faulty signals is much larger than for non-faulty signals). In the example shown, the power seemed to increase with time as the number of frames monitored increased. This is not surprising as the amplitude of vibrations also increased with time, due to the pump heating up and increased friction of the bearings when it was running for a longer time.

Refer to Figure 8–3. The mean angle of the BPF I pole was 1.6117 radians (513 Hz) and 1.7441 radians (555 Hz) for normal and faulty condition respectively. The mean angles of the BPF I poles in both cases were nearly the same. However the standard deviation of the angles of the BPF I poles ($\Delta\phi_k = 0.3094$) for the normal

condition was much larger than for the faulty case ($\Delta\phi_k = 0.097$). The standard deviation of the BPFIs angles can be used as an indicator for fault classification to distinguish between the non-faulty and defective cases as the difference between them is significant. The changes in the frequency content of the characteristic defect frequencies can be quantified by using AR modelling by looking at the corresponding changes in the angles of the BPFIs poles. Refer to Figure 8–4. The cumulative area traversed by the migratory faulty BPFIs poles (less than 0.01) was also markedly smaller than the case for the non-faulty condition (0.2). This is because when there is a BPFIs fault the pole corresponding to the BPFIs frequency remains at almost the same position, since there is vibration at BPFIs. When there is no BPFIs fault the so-called ‘critical pole’ is simply the pole closest to the BPFIs angular frequency, but since there is no strong vibration at that frequency the non-faulty BPFIs poles move considerably from frame to frame.

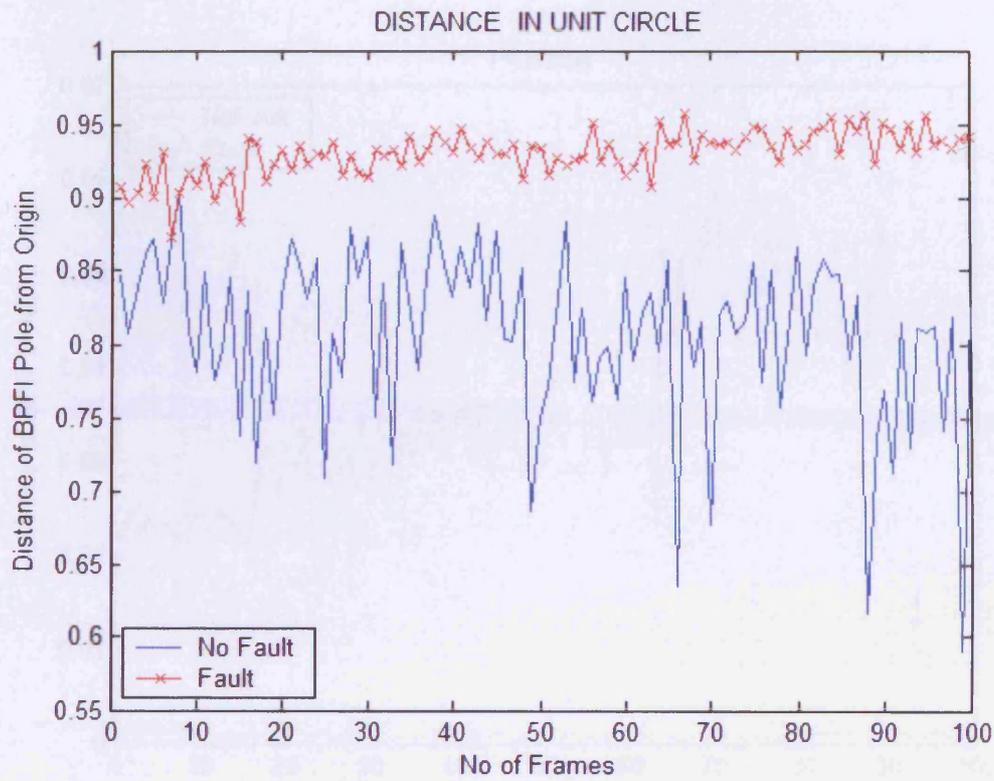


Figure 8-1 Distance of BPF1 pole from origin.

Monitoring the distance parameter λ . Pump running at 105 Hz. 100 frames of ADXL105 vibration data used each of 2s length. A 10th order AR model was used. The distance of the BPF1 pole for faulty condition is markedly higher than for normal condition.

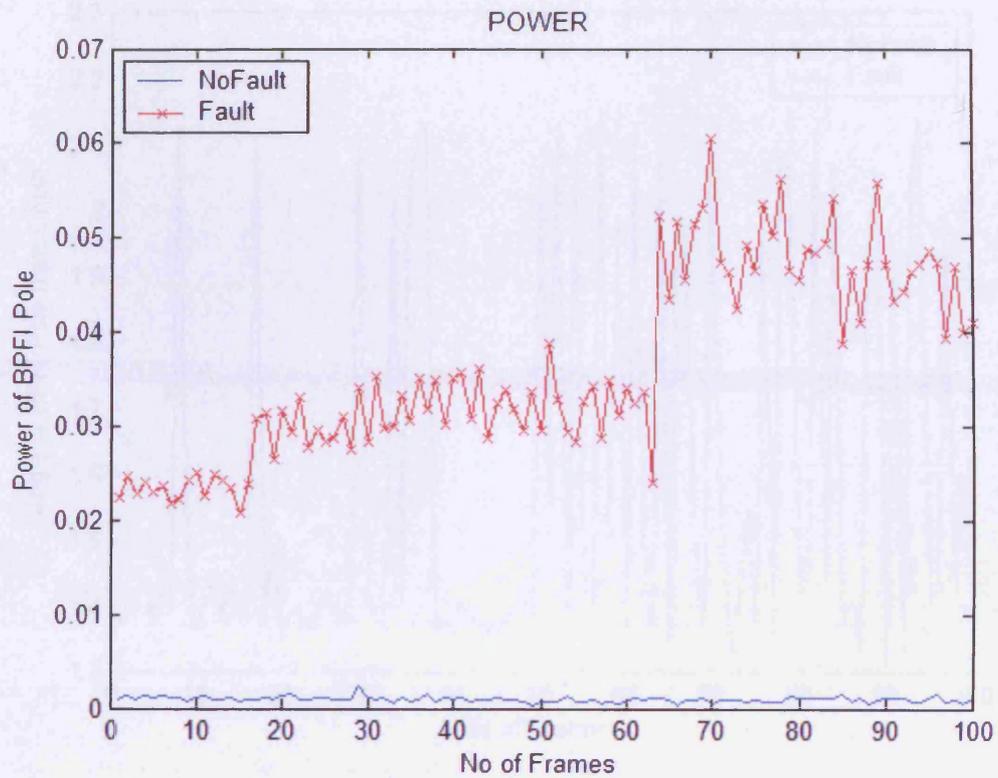


Figure 8-2 Power of BPF1 poles.

Monitoring the power parameters λ . Pump running at 105 Hz. 100 frames of ADXL105 vibration data used each of 2s length. A 10th order AR model was used. The power of faulty poles is much larger than for the non-faulty poles.

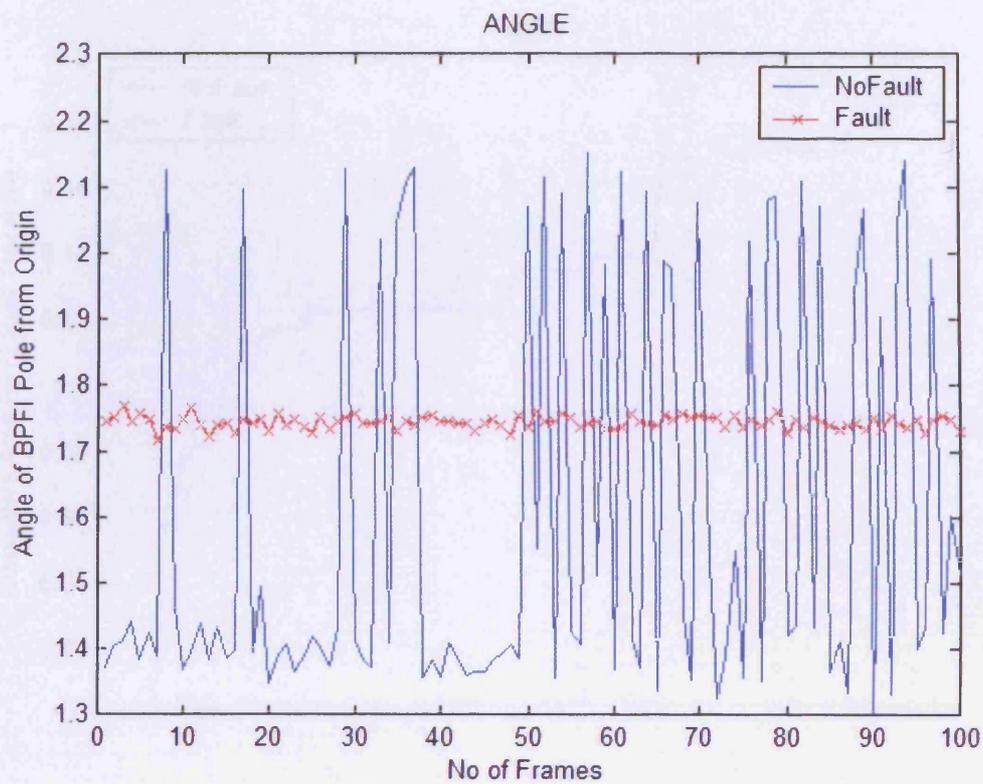


Figure 8-3 Angle BPF1 pole traverses.

Monitoring the angle parameters λ . Pump running at 105 Hz. 100 frames of ADXL105 vibration data used each of 2s length. A 10th order AR model was used. For normal condition, there was much variation in the BPF1 angles, but for faulty condition the variance was much less.

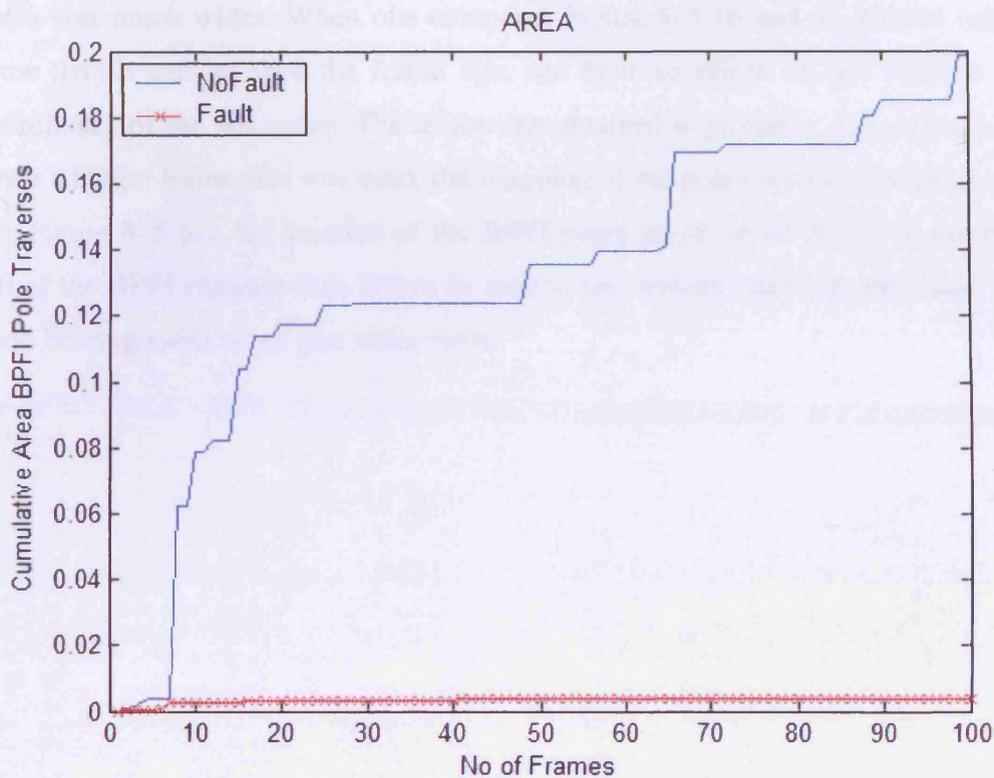


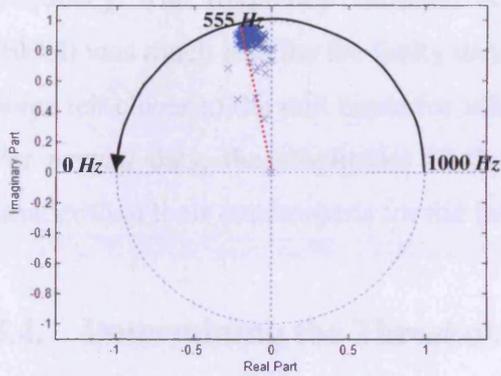
Figure 8-4 Cumulative area traversed by the BPF I poles.

Monitoring the area parameters λ . Pump running at 105 Hz. 100 frames of ADXL105 vibration data used each of 2s length. A 10th order AR model was used. The area mapped out for non-faulty condition was much larger than that for faulty condition.

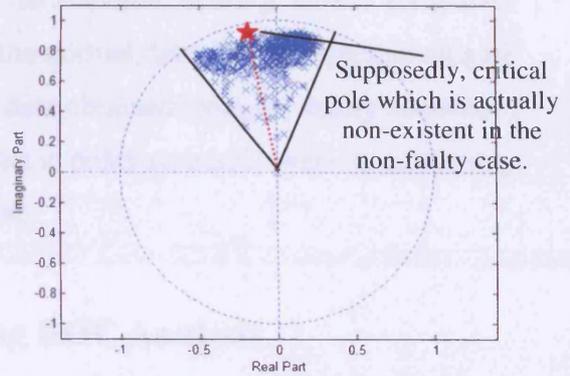
8.3. Testing the Effect of Frame Size

In order to model the changes in the vibratory signatures, the critical poles are obtained for 2 different frame sizes, for both normal and faulty conditions. Figure 8-5 (a and b) shows the BPF I poles plotted for a frame size of 2000 samples. Figure 8-5 (c and d) shows the BPF I poles plotted for a frame size of 5000 samples. The BPF I pole was chosen as the dominant pole, as the fault condition studied was the case of a pump fitted with a bearing with an inner race fault. The AR identification method placed the faulty BPF I critical poles centred around 555 Hz. There was a significant difference in the spread of the poles for the cases of damaged and undamaged bearings respectively.

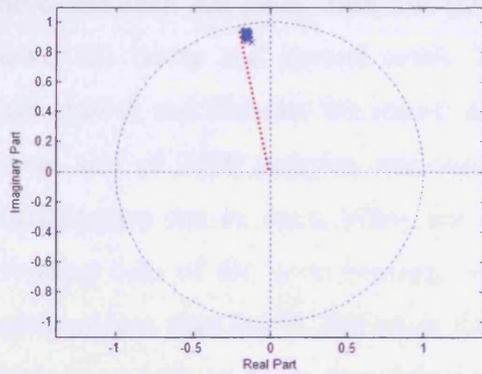
For the faulty condition, the spread of the poles was more concentrated and nearer to the unit circle for both frame sizes. For the non-faulty BPFIs, the spread of the poles was much wider. When one compares Figure 8–5 (b and d), plotted using the same data it can be seen the frame size can have an effect on the location of the coordinates of the AR poles. The difference obtained is probably due to the fact that when a bigger frame size was used, the mapping of the poles seems to be more robust. For Figure 8–5 (d), the location of the BPFIs poles seems to be either on the right or left of the BPFIs characteristic defect. In reality, the demodulated vibration data from a good bearing looks more like white noise.



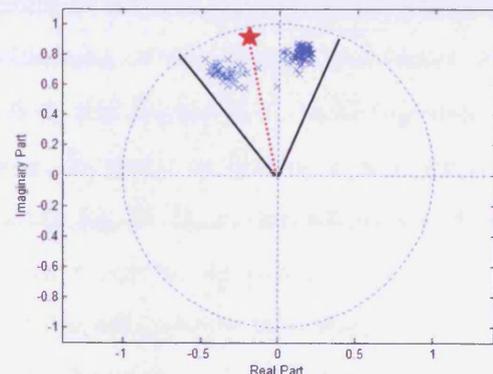
(a) Frame size of 2000 samples. Faulty poles.



(b) Frame size of 2000 samples. Non-faulty poles.



(c) Frame size of 5000 samples. Faulty poles.



(d) Frame size of 5000 samples. Non-faulty poles.

Figure 8-5 Distribution of poles from a pump running in normal conditions (right) and for a case with a bearing which has an inner race fault (left).

The dotted red line denotes BPF frequency of 555 Hz. Note clustering of poles near unit circle for BPF frequency at approximately 555 Hz for faulty data. Spread of poles is much larger for normal conditions.

and there is no actual significant peak and hence no corresponding BPF pole in that sense. The definition of the critical pole in this context is 'the closest pole to the characteristic defect frequency' (the BPF pole in this case) and the method has located the poles which happen to be closest, in angular frequency, to the BPF frequency. The frequency variation of the characteristic bearing defect frequency (BPF) was much less for the faulty data than the normal data. The complex conjugate poles fell closer to the unit circle for vibration data obtained from the faulty condition. For normal data, the amplitudes of the dominant poles (moduli) were considerably smaller than their counterparts for the faulty case.

