MACHINE CONDITION DIAGNOSTICS BASED ON ENHANCED SYSTEM IDENTIFICATION TECHNOLOGIES

YUANDONG XU

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of HUDDERSFIELD
Inspiring global professionals

July 2020

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions

Abstract

In modern industries, condition monitoring is considered as the most promising technology to ensure the productivity, reliability and safety of machines. This thesis focuses on developing effective and efficient approaches to monitor the changepoint of rotating machines at an early stage based on the resonant modulation and demodulation techniques. Having had a literature review relating to the resonant modulation based machine condition monitoring, it has been identified that potential technical deficiencies in achieving accurate condition monitoring may lie in the lack of a good understanding of resonant modulation and effective fault detection and diagnosis, which often leads to difficulties in optimising post-processing signal processing approaches to extract accurate fault features.

Based on the hypothesis of linear systems, the resonant modulation in outputs of the systems vary under different input excitations (periodic, approximately periodic impulsive, and quasi-stationary). Different inputs can lead to different modulation characteristics. Performing an effective demodulation of these responses is the most beneficial part for utilising the resonant modulation methods in vibration-based condition monitoring. The characteristics of the vibration signals are expected to be effectively analysed by different signal processing approaches for early fault detection and diagnosis.

From the perspective of the system identification, the stochastic subspace identification (SSI) approach is employed to select the optimal central frequencies rather than selecting the most impulsive frequency band, which is targeted by the method similar to Kurtogram. Through simulation and experimental studies, it shows that the SSI is more robust to the strong background noise (SNR$\leq$-30dB) in selecting the optimal frequency bands than the conventional Kurtogram method. Based on the frequency bands selected by the SSI, a novel method, named ensemble average of autocorrelation signals (EAAS), was developed to demodulate the periodic impacts induced resonant responses. The simulation study shows that EAAS can identify the incipient bearing faults from the noisy vibration signals at a SNR less than -35dB and the performance of the EAAS is also verified by the experiments on ball bearings of induction motors. For the aperiodic impacts induced resonant modulation, the output responses are not periodic and are terminologically named as cyclostationary signals. Two novel approaches, ensemble average of

autocorrelated envelopes (EAAE) and phase linearisation based modulation signal bispectrum (PL-MSB), were developed to characterise the cyclostationary responses due to the nonstationary impacts. The simulation studies demonstrated that the proposed EAAE and PL-MSB can handle the extremely poor signals (SNR≤-30dB) to extract the strong fault signatures for the purpose of accurate diagnosis of bearing faults. The experimental studies on tapered roller bearings show that the proposed two methods outperform the conventional envelope for early fault detection and diagnosis under an extremely low SNR and large random slippages.

Table of Contents

List of Figures

List of Tables

Dedications and Acknowledgements

This research work was carried out in the Centre for Efficiency and Performance Engineering (CEPE), School of Computing and Engineering (SCE), at the University of Huddersfield, UK. I would like to appreciate the financial support by the scheme of the vice chancellor research fee waiver from University of Huddersfield and the state scholarship from China Scholarship Council.

I am sincerely grateful to my main supervisor, Professor Fengshou Gu. He gave me the professional guidance, insightful comments and kind support during my whole PhD studies. His encouragement inspires me with enthusiasm to handle the challenges in research and life. My research would have been impossible without the aid and support of him. I could not have imagined having a better supervisor for my PhD research.

I would like to thank my co-supervisor, Prof Andrew D. Ball for his valuable suggestions on my research activities. His passion for research and immense knowledge in condition monitoring excite me to continuously conduct the research on machine condition monitoring.

I am profoundly grateful to all friends at CEPE for the discussions on modelling and simulation, for the arduous works before deadlines, and for all funs we have had in the last four years. My sincere thanks to the administrators and technicians in SCE.

Heartfelt thanks go to my parents Maoquan Xu and Shuyan Wang for their understanding and support throughout my life. Especially, my wife Xiaoli Tang gave me invaluable help during the PhD studies. Thanks to a wonderful wife like here and she has always been my strength throughout the times.

List of Abbreviations

ADC	Analogue-to-Digital Converter
AM	Amplitude Modulation
ARMA	Auto-Regressive Moving Average
CB	Conventional Bispectrum
CE	Conventional Envelope
CM	Condition Monitoring
DOF	Degree of Freedom
EAAE	Ensemble Average of Autocorrelated Envelopes
EAAS	Ensemble Average of Autocorrelation Signals
FFT	Fast Fourier Transform
FRF	Frequency Response Function
HFRT	High Frequency Resonance Technique
MSB	Modulation Signal Bispectrum
OMA	Operational Modal Analysis
ORS	On-Rotor Sensing
PL-MSB	Phase Linearisation based Modulation Signal Bispectrum
PM	Phase Modulation
PSD	Power Spectral Density
RMS	Root Mean Square
SD	Stabilization Diagram

SK	Spectral Kurtosis
SNR	Signal to Noise Ratio
SSI	Stochastic Subspace Identification
STFT	Short-Time Fourier Transform
TKEO	Teager Kaiser Energy Operator
TSA	Time Synchronous Averaging
VFD	Variable Frequency Drive
WT	Wavelet Transform
WPT	Wavelet Packet Transform

Publications from This Thesis

[1] **Y. Xu**, D. Zhen, J. X. Gu, K. Rabeyee, F. Chu, F. Gu, and A. D. Ball, 'Autocorrelated Envelopes for early fault detection of rolling bearings', *Mechanical Systems and Signal Processing*, vol. 146, p. 106990, Jan. 2021, doi: 10.1016/j.ymssp.2020.106990.

[2] **Y. Xu**, C. Fu, N. Hu, B. Huang, F. Gu, and A. D. Ball, 'A Phase Linearisation-based Modulation Signal Bispectrum for Analysing Cyclostationary Bearing Signals', *Structural Health Monitoring*, 2020, doi: 10.1177/1475921720949827.

[3] D. Zhen, J. Guo, **Y. Xu**, H. Zhang, and F. Gu, 'A Novel Fault Detection Method for Rolling Bearings Based on Non-Stationary Vibration Signature Analysis', *Sensors*, vol. 19, no. 18, p. 3994, Jan. 2019, doi: 10.3390/s19183994.

[4] K. Rabeyee, **Y. Xu**, F. Gu, and A. D. Ball, 'A novel wavelet thresholding method for vibration data denoising and diagnostic feature enhancement in condition monitoring', presented at the ICAC 2019 - 2019 25th IEEE International Conference on Automation and Computing, 2019, doi: 10.23919/IConAC.2019.8894986.

[5] K. Rabeyee, **Y. Xu**, S. Alabied, F. Gu, and A. D. Ball, 'Extraction of information from vibration data using double density discrete wavelet analysis for condition monitoring', presented at the 16th International Conference on Condition Monitoring and Asset Management, CM 2019, 2019, doi: 10.1784/cm.2019.110.

[6] L. Hu, **Y. Xu**, F. Gu, J. He, N. Hu, and A. Ball, 'Autocorrelation Ensemble Average of Larger Amplitude Impact Transients for the Fault Diagnosis of Rolling Element Bearings', *Energies*, vol. 12, no. 24, p. 4740, Jan. 2019, doi: 10.3390/en12244740.

[7] K. Rabeyee, X. Tang, **Y. Xu**, D. Zhen, F. Gu, and A. D. Ball, 'Diagnosing the change in the internal clearances of rolling element bearings based on vibration signatures', presented at the ICAC 2018 - 2018 24th IEEE International Conference on Automation and Computing: Improving Productivity through Automation and Computing, 2018, doi: 10.23919/IConAC.2018.8749121.

[8] **Y. Xu**, P. A. van Vuuren, X. Tan, F. Gu, and A. Ball, 'A robust method to detect faults of rolling bearings using ensemble average autocorrelation based stochastic subspace identification', in *COMADEM 2017 Proceedings*, Preston, UK, 2017.

[9] **Y. Xu**, L. Haiyang, D. Zhen, I. Rehab, F. Gu, and A. Ball, 'Fault detection of rolling bearings using an ensemble average autocorrelation based stochastic subspace identification', in *Proceedings of 24th International Congress on Sound and Vibration*, London, UK, 2017.

[10] **Y. Xu**, X. Tang, F. Gu, A. D. Ball, and J. X. Gu, 'Early detection of rolling bearing faults using an auto-correlated envelope ensemble average', in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–6, doi: 10.23919/IConAC.2017.8081993.

[11] X. Tang, **Y. Xu**, F. Gu, A. D. Ball, and G. Wang, 'Fault detection of rolling element bearings using the frequency shift and envelope based compressive sensing', in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–6, doi: 10.23919/IConAC.2017.8082063.

[12] O. Hamomd, S. Alabied, **Y. Xu**, A. Daraz, F. Gu, and A. Ball, 'Vibration based centrifugal pump fault diagnosis based on modulation signal bispectrum analysis', in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–5, doi: 10.23919/IConAC.2017.8082029.

Chapter 1 Introduction

This chapter gives a brief introduction to the research background associated with the research undertaken in this thesis. The background of machine condition monitoring is investigated, with a focus on the vibration-based condition monitoring technique, the most commonly used technique for monitoring the working conditions of the machines. Based on the investigation, the research motivation as well as the aims and objectives are summarised. Finally, the organisation of the whole thesis is explained.

1.1 Background of Condition Monitoring

Machines are expected to fully operate from the beginning of their working life right up until their failure in order to maximize the usage of their useful life. However, unexpected malfunctions in machines sometime result in production loss or even casualties. The instinct solution for machine failure is preventive maintenance, which means the maintenance is carried out periodically to prevent the occurrence of machine failures. This maintenance strategy is effective but not efficient due to extra cost and unnecessary resources by replacing the machine parts at set intervals.

In general, maintenance strategies can be classified into three groups: Reactive Maintenance, Preventive Maintenance, and Condition based Maintenance [1]. Here a brief introduction of three maintenance strategies is given as follows.

Reactive Maintenance. Reactive maintenance is the earliest maintenance strategy, in which machines run until break down and then the failure parts or systems can be fixed. The repair works are not scheduled, which means some industries using low cost and duplicate machines can use this maintenance strategy because the malfunction of one machine has little effect on the whole production line. However, in most industries the failure of critical equipment can result in significant loss of production and even people's lives if the maintenance is not conducted in time. The reactive maintenance usually requires more time to recover and high costs to repair.

Preventive Maintenance. Preventive maintenance also named as the time-based maintenance, is achieved by the periodic replacement or repair of machines or machine parts. Usually the time intervals are selected to be less than 1% to 2% of failure time. This maintenance strategy is planned in advance so that most machine failures can be avoided. However, unneglectable disadvantages are not only the occurrence of a small amount of failures but also the excessive maintenance work and replaced components.

***Condition based Maintenance*.** Condition based maintenance is to find the potential or early breakdown of a machine through condition monitoring. Thus, the maintenance can be carried out at the most appropriate time compared with the reactive maintenance and preventive maintenance. Condition based maintenance can lead to an optimum maintenance strategy because it is able to denote the current working conditions and also give the prediction of the remaining useful life of machines. However, the challenge for condition based maintenance is the high requirement of effective and efficient condition monitoring techniques.

The origin of Condition Monitoring (CM) can be traced to the earliest development of the machinery employed in different industries [2]. The methods of using human beings' senses (looking, listening, smelling and touching) are still valid even though more and more advanced

and sophisticated equipment is developed to monitor the health condition of machines. Condition monitoring has been widely adopted in many areas, such as power generation, manufacturing, mining, chemicals, transportation, etc. Practically, Condition Monitoring is carried out during the operation of machines, so that various measurements, including vibration, acoustics, acoustic emission, instantaneous currents, thermal variations, pressure and exhaust emissions, can be acquired. Based on this data, the health conditions of the machines can be determined by the acquired data via various signal processing methods, the maintenance decision can then be made based on the results of fault detection, diagnosis and prognosis.

1.2 Condition Monitoring Methods

Condition monitoring aims to acquire the current working condition of machines and denote the best time for the maintenance, hence minimizing the interruption to production. The commonly used techniques in condition monitoring are vibration/acoustics, oil analysis, process parameters, and thermal imaging.

Oil Analysis based Condition Monitoring. Lubrication oil is essential in the machines as it can significantly reduce friction forces and extend lifespan of machinery. Oil analysis is to take lubricant samples regularly and then detect the quantity, type, shape, size, material of the debris, the current oil viscosity, the additives and other contaminations. These parameters can show the type and location of defects according to the components' differences and the relative movement. Sometimes oil samples are not easy to obtain from the machines. Furthermore, equipment for oil analysis is quite expensive and the offline monitoring is usually not conducted in time, which makes oil analysis not reliable for most of the machines.

Process Parameters based Condition Monitoring. Process or performance parameters are an effective method for monitoring machine conditions because process parameters directly indicate the efficiency of certain machines, such as compressors and engines. If a malfunction occurs in the machine, the drop in outputs can be seen when comparing to the normal conditions. However process parameters based condition monitoring are not effective for transmission systems or other faults which has an insufficient negative effect on the machine efficiency.

Thermal Imaging based Condition Monitoring. Thermal imaging is becoming a more attractive way to obtain machine fault detection and diagnosis, owing to the fast development of advanced sensing instrument and computing capacity. Temperature information is highly correlated with the working conditions of machines. Heat usually comes from either friction or chemical reaction in most machines and consequently, temperature signals could be an effective indictor for machine health condition. Thermal imaging techniques extend the single point temperature into a two

dimensional heat distribution, which allows to more effectively monitor machines in a contactless way. However, this technology is not sensitive to the early faults of rotating machines, for instance fatigue pitting.

Vibration based Condition Monitoring. All working machines generate vibrations due to the dynamic forces no matter how good the machine conditions are. Despite the dynamics of the working principles, some instances, including manufacture errors, installation errors, load and speed oscillations, lubrication conditions, and fluid-structure interactions, makes vibration more pronounced. Such dynamic responses are often repeatable along the rotation of machines, and the periodicity of vibration signals can directly indicate the internal conditions, which makes the Fourier analysis significantly powerful in the frequency domain. This thesis focuses on the vibration based condition monitoring for achieving early fault detection and diagnosis.

Acoustics based Condition Monitoring. Airborne acoustic signals are basically the time-varying pressure waves. The variation of the varying pressure signals can be captured by microphones, which attracts increasingly more attention in machine condition monitoring due to the noncontact installation and wide monitoring range. Acoustic signals from operating machines are dependent on the environment and the distance of the measurement location from machines. Usually acoustic signals contain strong background noise and advanced signal processing methods are highly desired in acoustics based condition monitoring.

Instantaneous Angular Speed based Condition Monitoring. Rotating speeds of machines are a very important parameter to demonstrate the working conditions as the dynamic motions of machines are highly related to rotating speeds. Instantaneous angular speed (IAS) also called torsional vibration describes the dynamic oscillation along the torsional direction of the flexible rotor, which plays the critical role of rotating machines. IAS has been widely used in the reciprocating engines/compressor and gearboxes for effective and efficient condition monitoring.

Electrical Signals based Condition Monitoring. Induction motors are widely used as the power source in industries. Another frequently used technique in machine condition monitoring is based on stator current signals, which is known as motor current signature analysis (MCSA). The implementation of this technique does not need an extra transducer. The instantaneous current signals are sensitive to malfunctions or failure of the whole system.

Acoustic Emission based Condition Monitoring. Acoustic emission is the radiation of elastic waves due to the material deformation or crack formation. The rapid release of energy can be captured using a very high sampling rate by acoustic emission transducers from the surface of machines. The typically range in machine condition monitoring is between 100kHz and 1MHz.

Acoustic emission is an effective approach to machine condition monitoring and it has been widely used in fault detection and diagnosis of bearings, gearboxes and engines.

1.3 System Identification

Aiming to predict, control, monitor and improve the plant and machinery systems, a special technique named system identification utilises observed or experimental data to describe the systems by the mathematical language. Figure 1-1 depicts the schematic procedure of the system identification.

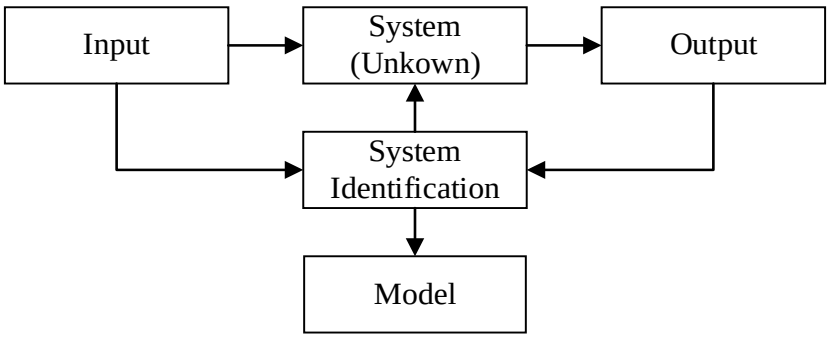


Figure 1-1 System identification schematic diagram [3]

The description of the relationship between input and output is the mathematical model that is the prime objective of the identification. System identification is a methodology for building mathematical models of dynamic systems using measurements of the system's input and output signals. The process of system identification requires that:

- Measure the input and output signals from the system in time or frequency domain.
- Select a model structure.
- Apply an estimation method to estimate value for the adjustable parameters in the candidate model structure.
- Evaluate the estimated model to see if the model is adequate for the application needs.

A dynamic model is a mathematical relationship between a system's input and output variables. Models of dynamic systems are typically described by differential or difference equations, transfer functions, state-space equations, and pole-zero-gain models. Generally, the model is divided into two parts: deterministic models depicted by mathematical description; stochastic models represented by both the statistical and mathematical description. Obviously, a precise model is beneficial for design, control, monitoring, diagnoses, prognosis and improvement.

According to the prior knowledge, system identification can be classified as white box, grey box and black box. A white box system means that it can be expressed as deterministic equations while a black box system can only be represented by inputs and outputs. Grey box systems overlap the white and black systems, which only have partial information of systems. However, most of

system identification is based on black box modelling by using historic data in mechanical engineering. Black box based system identification involves artificial intelligence, which has been developed quickly due to the updates of computational resources and theoretical research.

1.4 Research Motivation

The marketplace is of high competition and the machines must maintain high efficiency in case of the elimination. However, the health condition of machines varies with the time elapsing. Large quantities of information, such as acoustics, vibration, acoustic emission, temperature, pressure, lubrication oil, power supply, are captured and analysed during the operation. The data acquisition system, consisting of transducers, data acquisition devices and computers, plays an important role for collecting raw information of machines in condition monitoring. These techniques establish the basis of maintenance strategies because the present and past conditions of machine systems can be identified by the acquired signals. Only in this way can it ensure in real time that machines work in the appropriate conditions with high productivity and reliability.

Rotating machines play an extremely important role in machinery systems and rotating machines are widely used in domestic and industrial applications. Usually the rotating parts are supported by rolling element bearings or journal bearings. These bearings can guide the movement of rotors, reduce frictions between rotors and stators, and support internal or external torques. Suffering from hash working environments, bearings are usually the most vulnerable components in the rotating machines [4], for instance wind turbines, gas turbines, automobiles, trains, ships and aircrafts. The health conditions of bearings in these machines have a significant influence on the whole rotating machine system. Therefore, condition monitoring of bearings in these machines are highly desired for effective and reliable operation and production. Usually these rotating machines are working under very harsh conditions including high temperature, high pressure, heavy loads, and varying speeds. Figure 1-2 shows several bearing failure cases of industrial machines. The bearing fault detection and diagnosis was not successfully achieved in time, consequently the severe failures of bearings and even whole machines occurred. The harsh working conditions as well as the complex system construction lead to significant difficulties for condition monitoring due to the obtained signals being nonstationary and poor signal to noise ratio.

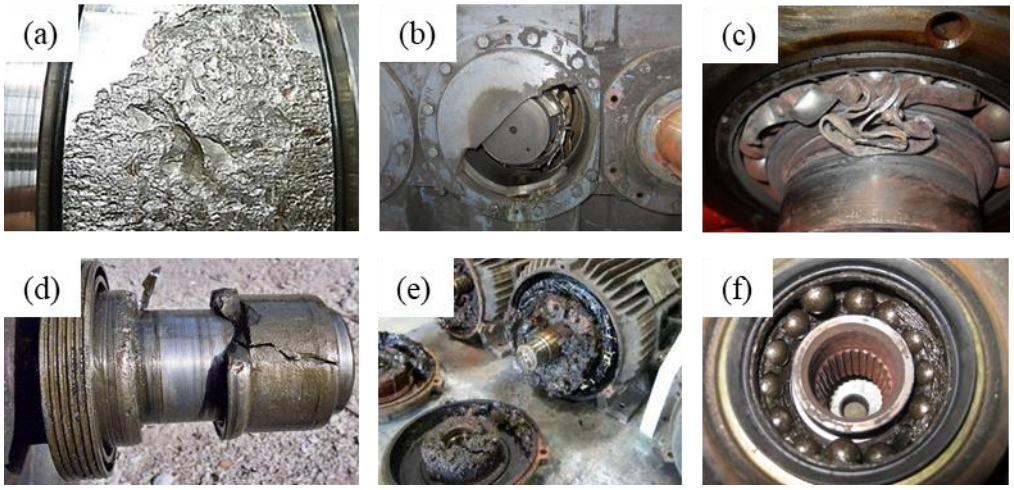


Figure 1-2 Bearing failure cases from industrial machines: (a) wind turbine; (b) gearbox; (c) pump; (d) train bogie; (e) motor; (f) automobile

The early fault detection and diagnosis of bearings allow ample time for planning and implementing maintenance operations. Rolling element bearings are designed to have long service life. Bearing defects are usually inevitable due to the large quantities of revolutions. Bearings work in various conditions and several factors can accelerate the defect development. More than 50% of bearing faults are induced by lubrication failure [5]. Corrosion is the second most common failure factor, and less than ideal mounting procedure is recognised as the third [6]. The other common causes of bearing failure are overload, misalignment, and installation errors. Cracks in microscale initially occur on micro-surfaces and then propagate to a localised fault in macroscale. The incipient fault of bearings is considered as the micro-scale cracks occurred on the micro-surfaces of bearing elements. The desired fault diagnosis is to identify the microscopic bearing defects, which then allows sufficient time for maintenance activities. The early fault detection and diagnosis aims to identify the bearing defects once the faults occur. However, the micro-scale cracks are very difficult to detect and diagnose due to the very weak fault signatures and strong background noise. The early fault detection and diagnosis [7] are extremely pursued in machine condition monitoring.