8.4. Determining the Threshold using ROC Analysis

ROC curve analysis was used to determine suitable threshold levels for the fault indicators such as AR poles distances from the centre of the unit circle and the AR poles power for faulty and normal cases. The distributions of the normal and faulty AR poles power and distance are shown in Figure 8-6 and Figure 8-7, as histograms. A frame size of 2000 samples was used for these. In terms of power, a near perfect classification can be seen. When the power (using Eq. [7.5]) of the AR poles of the vibration data of the good bearings were tracked, it can be seen they all had power values of less than 0.005. But when the power of the AR poles of the vibration data of the bearing with an inner race defect was tracked, the faulty poles' power had a wide distribution. They had values ranging from 0.015 to 0.065. There was no overlapping of the values, hence the power of the AR poles can be used as a 'perfect' fault classifier. For the distance of the AR poles from the centre of the unit circle, the normal poles had a distribution ranging from 0.4 to 0.95. The faulty poles had a distribution ranging from 0.81 to 0.97. There was some overlapping between the two classes. For instance, if 0.9 was chosen as the threshold level to determine if the poles corresponded to the case of damaged or non-damaged bearings, there is a chance that some normal poles with a distance of more than 0.9 are misclassified as faulty poles and vice-versa. However it was noted that the frame size used does have an effect on the amount of overlapping between the two classes of data and hence the sensitivity and specificity of the analysis. Figure 8-8 shows the distribution of normal and faulty

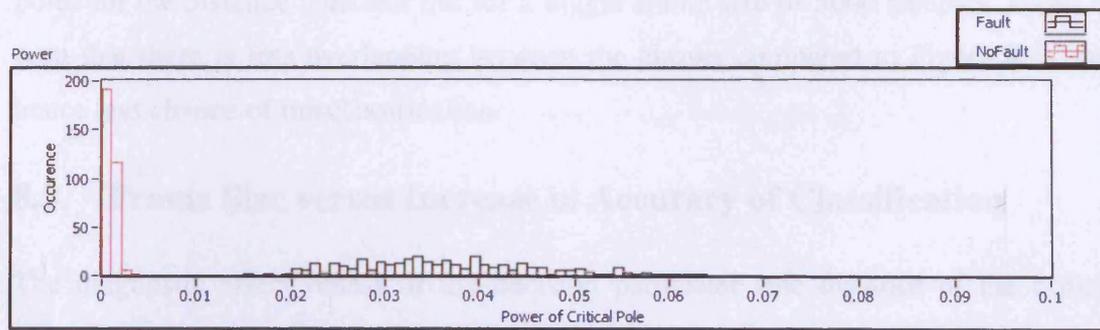


Figure 8-6 Distribution of Normal and Faulty Poles power parameter.

Pole distribution is separated into two normal distinct regions. No Overlapping of Normal and Faulty Poles. Note these were plotted for frame sizes of 2000 samples. No ROC analysis necessary.

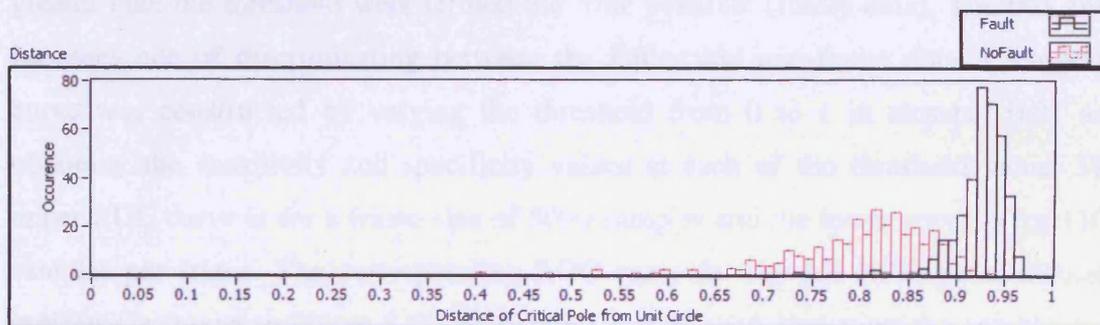


Figure 8-7 Distribution of Normal and Faulty Poles using distance from the centre of the unit circle as the decision parameter (2000 Samples).

Some overlapping of Normal and Faulty Poles. Note these were plotted for frame sizes of 2000 samples. If bigger sample sizes are used, less overlapping will occur.

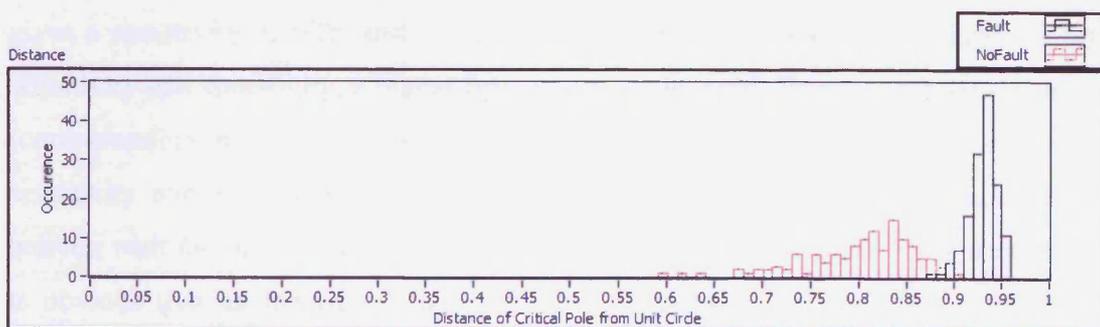


Figure 8-8 Distribution of Normal and Faulty Poles using distance from the centre of the unit circle as the decision parameter (5000 Samples).

Less overlapping of Normal and Faulty Poles. Note these were plotted for frame sizes of 5000 samples.

poles for the distance indicator but for a bigger frame size of 5000 samples. It can be seen that there is less overlapping between the classes compared to Figure 8–7 and hence less chance of misclassification.

8.5. Frame Size versus Increase in Accuracy of Classification

The diagnostic effectiveness of the decision parameter (the distance of the critical poles from centre of the unit circle) was estimated by constructing a curve of sensitivity against 1-specificity for windows of increasing frame sizes of 1100 to 5000 samples. The poles with a distance from the origin smaller than the threshold were termed the ‘true negative’ (non-faulty data). The poles with a distance from the origin greater than the threshold were termed the ‘true positive’ (faulty data). The task then becomes one of discriminating between the faulty and non-faulty data. Each ROC curve was constructed by varying the threshold from 0 to 1 in steps of 0.01 and obtaining the sensitivity and specificity values at each of the threshold value. The upper ROC curve is for a frame size of 5000 samples and the lower curve is for 1100 samples per frame. The corresponding ROC curve for the AR BPF I poles distance indicator is shown in Figure 8–9 . Each point on the curve represents the combination of true positives against false negatives estimated for a given threshold of the parameter. The aim is to maximise detection probability while minimizing false alarm rates. A good test curve is one for which sensitivity rises rapidly and 1- specificity hardly increases at all, until the sensitivity is high. Using a frame size of 2000 samples gives a sensitivity of 92% and a 1-specificity of 8%. If it is desired to increase both sensitivity and specificity, a bigger frame size can be used. Frames with 4000 samples (corresponding to a mere 2 s of data collection) produced excellent results, with sensitivity and specificity of 99%. This means that the undamaged and damaged bearing with the inner race defect can be distinguished with almost 100% accuracy. It is obvious that as the frame size increases the accuracy of the classification also increases. The sensitivity and specificity versus distance curves for a frame size of 4000 samples are shown in Figure 8–10. For example, selecting a distance value of 0.885 as threshold, leads to a sensitivity of 99% and a specificity of 99%.

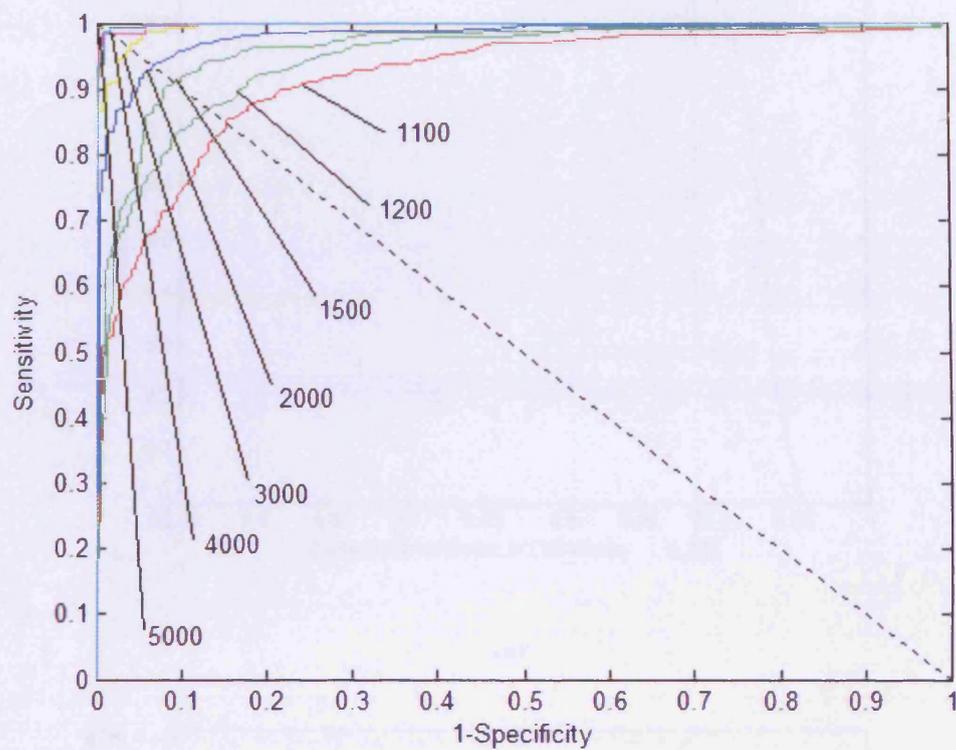
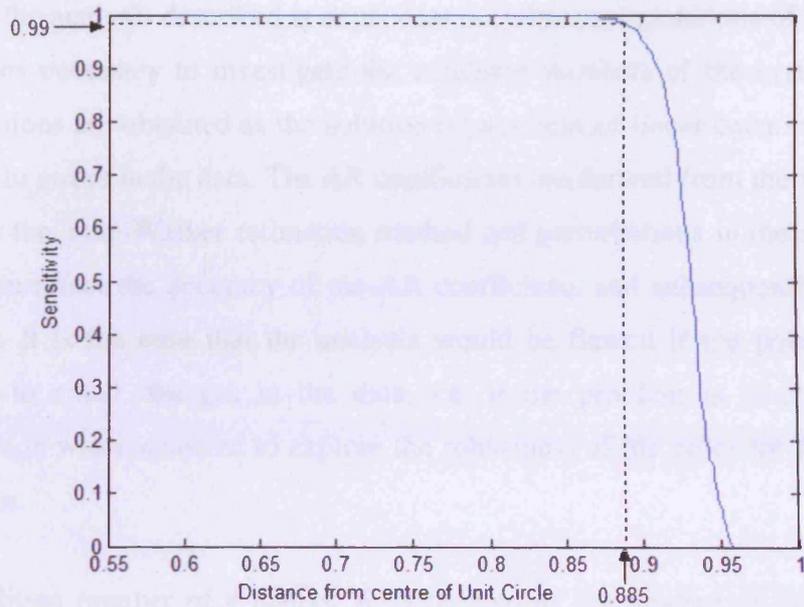
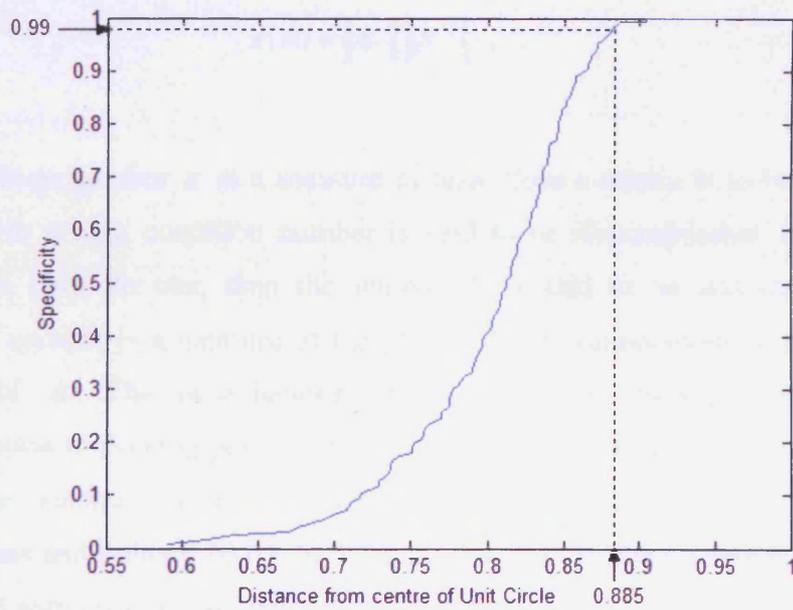


Figure 8-9 ROC curves for increasing frame sizes for the distance of the poles from the centre of the unit circle as a decision parameter.



(a)



(b)

Figure 8–10 Sensitivity (a) and Specificity (b) curves for a frame size of 4000 samples.

Note that distance from centre of unit circle was used as the decision parameter. Even when using the distance as the diagnostic parameter, one can obtain 99 % sensitivity and 99% specificity for frames of 2 seconds duration.

8.6. Condition Number of Critical Poles

Much of the analysis described is dependant on robust computations of pole locations. It becomes necessary to investigate the condition numbers of the system as the AR pole positions are obtained as the solution of a system of linear equations and will be sensitive to errors in the data. The AR coefficients are derived from the autocorrelation matrix in the Yule-Walker estimation method and perturbations in the autocorrelation matrix can affect the accuracy of the AR coefficients and subsequently the AR pole positions. It is the case that the analysis would be flawed if the poles were highly sensitive to small changes in the data, i.e. if the problem is ill-conditioned. An investigation was conducted to explore the robustness of the poles for faulty and non-faulty data.

The condition number of a matrix A is defined as the product of the norm of that matrix and the norm of its inverse.

$$\kappa(A) = \|A\| \|A^{-1}\|. \quad [8.1]$$

The condition number κ is a measure of how close a matrix is to being singular. A matrix with a high condition number is said to be ill-conditioned. If the condition number is close to one, then the matrix A is said to be well-conditioned. The condition number is a measure of the sensitivity of computations to perturbations in the matrix A . The perturbations could be due to measurement errors or to representation in floating point format. A tolerable limit that is commonly accepted is that if the condition number is below 100 then the matrix is not ill-conditioned (Ebrahimian and Baldick 2001). In this case, computations involving the matrix can be performed with greater confidence.

The eigenvalues, λ , or characteristic roots of a $k \times k$ square matrix A are the k solutions that satisfy Eq.[8.2].

$$|A - \lambda I| = 0 \quad [8.2]$$

where $|\cdot|$ is the matrix determinant and I is the identity matrix with the same dimension as A .

For computation of the condition numbers for the AR poles, the first step is to transform the vector of AR coefficients, \bar{A} , of each time sample of data to its canonical form or companion square matrix equivalent. From the known vector of AR coefficients of p^{th} order, the equivalent canonical matrix has the form

$$\tilde{A} = \begin{bmatrix} -a_p & 1 & 0 & \dots & 0 \\ -a_{p-1} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -a_1 & 0 & 0 & \dots & 1 \\ -a_0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [8.3]$$

The eigenvalues of the \tilde{A} matrix are equivalent to the AR poles. The 2-norm condition number of the \tilde{A} matrix for faulty and non-faulty data for increasing frame sizes was found using the 'cond' function in Matlab Toolbox and results are shown in Table 8-1. It can be seen that the pole computations can be performed with confidence regarding the numerical conditioning. Noting that the 2-norm condition number of a matrix is an upper bound on the sensitivity of the individual pole locations to small changes in the data, it can be confirmed that all poles are reasonably well conditioned and thus the paradigm of monitoring pole deviations is well founded. Such movement is due to engineering phenomena and not due to computational inaccuracy.

Table 8–1 Condition numbers for increasing frame sizes for Non-faulty and Faulty Data obtained using the ‘cond’ function in Matlab.

Sample size of Frame	Non-Faulty Data	Faulty Data
1100	12.06	13.24
1200	10.85	13.54
1500	13.29	17.99
2000	14.50	19.35
3000	16.48	26.73
4000	19.71	36.94
5000	26.60	36.62

It is interesting to note that the condition numbers of the faulty data were always more than for the non-faulty counterpart in this case, despite any size of the frame of data inferring that they were less well-conditioned than the non-faulty data. If the poles of an AR process are located near the unit circle, even small inaccuracies in the data can cause deviations in the AR parameters and this can cause the poles to move outside the unit circle, resulting in an unstable (and invalid) model. Double precision arithmetic was used to minimise chances of ill-conditioning due to the numerical floating-point representation. The condition numbers obtained were always less than 40, even for the faulty case, indicating that the level of ill-conditioning is tolerable.

Another trend observed was that, as the frame size was increased, the condition numbers for both types of data increased, indicating that they are more sensitive to perturbations in the data for bigger frame sizes. It is believed that there is a tendency to overfitting when bigger frame sizes are used. Even then the level of ill-conditioning is not large, indicating that larger sample sizes can be tolerated due to the increase in sensitivity rates of the detection capability as verified by the ROC analysis.

The smallest sample size that can be used with the AR pole based trajectory tool is 1100 samples. Using a smaller sample than this renders the AR model invalid as once the data is downsampled to fit into a smaller frequency range, the length of the downsampled is too small compared with AR model order, and the autocorrelations functions and consequently, the AR parameters cannot be effectively calculated.

8.7. The AR Pole based Trajectory Technique Versus the FFT-based Technique

The performance of the fault detection tool based on AR pole trajectory was compared with the FFT technique for an evaluation of its effectiveness at detecting those frames of data from the pump with the inner race fault. For the FFT technique, varying frame sizes of data obtained from the pump running in normal conditions and also for a pump running in the faulty condition (pump fitted with the bearing with inner race crack) was obtained and transformed into the frequency domain.

For a fair comparison to be made, the data was also demodulated using the Hilbert transform and downsampled to 2 *kHz* prior to the estimation of the maximum frequency peak. For the FFT-based technique, no windowing, 0% overlap and no averaging was used to obtain the frequency estimates. The amplitude of the maximum spectral peak in the frequency range from 0 to 1 *kHz* was tracked for both types of data and results are shown. Figure 8–11, Figure 8–12 and Figure 8–13 show the distribution of the maximum spectral peak for a frame size of 2000 samples, 5000 samples and 10000 samples respectively.

For a frame size of 2000 samples (Figure 8–11), the performance of fault classifier based on the FFT technique was very poor, especially for those data obtained from the faulty pump. The amplitude of the maximum peak of the faulty data varied from a minimum value of 0 up to a maximum value of 0.85. There was also a lot of overlapping between the faulty and non-faulty data. When the frame size was increased to 5000 samples (Figure 8–12), the spread of the maximum peak decreased and there was less overlapping. Only when the frame size was increased to 10000 samples, there was no overlapping between the faulty and non-faulty data and the spread of the faulty data became more deterministic by having a smaller spread of values from 0.04 to 0.4.

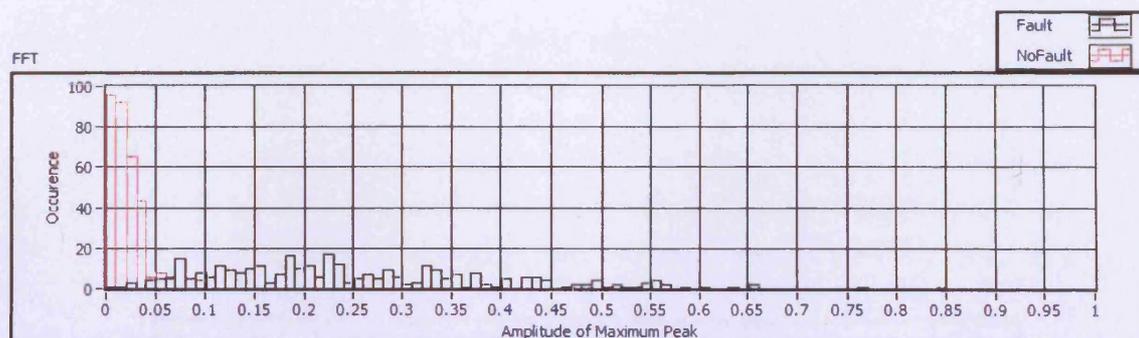


Figure 8-11 Distribution of Normal and Faulty peaks for the FFT-based technique plotted for frame size of 2000 samples.

Note lots of overlapping between faulty and non-faulty peaks and wide distribution of faulty peaks between 0 and 0.85.

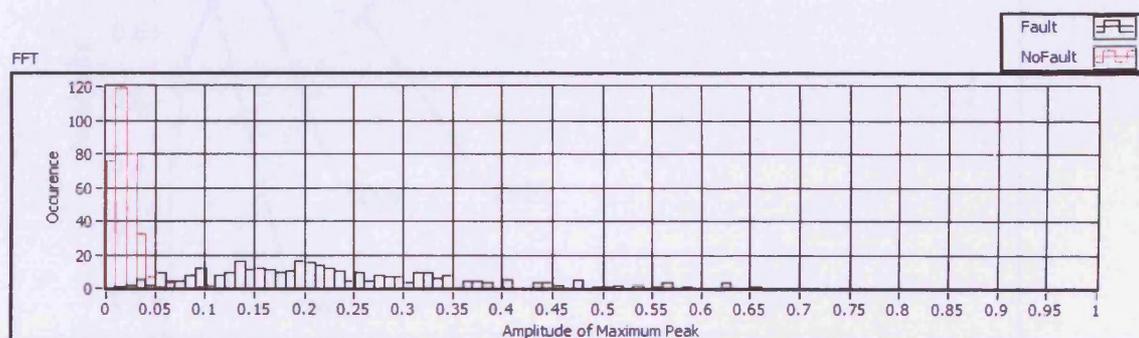


Figure 8-12 Distribution of Normal and Faulty peaks for the FFT-based technique plotted for frame size of 5000 samples.

Note less overlapping between faulty and non-faulty peaks compared to those plotted for a frame size of 2000 samples and a smaller distribution of faulty peaks between 0 and 0.66.

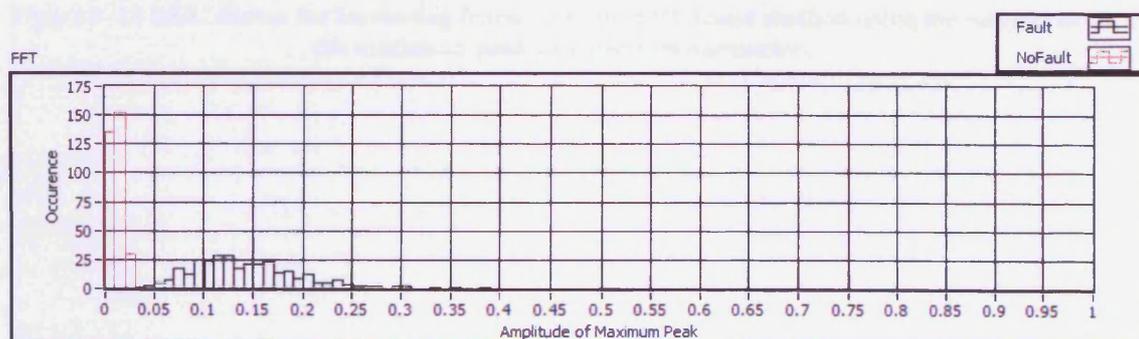


Figure 8-13 Distribution of Normal and Faulty peaks for the FFT-based technique plotted for frame size of 10000 samples.

Note no overlapping between faulty and non-faulty peaks. Faulty peaks and non-faulty peaks nearly separated into two distinct regions compared to those plotted for frame sizes of 2000 and 5000 samples and a smaller distribution of faulty peaks between 0.03 and 0.4.

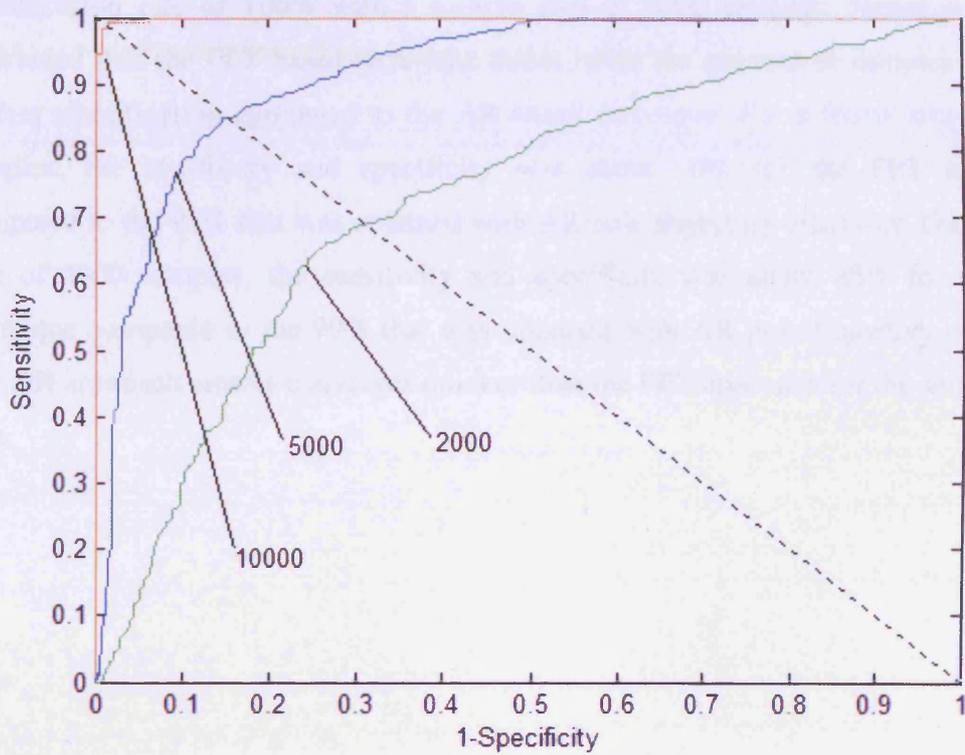


Figure 8-14 ROC curves for increasing frame sizes for FFT-based method using the magnitude of the maximum peak as a decision parameter.

The equivalent ROC curve was also obtained for the FFT technique to illustrate its diagnostic effectiveness and is shown in Figure 8–14. The upper ROC is for 10000 samples and the lower ROC curve is for 2000 samples. Only frames with 10000 samples produced almost perfect classification with a sensitivity and specificity of almost 100% for the FFT-based method. It should be noted the fault detection tool based on movement of AR poles in the z-domain had obtained a near perfect classification rate of 100% with a sample size of 5000 samples. Hence it can be concluded that the FFT-based technique needs twice the amount of samples for near perfect classification compared to the AR-based technique. For a frame size of 2000 samples, the sensitivity and specificity was about 70% for the FFT technique compared to the 92% that was obtained with AR pole trajectory classifier. For a frame size of 5000 samples, the sensitivity and specificity was about 85% for the FFT technique compared to the 99% that was obtained with AR pole trajectory classifier. The AR approach clearly converges quicker than the FFT approach for the same frame size.

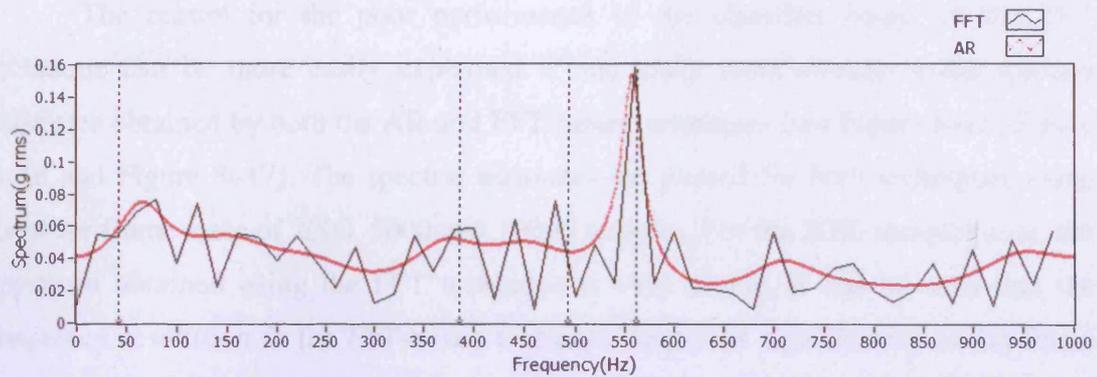


Figure 8-15 Faulty peak detected at 555 Hz for both AR and FFT-based techniques when a frame size of 2000 samples is used.

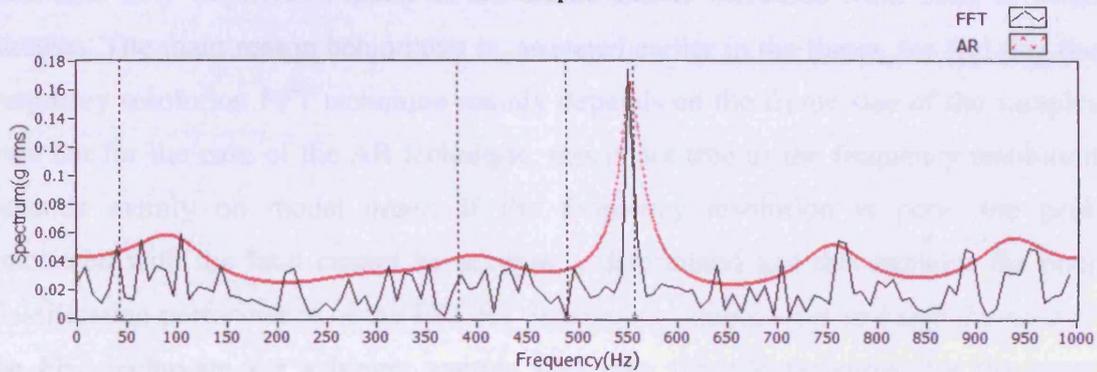


Figure 8-16 Faulty peak detected at 555 Hz for both AR and FFT-based techniques when a frame size of 5000 samples is used.

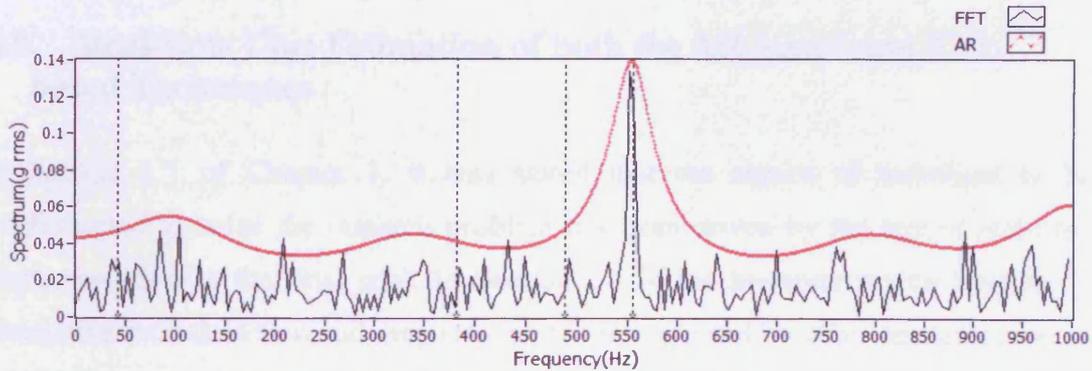


Figure 8-17 Faulty peak detected at 555 Hz for both AR and FFT-based techniques when a frame size of 10000 samples is used.

The reason for the poor performance of the classifier based on the FFT technique can be more easily explained if one looks more closely at the spectral estimates obtained by both the AR and FFT based techniques (see Figure 8–15, Figure 8–16 and Figure 8–17). The spectral estimates are plotted for both techniques using the three frame sizes of 2000, 5000 and 10000 samples. For the 2000 samples case, the spectrum obtained using the FFT technique is very coarse. It can be seen that the frequency resolution in the FFT-based technique improves significantly as the frame size is increased from 2000 to 5000 samples. The frequency resolution of AR-based technique only improves slightly as the frame size is increased from 2000 to 5000 samples. The main reason behind this is, as stated earlier in the thesis, the fact that the frequency resolution FFT technique mainly depends on the frame size of the samples used but for the case of the AR technique, this is not true as the frequency resolution depends mainly on model order. If the frequency resolution is poor, the peak associated with the fault cannot be accurately determined and this explains the poor classification performance of the FFT for the smaller sample sizes and also the need of the FFT technique for a bigger sample size than the AR technique for the same classification performance.

8.8. Real-time Cost Estimation of both the AR-based and FFT-based Techniques

In Section 1.1 of Chapter 1, it was stated that the choice of technique to be implemented to solve the research problem has been driven by the aim of real-time implementation as the final goal. In Section 2.4 of the literature review Chapter 2, alternative fault detection and diagnosis methods suggested by other researchers were reviewed but their main limitations were that they were rendered too complex and sophisticated to be implemented in a real-time environment.