1.5 Aim and Objectives of the Research

The aim of this research project is to investigate the resonant modulation in vibration signals from rotating machines for achieving incipient fault detection and diagnosis. The resonant modulation is a common and important part in vibration signals and the employment of the resonant modulation can lead to the effective condition monitoring of machines. To fully understand the resonant modulation and exploit the merits of the resonance in machines, this research is carried out by completing the following objectives:

(1) Review the research on resonant modulation and the corresponding signal processing methods for machine fault detection and diagnosis;
(2) Investigate the resonant modulation under various excitations and hence pave the way to the advanced demodulation methods for incipient fault diagnosis;
(3) Propose a system identification method to determine the optimal frequency bands other than using the methods to find the most impulsive frequency bands;
(4) Develop a signal processing method to demodulate deterministic resonant modulation signals induced by the periodic excitations;
(5) Develop a signal processing method to demodulate cyclostationary resonant modulation signals induced by the cyclostationary impacts.

1.6 Organisation of the Thesis

The remaining contents of the thesis are organised as follows:

Chapter 2 reviews the research activities of the demodulation approaches in machine condition monitoring and the publications on investigating resonant modulation in rotating machines.

Chapter 3 explains the resonant modulation in the linear system under the three types of inputs, which paves the way to develop effective demodulation methods for early fault detection and diagnosis.

Chapter 4 introduces the developed state space model of the bearing, describes the SSI method for selecting the optimal frequency bands and demonstrates the demodulation approach for extracting the periodic fault features.

Chapter 5 verifies the performance of the proposed methods by experimental studies on the ball bearings.

Chapter 6 proposes two novel methods to demodulate the cyclostationary resonant modulation signals effectively from the extremely poor SNR.

Chapter 7 examines the developed methods by using the vibration signals from tapered roller bearings.

Chapter 8 draws the conclusions of the research works and provides recommendations for further investigations on the resonant modulation based incipient fault detection and diagnosis.

Chapter 2 Literature Review of Demodulation Analysis in Condition Monitoring

This chapter reviews the commonly used demodulation approaches in machine condition monitoring. The demodulation techniques, including envelope analysis, Teager Kaiser energy operator, modulation signal bispectrum and spectral correlation, are reviewed to show the current research state on demodulation analysis in bearing fault detection and diagnosis. Large quantities of researchers make great contributions to detect the impulsive behaviour in vibration signals. The resonant modulation is not fully exploited for achieving early fault detection and diagnosis.

2.1 Demodulation Techniques based Fault Detection and Diagnostics

As introduced in the first chapter, condition monitoring is effective to detect and prognosticate malfunctions in machinery systems and many researchers have made great contributions to exploring condition monitoring techniques. The application of CM techniques can date back to the development of machines and at the beginning human senses were relied on to monitor the machine conditions [2]. Over time, technologies in this field have made great progress and many measurements, such as vibration, acoustics, acoustic emission, instantaneous current, temperature, pressure, and chemical parameters can be utilised to achieve machine condition monitoring successfully. Rotating machines promoted the industrial revolution and have played a dominant role since the last century [8]. As the rotating machines are extensively used, CM of these machines is significant to guarantee good performance. Though a variety of condition monitoring methods are available, the vibration based methods are considered to be the most widely applied technique for CM [9]. Many journal articles, conference papers and technical reports on this subject are published every year.

Modulation is a common phenomenon in rotating machines and the modulation gives a chance for detecting and diagnosing the defects as early as possible. The modulation phenomenon can enhance the weak fault information by carrying the low frequency information to a high frequency range. The meshing frequencies, natural frequencies, passing frequencies, and supply frequencies are potential carrier frequencies in collected signals from rotating machines. The modulation can make the fault information more pronounced in the high frequency range than that in the low frequency range. An early milestone method in the time domain is the development of the envelope analysis on bearing vibration signals by Darlow [10]. The original name of the envelope is “high frequency resonance technique”, which demodulates the resonant responses to find the bearing defects effectively. The envelope analysis is still popular in machine condition monitoring for demodulation analysis [11], [12]. The envelope analysis is a very good and robust method in vibration based fault detection and diagnosis. Usually the envelope is obtained from analytic signals by the Hilbert transform and the noise is inclusive during the calculation. Consequently the main problem of the envelope in early fault detection and diagnosis is that the envelope itself cannot suppress background noise, which makes the envelope is ineffective in early fault diagnosis. Another frequently used demodulator is Teager Kaiser Energy Operator (TKEO) [13], Kaiser formalised the energy operator for processing speech signals. The TKEO can denote the energy of the transient responses which can reveal the cyclic variation in rotating machines. The effectiveness of TKEO relies on the impacts induced transient responses, which is the typical resonant modulation in vibration signals. Bozchalooi and Liang [14] used wavelet transform to

suppress the random noise in vibration signals and then employed TKEO to demodulate the vibration signals for effective detection and diagnosis of gear faults. Liang and Bozchalooi [15] state that the TKEO can eliminate the frequency interference and demodulate the amplitude modulation without the requirement of the prior knowledge. They demonstrated the TKEO can achieve bearing fault detection and diagnosis at the SNR of -15 dB. Liu, Wang and Lu [16] used the Local Characteristic-scale Decomposition to obtain monocomponent signals and then the decomposed signals were demodulated by the TKEO. The monocomponent signals are desired for TKEO because of the less interference between multiple harmonics. Imaouchen et al. [17] proposed the ensemble empirical mode decomposition based frequency-weighted TKEO for detecting the bearing faults. The TKEO fails in the simulated case at a noise level of SNR-5dB. The TKEO approach is also a good method to demodulate the amplitude and frequency modulation vibration signals. One of the deficiencies of TKEO is similar to the envelope analysis, which is that it is not robust to strong background noise. Moreover, the TKEO is approximately the squared envelope of the differential of the signal [18]. In addition, Randall and Smith denote that the TKEO of a multi-component carrier frequencies is meaningless for the frequency modulation signals [18]. It is assumed that the target signal of TKEO is dominated by a mono-component carrier frequency and the TKEO of the raw signal may result in misleading results. Reference [18] gives an guide to implement the TKEO for machine fault detection and diagnosis.

The envelope based demodulation analysis fully exploits the merit of the high Signal to Noise Ratio (SNR) around the natural frequencies. The prior step for envelope analysis is the determination of the optimal frequency bands otherwise the envelope analysis cannot be guaranteed to show the reliable results. Concentrated on the determination of optimal frequency bands and envelope analysis, the large quantities of research were carried out to find a robust way to select good candidates for demodulation analysis. Since Antoni [19] studied Spectral Kurtosis (SK) thoroughly, Kurtogram [20] based on short-time fast Fourier transform (STFT) and Wavelet Transform (WT) has been explored by many researchers [21]. Gu [22] paid attention to the bearing fault severity diagnosis based on the kurtogram and envelope analysis. With the assistance of the Kurtogram, an optimal frequency band can be determined to achieve the best demodulation results by the envelope. The Kurtogram is struggling to find a good frequency band under the adverse impacts of strong background noise and aperiodic impulses, which are often the circumstances at the early fault stage of bearings. Barszcz and JabŁoński [23] proposed the protrugram, which calculates the kurtosis of the envelope spectrum other than the filtered time signals and the method can handle the vibration signals with poorer signal to noise ratios than Kurtogram. However, this method calculates the envelope spectrum before getting the optimal frequency band, which

actually obtains the diagnosis results in the middle. The protrugram is not a logic way to detect and diagnose bearing faults. Antoni [24] developed the infogram by replacing the kurtosis with negentropy and moreover, both time and frequency domain can share the same concept. The infogram is motivated by the protrugram and the infogram is also calculated from the envelope or squared envelope. These two methods share same deficiencies that they are not effective under a very poor signal to noise ratio. The improvement of these algorithms for determining optimal frequency bands attracts a large number of researchers [25]–[28]. The improved methods increase the robustness of the aforementioned signals but are hard to break through the problem of heavy noise mixed in the signals. Even with the proper frequency bands, the envelop analysis or TKEO are still ineffective to extract the fault signatures in the circumstances of incipient fault diagnosis because both envelope and TKEO do not have the ability of noise suppression.

Another famous demodulation technique belongs to the high order spectrum analysis. The applications of the bispectrum can effectively decouple the modulation components [29]–[31]. To improve the performance of conventional bispectrum for charactering the modulation signals, the Modulation Signal Bispectrum (MSB) is introduced by Gu [32] to demodulate fault features and suppress random noise from vibration signals to identify and quantify common faults of reciprocating compressors. Then, Tian [33], [34] and Rehab [35] applied the MSB based fault detector to identify the bearing faults at a very low SNR (less than -20dB). The MSB originates from the conventional bispectrum (CB) and takes into account the lower and upper sidebands simultaneously. A comparison between CB and MSB was carried out in [36] and it evidently shows that the MSB can give more accurate and sparse diagnostic results. However, the MSB methods are based on the Fourier transform and is designed to eliminate the stationary noise, which cannot effectively tackle the nonstationary signals generated by bearings due to the random slippage between bearing elements. Recently, Spectral Correlation has been developed to characterise cyclostationary signals by extracting cyclic energy features, and it has been applied successfully to vibration signals for detecting and diagnosing bearing problems. Antoniadis and Glossiotis used the framework of cyclostationary analysis to extract the periodical varying statistics in bearing vibration signals [37]. The squared envelope analysis can obtain a similar representation with the spectral correlation over all frequency range when processing cyclostationary signals [38]. Antoni [39] dedicated the cyclostationary processes in rotating machines which formed a guideline for characterising rotating machine signals. The cyclic spectral analysis can detect and diagnose bearing faults in poor quality signals and furthermore it can also indicate the fault severity [40]. Antoni [41] present the similarities and differences between the spectral correlation and the conventional spectral analysis and this reference also gives a guide for

implementing cyclic spectral analysis in practice. Raad [42] et al. proposed several efficient indicators to evaluate the cyclostationarity at different orders without computing the polyspectra, these indicators are verified and appear to be very promising to indicate the conditions of machines. The cyclic spectra analysis has attracted an increase in attention on machine condition monitoring Antoni [43] gives an elaborate tutorial on cyclostationary analysis for machine fault detection and diagnosis using 20 examples. Similarly a guideline is also given by Wang [44] for easily and quickly finding informative spectral frequency bands for generating enhanced envelope spectrum from spectral coherence. Although the Spectral Correlation is versatile, the computation consumes many resources. Antoni [45] developed a short-time Fourier transform (STFT) based fast computing method of Spectral Correlation. The spectral correlation is very effective at extracting energy oscillation information from cyclostationary signals, but these methods are unable to suppress the strong background noise. The capability of reducing noise is highly desired in early fault detection and diagnosis. To the best of the author's knowledge, scarce methods can achieve the bearing fault detection and diagnosis under the SNR of less than -30dB.

2.2 System Identification in Condition Monitoring

The application of system identification has been increasing significantly for several decades, which interests a variety of researcher. Zadeh [46] and Eykhoff [47] proposed that the basics of the system identification is to model the system. Ljung [48] defined that the system identification uses the input and output data to select one better model from another according to the certain criteria. All the definitions emphasise the importance of four elements: input data, output data, the model and the standard.

Usually, the methods can be divided into linear and nonlinear system identification according to the property of the system. For linear systems, the origination can be dated back to the late 1960s and the development is along these two directions. The first direction is named as prediction error framework, which means the system model comes from the minimum difference between the practical and simulated outputs. Åström [49], [50] et al. introduced the Maximum Likelihood principle to improve the modelling techniques. Ljung [51] summarised the linear system technique to be a similarity issue. Schoukens [52] used the Maximum Likelihood principle to identify the system in the frequency domain. The second direction, instrumental variable method, was developed almost at the same time, which is owing to the output based system identification of Ho and Kalman [53]. By further developing the output-only technique, Akaike [54], [55] proposed the white stochastic data driven identification method. Viberg [56] and Overchee [57] proposed a magical method, subspace identification, to promote the techniques. Bauer [58] made a survey on

the statistical features of subspace identification to reveal that many researchers had explored this in the last ten years. Generally, the strategies of system identification are nonparametric and parametric identification [59]. The nonparametric strategy means the identification is to estimate the impulse response function, frequency response function, correlation function or power spectral density while the other strategy is to identify a set of parameters associated with the system, both two strategies aim to develop a best model. Gevers [60] made a seminal overview on the system identification techniques. Linear system identification is a discipline that has evolved considerably during the last 30 years and now linear time-invariant system identification is considered to be a reliable source. The system identification techniques are widely used in the control system but they are not frequently applied in machine condition monitoring.

However, linear systems are an exception and nonlinear behaviours are generic in nature. The field of nonlinear system identification has been explored for a long time since Ibáñez [61], Masri and Caughey [62] studied the nonlinear methods.. Nonlinear system dynamics has been studied for a relatively long time and numerous methods have been developed. The methods are classified according to seven categories: linearization, time and frequency domain methods, modal methods, black-box modelling and structural model updating in reference [63]. For years, one way to study nonlinear systems was the linearization approach [64], many efforts have been spent in order to develop theories for the investigation of nonlinear systems identification. Caughey [65], [66] combined stochastic averaging method and the equivalent linearization method to give a good prediction of nonlinear oscillation. Iwan [67] and Mason [68] developed the linearization method based on Caughey's work. Rice [69] describes how the system governed by the non-linear differential equation can be identified. Obviously, the equivalent linearization method cannot identify the strong nonlinear systems. Hence, the concept, equivalent linear systems with random coefficients, was employed to predict the response of the strong nonlinear systems [70]–[73]. The time domain methods utilize the time series to identify the system, while frequency domain methods are correlated to the frequency response function or spectra. One of the milestones of time domain identification is the restoring force surface method, which is introduced by Masri and Caughey [62]. Since Box and Jenkins [74] used the linear variant method based on Auto-Regressive Moving Average (ARMA) to model and predict the system behaviours, Leontaritis and Billings [75], [76] transferred the ARMA to nonlinear field. Lacy and Berstein [77] presented the nonlinear SSI approach. The development of frequency domain methods is later than the time domain. Bendat [78] considered the higher order spectrum to identify the systems and Yasuda [79], [80] et al estimated the system parameters by harmonic balance method. Rice and Fitzpatrick [81], [82] identified both the single and multiple degree of freedom nonlinear systems through the

reverse path analysis. Another important method in frequency domain is the nonlinear identification through feedback of the output, which is presented by Adams and co-authors [83] and Haroon [84] et al who further developed this method. The combination of time and frequency methods also plays a vital role in the identification field. Staszewski [85] carried out a survey on the use of the wavelet analysis for the nonlinear phenomenon. Rosenberg [86], [87] introduced nonlinear normal mode and applied it in the vibration theory. Bellizzi [88] compared the nonlinear modes from the experiment and calculation to implement system identification. Nonlinear system identification is very active now and large quantities of scholars are making great efforts in this field.

The last part of system identification is the black box modelling. Recently, machine learning has greatly expanded the capabilities of empirical modelling and these new approaches in this field are referred to data driven modelling. As the name suggests, data driven modelling is to identify the black box systems by only using the input and output variables. Many black-box methods have been developed, such as polynomial regression [89], [90], neural networks [91]–[94], fuzzy logic [92], [95]–[97], support vector machine [98]–[100]. The black box modelling technique has attracted an increase in attention in machine condition monitoring.

System identification serves as an effective and efficient tool for understanding the system dynamics. The applications of system identification in machine condition monitoring are not as common as the control system. Wei, Jia and Liu [101] proposed a model based fault detection of railway vehicle suspensions by only using the acceleration signals. Jesueesk and Ellermann [102] utilised multiple Kalman filter to estimate the working conditions of the full-scale railway vehicle suspension system. Liu and co-authors [103] modelled the rail vehicle suspension system based on a recursive least square filter and lead to the effective fault diagnosis technique. Liu et al. [104] classified the correlation signal as several subset and then identified it by the stochastic subspace identification to monitor the suspension conditions. The employment of system identification in another perspective is the deployment of the resonance. In most cases, the resonant behaviour is not desired for machines and the resonance may cause the failure of machines and even catastrophic accidents. The natural frequencies of mechanical systems are the inherent properties, which are independent on the excitations. Generally a machine has numerous natural frequencies and only the first several frequencies are considered because the high orders of the modes locate in a very high frequency range. Although most machines are designed to escape the resonance, the excitation of the modes is not avoidable. The resonant behaviour is a profitable carrier in vibration signals from rotating machines, which gives a valuable chance for effectively monitoring the incipient failure. Wang [105] presents a resonance demodulation technique to extract the tooth

crack fault features from the residual signals, which are obtained by subtracting time synchronous averaging signals from raw vibration signals. The method is superior to the conventional meshing components and leads to a robust indicator of gear tooth cracks. The gear meshing resonance is studied further to demonstrate that the resonant demodulation can separate the compound localised faults on both gear tooth and rolling bearings [106]. The reason behind this is the bearing faults impact on bearing resonance that would modulate the meshing components because the input for resonant modulation in rolling bearings is exclusive of the meshing impacts. Wang, Chu and Han [107] developed a resonant modulation based filtering algorithm to extract the gear failure information from the strong background noise. Tong et al. [108] gave an insight of the meshing resonance in gearboxes and developed a meshing impact energy distribution approach to select the optimal frequency band for resonant demodulation. The resonance modulation is an important phenomenon in machine vibrations and the resonant demodulation gives the potential to achieve early fault detection and diagnosis. However, the effectiveness and efficiency of early fault detection and diagnosis significantly relies on emerging signal processing methods. Although large quantities of advanced methods have been developed, most of the research activities focus on the impulsive characteristics of the output vibration signals but neglect the system itself. For example, most of the research activities in bearing fault diagnostics make great contributions to locating the most impulsive frequency band of the vibration signals and little research has been found to investigate the influence from the transfer function of the bearing system. System identification is the interface between the real world and the mathematical world. This technique is of great help to understand the dynamics of machine systems. The system outputs are the interference with the responses, which makes the fault detection and diagnostics more difficult. In order to gain the insight of systems for the purpose of efficient fault detection and diagnoses, system identification is employed to further understand the dynamics of the machines. With the reference of system identification, CM can be more effective and more reliable in the field of fault detection and diagnoses. The bearing fault detection and diagnosis is based on the resonant modulation, which has been investigated for decades. Most of the research is focused on detecting and enhancing the impulsive behaviour, which is not feasible in early fault detection and diagnosis because the incipient faults induced impulsive responses are fully submerged into strong background noise. To solve the problem of early bearing fault detection and diagnosis, research on system identification enhanced resonant modulation is eagerly undertaken in this thesis. ,

Chapter 3 Modulation Mechanisms in Rotating Machines

As the input of the system varies from stationary to nonstationary due to the working principle of machines, this chapter starts from the general modulation signals and then discusses the resonant modulation responses under the excitations of the periodic, aperiodic, and quasi-stationary input. The modulation mechanism can pave the way to the effective approaches for the fault detection and diagnosis.

3.1 Modulation Signals

Modulation is a general phenomenon in the vibration signals of rotating machines. The modulation of vibration signals usually includes amplitude modulation, frequency modulation and phase modulation. Phase modulation is the angular deviation from the linearly phase of the carrier frequency and thus the derivation of the phase modulation is the frequency modulation. Consequently, the vibration signals would be discussed from the Amplitude Modulation (AM) and Phase Modulation (PM).

Generally, the amplitude of the carrier signal varying with time is known as the amplitude modulation. Figure 3-1 displays an example of the amplitude modulation signal. The instantaneous amplitude of the carrier signal in Figure 3-1 (b) varies as the modulating signal in Figure 3-1 (a), leads to the amplitude modulation signal in Figure 3-1 (c). The spectrum in Figure 3-1 (d) can show the components clearly.