From the studies reported by Schlindwein, (Schlindwein 1988; Schlindwein and Evans 1989; Schlindwein and Evans 1990), it was stated and concluded that the AR technique is slower than the FFT technique. Knowing this but with the available knowledge drawn from Chapter 2 (in Section 2.4.2) that AR modelling can work with smaller sample sizes and achieve far superior resolution capabilities compared to the FFT technique, an assumption was made in this research work throughout the thesis

8.8 Real-time Cost Estimation of both the AR-based and FFT-based Techniques 169

that AR modelling is more suitable for real-time applications. That assumption is validated and verified in this section by benchmarking the AR and FFT applications in terms of their real-time cost.

In this section, the computational cost of real time processing of the AR and FFT methods are analysed. Code was benchmarked using the sequence structure in LabVIEW as shown in Figure 8–18 using a timestamp obtained before the code had begun to run and getting a timestamp after the code had run to completion. The algorithms for the AR and FFT were inserted as the benchmark code and the amount of elapsed time that it takes for a VI to complete was calculated. The difference between the initial and final time given by the timestamps corresponds to the amount of elapsed time that it takes benchmarked code to complete. The VI was run a few times and the average time was taken for one instance of the benchmarked code to run was obtained. The time taken for the AR and FFT methods for the three frame sizes of 2000, 5000 and 10000 samples are shown in column (c) in Table 8–2.

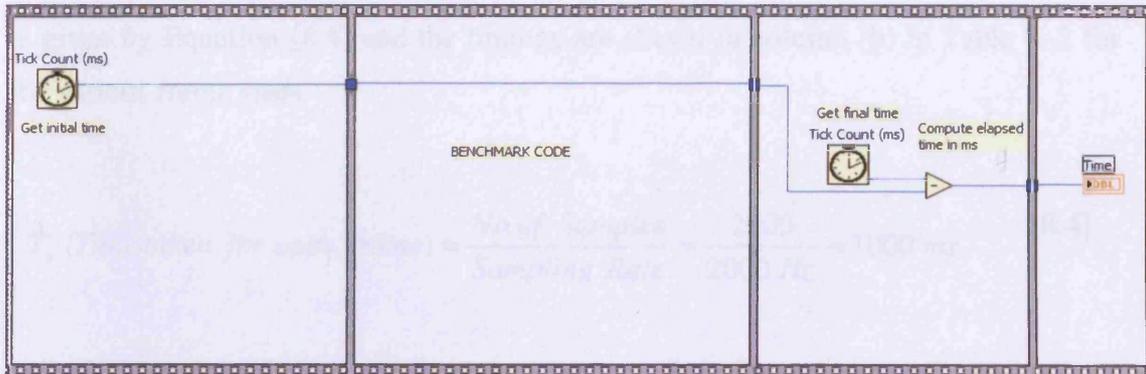


Figure 8–18 VI showing sequence structure used in LabVIEW to get execution time of code.

Table 8–2 Benchmarking the AR and FFT methods for 3 different frame sizes.

Method	Time taken for frame to be transmitted(ms) T_s	Processing Time of Method(ms) T_P	Total Time(ms) T_T
(a)	(b)	(c)	(d)
AR(2000)	1000	40	1040
AR(5000)	2500	86	2586
AR(10000)	5000	210	5210
FFT(2000)	1000	10	1010
FFT(5000)	2500	28	2528
FFT(10000)	5000	61	5061

8.8 Real-time Cost Estimation of both the AR-based and FFT-based Techniques 171

The times taken to transmit the various frames of data T_s at the sampling rate of 2 kHz is given by Equation [8.4] and the timings are shown in column (b) in Table 8–2 for the various frame sizes.

$$T_s \text{ (Time taken for each frame)} = \frac{\text{No of Samples}}{\text{Sampling Rate}} = \frac{2000}{2000 \text{ Hz}} = 1000 \text{ ms} \quad [8.4]$$

To find the basic processing requirement to acquire the vibration signal and process it to the fault classification stage, the time taken for frame to be transmitted T_s and the time taken for the frame to be processed T_p has to be summed, and the total time T_T gives an indication of the computational cost of the method. For the FFT algorithm, since no overlapping of windowing, no windowing and no averaging was used the processing power is proportional to $N \log_2 N$ operations where N is the sample size. For the AR method, the processing power is a function of the model order. Here to find the processing requirements, the model order was kept constant at 10.

The time taken for the AR method is a sum of 3 basic steps,

- 1) Time taken finding the autocorrelation functions
- 2) Time taken for estimation of the AR parameters using the Yule-Walker estimation method using the Levinson-Durbin recursion.
- 3) Time taken to find the roots of the characteristic polynomial of the AR coefficients which are the AR poles.

From the results shown in Table 8–2, for a frame size of 2000 samples, the processing time T_p of the AR method (40 ms) is nearly 4 times as long as the FFT method (10ms) to process data samples of the same size since the algorithms for performing AR modelling are computationally more demanding than the FFT

8.8 Real-time Cost Estimation of both the AR-based and FFT-based Techniques 172

approach. To acquire a frame size of 2000 samples, a sampling time T_s of 1000 ms is required by both methods. It should be noted that the processing time taken by each method T_p is only a fraction of the time required to acquire the necessary data. For the same frame size of 2000 samples, the total processing time T_T taken by the AR method is 1040 ms but the total processing time T_T taken by the FFT method is 1010 ms. Hence essentially, for the same frame size, the AR method takes a longer time to acquire and transmit.

From the analysis done in Section 8.7, it was concluded that AR method required 5000 samples for near perfect classification rate but the FFT method needs twice the same frame size of 10000 samples for the same diagnostic performance. Hence for an almost perfect classification rate, the total processing time T_p taken by the AR method is 2586 ms but the total processing time T_p taken by the FFT method is 5061ms. An important consideration here is that the amount of time that can be used to acquire data for a particular time frame far supersedes the time taken by each of the processing methods. This is fundamentally because as stated in Section 2.4.2.2 and shown with results in Section 8.7, the frequency resolution capabilities of the AR method are far superior to those of the FFT method and as a result of this, one can benefit by working with smaller sample sizes when working with the AR method for the same resolution. Consequently, savings on the time at acquiring the data can be achieved with the added benefit of better classification rates at smaller sample sizes. Based on the findings reported in this section, it can be concluded that the method of fault detection based on AR pole trajectory is more applicable for real-time implementation than the FFT-based method due to its suitability to work with smaller record lengths. Hence the assumption that the AR method is more suitable from real-time point of view is valid.

8.9. Summary

In this chapter, an automated assessment of bearing faults which relies on the AR pole positions in the complex z plane has been introduced. An AR model was used to estimate the power and distance of the poles at the characteristic bearing frequency,

with a higher power and a smaller distance indicative of faulty bearings. It was shown, with experimental results, that the AR poles move closer to the unit circle as the severity of the defects increases. The application of the proposed scheme enabled the detection of the damaged bearings with an inner race defect with a 100 % detection rate when a frame size of more than 4000 samples was used and the distance of the critical poles was chosen as the fault indicator. With a smaller sample size of 2000 samples, a lower detection rate of 92% was achieved. When the power of the migratory poles was used as the fault indicator, a 100 % detection rate was obtained even for small sample sizes such as a sample size of 2000 samples, indicating that this processing methodology is very effective even for short data records. The simplicity of the use of the classification parameters λ , the short length of data needed for the diagnostics (2 seconds) and the obtained specificity and sensitivity of the classification scheme reveal the relevance and usefulness of a model-based diagnostics system for the classification of normal *versus* faulty bearings.

An investigation was also conducted to investigate the stability of the critical poles by computing their condition numbers. The condition numbers for the 'critical' poles were of small values indicating that the system is well conditioned, even for poles close to the unit circle.

Prior classical research work done by Kay and Marple (Kay and Marple 1981) has stated AR modelling can work with a smaller sample sizes and achieve better resolution capabilities compared to the FFT technique. The findings reported in Section 8.7 have been shown to be in agreement with the literature cited in 2.4.2.2. The FFT-based method achieved almost perfect classification with a sensitivity and specificity of almost 100% only with a frame size of 10000 samples but the fault detection tool based on movement of AR poles in the z-domain had obtained the same classification rate from half the frame size of 5000 samples. The findings in Section 8.8 are also consistent with literature cited in Section 2.4.2. In terms of real-time applicability, the AR-based method was indeed found to be more suitable than the FFT-based because of its ability to work with smaller sample sizes.

In summary, this chapter has introduced a new method in the field of fault detection of rotating machinery based on the AR pole based trajectory method. Results

show that tracking the movement of the AR critical poles as a maintenance decision-making criterion can be used to differentiate pumps running in normal and defective working environments and hence can be successfully used for the real-time fault detection of dry vacuum pumps.

9. CONCLUSIONS AND FUTURE WORK

9.1. Conclusions

This thesis has demonstrated the high potential of the Autoregressive method to detect and characterise bearing faults from a “quasi-steady state” rotating machine and presented a case study of the detection of an inner race fault of a HV end bearing of a dry vacuum pump. The main work presented has shown the applicability of the AR modelling technique for vibration-based condition monitoring. Various facets of AR modelling were extensively examined to fine tune the AR model before proceeding to the final stage of fault classification. The following paragraphs summarise the main findings presented in this thesis. Further research is also described that could be oriented to confirm and enlarge these findings and to overcome limitations of the studies reported here.

Chapter 3 had described the methodology used to collect the data presented in this thesis and the performance of an iMEMs ADXL105 surface micromachined accelerometer was compared with that of a piezoelectric accelerometer to explore alternative, cost-effective solutions for acquiring reliable data for vibration-based fault detection and diagnostics. It was validated and showed that the main advantages of the ADXL105 accelerometer are its low cost, its ability to measure DC response and its better temperature stability. If proper mounting techniques were adopted, it was found that the accelerometer can be used successfully for machine diagnostics.

A detailed study was first conducted to analyse the feasibility and suitability of AR modelling as a fault detection tool for condition monitoring with applicability to the quasi-steady stationary vibration signals. The problem of AR model order selection for the vibration signals was investigated and illustrated with supporting results in Chapter 4. As there is no general rule for determining the model order and the model order depends highly on the signal being investigated, it is hoped that the methods introduced in the thesis, especially the proposition of the easy rule for order determination with particular reference to rotating machinery alleviates the difficulties faced in classical model order determination. This original finding is believed to be a

potential contribution to research areas that have been identified as relevant to condition monitoring with particular respect to “steady state” rotating machinery.

The problem of finding the optimum sample size to use for rotating machinery diagnostic was investigated in Chapter 5 and it was concluded that the variance of the prediction error of the AR models can be used as a ballpark indicator to determine the minimum sample size required and this minimum sample size was speed dependent. A general remark that can be made is that the vibration signals from higher speed machines required bigger model orders and bigger sample sizes to encapsulate their characteristic behaviour. It is reasonable to speculate that, since the Yule-Walker estimation method was used and this method is based on the solution of Levinson Durbin recursion on a set of linear equations using the ACF of the samples, using a larger sample size yields a more accurate estimate of the ACF. As the AR process is a linear regression model, a bigger number of samples per complete revolution of the signal was required to capture the dynamic characteristics of the bearing condition. This also implied higher AR model orders.

The effect of the sampling frequency on AR model selection was also investigated in this chapter. With supporting results, it has been verified that the length in time one has to trace back is more important than the sampling frequency in the context of AR model identification. It was concluded that the effect of increasing the sampling frequency was to increase the model order required.

Interestingly, it was found that AR modelling couples the benefits of both time domain analysis and frequency domain analysis. It has superior frequency resolution capabilities compared to the FFT approach for frequency estimation. It has the ability to work with shorter data records and the sampling rate required is just slightly above the Nyquist rate. Also, it does not have negative effects such as spectral leakage and the need for the use of antileakage windows. The use of shorter records and smaller sampling rates has great potential for implementation in real-time applications. Both the FFT and the AR technique work on the assumption that the signal to be modelled is stationary. From this aspect, the ability to use shorter fragments of data is preferred for frequency estimation as stationarity can be more reasonably assumed for shorter segments of data. This is an added advantage of the AR technique as it can achieve

qualitative frequency resolution improvement over the FFT estimation especially for short frames of data.

The pre-treatment of the faulty vibration signals using resonance demodulation was found to be useful as it enhances the SNR of the signal and removes the modulating carrier frequency (fundamental rotating speed of the pump), which can obscure weak impulses produced by the defective bearings. Though this work is not new and has been proven by other researchers, it was shown how resonance demodulation had been carried out in the context of this work in Chapter 6.

In summary, the main original contribution of the thesis was the development of a classification scheme based upon the position of the Autoregressive (AR) poles for tracking of the time varying behaviour of the spectral components of vibration data. The main facets of the novel AR pole trajectory tool were illustrated in Chapters 7 and 8. The task of condition monitoring and fault diagnosis of ball bearings is often cumbersome and labour intensive. Various techniques have been proposed in literature for rolling bearing fault detection and diagnosis. The challenge, however, is to extract, efficiently and accurately, features from signals acquired from these elements, particularly in the frequency domain.

The AR method has offered a unique way of detecting bearing faults by spectral decomposition and tracking of the AR poles in the complex z domain. The poles of the AR model were extracted as the feature relating to the characteristic bearing defect frequencies. As derived in the thesis, the time variant spectral parameters needed for diagnosis of the bearing condition can be extracted from parameters λ of the pole positions in the complex z -plane. The time varying distance, power and frequency components were monitored by tracking the movement of critical poles. This reduces the signal space tremendously, reducing the complexity of the classifier and aiding fault classification.

To test the efficacy of the scheme, the proposed method was measured with ROC analysis and applied to increasing frame sizes of vibration data. This ROC analysis showed that the proposed AR pole trajectory fault detection tool achieved high sensitivity and specificity rates, highlighting the relevance of the method for

monitoring purposes. The results of Chapter 8 showed that a sample size of 4000 samples per frame was sufficient for almost perfect detection and classification when the AR poles' distance from the centre of the unit circle was used as the fault indicator. The power of the migratory poles was an alternative perfect classifier which can be used as a fault indicator.

It was also shown that the equivalent FFT approach had required 10000 samples per frame for near perfect classification when the magnitude of the biggest peak was used as the fault classifier. This finding is in support with theory cited in Section 2.4.2.2 that the AR technique has a far superior frequency resolution capability compared to the FFT method for the same frame size used and this explains its better performance as a fault classifier when small frame sizes (record lengths) are used. The real-time computational cost of applying both techniques was also benchmarked in this chapter and it was shown that the AR method was more suitable for real-time implementation than the FFT-based method.

The work undertaken within this research makes a significant contribution in the field of condition monitoring of rotating machinery by introducing a new fault detection tool based on AR pole trajectory. All in all, as a model based diagnostic approach, it was found that parameterisation of the data signals as either AR coefficients or trajectory of the pole locations has great potential in terms of its applicability to fault prediction. The AR pole representation has allowed a more straightforward quantitative estimation of the spectral parameters and facilitated the understanding of the intrinsic spectral characteristics of the process and is particularly useful for fault detection.

9.2. Future Work

There are several directions in which this research could evolve. One direction in which further work could progress is towards the real-time implementation of the AR fault trajectory tool for industrial use. The experimental vibration signals used in the study were obtained from dry vacuum pumps under laboratory conditions. The performance of the AR fault trajectory tool needs to be tested in real industrial

environments. For purpose of the study conducted, the model order was fixed for the AR fault trajectory tool. For a real time-implemented system, there is the possibility of moving towards on-line order estimation, for modelling the healthy and faulty vibration signals for real-time feature extraction and classification, and even online training.

For most of the analysis done the pump was run at constant speed and the load was also almost constant. An improvement would be to construct a model which takes into account changes in load and speed. The effects of a variable machine speed on machine vibration and the implications for bearing fault detection have to be investigated further. These effects are important to understand because, when ignored they can significantly hinder the ability to detect bearing faults. The speed of the machine can potentially be one of the most significant factors affecting the machine vibration. If the machine is driven by a power electronic inverter, the speed is controlled directly by the drive. If the machine is controlled by an induction motor, speed is affected by load level. In either case, machine speed can change continuously and this relationship between speed and machine vibration has to be monitored. Variable machine speed can directly and nonlinearly alter the level of machine vibration. This is due to differences in mechanical damping and resonances at various machine speeds. While this effect is difficult to notice in healthy machines, it can become significant as the bearing health degrades. An additional effect that speed can exert is on the rate of development of bearing fault (Besancenez, Dron *et al.* 2001). Variations in speed can actually retard or temporarily mask the increase in machine vibration due to a bearing fault. This phenomenon has been observed in experimental trials as the bearing fault enters an advanced or more deteriorated stage. This can inadvertently make a machine appear healthy even though a bearing failure is imminent. This area of research clearly needs further study.

More data, including data from pumps with different faults (and including distributed faults) also needs to be collected in order to validate and expand the results already achieved. Due to time constraints and the intention to finish the PhD studies within the period of three years, the main analysis was restricted to using only one set of localised faulty signals - which was the bearing with the inner race crack. This type of fault belongs to the class of localised defects. Another set of faulty data set became

available towards the end of the project and this was the faulty bearing with an oversized ball element. But this set of faulty data set was not utilized in our studies as this type of fault belongs to the more complicated class of distributed defects. An oversized bore in the bearing affects the radial and axial clearances, introducing nonlinearity in the dynamic system. Distributed defects are more difficult to detect as their presence cannot be detected from the presence of peaks at characteristic bearing defect frequencies alone. An analytic model has to be constructed at the microscopic level to include the effects of the race surface waviness and non-linear Hertzian contact information (Harsha, Sandeep *et al.* 2003; Sopanen and Mikkola 2003a). Previous researchers have presented conflicting results with regards to the detection of distributed defects and this is one area which needs further exploration (Tandon and Choudhury 2000; Sopanen and Mikkola 2003b). More studies are necessary to clearly elucidate different faulty conditions due to distributed bearings in the dry vacuum pump as well. Some part of this research is being carried out by a fellow researcher in the team who is looking at other bearing faults apart from the inner race fault. Also, this researcher is looking at the transient behaviour of signals such as sound and acoustic emissions obtained from the dry vacuum pump.

There are some bearing data centres available online like the seeded fault test data centre of the Case Western Reserve University. Tests were not conducted on these data as the main idea was to study the development fault conditions in the dry vacuum pump. Also analysis based on artificially simulated faulty bearing signals was not adopted in our case as it is our opinion that the simulated signals are too simplistic and do not represent a 'true picture' of the condition of the bearings in the dry vacuum pump. The task of building a true mathematical model of the bearing system is a difficult one as various factors such as the resonance in the bearing components and structure, damping ratios in the propagating path from the source to the transducers and environmental conditions are difficult to emulate and estimate. For our case, it was unfortunate that the introduction of a seeded fault in bearings of the dry vacuum pump was a complicated, time-consuming and costly process, as skilled personnel are required to fit and assemble the pump with specialised tools. So we were restricted to a limited data set of faulty signals. However, it is important that records from other faulty bearings with outer race crack, cracks on the main rolling element and the bearing cage are also obtained to study the effectiveness of the tools proposed in this

study for other bearing fault conditions. In addition, the case of multiple fault conditions in a single bearing also needs to be considered for a more thorough investigation.

As final remarks, the research work done here is seen to be highly beneficial for the semiconductor industry, as dry vacuum pumps are critical components of the manufacturing process, and development of reliable and predictable automated real-time online fault condition monitoring techniques for monitoring the dry pumps can reap huge benefits in terms of financial and manpower savings. Initial results show that the AR fault trajectory tool provides a novel approach to the detection and diagnosis of bearing faults occurring in dry vacuum pumps and it is our sincere hope that the techniques proposed here will be employed by those involved in the industry to engineer effective solutions for real-time diagnosis of pump applications.

REFERENCES

- Abbaszadeh, K., M. Rahimian, H. A. Toliyat and L. E. Olson (2002). "Rail defect diagnosis using wavelet packet decomposition", *Proceedings of the IEEE Industry Applications Conference*, Pittsburgh, PA, USA, pp. 478-484.
- ADXL105 (1999). "Datasheet", Analog Devices Inc.
- Akaike, H. (1969). "Fitting autoregressive models for prediction", *Annals of the Institute of Statistical Mathematics*, 21, pp. 243-247.
- Akaike, H. (1974). "A new look at the statistical model identification", *IEEE Trans. Autom. Control*, 19, pp. 716-23.
- Akutsu, I., T. Matsuoka, M. Ozaki, T. Kyuko, S. Miyashita, T. Ozawa, M. Naka, H. Ohnishi, Y. Narahara and G. Horikoshi (2000). "A gradational lead screw dry vacuum pump", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 18(3), pp. 1045-1047.
- Alfredson, R. J. and J. Mathew (1985a). "Frequency domain methods for monitoring the condition of rolling element bearings." *Mechanical Engineering Transactions - Institution of Engineers, Australia*, 10(2), pp. 108-112.
- Alfredson, R. J. and J. Mathew (1985b). "Time domain methods for monitoring the condition of rolling element bearings." *Mechanical Engineering Transactions - Institution of Engineers, Australia*, 10(2), pp. 102-107.
- Andrade, F. A., I. Esat and M. N. M. Badi (2001). "A new approach to time-domain vibration condition monitoring: gear tooth fatigue crack detection and identification by the Kolmogorov-Smirnov test", *Journal of Sound and Vibration*, 240(5), pp. 909-919.
- Antoni, J. and R. B. Randall (2005). "On the use of the cyclic power spectrum in rolling element bearings diagnostics", *Journal of Sound and Vibration*, 281(1-2), pp. 463-468.
- Bachmann, P. and M. Kuhn (1990). "Evaluation of dry pumps vs rotary vane pumps in aluminum etching", *Vacuum*, 41(7-9), pp. 1825-1827.
- Baillie, D. C. and J. Mathew (1996). "Comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings", *Mechanical Systems & Signal Processing*, 10(1), pp. 1-17.
- Barret, R. (1993). "Low Frequency Machinery Monitoring: Measurement Considerations", *Wilcoxon Research, Application Note*.
- Batina, I., J. Jensen and R. Heusdens (2006). "Noise power spectrum estimation for speech enhancement using an autoregressive model for speech power spectrum dynamics", *Proceedings of the ICASSP 2006 - IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, pp. 1064-1067.
- Berges, H. P. (1987). "Increased reliability when pumping reactive gases, due to monitoring of vacuum pump and accessories", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 5(4), pp. 2616-2619.

- Besancenez, M., J. P. Dron, L. Rasolofondraibe and A. Pavan (2001). "Implementation of parametric spectrum analysis to the monitoring of a forming press", *European Journal of Mechanical and Environmental Engineering*, 46(1), pp. 37-44.
- Bez, E. and D. G. Guarnaccia (1990). "Operational experience with totally oil-free rough vacuum pumps", *Journal of Vacuum*, 41(7), pp. 1819-1821.
- Boardman, A., F. S. Schlindwein, A. P. Rocha and A. Leite (2002). "A study on the optimum order of autoregressive models for heart rate variability", *Physiological Measurement*, 23(2), pp. 325-36.
- Box, G. E. P., G. M. Jenkins and G. C. Reinsel (1994). "Time Series Analysis: Forecasting and Control", Prentice Hall.
- Broersen, P. M. T. (1985). "Selecting the Order of Autoregressive Models from Small Samples", *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*, 33, pp. 874-879.
- Broersen, P. M. T. (1990). "The prediction error of autoregressive small sample models", *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*, 38(5), pp. 858-860.
- Broersen, P. M. T. (1998). "Estimation of the accuracy of mean and variance of correlated data", *IEEE Transactions on Instrumentation and Measurement*, 47(5), pp. 1085-1091.
- Broersen, P. M. T. (2000a). "Facts and Fiction in Spectral Analysis", *IEEE Transactions on Instrumentation and Measurement*, 49(4), pp. 766-772.
- Broersen, P. M. T. (2000b). "Finite Sample Criteria for Autoregressive Order Selection", *IEEE Transactions on Signal Processing*, 48(12), pp. 3550-3558.
- Broersen, P. M. T. (2002). "Automatic spectral analysis with time series models", *IEEE Transactions on Instrumentation and Measurement*, 51(2), pp. 211-216.
- Broersen, P. M. T. (2006). "Automatic Autocorrelation and Spectral Analysis", Springer.
- Broersen, P. M. T., S. De Waele and R. Bos (2003). "Estimation of autoregressive spectra with randomly missing data", *Proceedings of the 20th IEEE Information and Measurement Technology Conference, May 20-22 2003, Vail, CO, United States*, pp. 1154-1159, Institute of Electrical and Electronics Engineers Inc.
- Broersen, P. M. T. and H. E. Wensink (1993). "On finite sample theory for autoregressive model order selection", *IEEE Transactions on Signal Processing*, 41(1), pp. 196-204.
- Chen, Y. D., R. Du and L. S. Qu (1995). "Fault Features of Large Rotating Machinery and Diagnosis using Sensor Fusion", *Journal of Sound and Vibration*, 188(2), pp. 227-242.
- Choudhury, A. and N. Tandon (2000). "Application of acoustic emission technique for the detection of defects in rolling element bearings", *Tribology International*, 33(1), pp. 39-45.

- Davis, R. P., R. A. Abreu and A. D. Chew (2000). "Dry vacuum pumps: A method for the evaluation of the degree of dry", *Proceedings of the 46th international symposium of the american vacuum society*, pp. 1782-1788.
- de Waele, S. and P. M. T. Broersen (2000). "Burg algorithm for segments", *IEEE Transactions on Signal Processing*, 48(10), pp. 2876-2880.
- Doscher, J. (1997). "Monitoring Machine Vibration with Micromachined Accelerometers", *Sensors*, 14(5), pp. 33-38.
- Doscher, J. (1999a). "ADXL105: A Lower-Noise, Wider-Bandwidth Accelerometer Rivals Performance of More Expensive Sensors", *Analogue Dialogue*, 33(6), pp. 27-29.
- Doscher, J. (1999b). "Using iMEMS Accelerometers in Instrumentation Applications", *Proceedings of the 45th International Instrumentation Symposium*.
- Dron, J. P., F. Bolaers and L. Rasolofondraibe (2004). "Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: Application to the detection of defects in ball bearings", *Journal of Sound and Vibration*, 270(1-2), pp. 61-73.
- Dron, J. P., S. F. Bolaers and I. Rasolofondraibe (2003). "Influence of the de-noising method by spectral subtraction on the diagnosis of rolling bearings", *European Journal of Mechanical and Environmental Engineering*, 48(1), pp. 3-12.
- Dron, J. P., L. Rasolofondraibe, F. Bolaers and A. Pavan (2001). "High-resolution methods in vibratory analysis: Application to ball bearing monitoring and production machine", *International Journal of Solids and Structures*, 38(24-25), pp. 4293-4313.
- Dron, J. P., L. Rasolofondraibe, C. Couet and A. Pavan (1998). "Fault detection and monitoring of a ball bearing benchtest and a production machine via autoregressive spectrum analysis", *Journal of Sound and Vibration*, 218(3), pp. 501-525.
- Du, Q. and S. Yang (2007). "Application of the EMD method in the vibration analysis of ball bearings", *Mechanical Systems and Signal Processing*, 21(6), pp. 2634-2644.
- Durbin, J. (1959). "The Fitting of Time-Series Models", *Rev. Inst. Internat. Statist.*, 28, pp. 229-249.
- Duval, P. (1987). "Will tomorrow's high-vacuum pumps be universal or highly specialized?" *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 5(4), pp. 2546-2551.
- Duval, P. (1989). "Selection criteria for oil-free vacuum pumps", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 7(3), pp. 2369-2372.
- Ebrahimian, R. and R. Baldick (2001). "State estimator condition number analysis", *IEEE Transactions on Power Systems*, 16(2), pp. 273-279.
- Edwards, B. (1993). "The Technology of Dry Vacuum Pumps", *Edwards High Vacuum International*, Publication No. I2-A526-20-895-R.
- Esch, J. (2002). "Prolog to sensorless control of induction motors", *Proceedings of the IEEE*, 90(8), pp. 1358.

- Fang, W., P. Willett and S. Deb (2000). "Condition monitoring for helicopter data", *IEEE International Conference on Systems, Man and Cybernetics.*, 1, pp. 224-9.
- Fougere, P. F., E. J. Zawalick and H. R. Radoski (1976). "Spontaneous line splitting in maximum entropy power spectrum analysis", *Physics of The Earth and Planetary Interiors*, 12(2-3), pp. 201-207.
- Gaubitch, N. D., D. B. Ward and P. A. Naylor (2006). "Statistical analysis of the autoregressive modeling of reverberant speech", *Journal of the Acoustical Society of America*, 120(6), pp. 4031-4039.
- Hablanian, M. H. (1981). "Performance of mechanical vacuum pumps in the molecular flow range", *Journal of Vacuum Science & Technology A*, 19(2), pp. 250-252.
- Hablanian, M. H. (1986). "Performance characteristics of displacement type vacuum pumps", *Journal of Vacuum Science & Technology A*, 4(3), pp. 286-292.
- Hablanian, M. H. (1988). "The emerging technologies of oil-free vacuum pumps", *Journal of Vacuum Science & Technology A*, 6(3), pp. 1177-1182.
- Harsha, S. P. (2006a). "Nonlinear dynamic analysis of a high-speed rotor supported by rolling element bearings", *Journal of Sound and Vibration*, 290(1-2), pp. 65-100.
- Harsha, S. P. (2006b). "Nonlinear dynamic response of a balanced rotor supported by rolling element bearings due to radial internal clearance effect", *Mechanism and Machine Theory*, 41(6), pp. 688-706.
- Harsha, S. P., K. Sandeep and R. Prakash (2003). "The effect of speed of balanced rotor on nonlinear vibrations associated with ball bearings", *International Journal of Mechanical Sciences*, 45(4), pp. 725-740.
- Harting, D. R. (1978). "Demodulated resonance analysis - A powerful incipient failure detection technique", *ISA Transactions.*, vol. 17(no. 1), pp. 35-40.
- He, Z., J. Zhao, Y. He and Q. Meng (1996). "Wavelet transform and multiresolution signal decomposition for machinery monitoring and diagnosis", *Proceedings of the IEEE International Conference on Industrial Technology Shanghai, China* pp. 724-727.
- Heng, R. B. W. and M. J. M. Nor (1998). "Statistical Analysis of Sound and Vibration Signals for Monitoring Rolling Element Bearing Condition", *Applied Acoustics*, 53(1-3), pp. 211-226.
- Ho, D. and R. B. Randall (2000). "Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signals", *Mechanical Systems and Signal Processing*, 14(5), pp. 763-788.
- Hoffman, A. J. and N. T. van der Merwe (2002). "The application of neural networks to vibrational diagnostics for multiple fault conditions", *Computer Standards & Interfaces*, 24(2), pp. 139-149.
- Howard, I. (1994). "A Review of Rolling Element Bearing Vibration - Detection, Diagnosis Prognosis", DST0-RR-0013, Research Report, Aeronautical and Maritime Research Laboratory AirFrames and Engines Division.

- Huang, H.-H. and H. P. B. Wang (1996). "Integrated monitoring and diagnostic system for roller bearings", *International Journal of Advanced Manufacturing Technology*, 12(1), pp. 37-46.
- Ifeachor, E. C. and B. W. Jervis (2003). "Digital Signal Processing: A Practical Approach", Pearson Education.
- Isermann, R. (1995). "Model base fault detection and diagnosis methods", *Proceedings of the 1995 American Control Conference* Seattle, WA, USA, pp. 1605-9, American Autom Control Council.
- Isermann, R. and P. Balle (1997). "Trends in the application of model based fault detection and diagnosis of technical processes", *Proceedings of the 13th World Congress, International Federation of Automatic Control. Vol.N. Fault Detection, Pulp and Paper, Biotechnology*, San Francisco, CA, USA, pp. 1-12, Pergamon, Oxford, UK.
- Ishizuka, K., H. Kato and T. Nakatani (2005). "Speech signal analysis with exponential autoregressive model", *Proceedings of the*, Philadelphia, PA, United States, pp. 225-228, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States.
- ISO (1995). "Mechanical vibration - Evaluation of machine vibration by measurements on non-rotating parts", International Standard of Organisation, ISO 10816.
- ISO (1996). "Mechanical vibration of non-reciprocating machines", International Standard of Organisation, ISO 7919.
- Jenkins, G. M. and D. G. Watts (1968). "Spectral analysis and its applications", San Francisco [Calif.]; London : Holden-Day.
- Johnsen, S. J. and N. Andersen (1978). "On power estimation in maximum entropy spectral analysis", *Geophysics*, 43(4), pp. 681-90.
- Junsheng, C., Y. Dejie and Y. Yu (2006). "A fault diagnosis approach for roller bearings based on EMD method and AR model", *Mechanical Systems and Signal Processing*, 20(2), pp. 350-362.
- Kay, S. M. (1988). "Modern Spectral Estimation-Theory and Application", Prentice Hall.
- Kay, S. M. and S. L. Marple (1981). "Spectrum analysis-a modern perspective", *Proceedings of the IEEE*, 69(11), pp. 1380-1419.
- Keeton, P. I. J. and F. S. Schlindwein (1998). "Spectral broadening of clinical Doppler signals using FFT and autoregressive modelling", *European Journal of Ultrasound*, 7(3), pp. 209-218.
- Kelly, S., D. Burke, P. de Chazal and R. Reilly (2002). "Parametric models and spectral analysis for classification in brain-computer interfaces", *Proceedings of the DSP 2002 - 14th International Conference on Digital Signal Processing*, vol.1, pp. 307-310.
- Konishi, S. and K. Yamasawa (1999). "Diagnostic system to determine the in-service life of dry vacuum pumps [used in LPCVD semiconductor fabrication facility]", *IEE Proceedings-Science, Measurement and Technology*, 146(6), pp. 270-6.