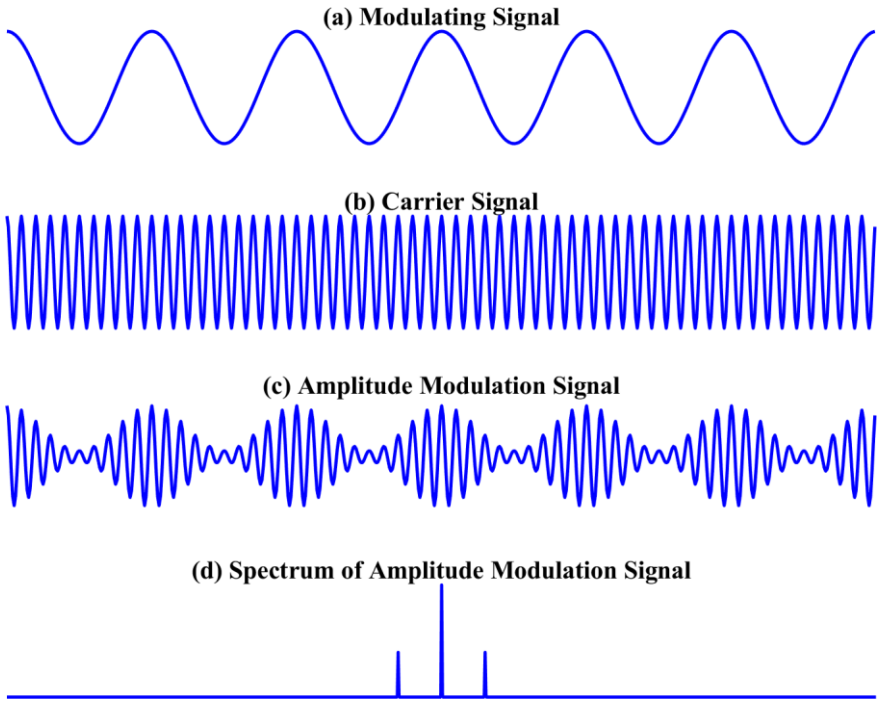


Figure 3-1 Example of amplitude modulation

Another type of modulation signals is the phase modulation and the instantaneous phase of the carrier signal which does not linearly increase with the time but varies regularly. The example of a phase modulation signal is shown in Figure 3-2. As shown in Figure 3-2 (c), the phase modulation signal shows a different manner and the spectrum in Figure 3-2 (d) shows the phase modulation signal has numerous sidebands with the interval of modulating frequency.

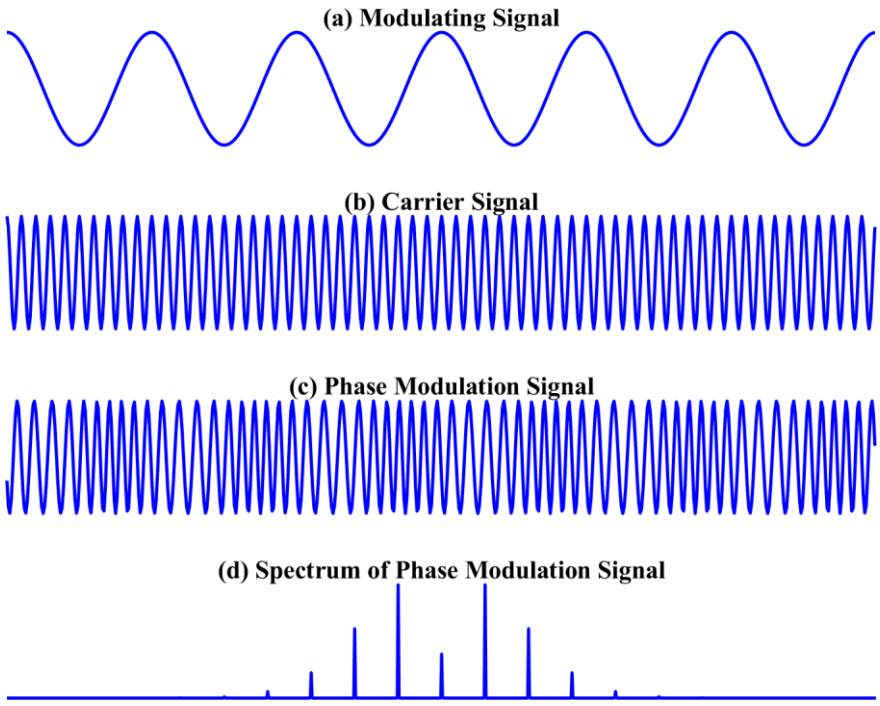


Figure 3-2 Example of phase modulation

3.2 Modulation in Linear Systems

The modulation mechanism in rotating machines is based on the definition of the AM and PM. Usually the vibration signals are collected on the surface of the machine surfaces and the responses collected by the accelerometers which actually pass the system transfer function and are briefly demonstrated by Figure 3-3.

Figure 3-3 Responses of mechanical systems

Consequently, the vibration signals are dependent on the transfer function and the input excitations. With the hypothesis of linear systems, the response $y(n)$ of a linear system to the input $x(n)$

$$y(n) = \mathcal{H}\{x(n)\} \tag{3.1}$$

where, $\mathcal{H}\{\cdot\}$ is the transfer function of the system.

The Equation (3.1) can be rewritten as a convolution between the impulse response function $\mathrm{h(n)}$ and the input $x(n)$.

$$y(n) = h(n) * x(n) \tag{3.2}$$

Here the discussion of the modulation mechanisms is based on the example a linear system with three Degree of Freedom (DOF) and the Frequency Response Function (FRF) of the system is shown in Figure 3-4.

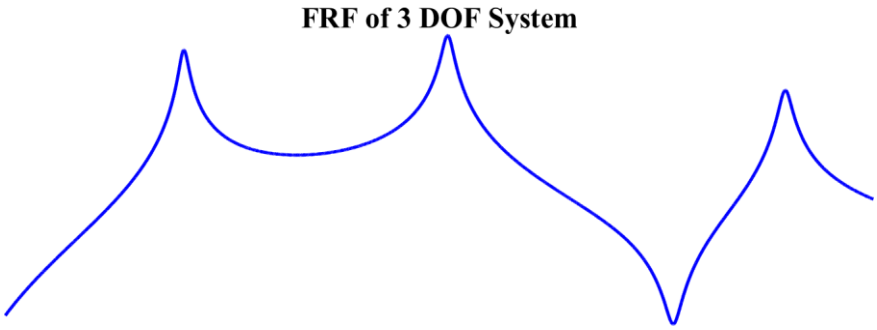


Figure 3-4 FRF of the 3 DOF system

Once the input matches with the natural frequencies of the system, the cumulation of the energy can lead to the great amplitude around the natural frequencies giving the mechanical resonance. In most cases, the resonance is not a desired behaviour for machines and the resonance may cause the failure of machines and even catastrophic accidents. The natural frequencies of mechanical systems are the inherent properties, which are independent on the excitations. Generally a machine has numerous natural frequencies and only the first several frequencies are considered because the high orders of the modes locate in a very high frequency range. Although most of machines are designed to escape the resonance, the excitation of the modes is rarely avoidable. Especially the defects induced input variation may excite the system modes. Therefore, the resonant responses give a valuable chance to enlarge the weak fault information of machines for effective condition monitoring.

As shown in Figure 3-5, the input of the rotating mechanical system can be divided into three groups. For rotating machines under constant speeds and loads, the input of the system can be classified into periodic and cyclostationary signals. The periodic signal is part of the stationary signals. and the cyclostationary signals are within the range of nonstationary signals. The cyclostationary signals have high order statistical properties and vary cyclically with time. The cyclostationary signals can be classified into the impulsive and quasi-stationary signals. The impulsive cyclostationary signals usually indicate localised early faults of rolling element bearings. The quasi-stationary signals relate more to the tribological responses from the interaction between oil films and micro-surface asperities. In this thesis, the periodic and aperiodic impulsive excitations are the main investigation contents in the fault detection and diagnosis of rolling element bearings.

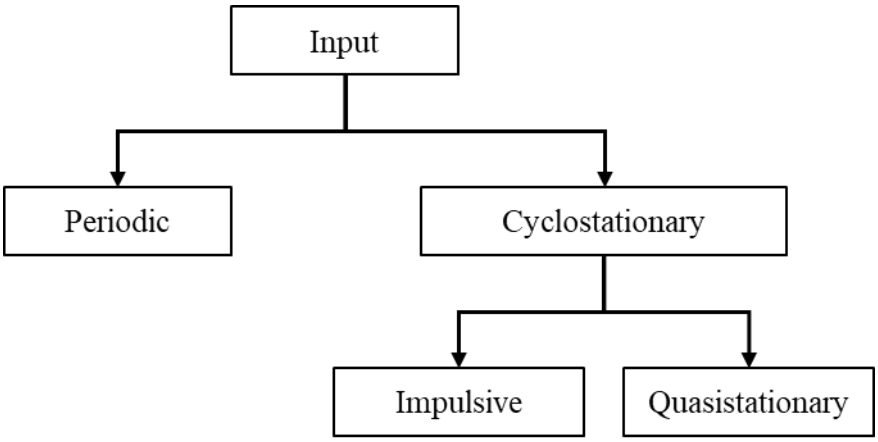


Figure 3-5 Input types in rotating machines

3.2.1 Periodic Input

In the first case, the periodic input is discussed here. The periodic input is quite straightforward for understanding the modulation mechanism. The periodic impacts generated by the rotation of rotors are far from perfect in illustrating sinusoid waveform, hence these periodic signals are composed by a series of harmonics explained by the Fourier theory.

$$\begin{aligned} x(n) &= a_0 + \sum_{k=1}^{N}\left(a_k \cos\frac{2\pi k}{L}n + b_k \sin\frac{2\pi k}{L}n\right) \\ &= a_0 + \sum_{k=1}^{N}\left[\sqrt{{a_k}^2 + {b_k}^2}\cos\left(\frac{2\pi k}{L}n - arctan\frac{b_k}{a_k}\right)\right] \end{aligned} \tag{3.3}$$

Figure 3-6 (a) shows the example of a periodic signal in the frequency domain. Usually the amplitude of the harmonics in vibration signals decrease. In rare circumstances, the high order harmonics could have high amplitude, it depends on the characteristics of the input forces or torques. If the target system has the characteristics of the FRF in Figure 3-6 (b), the output obtained from external surface is displayed in Figure 3-6 (c), which has been transformed into the frequency domain.

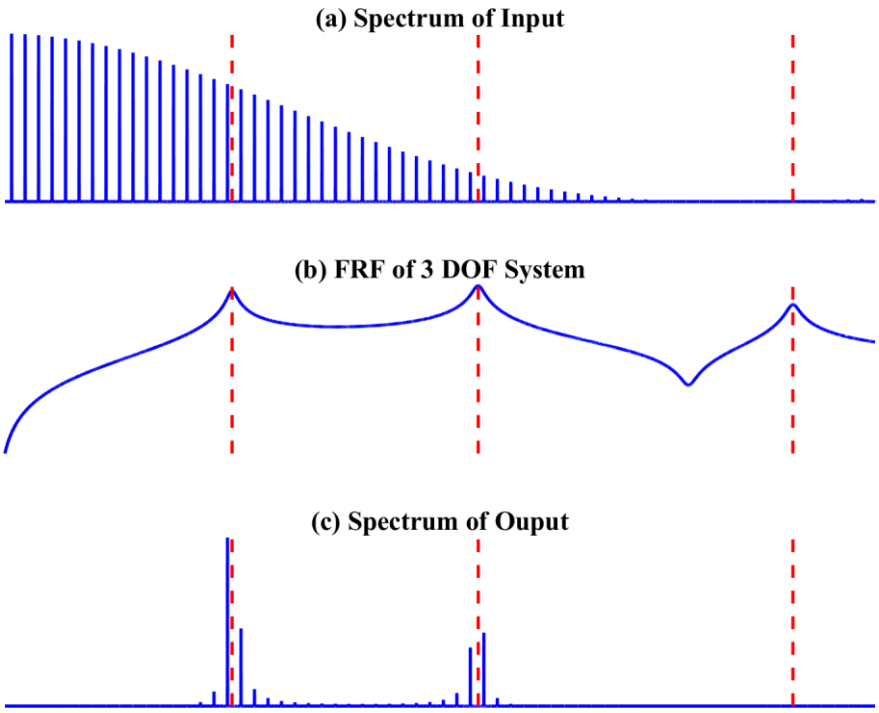


Figure 3-6 Periodic input based modulation

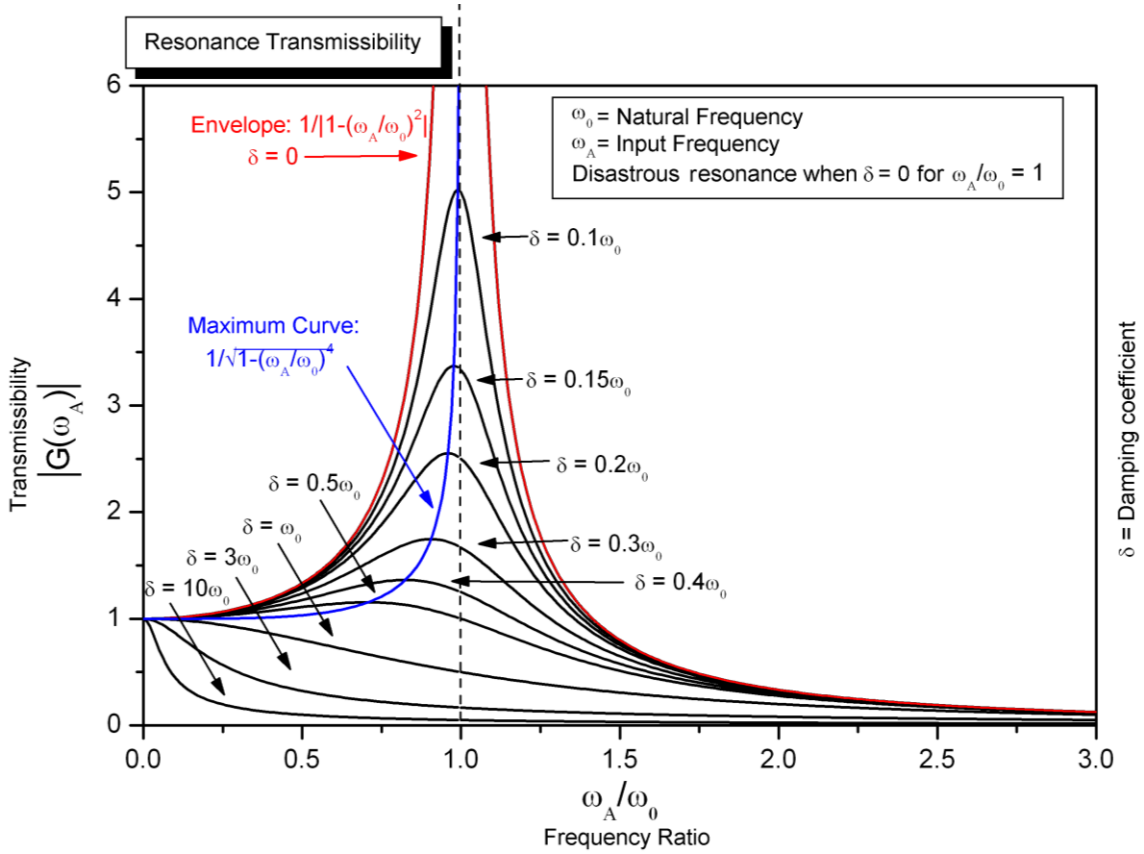


Figure 3-7 Resonance transmissibility [109]

The FRF denotes the transmissibility of the system in a certain degree. As shown in Figure 3-7, although the harmonics in the low frequency range have high amplitude, the low transmissibility of the system results in the low amplitude output. When the harmonics are close to the resonant

frequency, the responses show a tendency towards the pronounced amplitude. The vibration signals in the resonant area result in the modulation phenomenon. The natural frequencies become the carrier frequency and the periodic fault frequency denotes the modulating frequency. Not all the natural frequencies are a good carrier for the modulation signals because the harmonics in certain natural frequencies are nearly zero and unable to excite the resonant behaviour. For instance, the third mode in the Figure 3-6 is not obvious in the output while the first two natural frequencies are the good candidates. The energy in this output responses is focused in the range of two resonant frequencies.

The components in low frequency range are usually contaminated by the severe background noise, and are difficult to extract. The resonant area can enhance the periodic input and the demodulation from the limited frequency band can obtain the features with high signal to noise ratio (SNR). Although the demodulation analysis upon the resonant modulation has already utilised the most beneficial part of the vibration signals, the incipient faults such as a tiny fatigue pitting on the gear tooth surface results in very weakly impulsive impacts. These periodic impacts within the resonant zones are difficult to distinguish from the strong background noise. The high capacity of noise suppression is strongly desired for effective fault detection and diagnosis in an early stage.

3.2.2 Approximately Periodic Impulsive Input

The periodic input is an ideal scenario in mechanical dynamics. In most cases, the input of the mechanical system is far from the perfect periodic signal. These excitation sources are about periodic, terminologically named as the cyclostationary signals. The most typical case is the bearing vibration signals. If bearings run at purely rotation, the localised defects can generate periodic signals. However, the bearings are not working as designed and the rolling elements running in and out of the loading zone cause some random slippages. The random slippage destroys the periodicity of the excitation forces and results in a series of approximately periodic impacts. The system input and output in the time domain are shown in Figure 3-8. The temporal signal seems periodic but it is actually a nonstationary signal.

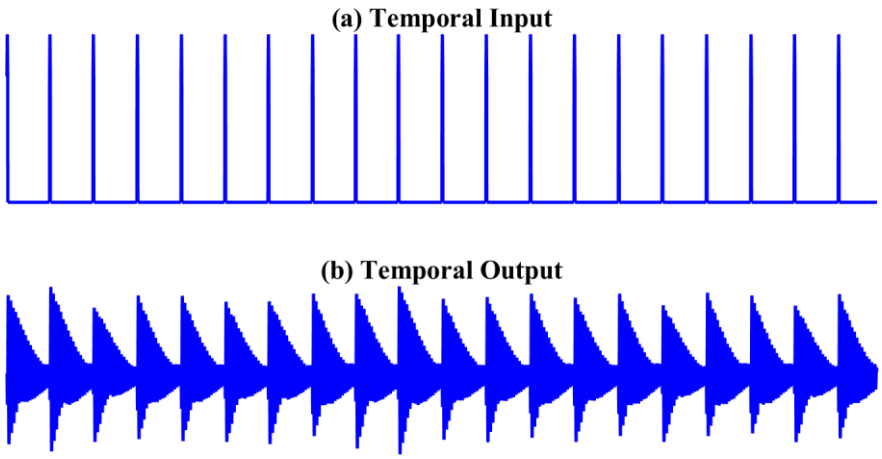


Figure 3-8 Aperiodic impulsive signals in the time domain

The characteristics of the signal can be demonstrated more clearly in the frequency domain. The spectrum of the input and output signals are shown in Figure 3-9. If the nonstationary signal is explained by the Fourier transform, the amplitude of high order harmonics significantly decreases. The responses are shown in Figure 3-9 (c) and the discrete components in the spectrum disappears due to the impulsive cyclostationary input. The components around the first and second natural frequencies are taking the main energy of the vibration signals, which doesn't show sparse frequencies but two clusters in the frequency domain.

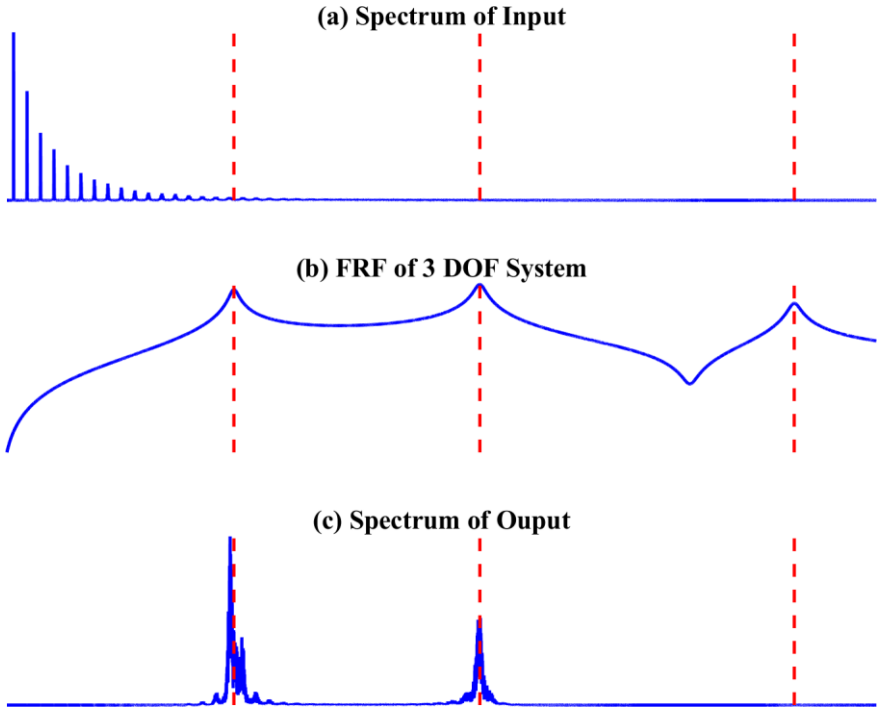


Figure 3-9 Aperiodic impulsive signals in the frequency domain

The first order statistics don't vary periodically. The second order statistic parameters, for example spectral correlation analysis, can disclose the discrete cyclic frequencies. The integral of the autocorrelation functions along the frequency axis can reveal the cyclostationarity of the vibration

signals. The envelope analysis has been verified to be equivalent to envelope analysis in a certain degree[110], which is much more straightforward than the spectral correlation analysis. The envelope spectrum of the vibration around the second natural frequency is shown in the Figure 3-10.