- Lessard, P. A. (2000). "Dry vacuum pumps for semiconductor processes: Guidelines for primary pump selection", *Proceedings of the 46th international symposium of the american vacuum society*, pp. 1777-1781.
- Lim, J. Y., W. Cheung and K. H. Chung (2004). "Non-destructive characteristics evaluation for low vacuum dry pumps in the semi-conductor manufacturing process line", *Proceedings of the*, Jeju Island, South Korea, pp. 2345-2350, Trans Tech Publications Ltd, Zurich-Ueticon, CH-8707, Switzerland.
- Linkens, D. A. (1979). "Empirical Rules for the Selection of Parameters for Auto-Regressive Spectral Analysis of Biomedical Rhythms", *Journal of Signal Processing 1*, pp. 243-258.
- Mainardi, L. T., A. M. Bianchi, G. Baselli and S. Cerutti (1995). "Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameters", *IEEE Transactions on Biomedical Engineering*, 42(3), pp. 250-259.
- Maiwald, D., J. W. Dalle Molle and J. F. Bohme (1993). "Model identification and validation of nonstationary seismic signals", *Proceedings of the*, pp. 319-322.
- Makhoul, J. (1975). "Linear Prediction: A Tutorial Review", 63(4), pp. 561-580.
- Marcek, D. (2000). "Forecasting of economic quantities using fuzzy autoregressive models and fuzzy neural networks", *Neural Network World*, 10(1-2), pp. 147-155.
- Martin, H. R. and F. Honarvar (1995). "Application of Statistical Moments to Bearing Failure Detection", *Applied Acoustics*, 44, pp. 67-77.
- Mathew, J. (1987). "Machine Condition Monitoring using Vibration Analyses." *Acoustics Australia*, 15(1), pp. 7-13.
- Mathew, J. (1989). "Monitoring the vibrations of rotating machine elements. An overview", *Proceedings of the Twelfth Biennial Conference on Mechanical Vibration and Noise, Presented at the 1989 ASME Design Technical Conference, Sep 17-21 1989*, Montreal, Que, Can, pp. 15-22, Publ by ASME, New York, NY, USA.
- Mathew, J. and R. J. Alfredson (1983). "Condition monitoring of rolling element bearings using vibration analysis." *Proceedings of the Winter Annual Meeting - American Society of Mechanical Engineers.*, Boston, Mass, USA, pp. 7, ASME, New York, NY, USA.
- Mathew, J. and R. J. Alfredson (1984). "Condition monitoring of rolling element bearings using vibration analysis." 106(3), pp. 447-453.
- Mathew, J. and R. J. Alfredson (1986). "Condition monitoring of journal bearings using vibration and temperature analysis." *Proceedings of the Papers Presented at the International Conference on Condition Monitoring.*, Brighton, Engl, pp. 21-33, BHRA, Cranfield, Engl.
- Mathew, J. and R. J. Alfredson (1987). "Monitoring of gearbox vibration operating under steady state conditions." *Proceedings of the Mechanical Signature Analysis: Machinery Vibration, Flow-Induced Vibration, and Acoustic Noise Analysis. Presented at the 1987 ASME Design Technology Conferences - 11th*

- Biennial Conference on Mechanical Vibration and Noise.*, Boston, MA, USA, pp. 47-54, ASME, New York, NY, USA.
- Mathew, J., B. T. Kuhnell and M. Y. H. Yii (1987a). "Low cost automated vibration monitoring and analysis." *Proceedings of the Mechanical Signature Analysis: Machinery Vibration, Flow-Induced Vibration, and Acoustic Noise Analysis. Presented at the 1987 ASME Design Technology Conference - 11th Biennial Conference on Mechanical Vibration and Noise.*, Boston, MA, USA, pp. 141-146, ASME, New York, NY, USA.
- Mathew, J., B. T. Kuhnell and M. Y. H. Yii (1987b). "Predictive maintenance using a low cost automated vibration monitoring system." *Proceedings of the Maintenance Engineering Conference 1987: Effective Maintenance, the Road to Profit.*, Melbourne, Aust, pp. 27-31, Inst of Engineers, Australia, Barton, Aust.
- Maurice, L., P. Duval and G. Gorinas (1979). "Oil backstreaming in turbomolecular and oil diffusion pumps", *Journal of Vacuum Science and Technology*, 16(2), pp. 741-745.
- 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.
- McCormick, A. C. and A. K. Nandi (1998). "Cyclostationarity in rotating machine vibrations", *Mechanical Systems & Signal Processing*, 12(2), pp. 225-242.
- McCormick, A. C., A. K. Nandi and L. B. Jack (1998). "Application of periodic time-varying autoregressive models to the detection of bearing faults", *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 212(6), pp. 417-428.
- McFadden, P. D. and J. D. Smith (1984a). "Model for the vibration produced by a single point defect in a rolling element bearing", *Journal of Sound and Vibration*, 96(1), pp. 69-82.
- McFadden, P. D. and J. D. Smith (1984b). "Vibration monitoring of rolling element bearings by the high-frequency resonance technique -- a review", *Tribology International*, 17(1), pp. 3-10.
- McFadden, P. D. and J. D. Smith (1985). "The vibration produced by multiple point defects in a rolling element bearing", *Journal of Sound and Vibration*, 98(2), pp. 263-73.
- McFadden, P. D. and M. M. Toozhy (2000). "Application of Synchronous Averaging to Vibration Monitoring of Rolling Element Bearings", *Mechanical Systems and Signal Processing*, 14(6), pp. 891-906.
- McInerny, S. A. and Y. Dai (2003). "Basic vibration signal processing for bearing fault detection", *IEEE Transactions on Education*, 46(1), pp. 149-56.
- Mechefske, C. K. (1993a). "Machine condition monitoring: part 1 - optimum vibration signal lengths", *British Journal of Non-Destructive Testing*, 35(9), pp. 503-507.

- Mechefske, C. K. (1993b). "Machine condition monitoring: part 2 - the effects of noise in the vibration signal", *British Journal of Non-Destructive Testing*, 35(10), pp. 574-579.
- Mechefske, C. K. and J. Mathew (1993). "Parametric spectral estimation to detect and diagnose faults in low speed rolling element bearings: Preliminary investigations", *Mechanical Systems and Signal Processing*, 7(1), pp. 1-12.
- Myerson, E. B. (2000). "Effectively troubleshoot dry vacuum pumps", *Chemical Engineering Process*, 96(9), pp. 49-52.
- Nandi, S. and H. A. Toliyat (1999). "Condition monitoring and fault diagnosis of electrical machines-a review", *Proceedings of the 34th Annual Meeting of the IEEE Industry Applications*, Phoenix, AZ, USA, pp. 197-204.
- Ocak, H. and K. A. Loparo (2005). "HMM-based fault detection and diagnosis scheme for rolling element bearings", *Transactions of the ASME. Journal of Vibration and Acoustics*, 127(4), pp. 299-306.
- Oppenheim, A. V. and R. W. Schaffer (1975). "Digital Signal Processing". New Jersey, Prentice-Hall.
- Pardey, J., S. Roberts and L. Tarassenko (1996). "A review of parametric modelling techniques for EEG analysis", *Medical Engineering Physics*, 18(1), pp. 2-11.
- Parzen, E. (1975). "Multiple time series: determining the order of approximating autoregressive schemes", *Buffalo, NY: Statistical Sciences Division, State University of New York*, No 23.
- Peck, J. P. and J. Burrows (1994). "On-line condition monitoring of rotating equipment using neural networks", *ISA Transactions*, 33(2), pp. 159-164.
- Poulimenos, A., M. Spridoonakos and S. Fassois (2006). "Parametric Time-Domain Methods for Non-Stationary Random Vibration Identification and Analysis: An Overview and Comparison", *Proceedings of the ISMA2006-International Conference on Noise and Vibration Engineering*, Leuven, Belgium, pp. 2885-2905.
- Prashad, H., M. Ghosh and S. Biswas (1985). "Diagnostic Monitoring of Rolling-Element Bearings by High-Frequency Resonance Technique", *ASLE Transactions*, 28(4), pp. 439-448.
- Prasolov, A. V. (2004). "On identification of autoregressive model with a lag", *Proceedings of the Modelling and Simulation*, Marina Del Rey, CA, United States, pp. 7-12, Acta Press, Anaheim, CA, United States.
- Proakis, J. G. and D. G. Manolakis (1995). "Digital Signal Processing: Principles, Algorithms, and Applications", Prentice Hall.
- Randall, R. B. (2001). "Editorial for Special Edition on Gear and Bearing Diagnostics", *Mechanical Systems and Signal Processing*, 15(5), pp. 827-829.
- Randall, R. B. (2004). "Detection and diagnosis of incipient bearing failure in helicopter gearboxes", *Engineering Failure Analysis*, 11(2), pp. 177-190.
- Randall, R. B., J. Antoni and S. Chobsaard (2000). "A comparison of cyclostationary and envelope analysis in the diagnostics of rolling element bearings", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3882-3885

- Randall, R. B., J. Antoni and S. Chobsaard (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*, 15(5), pp. 945-962.
- Rissanen, J. (1984). "Universal coding, information prediction and estimation", *IEEE Trans. Inf. Theory*, 30, pp. 629-36.
- S. Lawrence Marple Jr. (1987). "Digital Spectral Analysis with Applications", Prentice-Hall, Inc.
- Sabin, E. (1995). "Vibration analysis of dry pumps", *Semiconductor International*, 18(8), pp. 249-250.
- Sawalhi, N., R. B. Randall and H. Endo "The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis", *Mechanical Systems and Signal Processing*, In Press, Corrected Proof.
- Schindwein, F. S. and D. H. Evans (1989). "A Real-Time Autoregressive Spectrum Analyser for Doppler Ultrasound Signals", *Ultrasound in Med & Biol*, 15(3), pp. 263-272.
- Schindwein, F. S. and D. H. Evans (1990). "Selection of the order of autoregressive models for spectral analysis of doppler ultrasound signals", *Ultrasound in Medicine & Biology*, 16(1), pp. 81-91.
- Schindwein, F. S., Evans, D.H (1988). "Real Time Spectral Analysis of Doppler Signals Using a Digital Signal Processor", *Physics in Medical Ultrasound II*, Report No 57.
- Shannon, C. E. (1998). "Communication in the presence of noise", *Proceedings of the IEEE*, 86(2), pp. 447-457.
- Sheen, Y.-T. (2007). "An analysis method for the vibration signal with amplitude modulation in a bearing system", *Journal of Sound and Vibration*, 303(3-5), pp. 538-552.
- Shiroishi, J., Y. Li, S. Liang, T. Kurfess and S. Danyluk (1997). "Bearing condition diagnostics via vibration and acoustic emission measurements", *Mechanical Systems and Signal Processing*, 11(5), pp. 693-705.
- Smail, M., M. Thomas and A. Lakis (1999). "ARMA Models for Modal Analysis: Effect of Model Orders and Sampling Frequency", *Journal of Mechanical Systems and Signal Processing* 13(6), pp. 925-941.
- Smith, J. D. (1982). "Vibration monitoring of bearings at low speeds", *Tribology International*, 15(3), pp. 139-144.
- Sopanen, J. and A. Mikkola (2003a). "Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: Theory", *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 217(3), pp. 201-211.
- Sopanen, J. and A. Mikkola (2003b). "Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 2: Implementation and results", *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 217(3), pp. 213-223.

- Stammers, C. W. (1989). "Vibration and wear detection in rotating machinery by acoustic analysis", *Applied Acoustics*, 28(3), pp. 213-219.
- Tandon, N. and A. Choudhury (1997). "An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect", *Journal of Sound and Vibration*, 205(3), pp. 275-292.
- Tandon, N. and A. Choudhury (1999). "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings", *Tribology International*, 32(8), pp. 469-480.
- Tandon, N. and A. Choudhury (2000). "Theoretical model to predict the vibration response of rolling bearings in a rotor bearing system to distributed defects under radial load", *Journal of Tribology, Transactions of the ASME*, 122(3), pp. 609-615.
- Tandon, N. and B. C. Nakra (1992). "Vibration and acoustic monitoring technique for the detection of defects in rolling element bearings- a review", *The Shock and Vibration Digest* 24(3), pp. 3-11.
- Tandon, N., K. M. Ramakrishna and G. S. Yadava (2007). "Condition monitoring of electric motor ball bearings for the detection of grease contaminants", *Tribology International*, 40(1), pp. 29-36.
- Tandon, N., G. S. Yadava and K. M. Ramakrishna (2007). "A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings", *Mechanical Systems and Signal Processing*, 21(1), pp. 244-256.
- Tang, K. Z., K. K. Tan, C. W. de Silva, T. H. Lee, K. C. Tan and S. Y. Soh (2001). "Application of vibration sensing in monitoring and control of machine health", *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Como, Italy pp. 377-382.
- Thanagasundram, S., Y. Feng, I. S. M. A. Rayan and F. S. Schlindwein (2005). "A case study of Auto-Regressive Modelling and Order Selection for a Dry Vacuum Pump", *Proceedings of the ICSV12-Twelfth International Congress on Sound and Vibration*, Lisbon, Portugal.
- Thanagasundram, S., K. R. Gurung, Y. Feng and F. S. Schlindwein (2006). "AR Pole Trajectory in Condition Monitoring Studies", *Proceedings of the ICSV13-13th International Congress on Sound and Vibration*, Vienna, Austria.
- Thanagasundram, S. and F. S. Schlindwein (2006a). "Autoregressive based diagnostics scheme for detection of bearing faults", *Proceedings of the ISMA2006-International Conference on Noise and Vibration Engineering*, Leuven, Belgium, pp. 3531-3546.
- Thanagasundram, S. and F. S. Schlindwein (2006b). "Autoregressive Order Selection for Rotating Machinery", *The International Journal of Acoustics and Vibration*, 11(3), pp. 144-154.
- Thanagasundram, S. and F. S. Schlindwein (2006c). "Comparison of integrated micro-electrical-mechanical system and piezoelectric accelerometers for machine condition monitoring", *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 220(8), pp. 1135-1146.

- Thanagasundram, S. and F. S. Schlindwein (2006d). "A Fault Detection Tool Using Analysis From an Autoregressive Model Pole Trajectory", *Journal of Sound and Vibration*, Accepted for publication with minor revisions.
- Thanagasundram, S. and F. S. Schlindwein (2006e). "Finding the required sample size for vibrational analysis using autoregressive modelling techniques", *ASME Journal of Vibration and Acoustics*, In Review.
- Thanagasundram, S. and F. S. Schlindwein (2007). "A Fault Detection Tool Using AR Pole Trajectory", *Proceedings of the WCEAM CM 2007- Second World Congress on Engineering Asset Management (EAM) and The Fourth International Conference on Condition Monitoring*, Harrogate, UK, pp. 1887-1896.
- Thrane, N. (1984). "The Hilbert transform", *Bruel and Kjaer Technical Review*(3), pp. 3-15.
- Troup, A. P. and N. T. M. Dennis (1991). "Six years of ``dry pumping'': A review of experience and issues", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 9(3), pp. 2048-2052.
- Troup, A. P. and D. Turrell (1989). "Dry pumps operating under harsh conditions in the semiconductor industry", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 7(3), pp. 2381-2386.
- Venkatesan GT, Danlu Zhang, Haveh M, T. AH and B. KM. (1999). "Signal processing for fault monitoring using acoustic emissions", *EU-International Journal of Electronics & Communications*, 53(6), pp. 333-8.
- Vogelsang, H., B. Verhulsdonk, M. Turk and G. Hornig (1999). "Pulsation problems in rotary lobe pumps", *World Pumps*(389), pp. 6.
- Wang, J. (2006). "Vibration-based fault diagnosis of pump using fuzzy technique", *Journal of International Measurement Confederation*, 39(2), pp. 176-185.
- Wang, W. (2001a). "Identification of gear mesh signals by kurtosis maximisation and its application to CH46 helicopter gearbox data", *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, pp. 369-372.
- Wang, W. (2001b). "Linear model identification for gear fault detection using higher order statistics and inverse filter criteria", *Proceedings of the Sixth International Symposium on Signal Processing and its Applications* vol.1, pp. 371-4.
- Wang, W. (2001c). "Linear model identification for gear fault detection using higher order statistics and inverse filter criteria", *Proceedings of the Sixth International Symposium on Signal Processing and its Applications* Kuala Lumpur, Malaysia pp. vol.1 371-374
- Wang, W. and A. K. Wong (1999). "Some new signal processing approaches for gear fault diagnosis", *Proceedings of the Fifth International Symposium on Signal Processing and its Applications* pp. 587-590
- Watts, G. M. J. a. D. G. (1968). "Spectral analysis and its applications", Holden-Day, San Francisco.

- Wong, W., L. Laurenson, R. G. Livesey and A. P. Troup (1988). "An evaluation of the composition of the residual atmosphere above a commercial dry pump", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 6(3), pp. 1183-1186.
- World Pumps (1999). "Smart Pump is quick to repay initial outlay", *World Pumps*, N 392 (May 1999), pp. 9-11(3).
- World Pumps (2002). "Early fault detection - An overview", *World Pumps*, N 428 (May 2002), pp. 54-57.
- World Pumps (2005). "Pump controller with fieldbus communications", *World Pumps*, Issue 471 (Dec 2005), pp. 9.
- World Pumps (2006). "Advances in ultrasonic controllers— new energy-saving solutions for pump control", *World Pumps*, N 476(May 1999), pp. 44-46.
- Wycliffe, H. (1987a). "Mechanical high-vacuum pumps with an oil-free swept volume", *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 5(4), pp. 2608-2611.
- Wycliffe, H. (1987b). "Rotary pumps and mechanical boosters-as used on today's high vacuum systems", *Journal of Vacuum*, 37(8), pp. 603-607.
- Wyner, A. D. and S. Shamai (1998). "Introduction to 'Communication in the Presence of Noise' by C. E. Shannon", *Proceedings of the IEEE*, 86(2), pp. 442-446.
- Yang, H., J. Mathew and L. Ma (2005). "Fault diagnosis of rolling element bearings using basis pursuit", *Mechanical Systems and Signal Processing*, 19(2), pp. 341-356.
- Zakrzewski, E., P. L. May and B. S. Emslie (1988). "Developments in vacuum pumping systems based on mechanical pumps with an oil free swept volume", *Journal of Vacuum* 38(8-10), pp. 757-760.
- Zarei, J. and J. Poshtan (2007). "Bearing fault detection using wavelet packet transform of induction motor stator current", *Tribology International*, 40(5), pp. 763-769.
- Zhang, S., J. Mathew, L. Ma and Y. Sun (2005). "Best basis-based intelligent machine fault diagnosis", *Mechanical Systems and Signal Processing*, 19(2), pp. 357-370.
- Zheng, G. T. and W. J. Wang (2001). "A new cepstral analysis procedure of recovering excitations for transient components of vibration signals and applications to rotating machinery condition monitoring", *Transactions of the ASME. Journal of Vibration and Acoustics*, 123(2), pp. 222-9.
- Zhuge, Q., Y. Lu and S. Yang (1990). "Non-stationary modelling of vibration signals for monitoring the condition of machinery", *Mechanical Systems and Signal Processing*, 4(5), pp. 355-365.

A APPENDIX

The working principles of the Roots and Claw Dry Vacuum Pump

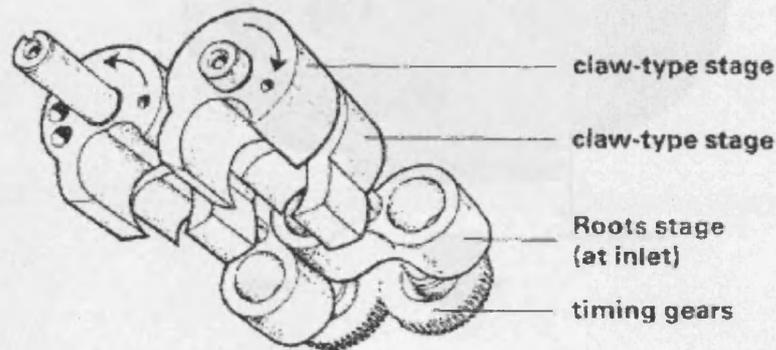


Figure A-1 The Root and Claw mechanism.

A perspective view illustrating the relationship between the rotors. One of the Claw stages is mounted in reverse orientation to the other illustrating the reverse Claw mechanism. Note that this pump has only 2 stages of Claws whereas the iGX100 pump used in the study has 4 stages of Claws. Reproduced from (Wycliffe 1987a).

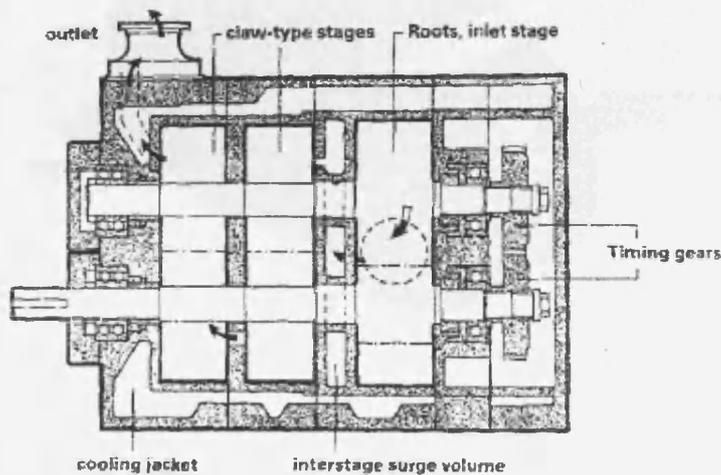


Figure A-2 Cross section of a three-stage pump, with one Roots and two Claw type stages.

Note that this pump has only 2 stages of Claws whereas the iGX100 pump used in the study has 4 stages of Claws Reproduced from (Wycliffe 1987a).

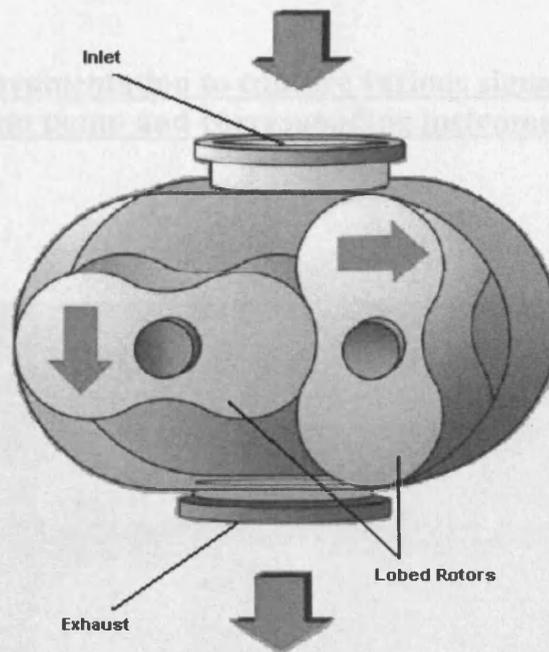


Figure A-3 Root Stage. Reproduced with permission from (Edwards 1993)

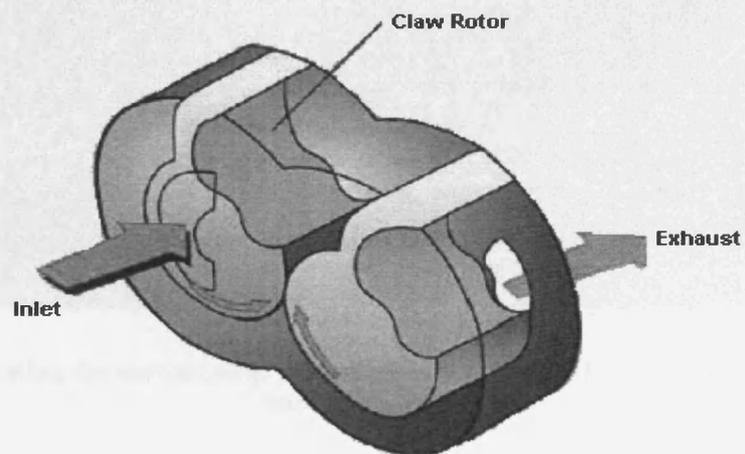


Figure A-4 Claw Stage. Reproduced with permission from (Edwards 1993)

Figure showing instrumentation to capture various signals from the iGX100 Dry Vacuum pump and corresponding instrumentation list

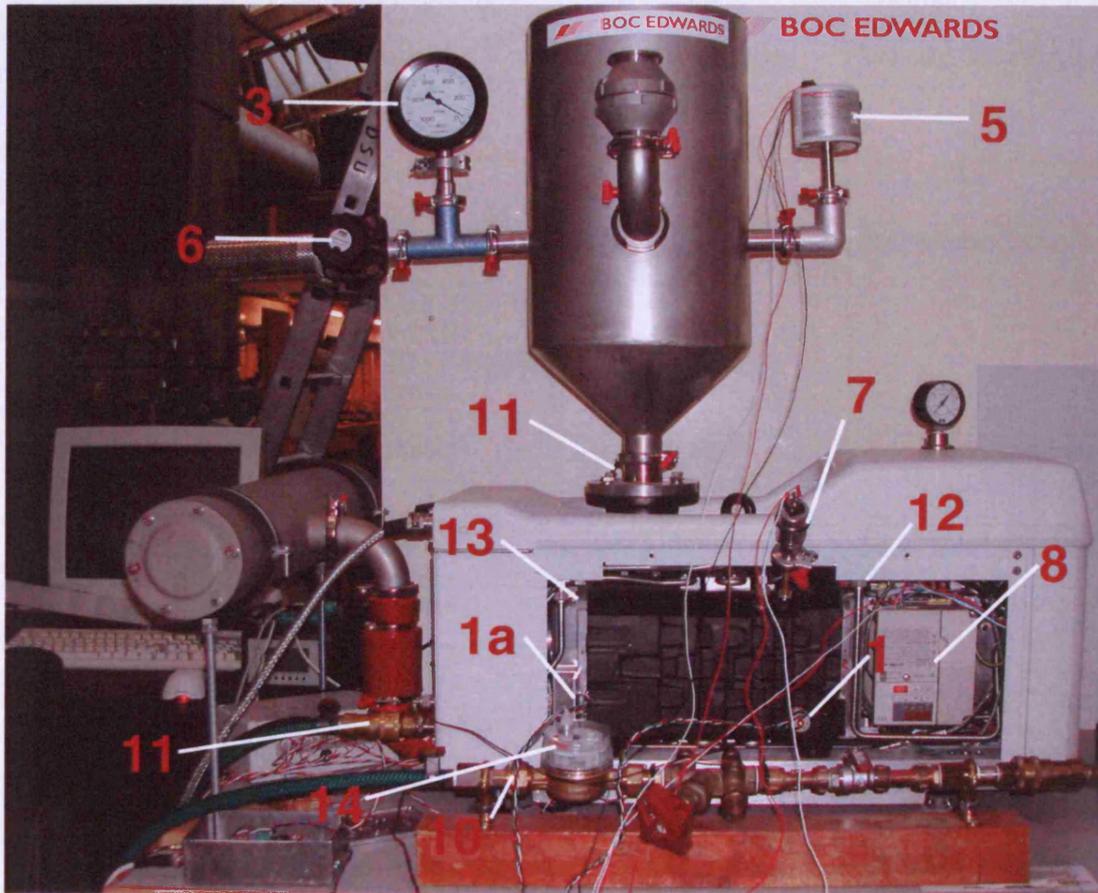


Figure A-5 Monitoring the various pump parameters of the iGX100 dry vacuum pump in our laboratory.

APPENDIX A

Key	Signal	Type	Fitted Location
1	Vibration	Accelerometer <u>Brüel and Kjaer type 4370-VS</u>	As required
1a		Accelerometer <u>ADXL105 iMEMs</u>	
2	Exhaust Pressure	<u>ASM Gauge</u>	Thermal Break Purge gas port
3	Inlet Pressure (Vacuum)	<u>CG16k Dial Gauge</u>	Vacuum Vessel Inlet
4		<u>Pirani gauge</u>	
5		<u>Barocel Gauge</u>	
6	Air mass flow transducer	Flow rates > 5slpm only	Lucas 4AM
7	Gear tooth proximity	<u>Magnetic Pick-up (RS part number 304-172)</u>	Rotor Drive Gear
8	<u>Yasakawa Varispeed 606V7 Inverter:</u> Current, DC Voltage Power Torque, Voltage		
9	Temperature 1	<u>Type IC LM35 (Farnell item no. 409-080)</u>	Cooling Water inlet
10	Temperature 2	Type IC LM35	Cooling Water outlet
11	Temperature 3	Type IC LM35	Atmosphere
12	Temperature 4	Type IC LM35	Pump LV Bearing Section (external)
13	Temperature 5	Type IC LM35	Pump HV Bearing Section (external)
14	Coolant Mass Flow	Pulsed Output Single Jet Turbine Water Meters. <u>RS No 399-5018</u>	Coolant Inlet

Table A-1 Instrumentation list of the iGX100 Dry Vacuum Pump.

Note in Figure A-5 that sensors 4 and 7 were not fitted onto the pump when the picture was taken shown but can be mounted on the pump if required.

A typical ceramic Barden ball bearing mounted on the HV and LV ends of the dry vacuum pump



Figure A-6 Barden Bearing Ceramic Bearing at High Vacuum (HV) and Low Vacuum (LV) ends of iGX100 Dry Vacuum Pump.

Parts of a Bearing



Figure A-7 Bearing Components-Inner Race , Outer Race and the Ball Bearings.

B APPENDIX

A bearing with an inner race crack

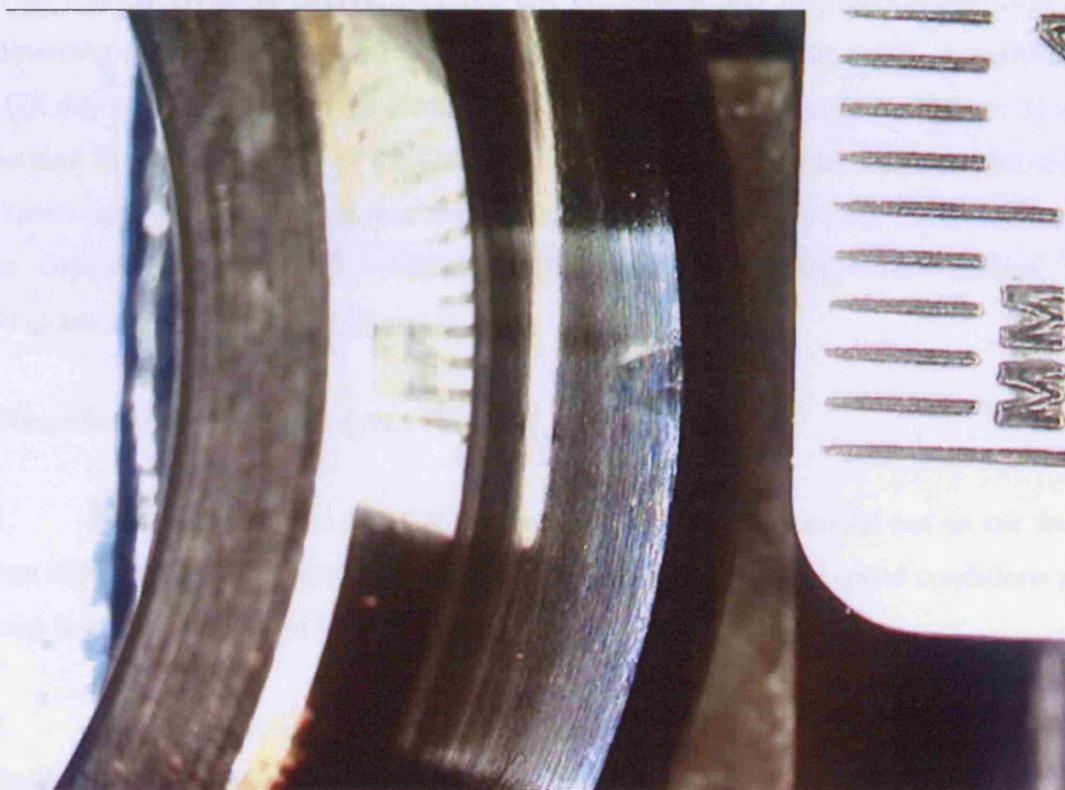


Figure A-8 Crack on inner race of bearing. Crack was approximately 2 mm wide and 2 mm deep.

B APPENDIX

The Experimental Platform

This section gives an overview of the test equipment and instrumentation used for obtaining the test signals used for our experimental phase of the work. A multistage iGX dry vacuum based on the Roots and Claws principle (described in Chapter 2) and located in the laboratories of the University of Leicester was used as the source of the “quasi-steady state” rotating machinery. The schematic of the pump, the sensors used in capturing the data as well as set-up of the data acquisition system are diagrammatically shown in Figure B-1.

The collection of data was done in stages.