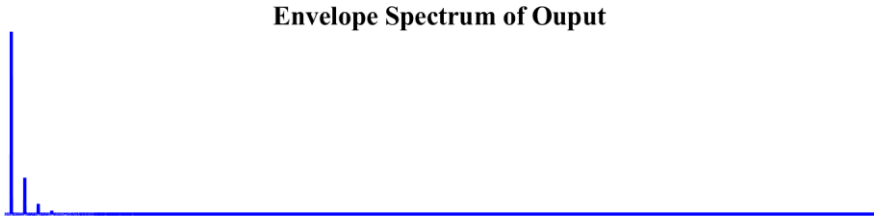


Figure 3-10 Envelope spectrum of the output

The vibration signals obtained from a rolling bearing with a small localised defect are typically impulsive cyclostationary. The localised defect on bearings at early stages generates a series of weak impacts induced by the rolling elements passing the fault[1]. The nonstationary inputs result in the nonstationary output. The responses around the natural frequencies in the example is noise-free and in practical conditions, the transmission path is long and the responses acquired are significantly low SNR. The noise reduction is expected in effective condition monitoring approaches.

3.2.3 Quasi-stationary Input

Another typical input in machines is the tribological behaviour. The lubrication joint in machines are designed to build a perfect oil film so that the mating surfaces can be fully separated and also the oil film is not too thick to consume the extra energy for overcoming the friction.

Properties of surface topography have a significant influence on friction and lubrication. Surface profiles in the lubricated conjunctions are the random deviation from the nominal surface. The two dimensional topography have two main properties, including roughness and waviness. Roughness is the fluctuation of the micro-surface, which is usually represented by the distribution of asperities with varying amplitudes and spacing. Asperities are considered as the main excitation sources in tribological responses. Waviness is also the fluctuation of the surface but it describes the profile with a long wavelength. Waviness on surfaces usually relates to the unstable machining process, for instance the vibration or even chatter.

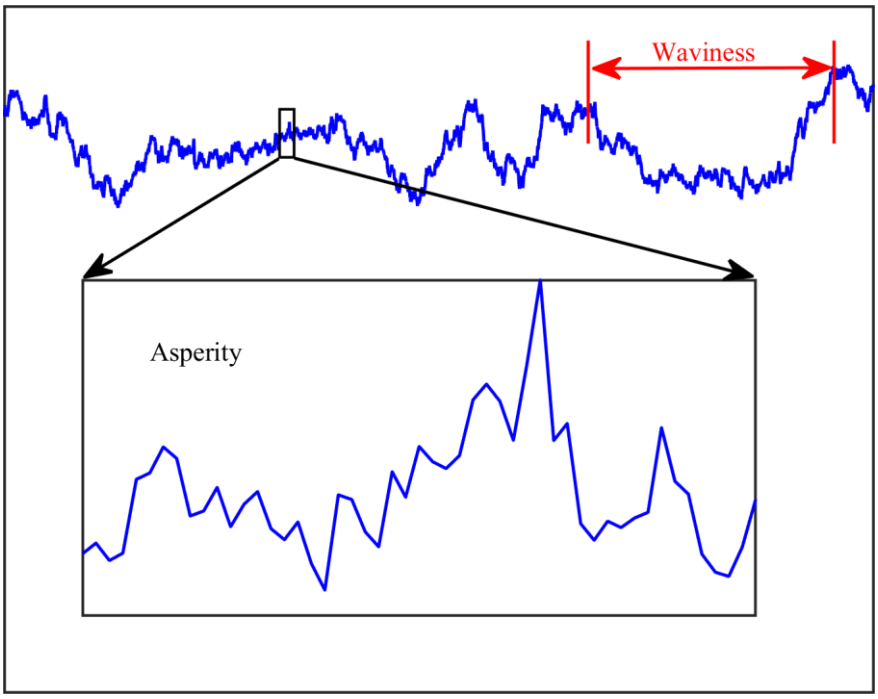


Figure 3-11 Surface profile: waviness and asperities

Suppose the surfaces textures are both homogeneous and isotropic, the interaction between the random asperities and oil film generates stationary excitations. Due to the waviness in the profile, the random stationary excitations become cyclostationary signals, which is demonstrated by the example in Figure 3-12 (a). The cyclostationary signal in Figure 3-12 (a) is an amplitude modulated white noise by a sinusoidal waveform. The output of the system excited by the input is displayed in Figure 3-12 (b), which is still a cyclostationary signal.

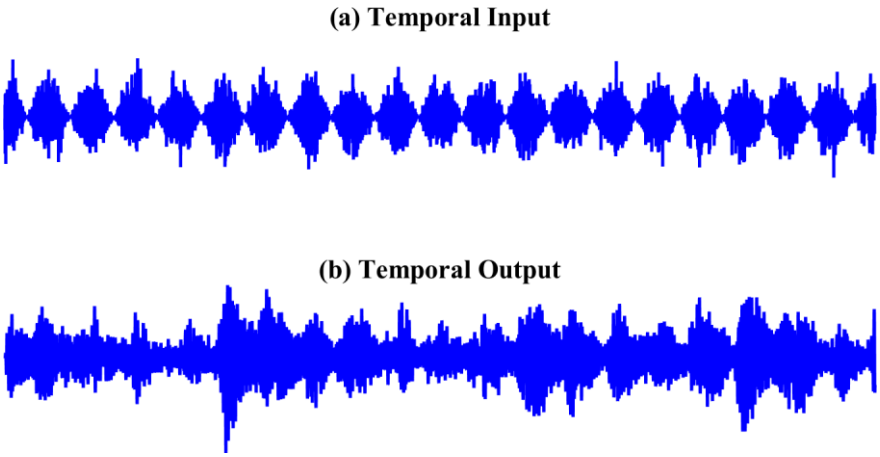


Figure 3-12 Quasi-stationary signals in the time domain

To demonstrate the characteristics of the modulation signals, the input and output signals are converted into the frequency domain by the Fourier transform. The input signal in Figure 3-13 (a) shows a flat spectrum and the cyclic properties cannot be identified by the Fourier analysis. By multiplying with the FRF of the 3 DOF system, the output response is shown in Figure 3-13 (c). The resonant modulation based on the natural frequencies displays three clusters in the whole

spectrum. The other frequency bins are not pronounced and hence easily submerged into the strong background noise.

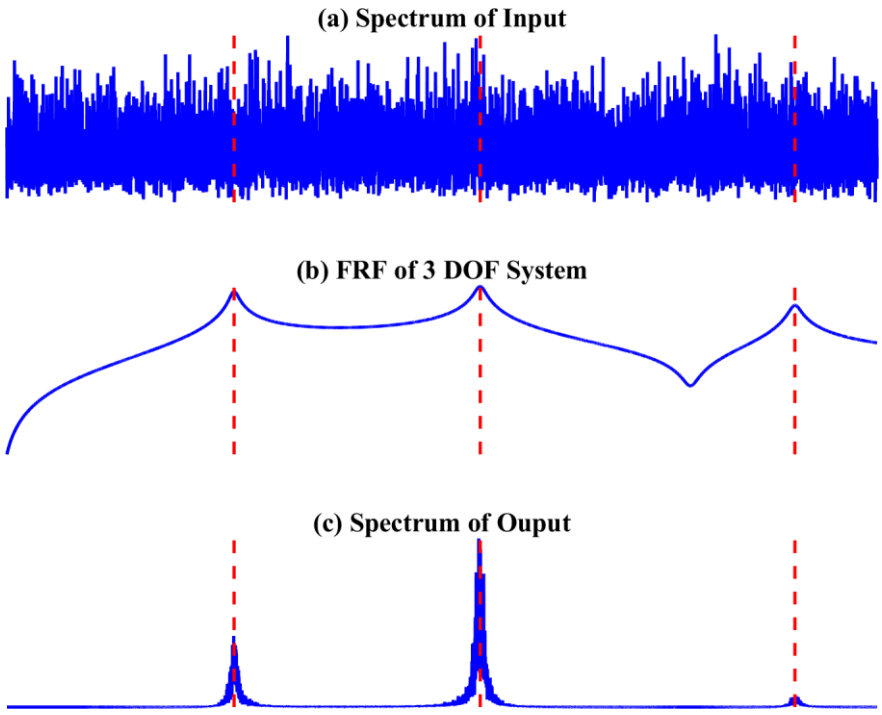


Figure 3-13 Quasi-stationary signals in the frequency domain

The cyclostationary excitations are induced by the tribological behaviour and therefore the strength of the resonance can indicate the lubrication conditions if the responses obtained have a high signal to noise ratio. The properties of the quasi-stationary outputs and random noise are difficult to distinguish, hence noise reduction is not easily achieved when monitoring the lubrication conditions.

3.3 Summary

This chapter describes the resonant modulation in rotating machines. Based on the hypothesis of a linear system, the output of the systems vary due to the different input excitations. Briefly the input can be divided into three types: periodic, aperiodic impulsive, and quasi-stationary. Consequently, the responses from the system are of different modulation characteristics. The effective demodulation analysis of these responses is the most beneficial part for utilising the resonant modulation in vibration based condition monitoring. The characteristics of the vibration signals are expected to be effectively analysed by different signal processing approaches for the early fault detection and diagnosis. This thesis is based on the periodic and aperiodic impulsive signals and then three novel signal processing methods are developed for achieving the condition monitoring of rolling element bearings at a very early stage, which means that the developed methods can detect and diagnose bearing faults under an extremely low SNR (<-30dB).

Chapter 4 Simulation Study of Ball Bearing Fault Detection and Diagnosis

This chapter focuses on investigating resonant modulation induced by the periodic impacts through the simulation studies. The state space model of the rolling element bearings was established to simulate the dynamic responses of rolling bearings. The stochastic subspace identification (SSI) was introduced to serve as the method to determine the optimal central frequencies. The robustness of the SSI method was examined by large quantities of Gaussian and non-Gaussian noise. At the same time, a signal processing method, ensemble average of autocorrelation signals (EAAS), was developed to demodulate the resonant modulation for detecting and diagnosing the bearing faults at early stage. Different levels of white Gaussian noise were added to evaluate the performance of the proposed demodulation method. The capability of the SSI and EAAS is benchmarked by the Kurtogram and envelope analysis, respectively.

4.1 Model Development

4.1.1 Characteristic Frequencies of Bearing Faults

As demonstrated in Figure 4-1, a rolling element bearing is comprised of an inner race, an outer race, rolling elements and a cage which holds the rolling elements in a given relative position. If a spall occurs on one of the races, a periodic impulse will be excited by the rolling elements, which presents the fault characteristics determined by the dimensions of the bearing, the fault parts, and the rotating speed. The characteristic fault frequencies of rolling element bearings are determined by the shaft rotating speeds and the bearing dimensions, which is deduced from the relative speeds between bearing elements. For a fixed outer race bearing, Equations (4.1) to (4.4) can be used to calculate the theoretical fault frequencies [22].

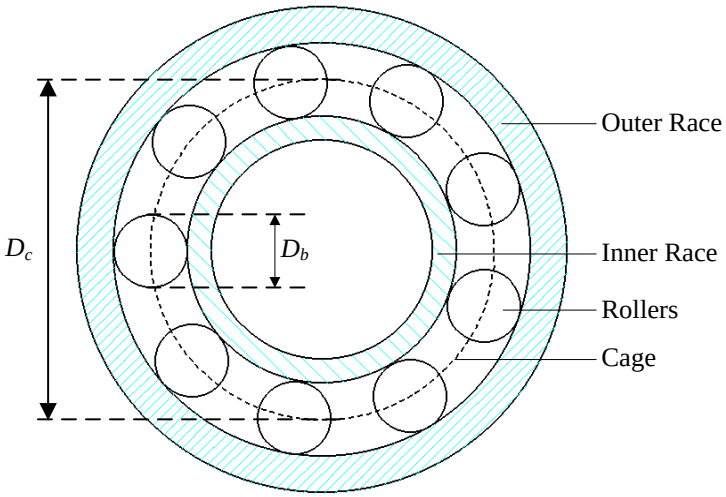


Figure 4-1 Schematic diagram of the rolling element bearing

Characteristic frequency of outer race faults:

$$f_o = \frac{N_r}{2} f_r (1 - \frac{D_b}{D_c} cos\, \varphi) \quad (4.1)$$

Characteristic frequency of Inner race faults:

$$f_i = \frac{N_r}{2} f_r (1 + \frac{D_b}{D_c} cos\, \varphi) \quad (4.2)$$

Characteristic frequency of roller faults:

$$f_b = \frac{D_c}{2D_b} f_r (1 - (\frac{D_b}{D_c} cos\, \varphi)^2) \quad (4.3)$$

Characteristic frequency of cage faults (often called the fundamental train frequency):

$$f_{cage} = \frac{1}{2} f_r (1 - \frac{D_b}{D_c} cos\, \varphi) \quad (4.4)$$

where, N_r is number of balls, f_r is the shaft rotating frequency, D_b is the roller diameter, D_c is the pitch circle diameter, and φ is the contact angle. Practically, the characteristic fault frequencies

for a certain bearing has only one critical variable, the rotating speed of the shaft as the dimensions of a certain bearing are determined.

4.1.2 State Space Model of Bearing Vibration

Vibration induced by the rolling elements passing raceway defects is usually of amplitude modulation signal [111]. The impacts from the localised faults are considered as a series of impulses. If the rotating speed of the shaft is constant, the impacts become a periodic signal. These impacts usually spread over a wide frequency range, which can excite the resonance of bearing systems and hence result in modulation phenomenon between periodic impacts and system resonances. The dynamic responses of rolling element bearings in this thesis are simulated by a state space model with three degree of freedom (DOF). Figure 4-2 (a) shows the lumped model with three DOF for simulating the dynamic responses. The model in Figure 4-2 (a) can be concisely depicted by the model in Figure 4-2 (b), which is a general representation of lumped models.

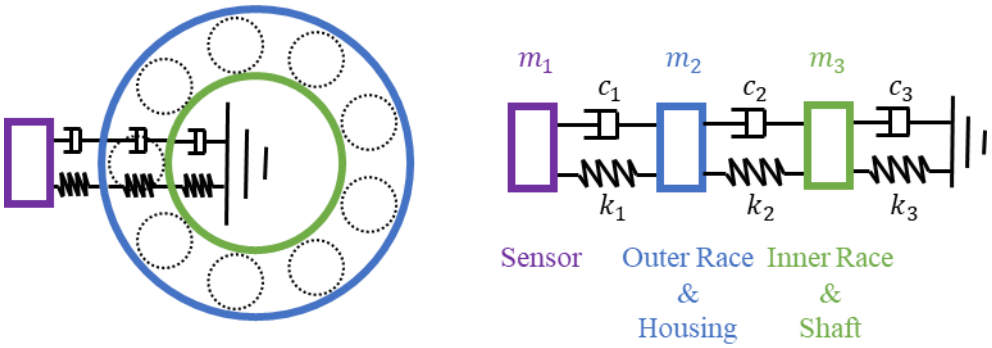


Figure 4-2 Schematic diagram of the bearing model: (a) schematic bearing; (b) simplified lumped model

The bearing system can be simplified as a three DOF model: the sensor, the stationary part (the bearing housing and the outer race), and the rotating part (the shaft and the inner race). Consequently, this three DOF model can be expressed by a common vibration equation as

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{C}\dot{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{x} = \boldsymbol{F} \tag{4.5}$$

where, $\boldsymbol{x} = [x_1 \quad x_2 \quad x_3]^T$ is the displacement of each DOF; $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ are the mass matrix, damping matrix and stiffness matrix respectively. The expresses of these system matrixes are shown as follows:

$$\boldsymbol{M} = \begin{bmatrix} m_1 & & \\ & m_2 & \\ & & m_3 \end{bmatrix} \tag{4.6}$$

$$\boldsymbol{C} = \begin{bmatrix} c_1 & -c_1 & \\ -c_1 & c_1 + c_2 & -c_2 \\ & -c_2 & c_2 + c_3 \end{bmatrix} \tag{4.7}$$

$$\boldsymbol{K} = \begin{bmatrix} k_1 & -k_1 & \\ & k_1 + k_2 & -k_2 \\ & -k_2 & k_2 + k_3 \end{bmatrix} \tag{4.8}$$

Suppose the rolling elements purely roll on the raceways and consequently, the impacts of rolling elements passing the defects become the main inputs of the bearing model. Hence, the forces **F** on the bearing model are

$$\boldsymbol{F} = [0 \quad f(t) \quad -f(t)]^T \tag{4.9}$$

where, $f(t)$ is the impact forces induced by the collisions between rolling elements and raceways. In most circumstances, the outer ring of bearings is fixed, and the inner ring is the rotating part. If the defect occurs on the outer ring, the impact forces are nearly constant, which can be expressed as

$$f(t) = \sum_{k=1}^{+\infty} A_f \delta\left(t - kT_f\right) \tag{4.10}$$

where, A_f is the amplitude of the impact forces from the rolling element passing defects; T_f is the characteristic time cycle of outer race faults; k denotes the number of the successive impacts during the bearing rotation. $\delta(\cdot)$ is the delta function, which is defined as

$$\delta(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{4.11}$$

It is well known that the load distribution of bearings is not even along with the circumference and the loaded zone is shown in the Figure 4-3.

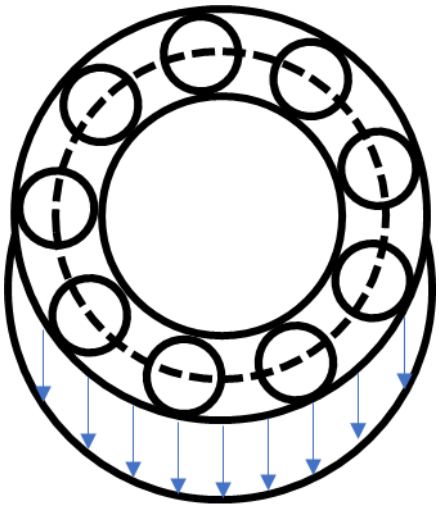


Figure 4-3 Loaded zone of rolling element bearings

Consequently, the impacts induced by the inner race faults would not be constant and will vary according to the relative position to the loaded zone. The variation of the impact strength is described by the rotation of the inner race, which is actually the same as the shaft rotation. The impact forces can be expressed as

$$f(t) = \sum_{k=1}^{+\infty} A_f \left| cos\left(2\pi \frac{f_r}{2} t + \varphi_f\right)\right| \delta\left(t - kT_f\right) \tag{4.12}$$

where, f_r is rotating frequency of the inner ring; φ_f is the initial phase upon the collision position to the loaded zone.

Therefore, the impact forces can be summarised as

$$f(t) = \begin{cases} \sum_{k=1}^{+\infty} A_f \delta\left(t - kT_f\right), & outer\ race\ faults \\ \sum_{k=1}^{+\infty} A_f \left| cos\left(2\pi \frac{f_r}{2} t + \varphi_f\right)\right| \delta\left(t - kT_f\right), & inner\ race\ faults \end{cases} \tag{4.13}$$

The state space model is a common representation of a linear system by first-order differential equations. It is a time domain approach to conveniently and compactly describe system dynamics with multiple inputs and outputs. Owing to the merits, the state space model is widely used in many different areas. This method is employed to establish the bearing dynamic model. The general expression of state space models is denoted as

$$\begin{cases} \dot{x}(t) = \boldsymbol{A_s} x(t) + \boldsymbol{B_s} u(t) \\ y(t) = \boldsymbol{C_s} x(t) + \boldsymbol{D_s} u(t) \end{cases} \tag{4.14}$$

where, $x(t)$ is the state vector; $y(\mathrm{t})$ is the output vector; $u(\mathrm{t})$ is the input vector; $\mathbf{A_s}$ is the system matrix; $\boldsymbol{B_s}$ is the input matrix; $\boldsymbol{C_s}$ is the output matrix; and $\boldsymbol{D_s}$ is the feedthrough matrix.

The Equation (4.14) is a common expression and in the bearing model, the feedthrough matrix is omitted as there is no feedback in the bearing dynamics. The parameters in the state space model are

$$\boldsymbol{A_s} = \begin{bmatrix} \mathbf{0} & \boldsymbol{I} \\ -\boldsymbol{M^{-1}K} & -\boldsymbol{M^{-1}C} \end{bmatrix} \tag{4.15}$$

$$\boldsymbol{B_s} = \begin{bmatrix} \mathbf{0} \\ -\boldsymbol{M^{-1}} \end{bmatrix} \tag{4.16}$$

$$\boldsymbol{C_s} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{4.17}$$

All the parameters used in the state space model of bearing vibrations are listed in the Table 4-1. This model is simplified from Rehab's thesis [112] and the parameters used in this simulation are also based on the Rehab's thesis.

Table 4-1 Model parameters

Parameters	Values
m_1	0.1kg
m_2	2.0kg
m_3	1.0kg
k_1	2.0e8N/m
k_2	3.6e9N/m
k_3	1.0e9N/m
c_1	89.4Ns/m
c_2	1.0e3Ns/m
c_3	1.1e3Ns/m

Based on the parameters used in the model, the modal characteristics can be shown by the Frequency Response Functions (FRFs). It can be seen that the natural frequencies of the bearing model are 3468Hz, 7117Hz and 11781Hz, which are potential carrier frequencies of the bearing defects induced modulation signals.