1. From the period of Jan 2004 to Feb 2005, tests were carried out on the fault-free iGX100 pump and test data were collected for various pump speed conditions and load factors in controlled laboratory conditions.
2. In May 2005, artificially induced bearing fault samples were supplied from Barden Bearings and these faulty bearings were returned with the original pump to the BOC facility at Shoreham for refitting. The first faulty bearing to be fitted was the bearing with a simulated inner race crack. This faulty bearing was fitted on the high

APPENDIX B

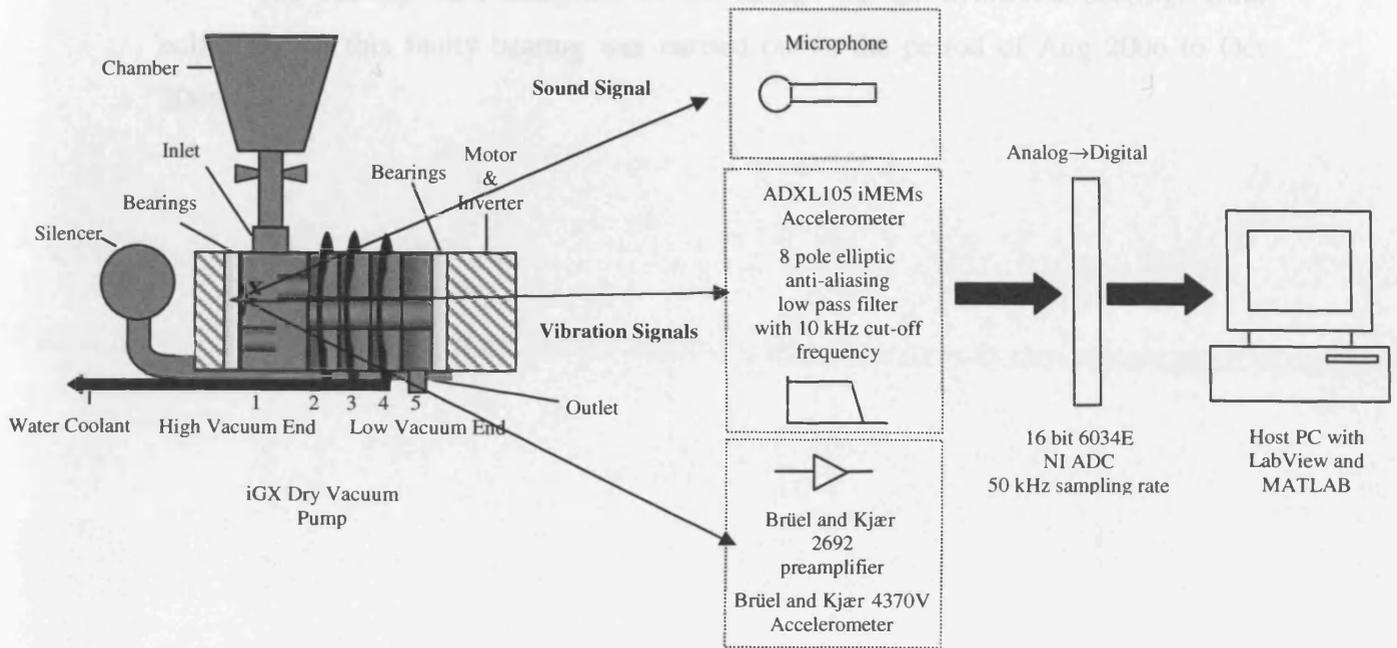


Figure B-1 Schematic of the complete data acquisition system.

The ADXL105 and Brüel and Kjær 4370V accelerometers were mounted radially on point marked X on the dry vacuum pump, near the high vacuum end. The microphone was attached in the vicinity on the high vacuum end.

APPENDIX B

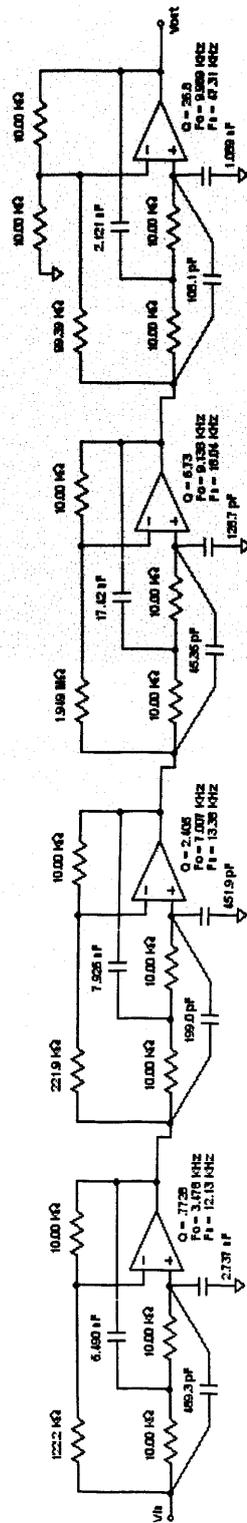
vacuum end of the pump and returned. Data collection was carried in the period of June 2006 to the end of Feb 2006.

3. The second fault condition to be tested was an oversized bearing. Data collection for this faulty bearing was carried out in the period of Aug 2006 to Oct 2006.

8 Pole Elliptic Low-pass Filter

8th Order Low Pass Elliptic

Pass Band Frequency = 10.00 KHz
 Stop Band Ratio = 1.2
 Stop Band Frequency = 12.00 KHz
 Pass Band Ripple = 1.000 dB
 Stop Band Attenuation = 71.76 dB



Rev J41 19 13/06/03 2304

Figure B-2 Schematic showing the 8 pole elliptic low-pass filter with a cutoff frequency of 10 kHz. The filter has a positive Single Amplifier Biquad (SAB) configuration.

Frequency Response of Low-pass Elliptic Filter

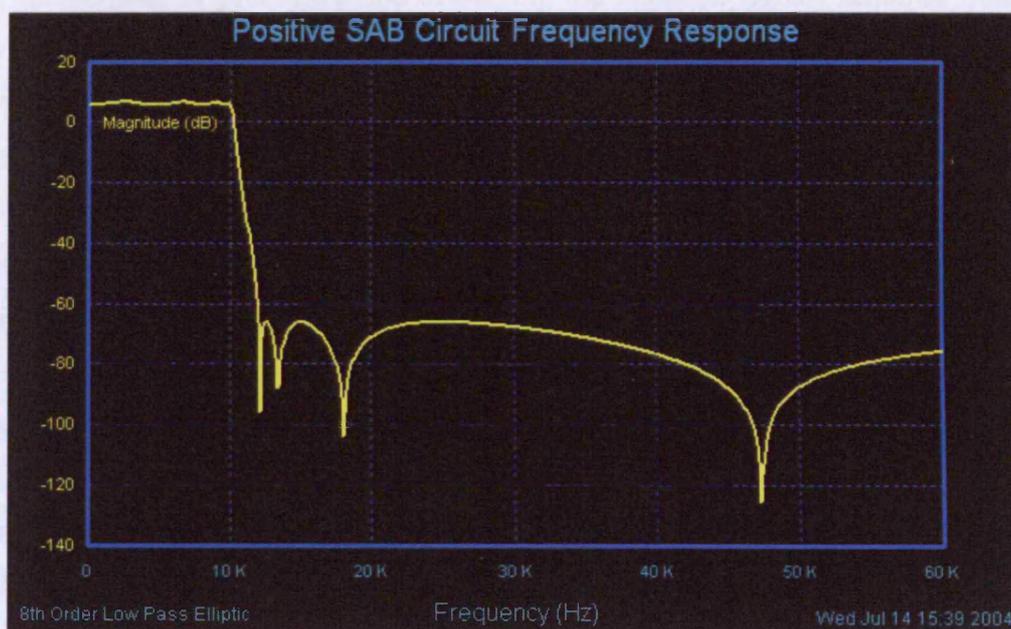


Figure B-3 Frequency Response of 8th Order Elliptic Filter.

Condition Monitoring VI

For the purpose of processing and displaying the vibration signals acquired from the dry vacuum pump, a condition monitoring system was developed using LabVIEW and MATLAB signal processing packages on a 2.4 GHz Pentium-based microcomputer. The main program was implemented as a Virtual Instrument (VI) using the LabVIEW programming language. LabVIEW is a graphical programming language that uses icons instead of lines of text to create applications. In contrast to text-based programming languages, where instructions determine program execution, LabVIEW uses dataflow programming, where the flow of data determines execution. LabVIEW programs are called Virtual Instruments or VIs, because their appearance and operation imitate physical instruments, such as oscilloscopes and multimeters. LabVIEW contains a comprehensive set of tools for acquiring, analyzing, displaying, and storing data, as well as tools to help in troubleshooting code. The VI system presents an effective and user-friendly human-machine interface for on-line condition monitoring, which is of critical importance to real-time fault diagnosis. Matlab script blocks were easily integrated into the LabVIEW VI environment and the signal processing capabilities of both LabVIEW and MATLAB were utilized simultaneously for our signal processing requirements.

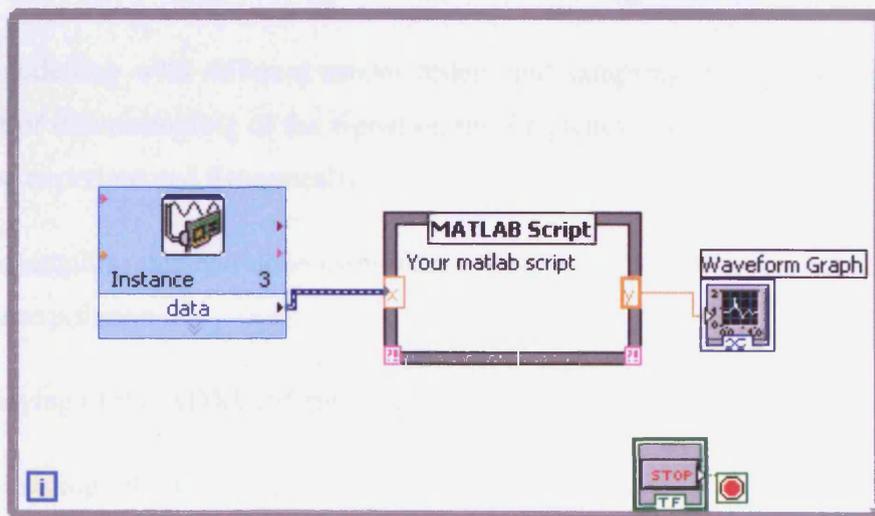


Figure B-4 Using MATLAB Script nodes in the LABVIEW Environment

The generation of the Fourier and AR spectra and the models for the AR technique and other related investigations were carried out using this condition

APPENDIX B

monitoring VI. For collection of experimental data, the speed of the pump was varied at different loading factors and for different sampling rates, both in the faulty and non-faulty conditions, and the test data samples were collected as required. These data were stored in LabVIEW Measurement (LVM) file format for subsequent analysis.

The data acquisition setup and the condition monitoring system of the dry vacuum pump processes three types of sensor signals. The first is the vibration signal from the ADXL105 accelerometer obtained from the dry vacuum pump from the HV end, the second is the vibration signal from the Brüel and Kjær accelerometer and the third is the ADXL105 accelerometer signal obtained from the dry vacuum pump from the LV end.

The main features of the condition monitoring VI are:

1. Displaying of the raw vibration signatures from both accelerometers in the time domain.
2. Displaying of the raw vibration signatures from both accelerometers in the frequency domain.
3. Spectral analysis using both the FFT and AR modelling techniques.
4. AR modelling with different model orders and sampling rates of the signal. The effect of downsampling of the signal on the frequency resolution of the AR spectra can be experimented dynamically.
5. Downsampling that can be done using four different methods - linear, cubic, spline or FIR interpolation
6. Displaying of the ADXL105 temperature for calibration purposes.
7. Data logging vibration data.
8. Derivation of the bearing defect frequencies online with bearing parameters.
9. Estimating the poles of the AR model.
10. Detecting peaks in the vibration spectra automatically using user set criteria.

APPENDIX B

11. Calculating power in interesting frequency bands.
12. Pre-processing of the vibration signals using resonance demodulation technique prior to bearing defect analysis.

The key benefits of the condition monitoring VI are:

- ✓ User friendliness of the Graphical User Interface.
- ✓ Online data acquisition, processing and display.
- ✓ Easy configuration of the analogue input channels.
- ✓ Signal storage for offline processing.
- ✓ Integration of many signal processing functions as a single module.
- ✓ Real-time performance.
- ✓ Easy architecture for future system expansion.

By developing and deploying an online real-time fault diagnostics system as a LaBVIEW VI, a fully automated condition monitoring software of the dry vacuum pump system has been set up.

Circuit for the iMEMs ADXL105 accelerometer

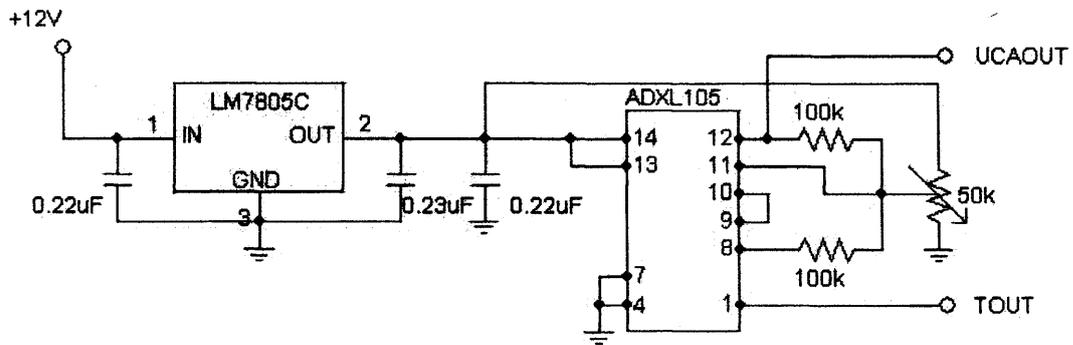


Figure B-5 ADXL105 Accelerometer Circuit.

Figure showing the ADXL105 prototype built

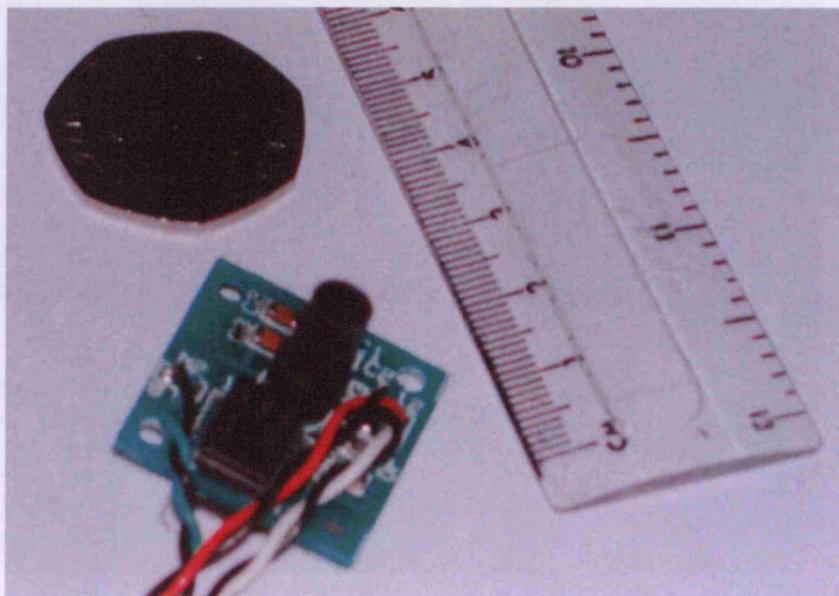


Figure B-6 ADXL105 PCB with Surface Mount Components

Both accelerometers mounted on pump

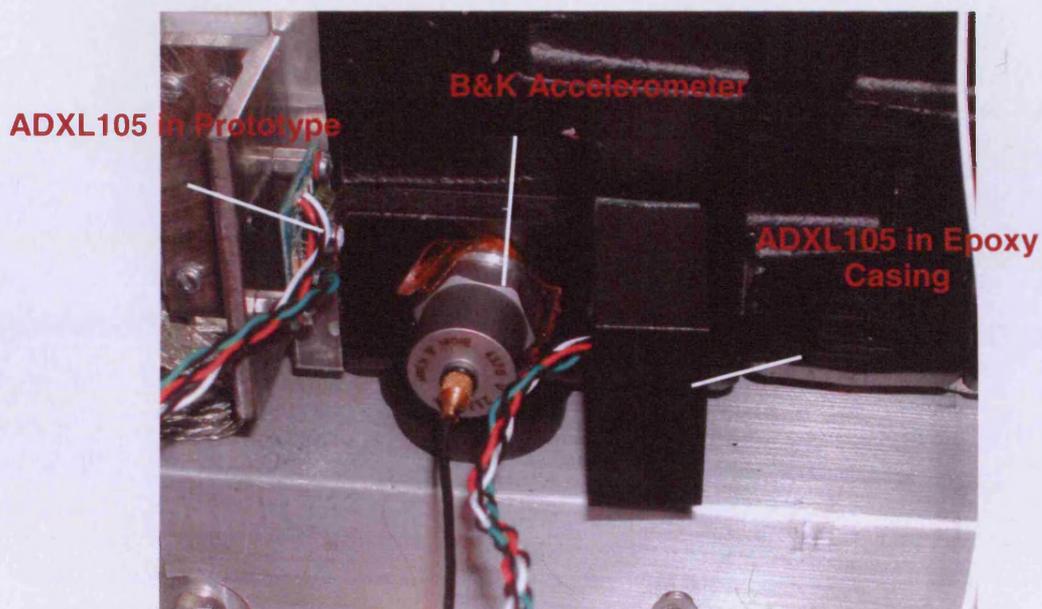


Figure B-7 ADXL105 and B&K accelerometers mounted on high vacuum end of dry vacuum pump.