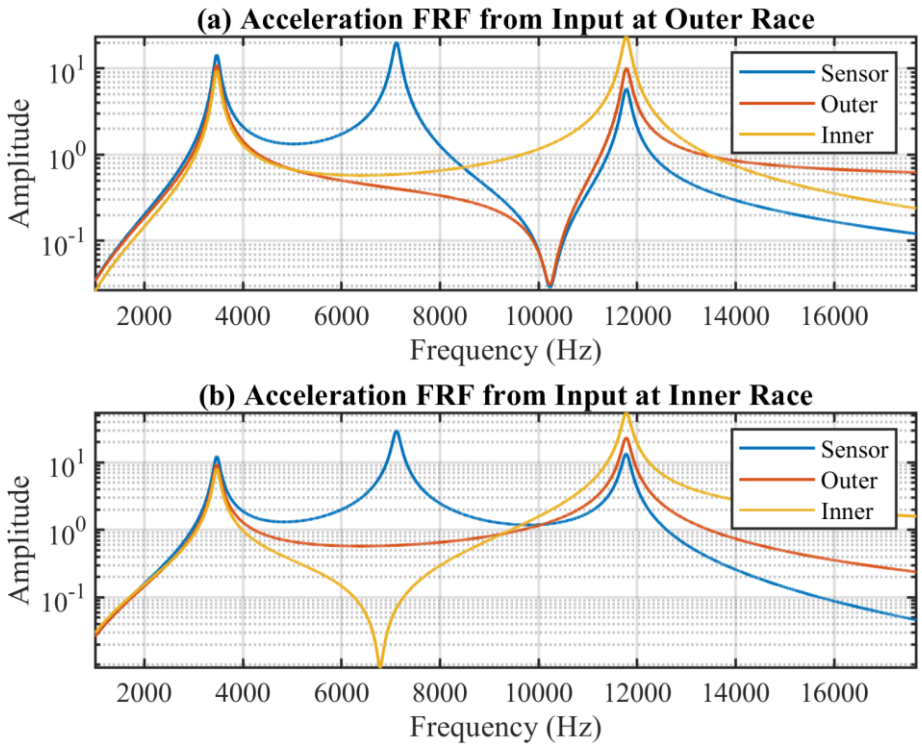


Figure 4-4 FRFs of the bearing model

4.2 Method to Determine Optimal Frequency Bands

The modulation characteristics in the rolling element bearing systems are that the resonant responses are modulated by the periodic impacts of rolling elements passing the localised defects. The modulation characteristics lead to the low frequency components shifting from low frequency range to high frequency range [113]. The High Frequency Resonance Technique (HFRT), later called envelope analysis, was developed by Darlow et al. [10] and demonstrated an outstanding demodulation capability. Envelope analysis in the whole frequency band includes some noise, which does not guarantee the optimal detection and diagnosis of the localised defects. To effectively demodulate the fault signatures from the vibration signals, the determination of the optimum frequency bands become the primary problem.

Based on the recursive impulse characteristics, numerous researchers are attracted to develop various approaches to select optimal frequency bands. The most commonly used method is Kurtogram, which aims to point out the band pass filtered signals with the most obvious impulse events. Most of the research works focus on finding the frequency bands which contains the impulse behaviour. Although these research activities have achieved significant outcomes in bearing fault diagnostics, the modulation characteristics are not considered during the development of the approaches. Therefore, the system identification based approach is developed in this thesis to find the optimal carrier frequencies for further demodulation analysis. There is a specific research topic in the field of system identification, which is named as Operational Modal Analysis (OMA). OMA can extract the modal parameters only from the outputs of the system. Stochastic Subspace Identification (SSI), used for over a decade, is a high-performance method to identify the resonance behaviours of the system, hence this method is introduced to determine the optimal frequency bands for further demodulation analysis. The procedure of identification is deduced from the first principle equation [114], [115]. Generally, a n degrees of freedom (DOF) system can be defined as

$$\boldsymbol{M}_s\ddot{x}_s(t) + \boldsymbol{C}_s\dot{x}_s(t) + \boldsymbol{K}_s x_s(t) = \boldsymbol{F}_s(t) = \boldsymbol{B}_s u(t) \qquad (4.18)$$

where, $\boldsymbol{M_s}$, $\boldsymbol{C}_s$and $\boldsymbol{K}_s \in \boldsymbol{R}^{n\times n}$ are the mass, damping and stiffness matrixes respectively, $\ddot{x}_s(t)$, $\dot{x}_s(t)$, and $x_s(t) \in \boldsymbol{R}^{n\times 1}$ are the acceleration, velocity and displacement vectors, $F_s(t) \in \boldsymbol{R}^{n\times 1}$ is the external force which is factorised to be an input space matrix $\boldsymbol{B_s} \in \boldsymbol{R}^{n\times m}$ and an input time vector $u(t) \in \boldsymbol{R}^{m\times 1}$.

With the following definitions

$$x(t) = \begin{bmatrix} x_s(t) \\ \dot{x}_s(t) \end{bmatrix} \qquad (4.19)$$

Equation (4.18) can be transformed to a state equation

$$\dot{x}(t) = \boldsymbol{A}_c x(t) + \boldsymbol{B}_c u(t) \tag{4.20}$$

with

$$\begin{aligned} \boldsymbol{A}_c &= \begin{bmatrix} \boldsymbol{0}_{n\times n} & \boldsymbol{I}_{n\times n} \\ -\boldsymbol{M}_s^{-1}\boldsymbol{C}_s & -\boldsymbol{M}_s^{-1}\boldsymbol{K}_s \end{bmatrix} \\ \boldsymbol{B}_c &= \begin{bmatrix} 0_{n\times m} \\ \boldsymbol{M}_s^{-1}\boldsymbol{B}_s \end{bmatrix} \end{aligned} \tag{4.21}$$

where, $\boldsymbol{A}_c \in \boldsymbol{R}^{2n\times 2n}$ is the state matrix and $\boldsymbol{B}_c \in \boldsymbol{R}^{2n\times m}$ is the input matrix.

Practically, the measurements are finite and only l points of the system are monitored, so the observation equation can be assumed to be

$$y(t) = \boldsymbol{C}_a \ddot{x}_s(t) + \boldsymbol{C}_v \dot{x}_s(t) + \boldsymbol{C}_d x_s(t) \tag{4.22}$$

where, $y(t) \in \boldsymbol{R}^{l\times 1}$ is the outputs, and $\boldsymbol{C}_a$, $\boldsymbol{C}_v$,$\boldsymbol{C}_s \in \boldsymbol{R}^{l\times n}$ are the output matrices of acceleration, velocity and displacement.

Then, Equation (4.22) can be rewritten as

$$y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \tag{4.23}$$

with

$$\begin{aligned} \boldsymbol{C} &= [\boldsymbol{C}_d - \boldsymbol{C}_a\boldsymbol{M}_s^{-1}\boldsymbol{K}_s \quad \boldsymbol{C}_v - \boldsymbol{C}_a\boldsymbol{M}_s^{-1}\boldsymbol{C}_s] \\ \boldsymbol{D} &= \boldsymbol{C}_a\boldsymbol{M}_s^{-1}\boldsymbol{B}_s \end{aligned} \tag{4.24}$$

where, $\boldsymbol{C} \in \boldsymbol{R}^{l\times 2n}$ is the output matrix and $\boldsymbol{D} \in \boldsymbol{R}^{l\times m}$ is the direct transmission matrix.

To sum up, a continuous time deterministic state space model shows as follows,

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}_c x(t) + \boldsymbol{B}_c u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \end{cases} \tag{4.25}$$

In practice, the data sets are acquired at a discrete time interval Δt, therefore, a discrete state space model is assumed to be

$$\begin{cases} x(k+1) = \boldsymbol{A}x(k) + \boldsymbol{B}u(k) \\ y(k) = \boldsymbol{C}x(k) + \boldsymbol{D}u(k) \end{cases} \tag{4.26}$$

where, $x(k)$ is the state vector at $k\Delta t$, $\boldsymbol{A} = e^{\boldsymbol{A}_c\Delta t}$ is the discrete state matrix and $\boldsymbol{B} = [\boldsymbol{A} - I]\boldsymbol{A}_c^{-1}\boldsymbol{B}_c$ is the discrete input matrix.

The stochastic noise is virtually inevitable and the input of a stochastic process is not easy to measure. The characteristics of the input are the same as the noise. Combining the stochastic input and noise, a stochastic and discrete identification can be derived to be presented as [115]

$$\begin{cases} x(k+1) = \boldsymbol{A}x(k) + w(k) \\ y(k) = \boldsymbol{C}x(k) + v(k) \end{cases} \tag{4.27}$$

with

$$E\left[\begin{pmatrix} w(j) \\ v(j) \end{pmatrix} (w(k)^T \quad v(k)^T)\right] = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{S} \\ \boldsymbol{S}^T & \boldsymbol{R} \end{pmatrix} \delta_{jk} \geq 0 \tag{4.28}$$

where, $x(k) \in \boldsymbol{R}^{2n\times 1}$ and $y(k) \in \boldsymbol{R}^{l\times 1}$ are respectively the state and the output of the system at discrete time k; $w_k \in \boldsymbol{R}^{2n\times 1}$ and $v_k \in \boldsymbol{R}^{l\times 1}$ are unmeasurable signals, which are assumed that: (1) The mean value is zero; (2) The signals are stationary; (3) The signals obey the Gaussian distribution. $\boldsymbol{A} \in \boldsymbol{R}^{2n\times 2n}$ is the system matrix; $\boldsymbol{C} \in \boldsymbol{R}^{l\times 2n}$ is the output matrix, the matrixes $\boldsymbol{Q} \in \boldsymbol{R}^{2n\times 2n}$, $\boldsymbol{S} \in \boldsymbol{R}^{2n\times l}$and $\boldsymbol{R} \in \boldsymbol{R}^{l\times l}$ are the corresponding covariance.

The state covariance matrices are defined as

$$\begin{aligned} \sum \quad &= E[x(k+1)x^T(k+1)] \\ &= E[(\boldsymbol{A}x(k) + w(k))(\boldsymbol{A}x(k) + w(k))^T] \\ &= \boldsymbol{A}E[x(k)x^T(k)] + E(w(k)w^T(k)) \\ &= \boldsymbol{A}\sum\boldsymbol{A}^T + \boldsymbol{Q} \end{aligned} \tag{4.29}$$

Where, $\sum$ is independent from time because of the features of the linear time-invariant (LTI) system.

The output covariance matrices are defined as

$$\boldsymbol{\Lambda}_i = E[y(k+i)y^T(k)] \tag{4.30}$$

Where, i is the time delay.

The state-output covariance is defined as

$$\begin{aligned} \boldsymbol{G} \quad &= E[x(k+1)y^T(k)] \\ &= E[(\boldsymbol{A}x(k) + w(k))(\boldsymbol{C}x(k) + v(k))^T] \\ &= \boldsymbol{A}\sum\boldsymbol{C}^T + \boldsymbol{S} \end{aligned} \tag{4.31}$$

Finally, the kernel equation of covariance based SSI can be deduced as

$$\boldsymbol{\Lambda}_i = \boldsymbol{C}\boldsymbol{A}^{i-1}\boldsymbol{G} \tag{4.32}$$

Before implementing the identification method, some definitions are introduced. In consideration of the stochastic inputs, the reference outputs are vital in the identification process. Provided that l outputs are monitored, r references will be selected

$$y(k) = \begin{bmatrix} y^{ref}(k) \\ y^{\sim ref}(k) \end{bmatrix} \tag{4.33}$$

where, $y^{ref}(k) \in \boldsymbol{R}^{r\times 1}$ is the references and $y^{\sim ref}(k) \in \boldsymbol{R}^{(l-r)\times 1}$ is the remaining elements.

The measured outputs of the system are formed into a block Hankel matrix, which is shown as

$$H = \frac{1}{\sqrt{j}}\begin{bmatrix} y^{ref}(0) & y^{ref}(1) & \cdots & y^{ref}(j-1) \\ y^{ref}(1) & y^{ref}(2) & \cdots & y^{ref}(j) \\ \cdots & \cdots & \cdots & \cdots \\ y^{ref}(i-1) & y^{ref}(i) & \cdots & y^{ref}(i+j-2) \\ \hline y(i) & y(i+1) & \cdots & y(i+j-1) \\ y(i+1) & y(i+2) & \cdots & y(i+j) \\ \cdots & \cdots & \cdots & \cdots \\ y(2i-1) & y(2i) & \cdots & y(2i+j-2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}^{ref}_{0|i-1} \\ \hline \boldsymbol{Y}_{i|2i-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}^{ref}_{p} \\ \hline \boldsymbol{Y}_{f} \end{bmatrix} \begin{matrix} \updownarrow ri \ \text{"past"} \\ \updownarrow li \ \text{"future"} \end{matrix} \qquad (4.34)$$

In addition, if the outputs are ergodic, the outputs covariance can be further estimated by

$$\hat{\boldsymbol{\Lambda}}_i = \frac{1}{j}\sum_{k=0}^{j-1} y(k+i)y^T(k) \qquad (4.35)$$

Then, the Toeplitz matrix consisting of output covariance is expressed as follows

$$\boldsymbol{T}_{1|i} = \boldsymbol{Y}_f\boldsymbol{Y}_p^T = \begin{bmatrix} \hat{\Lambda}_i & \hat{\Lambda}_{i-1} & \cdots & \hat{\Lambda}_1 \\ \hat{\Lambda}_{i+1} & \hat{\Lambda}_i & \cdots & \hat{\Lambda}_2 \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\Lambda}_{2i-1} & \hat{\Lambda}_{2i-2} & \cdots & \hat{\Lambda}_i \end{bmatrix} \qquad (4.36)$$

The Toeplitz matrix can be decomposed into two parts

$$\boldsymbol{T}_{1|i} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \cdots \\ \boldsymbol{CA}^{i-1} \end{bmatrix} [\boldsymbol{A}^{i-1}\boldsymbol{G} \quad \cdots \quad \boldsymbol{AG} \quad \boldsymbol{G}] = \boldsymbol{O}_i\boldsymbol{C}_i \qquad (4.37)$$

Both observability matrix $\boldsymbol{O}_i \in \boldsymbol{R}^{li\times n}$ and controllability matrix $\boldsymbol{C}_i \in \boldsymbol{R}^{n\times ri}$ can be obtained after singular value decomposition (SVD) of the Toeplitz matrix

$$\boldsymbol{T}_{1|i} = \boldsymbol{USV}^T = [\boldsymbol{U}_1 \quad \boldsymbol{U}_2]\begin{bmatrix} \boldsymbol{S}_1 & \\ & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \end{bmatrix} = \boldsymbol{U}_1\boldsymbol{S}_1\boldsymbol{V}_1^T \qquad (4.38)$$

where, $\boldsymbol{U} \in \boldsymbol{R}^{li\times li}$ and $\boldsymbol{V} \in \boldsymbol{R}^{ri\times ri}$ are orthonormal matrices; $\boldsymbol{S} \in \boldsymbol{R}^{li\times ri}$ is a diagonal matrix consisting of the singular values in a descending order. Hence, the observability and controllability matrices can be known as

$$\begin{aligned} \boldsymbol{O}_i &= \boldsymbol{U}_1\boldsymbol{S}_1^{1/2} \\ \boldsymbol{C}_i &= \boldsymbol{S}_1^{1/2}\boldsymbol{V}_1^T \end{aligned} \qquad (4.39)$$

Besides, another Toeplitz matrix $\boldsymbol{T}_{2|I+1}$containing the information of the state matrix *A* is easy to obtain according to the definition

$$\boldsymbol{T}_{2|I+1} = \begin{bmatrix} \boldsymbol{CA} \\ \boldsymbol{CA}^2 \\ \cdots \\ \boldsymbol{CA}^i \end{bmatrix} [\boldsymbol{A}^{i-1}\boldsymbol{G} \quad \cdots \quad \boldsymbol{AG} \quad \boldsymbol{G}] = \boldsymbol{O}_i\boldsymbol{AC}_i \qquad (4.40)$$

Since the observability and controllability matrix are clear, the state matrix $\boldsymbol{A}$ can be deduced from Equation (4.40).

$$\boldsymbol{A} = \boldsymbol{O}_i^{\dagger}\boldsymbol{T}_{2|i+1}\boldsymbol{C}_i^{\dagger} = \boldsymbol{S}_i^{-1/2}\boldsymbol{U}_1^T\boldsymbol{T}_{2|i+1}\boldsymbol{V}_1\boldsymbol{S}_1^{-1/2} \tag{4.41}$$

where $(\bullet)^{\dagger}$ is the pseudo inverse of a matrix. As the state matrix is obtained, the system identification goes into the post-processing stage. The characteristics of the system is related to the eigenvalues, which is shown as

$$\boldsymbol{A} = \boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1} \tag{4.42}$$

Where, $\boldsymbol{\Psi}$ is the matrix constituted of eigenvectors and $\boldsymbol{\Lambda} = diag(\lambda_q)$ is the diagonal matrix that is only containing the complex eigenvalues λ_q. Considering the relation between discrete and continuous system $\boldsymbol{A} = e^{\boldsymbol{A}_c\Delta t}$, the eigenvalues and eigenvectors of the continuous system are

$$\begin{cases} \boldsymbol{\Psi}_c = \boldsymbol{\Psi} \\ \lambda_c = \frac{ln(\lambda_q)}{\Delta t} \end{cases} \tag{4.43}$$

Finally, the resonant frequency of the system can be calculated theoretically by

$$f = \frac{\sqrt{(\lambda_c^R)^2 + (\lambda_c^I)^2}}{2\pi} \tag{4.44}$$

Where, λ_c^R and λ_c^I are the real and imaginary part of the complex values λ_c respectively.

The natural frequencies identified from faulty bearing vibration signals are the potential central frequencies of the optimal bands. However, the approach of SSI is easily influenced by many factors, for instance the background noise, the nonlinear characteristics, the nonstationary components, and so on. The undesired factors can lead to the appearance of fake modes, which gives misleading results. Therefore, a method based on the system orders was developed to distinguish real modes. It is well known that the core issue of the system identification is the determination of the correct model order. Theoretically a system has an infinite number of modes but usually only the modes within a certain frequency range or in a certain frequency band are considered in the practical analysis. In order to acquire all the desired modal parameters of the target system, the orders chosen are usually higher to decrease the bias [116]. During the process of identification, the system order increases to select the modes according to the quality criteria, which is the method of Stabilization Diagram (SD). The SD is proved to be a promising tool for the identification of real modes. Consequently, the excited modes in the output signals can be accurately identified and then used for the further demodulation analysis.

Vibration signals from a rolling element bearing with a localised fault contain a series of transient responses and the time interval between the transient responses is the desired information in the diagnostic analysis. The Kurtogram aims to locate the transient impulsive signals based on the kurtosis of signals in the time domain that are filtered by the Short-Time Fourier Transform (STFT)

or Wavelet Packet Transform (WPT). It has been proved useful in the detection and diagnosis of bearing faults. The Kurtogram comes from the spectral kurtosis and it gives a more vivid representation of the two dimensional results. Although Kurtogram is effective under some conditions, the performance is not satisfied in the presence of a low signal-to-noise ratio or non-Gaussian noise, which is a common situation for the incipient faults. In this research study, the Kurtogram is employed to benchmark the SSI based optimal frequency band selection. The comparison of these two methods will be shown in both simulation and experiment studies.

4.3 Method of Ensemble Average of Autocorrelation Signals

Envelope analysis is a popular method in detection and diagnostics of bearing faults. Vibration is the commonly used signal to monitor the bearings but the vibration features, associated with bearing faults, are often submerged by the heavy background noise. Although the envelope analysis fully exploits the merits of high signal to noise ratio around the resonant frequencies, it reaches its maximum potential when tackling the heavily contaminated signals. Therefore, the noise reduction is an essential capability for incipient fault detection and diagnosis.

In reference [117], autocorrelation functions are used to enhance the natural frequencies in real-time so that the bearing fault detection can adapt the changes of significant natural frequencies, corresponding to the change in conditions. Raheem [118] used wavelet transform to increase the signal to noise ration and then calculated autocorrelation functions of wavelet coefficients to extract bearing faults in the time domain. At the same time, Rafiee [119] and co-authors present a novel time–frequency-based feature recognition system for gear fault diagnosis using autocorrelation of continuous wavelet coefficients. Recently, autocorrelation is employed to guide the time synchronous average for fault detection of gearboxes in wind turbines [120]. A novel method, Ensemble Average of Autocorrelation Signals (EAAS), is developed to suppress the random noise and enhance fault signatures for diagnosing bearing defects at an early stage.

4.3.1 Properties of Autocorrelation Analysis

4.3.1.1 Signal Period and Phase

Autocorrelation analysis describes the similarity of the signal with its own delayed version, which is helpful in analysing periodic signals. It is an important concept in signal processing due to the merits of periodic enhancement and noise suppression. A periodic signal function $x(n)$ with a period $2L$, can be expanded into a Fourier series, which can be expressed as follows

$$\begin{aligned} x(n) &= a_0 + \sum_{k=1}^{N}\left(a_k \cos\frac{2\pi k}{L}n + b_k \sin\frac{2\pi k}{L}n\right) \\ &= a_0 + \sum_{k=1}^{N}\left[\sqrt{{a_k}^2 + {b_k}^2}\cos\left(\frac{2\pi k}{L}n - tan^{-1}\frac{b_k}{a_k}\right)\right], N \geq 1 \end{aligned} \quad (4.45)$$

The autocorrelation function of the signal $x(n)$ is proposed to be

$$\begin{aligned} R_{xx}(\tau) &= \frac{1}{2L}\sum_{n=-L}^{L-1} x(n)x(n+\tau) \\ &= \frac{1}{2L}\sum_{n=-L}^{L-1} \left[a_0 + \sum_{k=1}^{N}\left(a_k \cos\frac{2\pi k}{L}(n+\tau) + b_k \sin\frac{2\pi k}{L}(n+\tau)\right)\right] \cdot \\ &\quad \left[a_0 + \sum_{k=1}^{N}\left(a_k \cos\frac{2\pi k}{L}n + b_k \sin\frac{2\pi k}{L}n\right)\right] \\ &= a_0^2 + \sum_{k=1}^{N}\left(\frac{a_k^2+b_k^2}{2}\cos\frac{2\pi k}{L}\tau\right) \end{aligned} \quad (4.46)$$

From Equation (4.46), it can be seen that

- The autocorrelated signal of the period signal is also periodic and moreover, the period is same as that of the original signal. It can disclose the self-similarity of the original signals.
- The initial phase of the autocorrelation signal is always 0, which means that the initial phase of an autocorrelated signal is independent of the raw signal. In other words, the autocorrelation signals discard phase information and return only the power, which is therefore an irreversible operation.

4.3.1.2 Noise Suppression

Another great feature of the autocorrelation signal is that white noise decays quickly to zero with the increase of time lags. A vibration signal from faulty bearings usually consists of the periodic fault signal and background noise, which can be described as

$$y(n) = x(n) + w(n) \quad (4.47)$$

where, $x(n)$ is the periodic signal generated by the fault impact series, and $w(n)$ represents white noise. Consequently, the autocorrelation signal of $y(n)$ [121] can be expressed

$$\begin{aligned} R_{yy}(\tau) &= \frac{1}{2L}\sum_{n=-L}^{L-1} y(n)y(n+\tau) \\ &= \frac{1}{2L}\sum_{n=-L}^{L-1}[x(t)+w(t)][x(t+\tau)+w(t+\tau)] \\ &= R_{xx}(\tau) + R_{xw}(\tau) + R_{wx}(\tau) + R_{ww}(\tau) \end{aligned} \quad (4.48)$$

Being random, the autocorrelation of white noise is zero. The periodic signal $x(n)$ and white noise $w(t)$ are independent and thus, $R_{xw}(\tau)$ and $R_{wx}(\tau)$ are approximately zero. $R_{xw}(\tau)$ and $R_{wx}(\tau)$ can be omitted. The autocorrelation signal of $y(n)$ becomes

$$R_{yy}(\tau) \approx R_{xx}(\tau) \quad (4.49)$$

This is the reason why autocorrelation can achieve noise cancellation.

4.3.1.3 Inherent Synchronousness

Autocorrelation signals can retain periodic components and increase signal to noise ratio (SNR). As explained above, the calculation of the autocorrelation function discards phase information, and consequently, the autocorrelation functions of the signal segments are synchronous with a new phase of zero. According to Equation (4.50), the autocorrelation signal $R_{yy}(\tau)$ is independent of the time variable n in the raw signal $y(n)$.

$$\begin{aligned} R_{yy}(\tau) &= \frac{1}{2L}\sum_{n=-L}^{L-1} y(n)y(n+\tau) \\ &= \frac{1}{2L}\sum_{n=-L}^{L-1} y\left(n+n_{delay}\right)y\left(n+n_{delay}+\tau\right) \end{aligned} \quad (4.50)$$

Multiple autocorrelation signals calculated from different samples of a continuous periodic signal $y(n)$ are inherently synchronous because the initial phase of all autocorrelation signals is zero. The average of these signals becomes a straightforward realisation of Time Synchronous Averaging (TSA) in the lag domain. This property is important in the following EAAS method.

4.3.1.4 FFT based Fast Computation

The autocorrelation function of a stationary signal $x_N(n)$ with a length of N is expressed as follows

$$R_{xx}(\tau) = E\{x_N(n+\tau)x_N(n)^*\} \quad (4.51)$$

where, $E\{\cdot\}$ is the expectation and * is the complex conjugate. The Power Spectral Density (PSD) of a signal is equal to the Fourier Transform of the autocorrelation according to Wiener-Khinchin theorem. It allows fast computing of the autocorrelation using the Fast Fourier Transform (FFT), which can be achieved by the following equations.

$$X(n) = \sum_{k=0}^{N-1} x_N(n)e^{-i\frac{2\pi nk}{N}} \quad (4.52)$$

$$R_{xx} = \frac{1}{N}\sum_{k=0}^{N-1}[X(n)X(n)^*]e^{i\frac{2\pi nk}{N}} \quad (4.53)$$

In the practical application, the autocorrelation function of the discrete signal $x_N(n)$ is calculated based on a proper zero padding, which avoids the circularity problem. The signal $x_N(n)$ is zero padded to have a length of $2N-1$ and then, the zero padded signal $x_{2N-1}(n) = [x_N(n), \underbrace{0,0,\dots 0}_{N-1}]$ is transformed by the FFT.

$$X_{2N-1}(n) = \sum_{k=0}^{2N-1-1}[x_{2N-1}(n)]e^{-i\frac{2\pi nk}{2N-1}} \quad (4.54)$$

$$\hat{R}_{xx} = \frac{1}{2N-1}\sum_{k=0}^{2N-1-1}[X_{2N-1}(n)X_{2N-1}(n)^*]e^{i\frac{2\pi nk}{2N-1}} \quad (4.55)$$

After the inverse FFT, the autocorrelation array $\hat{R}_{xx}$ is rearranged to map the values to the lags from $-(N-1)$ to $+(N-1)$ by using Equation (4.56).

$$R_{xx}(\tau) = \left[\hat{R}_{xx}(N+1:2N-1); \hat{R}_{xx}(1:N)\right] \quad (4.56)$$

Autocorrelation becomes suitable for the online fault detection and diagnostics owing to the fast computing approach.

For discrete time signals, the autocorrelation function is defined to be either biased or unbiased

$$r_{xx}(l) = \begin{cases} \frac{1}{N-|m|}\sum_{n=1}^{N-m+1} x(n)x(n+m-1), & Unbiased \\ \frac{1}{N}\sum_{n=1}^{N-m+1} x(n)x(n+m-1), & Biased \end{cases} \quad (4.57)$$

4.3.2 Ensemble Average of Autocorrelation Signals (EAAS)

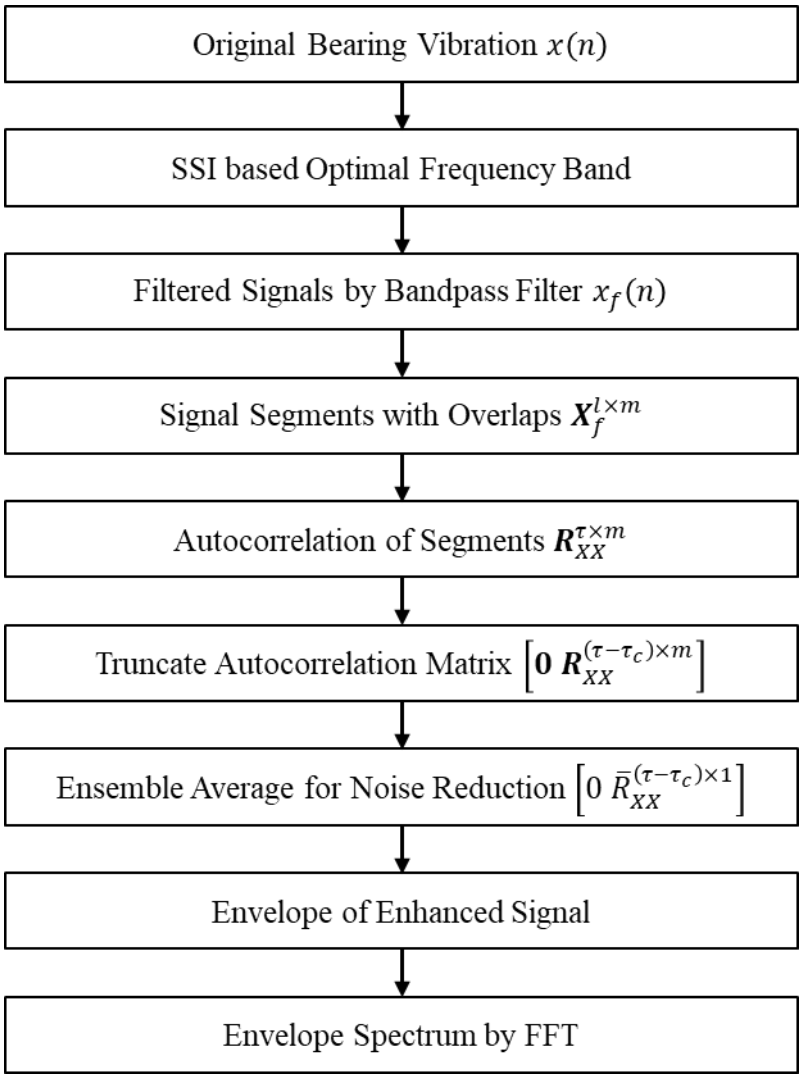


Figure 4-5 Flow chart of EAAS

The approach of Ensemble Average of Autocorrelation Signals (EAAS) is developed to suppress the noise and enhance the periodic information in vibration signals based on the properties of the autocorrelation function. Incipient fault signatures of bearings are often submerged within the

strong background noise. The EAAS is derived by considering the characteristics of periodicity enhancement and synchronicity of signal segments. The flow chart in the Figure 4-5 shows key steps to implement the EAAS.

In detail, the steps of EAAS are as follows:

Step 1. The raw vibration signal $x(n)$ with a length of N is processed by the Stochastic Subspace Identification to identify the natural frequencies, the natural frequencies can then be evaluated to find the optimal frequency bands.

Step 2. The filtered signal $x_f(n)$ is obtained by implanting a Fourier bandpass filter based on the selected frequency band from the raw vibration signal $x(n)$.

Step 3. The signal vector is re-organised to be a buffer matrix that comprises of short segments with overlaps. The re-organisation of signals is shown in Equation (4.58). The filtered signal with a length of N is rearranged to $m+1$ short segments with a length of l and an overlap of $l-k+1$. If the last segment is less than l, the segment is zero padded to l.

$$\boldsymbol{X}_f^{l\times m}=\begin{bmatrix} x_f(1) & x_f\,(k) & \cdots & x_f\left(k+(m-2)(k-1)\right) & \vdots \\ \vdots & \vdots & \cdots & \vdots & x_f\,(N) \\ x_f\,(k) & x_f\big(k+(k-1)\big) & \cdots & x_f\left(k+(m-1)(k-1)\right) & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_f\,(l) & x_f\left(l+(k-1)\right) & \cdots & x_f\left(l+(m-1)(k-1)\right) & 0 \end{bmatrix} \quad (4.58)$$

The segment length l is determined by the demand of frequency resolutions in the diagnostic results. In this thesis, the resolution of 1Hz is used to distinguish the characteristic fault frequencies and the length l is double of the sampling frequency, which will be explained in the next step. The overlap ratio $(l-k+1)/l$ is usually recommended to be 0.5 to 0.7 by experience, which increase the average times in the ensemble average operation. A too high overlap ratio does not contribute very much to the optimisation of the final results, which is not meaningful in the practical signal analysis.

Step 4. The segments are used to calculate the unbiased autocorrelation signals, which converts the time domain signal into the lag domain. The lag length directly determines the final frequency resolution. In this thesis, the resolution of 1Hz is used and the lag length is the same as the sampling frequency to ensure a high frequency resolution for fault detection. Moreover, to increase the accuracy of the autocorrelation estimation, the length l of the signal segments is chosen to be double that of autocorrelation lag τ.

$$\boldsymbol{X}_f^{l\times m} \Rightarrow \begin{bmatrix} R_{xx}^1(1) & R_{xx}^2(1) & \cdots & R_{xx}^m(1) \\ R_{xx}^1(2) & R_{xx}^2(2) & \cdots & R_{xx}^m(2) \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^1(\tau) & R_{xx}^2(\tau) & \cdots & R_{xx}^m(\tau) \end{bmatrix} = \boldsymbol{R}_{XX}^{\tau\times m} \tag{4.59}$$

Step 5. The beginnings of the autocorrelation signals contain greater amount of noise due to the limited signal lag. To eliminate the influence of extreme values at the beginnings, the autocorrelation signals of each segment only retain the lags which are greater than τ_c and the others are assigned to be zero. The τ_c is empirically selected to be about one tenth of the lag τ.

$$\boldsymbol{R}_{XX}^{\tau\times m} \Rightarrow \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^1(\tau_c+1) & R_{xx}^2(\tau_c+1) & \cdots & R_{xx}^m(\tau_c+1) \\ R_{xx}^1(\tau_c+2) & R_{xx}^2(\tau_c+2) & \cdots & R_{xx}^m(\tau_c+2) \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^1(\tau) & R_{xx}^2(\tau) & \cdots & R_{xx}^m(\tau) \end{bmatrix} \tag{4.60}$$

Step 6. The ensemble average of the autocorrelation segments can be achieved owing to the synchronised phase, which is denoted as follows.

$$\bar{\boldsymbol{R}}_{XX}^{\tau\times 1} = E[\boldsymbol{R}_{XX}^{\tau\times m}] \tag{4.61}$$

Step 7. The envelope of the enhanced signal $\bar{\boldsymbol{R}}_{XX}^{\tau\times 1}$ is then calculated from the analytic signals obtained by the Hilbert transform.

Step 8. The envelope spectrum is obtained by the Fourier transform, which makes the fault detection and diagnosis easily confirmed by the sparse representation in the frequency domain.

White noise decays to zero quickly in the procedure of autocorrelation calculation and the average is evidently useful to increase the signal-to-noise ratio. The combination of the two techniques leads to an outstanding noise elimination algorithm. Furthermore, the proposed method is achieved by calculating the autocorrelation of the short segments, which can be applied in the buffer based online data acquisition system and hence, leads to accurate and efficient real time fault diagnosis.

4.4 Demodulation Analysis of Outer Race Fault Signals

This section describes the performance of the aforementioned methods based on the developed bearing model. Two main steps of the bearing fault detection and diagnosis are the determination of the optimal frequency bands and the demodulation analysis. The performance of fault diagnosis methods is susceptible to background noise, either Gaussian or non-Gaussian. Therefore, the robustness of the developed and conventional methods is investigated under the influence of different levels of noise.

4.4.1 Determination of Optimal Frequency Bands

4.4.1.1 Effect of Gaussian Noise

The influence of random noise cannot be neglected in the process of fault detection and diagnosis because the ambient working conditions of rotating machines generate large quantities of noise. The white noise is the most challenging problem in machine condition monitoring, especially in incipient fault detection. The bearing vibration signals are inevitably contaminated by noise and in most circumstances, the noise is considered to obey the normal distribution. CM pursues to detect the faults as early as possible, which then allows ample time to implement the maintenance operations. Different levels of white Gaussian noise were added into the simulated signals so that it can be used to evaluate the proposed approach performance in incipient fault detection and diagnostics. Signal to Noise Ratio (SNR) is commonly used to quantify the strength of the white Gaussian noise, which is defined as

$$SNR = 10\,log_{10}\left(\frac{P_s}{P_n}\right) = 20\,log_{10}\left(\frac{RMS_s}{RMS_n}\right) \qquad (4.62)$$

where, P_s and P_n are the power of the noise free signal and the noise respectively; RMS_s and RMS_n are the root mean square values of the noise free signal and the noise respectively.

Figure 4-6 shows the time waveform of the simulated signals contaminated by six levels of white Gaussian noise. The publications on bearing condition monitoring can achieve the fault diagnosis at the SNR less than -20dB. To verify the effectiveness of the proposed method, the amount of noise added into the signal is from -10dB to -35dB with an even interval of -5dB. The periodic impacts induced by bearing faults are submerged into the strong background noise in all simulated cases, which gives a fair examination of the proposed method in early fault detection and diagnosis of rolling element bearings. From the six subfigures in Figure 4-6, it can be seen that the maximum values of the noisy signals vary from 200m/s^2 to 3000 m/s^2, which are 2 to 30 times of the amplitude of the faulty signal.

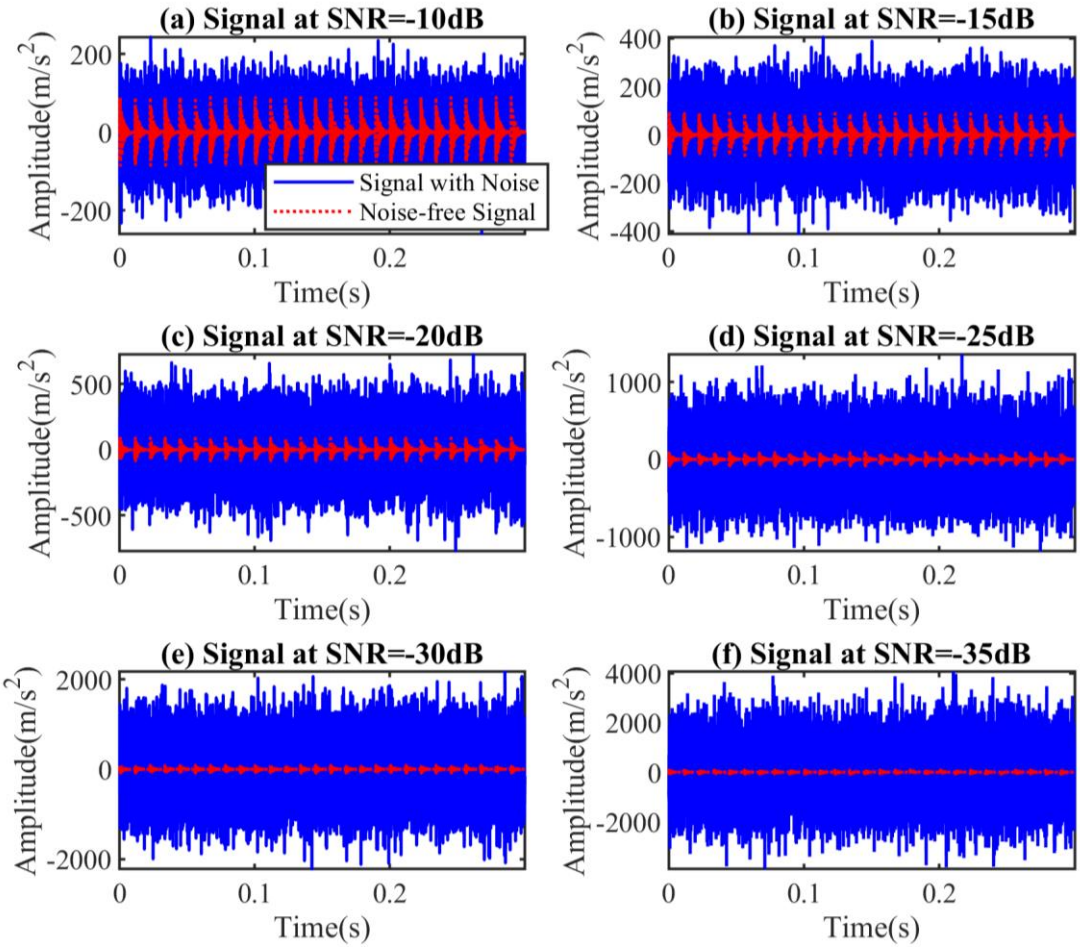


Figure 4-6 Time waveform of the simulated outer race fault signals at SNR of : (a) -10dB, (b) -15dB, (c) -20dB, (d) -25dB, (e) -30dB, (f) -35dB

Apart from the time domain analysis, the Fourier transform based frequency domain analysis is given in the Figure 4-7. The amplitude spectra can have a more impressive representation regarding the demodulation analysis. The noise floor in the spectrum increases along with the decrease of SNRs and the resonant areas in the vibration signals are submerged in the noise gradually. If the sparse components around the natural frequencies are above the noise carpet, the envelope analysis based on the optimal frequency bands can lead to the successful detection and diagnosis of the bearing defects, which is the reason why the envelope based demodulation analysis in the high frequency range is prior to the direct identification of the theoretically fault frequencies in the low frequency range. The experimental signals in the low frequency range contains large quantities of noise including vibration from the nearby machines, rotor imbalance, electrical noise. The fault information in the low frequency range of the amplitude spectrum is nearly covered by the undesired vibration and noise. Consequently, the detection and diagnosis in the low frequency domain becomes extremely difficult. The resonance modulation gives the opportunity of early fault detection and diagnosis owing to the high transmissibility of the bearing

system. The resonant responses usually have large amplitudes, which leads to a high SNR around the natural frequencies. However, the incipient defects induced fault information is usually very weak and hence, the faulty signatures are fully submerged into the background noise, which is similar with the cases shown in the Figure 4-7 (e) and (f). The poor SNR of the vibration signals makes the determination of the optimal frequency bands and the demodulation analysis very challenging.

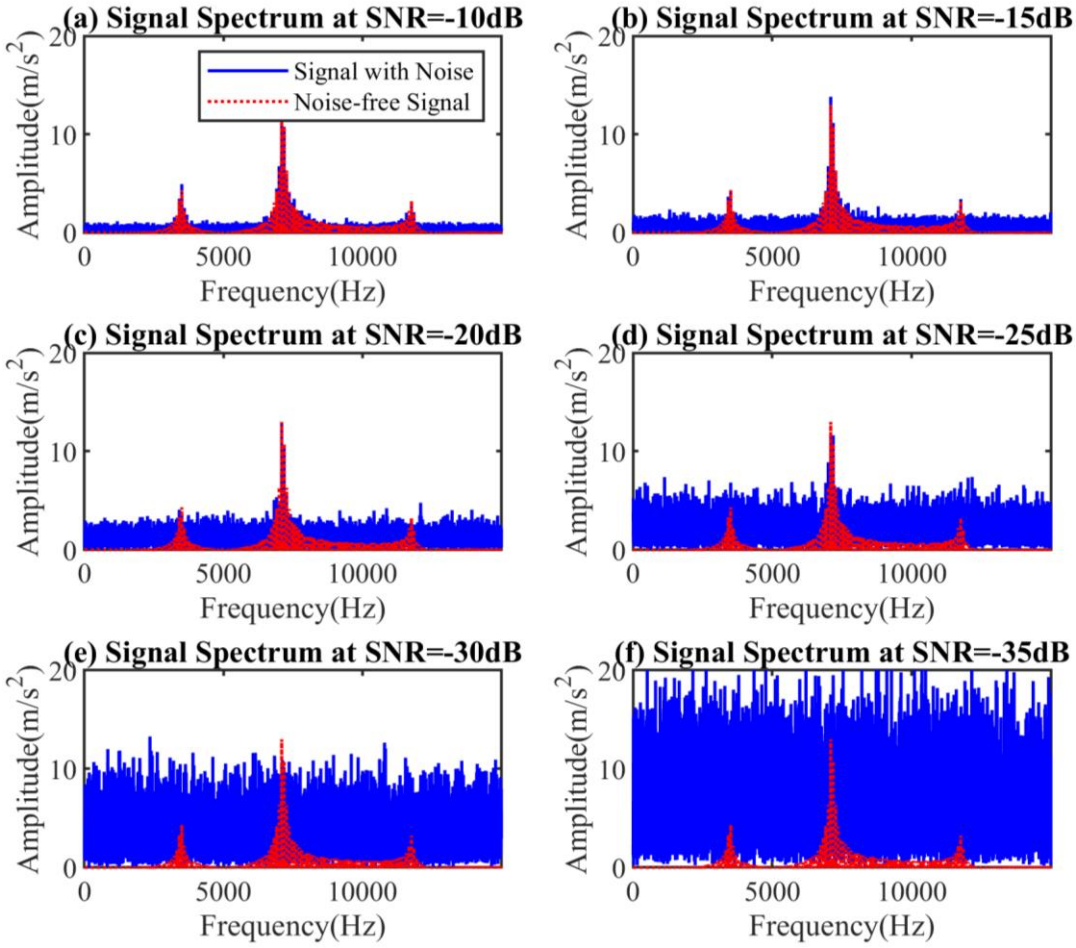


Figure 4-7 Spectra of the simulated outer race fault signals at SNR of : (a) -10dB, (b) -15dB, (c) -20dB, (d) -25dB, (e) -30dB, (f) -35dB

Considering the modulation mechanisms of the bearing fault vibration signals, a system identification based optimal frequency bands selection approach, Stochastic Subspace Identification, is proposed to find the optimal frequency bands in the presence of large quantities of background noise. The performance of the SSI to determine the demodulation frequency band is benchmarked by the famous Kurtogram.

The Kurtogram of the outer race fault signal at the SNR of -10dB is shown in the Figure 4-8. From the developed bearing model, the natural frequencies of the system are 3468Hz, 7117Hz and 11781Hz, which are the candidates of the optimum centre frequency. Based on the FRF of the

bearing model and the spectrum of the simulated signal, the frequency band centred at 7117Hz is theoretically the best option for the demodulation analysis. The Kurtogram demonstrates that the optimal frequency band is around 7031Hz with a bandwidth of 1563Hz, which locates at the fifth level of the Kurtogram.

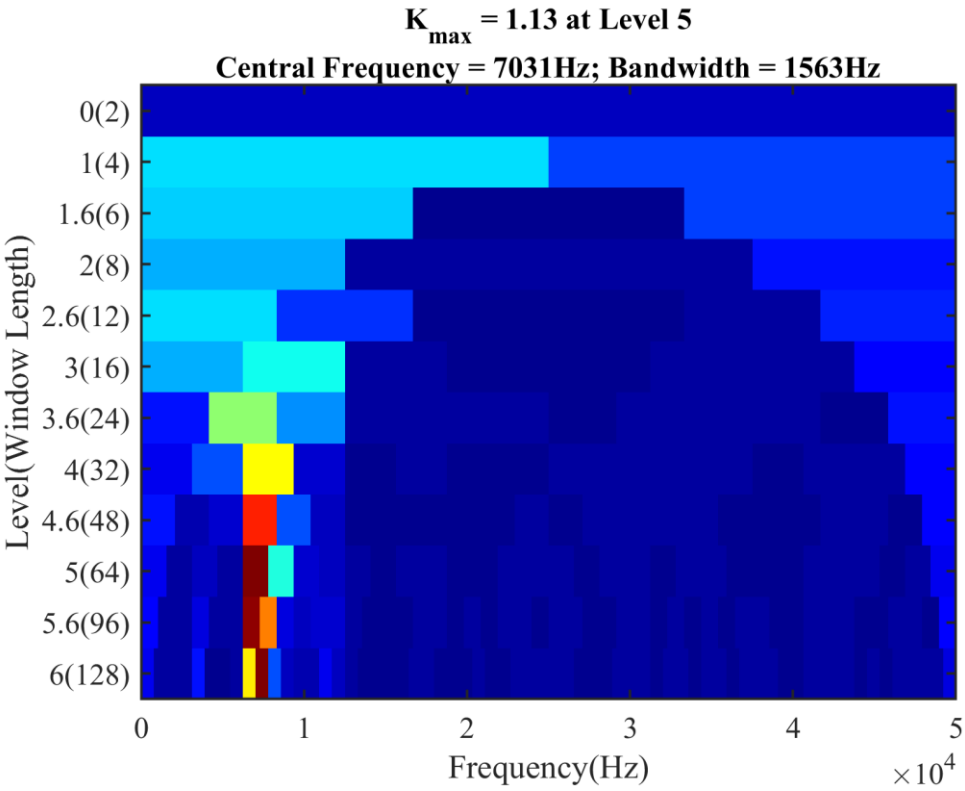


Figure 4-8 Kurtogram for the case of outer race faults at SNR -10dB

The Stabilisation Diagram used in the SSI gives the manifest description of the modes identified in the OMA. These three criteria in the Equation (4.63) to (4.65) show that three criteria are implemented by the threshold value of the variation [104]. The modal frequency stability is expressed as

$$|f_{i+1} - f_i| \leq 100 \qquad (4.63)$$

where, f_i and f_{i+1} are the natural frequencies at i_{th} and $(i+1)_{th}$ order. The modal shape stability is the coherence between the modal shapes at i_{th} order and $(i+1)_{th}$ order, which can be shown as

$$\left(1 - \frac{|\phi_{i+1}^{*}\phi_i|^2}{(\phi_{i+1}^{*}\phi_{i+1})(\phi_i^{*}\phi_i)}\right) < 1\% \qquad (4.64)$$

where, ϕ_i and ϕ_{i+1} are the modal shapes at i_{th} and $(i+1)_{th}$ order. The damping ratio stability is the difference of damping ratio at adjacent orders, which is expressed as

$$|\xi_{i+1} - \xi_i| < 0.2\% \qquad (4.65)$$

where, ξ_i and ξ_{i+1} are the damping ratios at i_{th} and $(i+1)_{th}$ order. With the increase of the system orders, the modal information extracted by the SSI is drawn in the Figure 4-9 (a). According to three criteria, including the stability of the modal frequencies, modal shapes and damping ratios, the modes meeting these criteria are plotted in the Figure 4-9 (b).

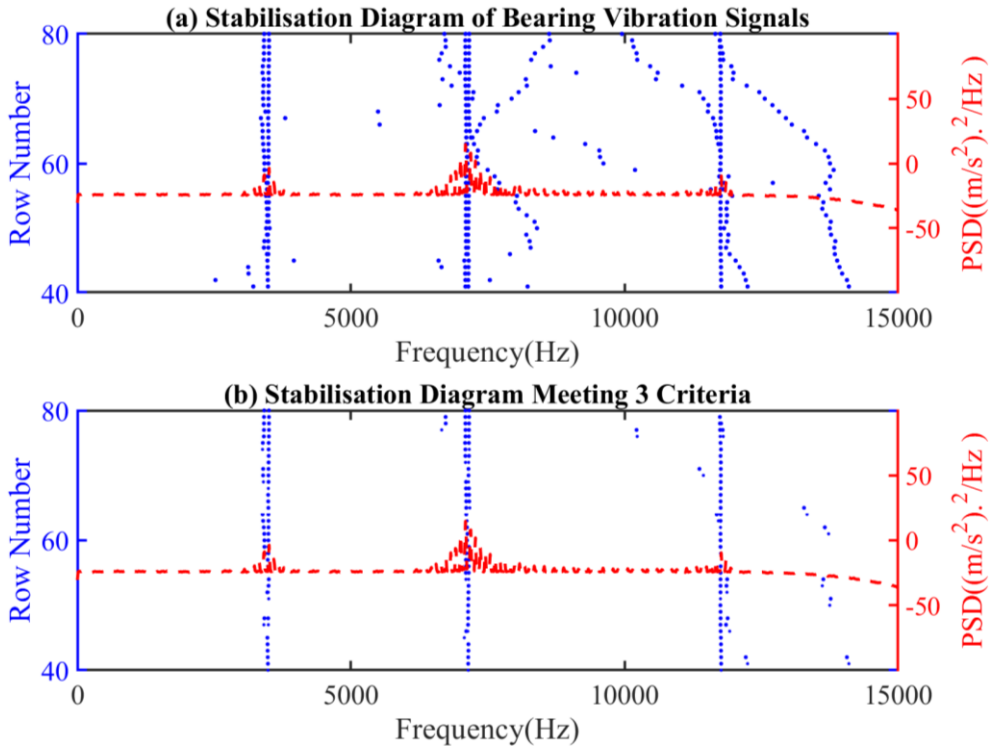


Figure 4-9 Stabilisation Diagram of SSI for the case of outer race faults at SNR -10dB: (a) all modes identified; (b) modes meeting three criteria

From the Stabilisation Diagram of the SSI method, three stable modes are identified, which are the exact three modes of the simulated bearing model. A Normalised Rate is developed to show the recurrence rates of the modes along with the order increase. The Normalised Rate of every frequency bin is defined as

$$R_N(f) = \frac{\sum_{j=N_s}^{N_e} f(j)}{N_s - N_e + 1} \times 100\% \qquad (4.66)$$

where, f is the frequency in the frequency range; N_s is the beginning order of the identified system; and N_e is the maximum order of the identified system. The Normalised Rate is an important indicator that can distinguish the real modes from the mixture of the real and fake modes in the identification. Figure 4-10 displays the Normalised Rates of these three modes and the indicators are almost one, which demonstrates the robustness of the mode identified by the SSI approach. The Normalised Rate is affected by the SNR of the signals. If the SNR of the output signals is high, the Normalised Rates of the real modes will be close to 100%. Consequently, the Normalised Rate is a potential indicator for the optimum centre frequency from the modes identified. The problem

for the SSI method based determination of the optimal demodulation analysis is that the SSI cannot indicate the bandwidth of the demodulation analysis. According to the engineering experience, the appearance of the characteristic fault frequency of the bearings and its first three harmonics can be used to confirm the detection and diagnosis of the defects solidly. Therefore, the bandwidth is determined to be three and half times of the desired characteristic fault frequency on both sides of the central frequency.

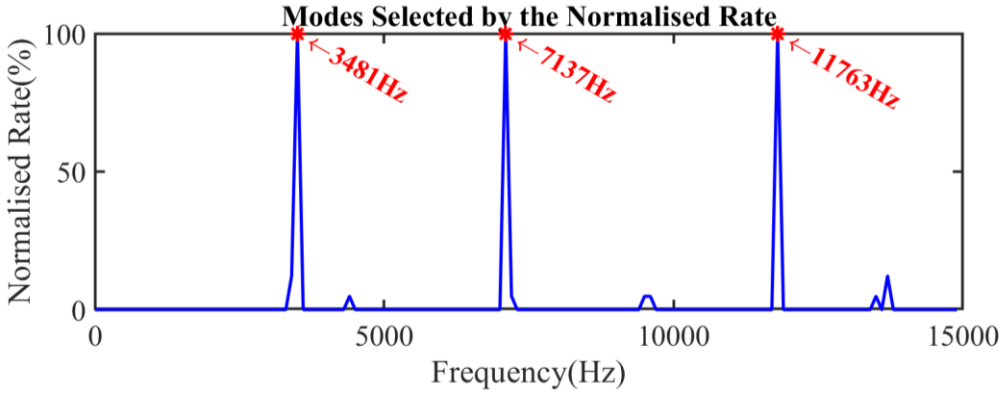


Figure 4-10 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -10dB

The case of the bearing signals at the SNR of -10dB is considered as high-quality measurements in the early fault detection and diagnosis. If the bearing faults are at the early stage, the fault size is tiny and the background noise including mechanical and electrical noise is much stronger than the fault induced signals, and hence, the SNR of the obtained signals are much lower than the simulated -10dB. In the simulation study, the signal quality can be controlled very well. The simulated cases at lower SNR was capable of examining the performance of the proposed method.

The second simulation case of the outer race faults is the signals at the SNR of -15dB, which means the Root Mean Square (RMS) value of the added noise is more than 1.7 times of the last -10dB case. Figure 4-11 and Figure 4-12 show that both methods can indicate the optimal central frequency. The Kurtogram in Figure 4-11 denotes that the optimal frequency band is around 7422Hz with a bandwidth of 781.3Hz, which locates at the sixth level of the Kurtogram. The Normalised Rate in Figure 4-12 shows that the natural frequency 7125Hz has the highest rate, which means the signal in this frequency band has a higher SNR than the other natural frequency based frequency bands.

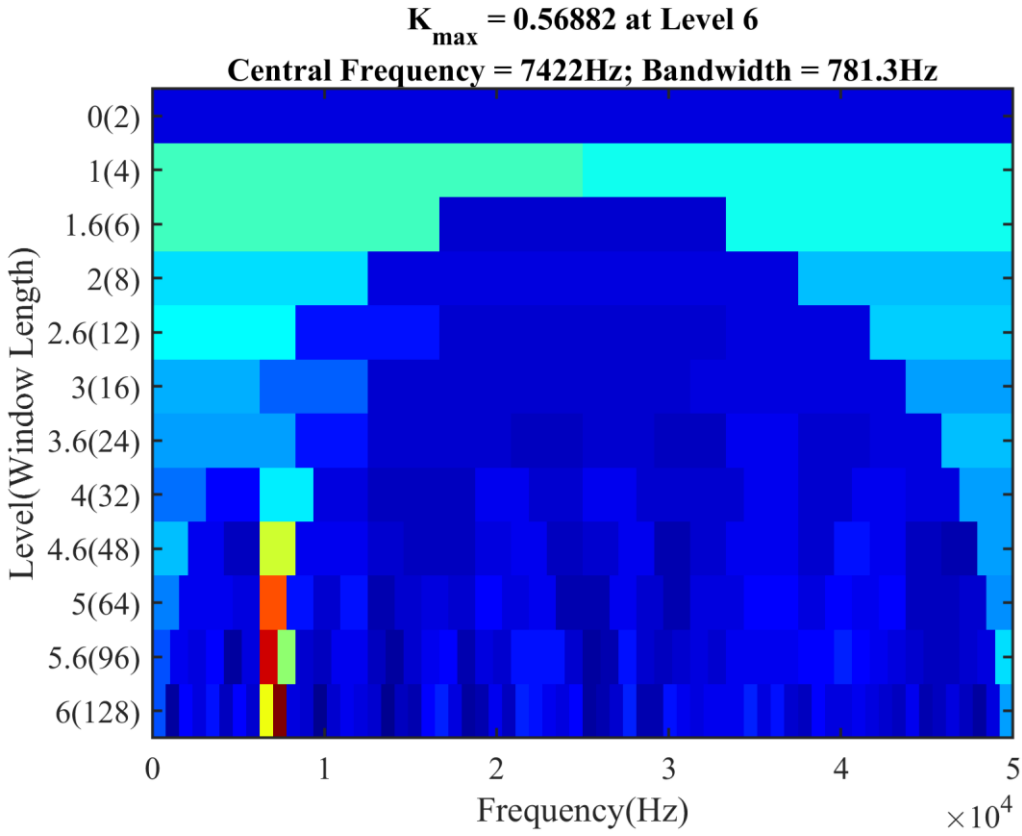


Figure 4-11 Kurtogram for the case of outer race faults at SNR -15dB

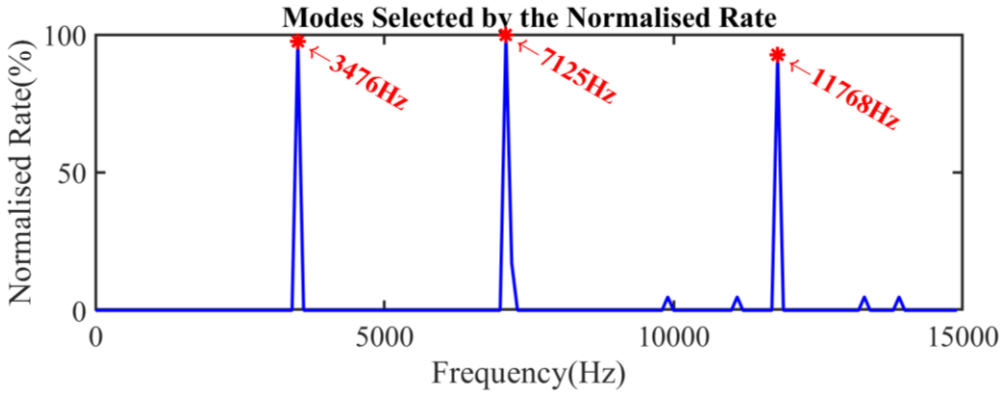


Figure 4-12 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -15dB

The third simulation case of the outer race faults is the signal at the SNR of -20dB and the amplitude of noise can reach up to 600m/s^2. In this case, the noise level for Kurtogram analysis is too high to obtain the accurate frequency band, which is shown in the Figure 4-13. The Kurtogram recommends the frequency band around 12.5kHz with a bandwidth of 25kHz at the first level. The frequency band is too wide and it is not meaningful in the envelope analysis. This result denotes that the defects induced vibration signals within all the frequency bands in the Kurtogram are submerged so that the impulsive fault characteristics cannot be recognised. The Normalised Rate in the Figure 4-14 shows that the natural frequency 7113Hz has the highest rate, which means the

signal in this frequency band has a higher SNR than the other natural frequencies based frequency bands.

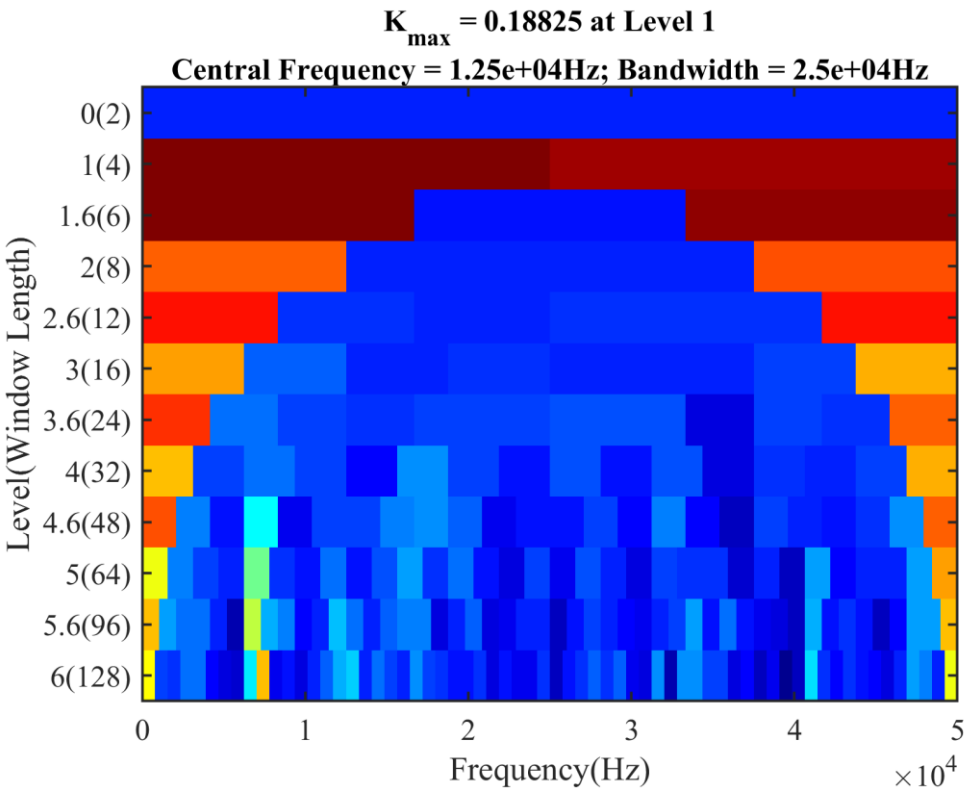


Figure 4-13 Kurtogram for the case of outer race faults at SNR -20dB

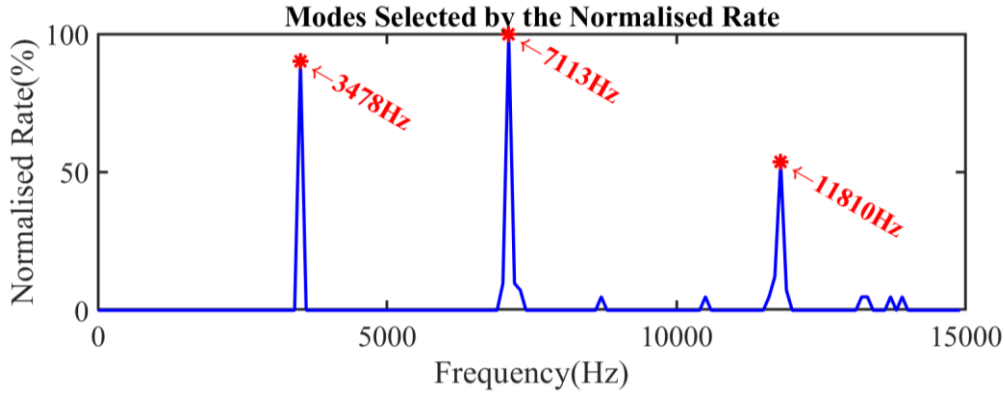


Figure 4-14 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -20dB

In the three following cases, the Kurtogram results in Figure 4-15, Figure 4-17 and Figure 4-19 are unable to select the optimal frequency bands. It chooses a quite wide frequency band and the bandwidth is more than 15kHz. Such a wide frequency band is not meaningful in the demodulation analysis, especially for the early fault detection and diagnosis because a wide frequency band includes too much noise, leading to the unsuccessful detection and diagnosis. In contrast, the SSI based approach can robustly identify the optimal central frequencies. The results of Normalised

Rates in the Figure 4-16, Figure 4-18, and Figure 4-20 show the central frequencies of the optimal frequency band in these three cases are the second order mode of the bearing model. Furthermore, the rates decrease along with the increase of background noise, which is acceptable as the theoretical expectation. The frequency bands recommended by the SSI are centred at about 7110Hz, which are the best candidate for the demodulation analysis. According to the spectrum, the selected mode is the most easily excited mode owing to the high transmissibility of the system resonance.

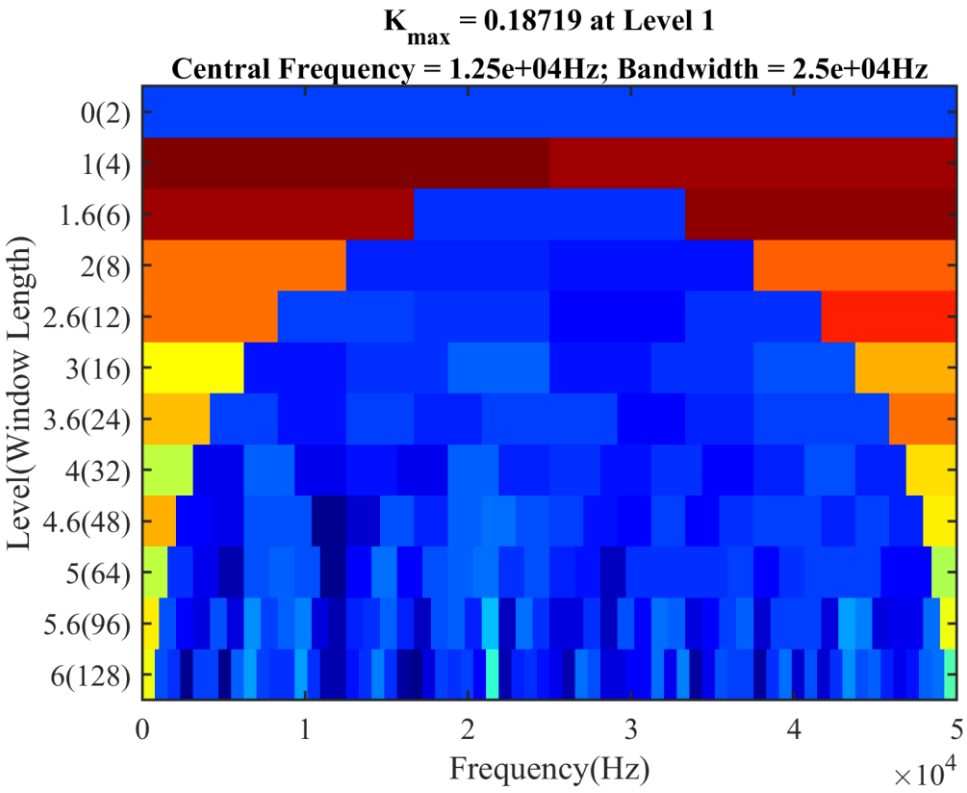


Figure 4-15 Kurtogram for the case of outer race faults at SNR -25dB

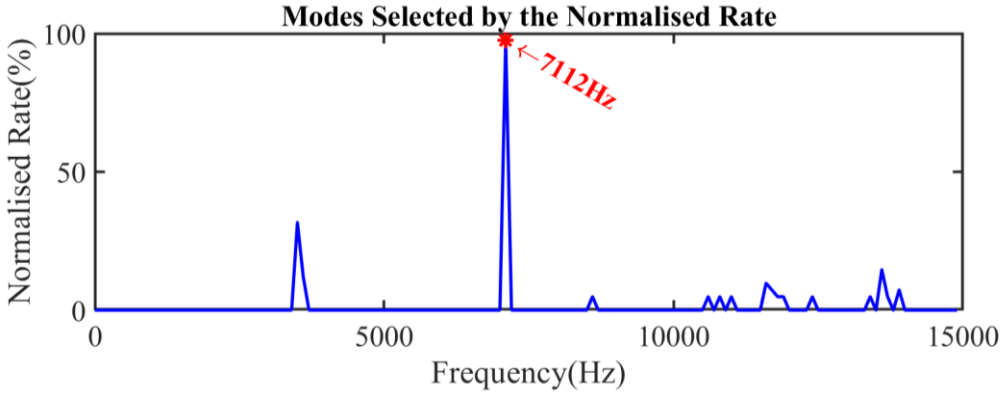


Figure 4-16 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -25dB

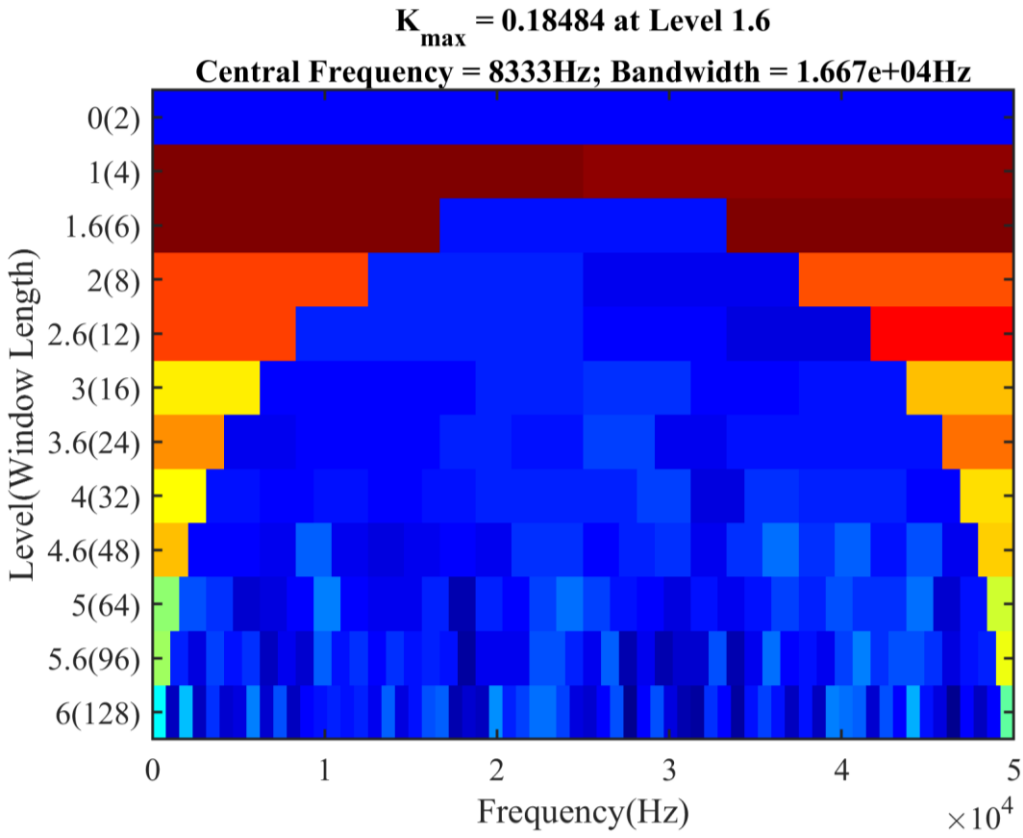


Figure 4-17 Kurtogram for the case of outer race faults at SNR -30dB

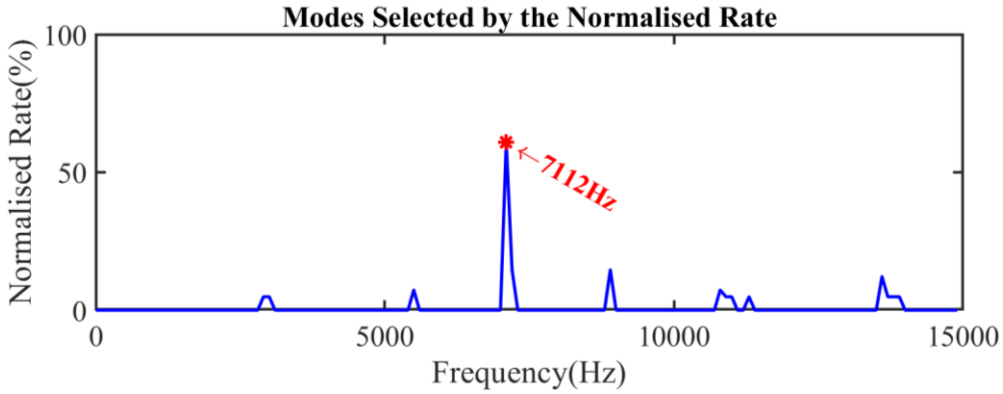


Figure 4-18 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -30dB

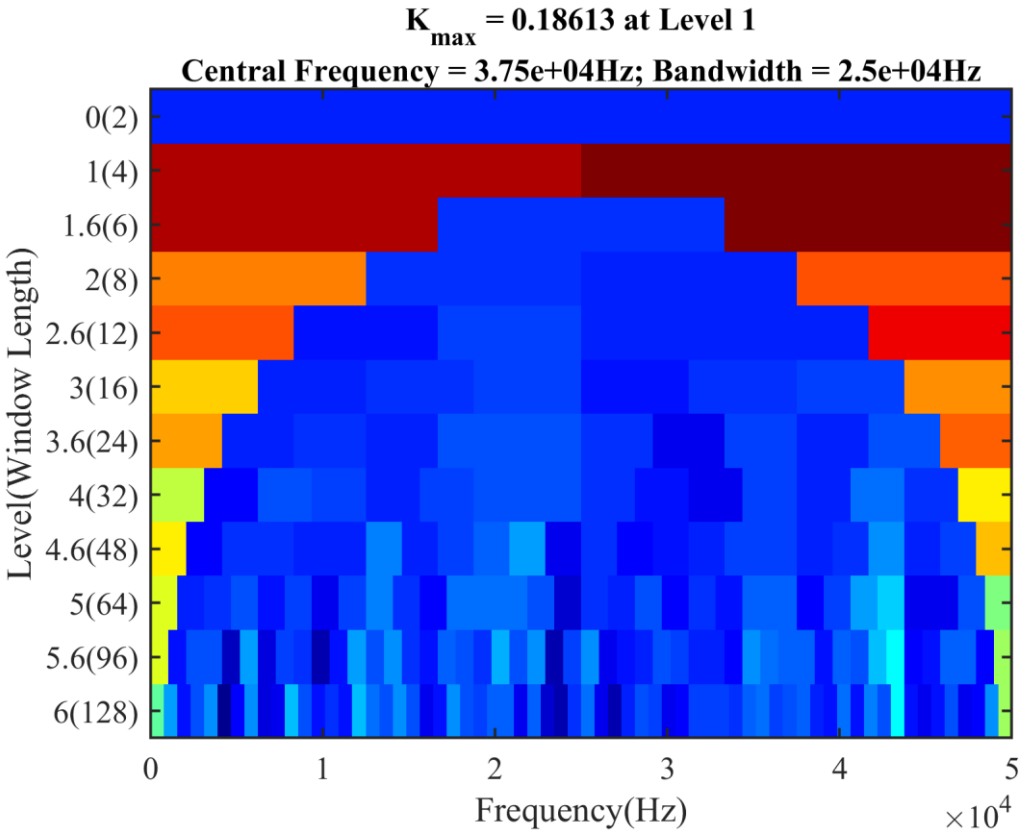


Figure 4-19 Kurtogram for the case of outer race faults at SNR -35dB

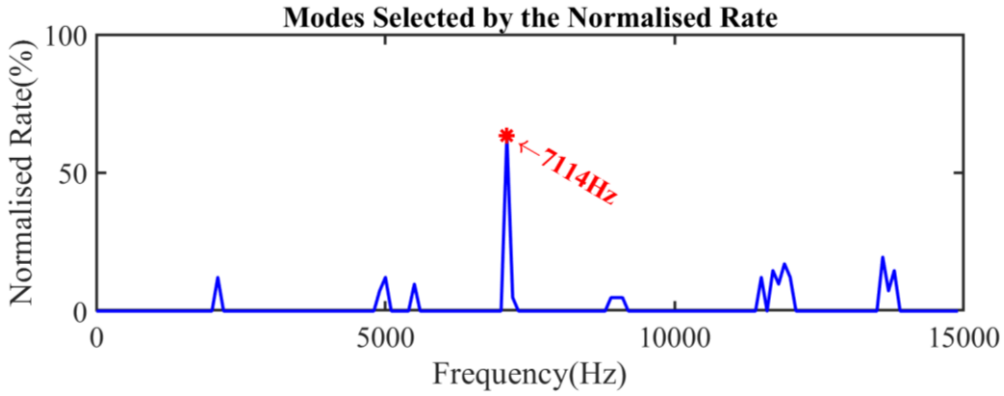


Figure 4-20 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -35dB

4.4.1.2 Effect of Non-Gaussian Noise

Although the white Gaussian noise is widely used in the simulation study, it cannot represent the background noise in the experimental conditions. The non-Gaussian noise, such as the random impulses, has significant influence upon the diagnostic results. The effects from this kind of non-Gaussian noise are investigated in this section. A series of aperiodic impulses were added into the simulated signals and furthermore, a small amount of white Gaussian noise was added

simultaneously to simulate the practical working conditions of the rolling bearings. The aperiodic impulses are generated by the random function and the amplitude of the large random impulses are set at double the maximum value of the signals with Gaussian noise. Similar to the generally defined SNR, the influence of the aperiodic impulses is measured by a new definition of the impulse SNR (SNR_i), which is expressed as follows.

$$SNR_i = 20\,log_{10}\left(\left|\frac{A_s}{A_i}\right|\right) \qquad (4.67)$$

where, A_s is the peak value of the noise-free signal and A_i denotes the maximum value of the large random impulses. Two cases with the aperiodic impulses are simulated to evaluate influences of the random impulses on the determination of the most informative frequency bands. The bearing signal is same as that used in the study of white Gaussian noise. The random impulses are generated by the Matlab function, *randi*, which obeys a uniform distribution. The amount of random impulses is ten per second, or in another word 0.01% of the total samples. Two cases including the pure signals and noisy signals are involved in the research of the negative effects from the non-Gaussian noise. The maximum values come from either the peak value of the periodic signals or a localised value with a large amplitude from the white Gaussian noise.

Figure 4-21 shows the temporal waveforms in the two non-Gaussian noise cases. In the Figure 4-21 (a), only several aperiodic impulses were displayed and the signal does not contain Gaussian noise which aims to show the inefficiency of the approaches only detecting the impulsive characteristics. Large quantities of methods were developed to find the frequency band that contains the most impulsive behaviour and these methods are usually not from the viewpoint of the modulation mechanism. This particular case can give an accurate picture of the deficiency of the conventional Kurtogram. However, this case does not take account of white noise and is not reasonable. Consequently, another case of the signal with both white Gaussian and non-Gaussian noise in Figure 4-21 (b) was given to show the reasonable comparison of the Kurtogram and SSI method. In this case, a decent amount of white Gaussian noise is generated to submerge the faulty bearing signals, making the SNR of the signal -10dB. At this noise level, the Kurtogram is able to detect the optimal frequency band if only Gaussian noise presents in the simulated signal. To compare the robustness of the proposed SSI and the Kurtogram, a series of random impulses at the rate of 0.01% was created as the additive non-Gaussian noise. As explained previously, the amplitude of these aperiodic impulses are two times the maximum value of the bearing signals with the white Gaussian noise.

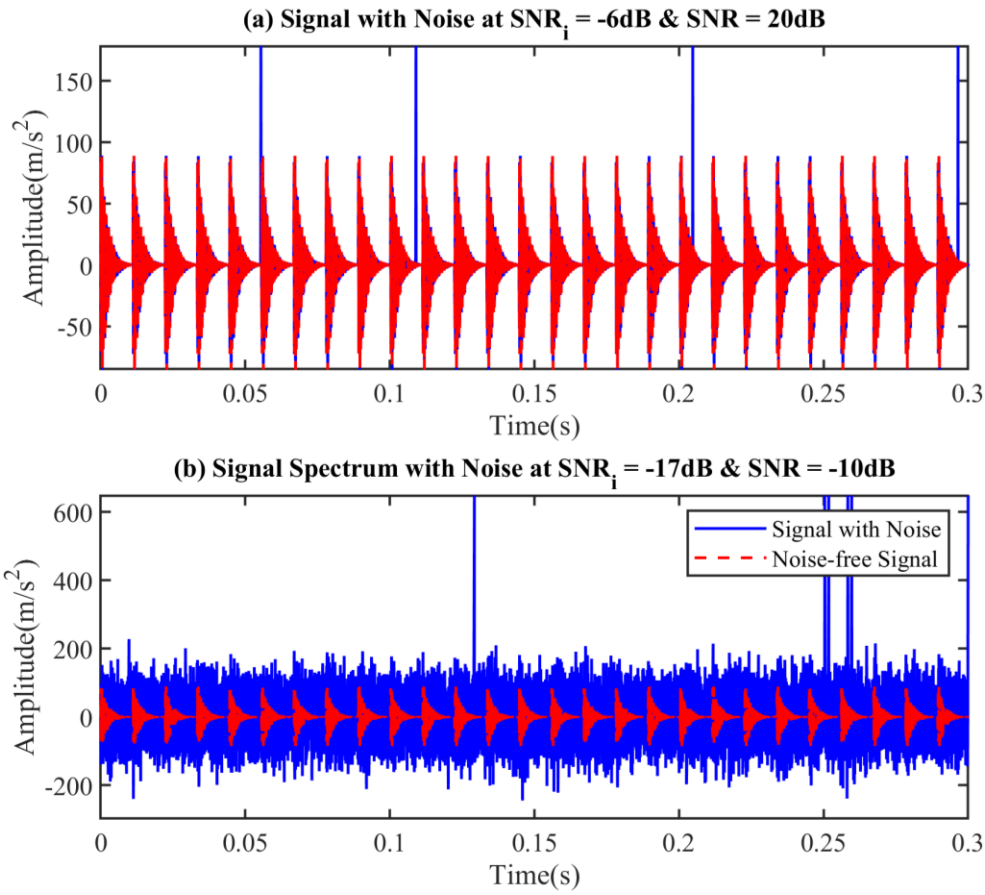


Figure 4-21 Time waveform of the simulated outer race fault signals with non-Gaussian noise at: (a) SNR_i -6dB and SNR 20dB, (b) SNR_i -17dB and SNR -10dB

Figure 4-22 gives a manifest and sufficient representation of the simulated signals in the frequency domain. The faulty signals are shown by the dashed red lines and shows the resonant modulation in the vibration signal from rolling element bearings. The limited number of impulses have little influence on the amplitude spectrum because the energy of the impulse signal is dispersed into the whole frequency range, and the effects become very small in the amplitude spectrum. As shown in Figure 4-22 (a), the spectra of the noise-free and noisy signals are quite difficult to distinguish. In the following simulation cases, the main difference in the amplitude spectra is on account of the white Gaussian noise that has an equal intensity in the frequency bins.

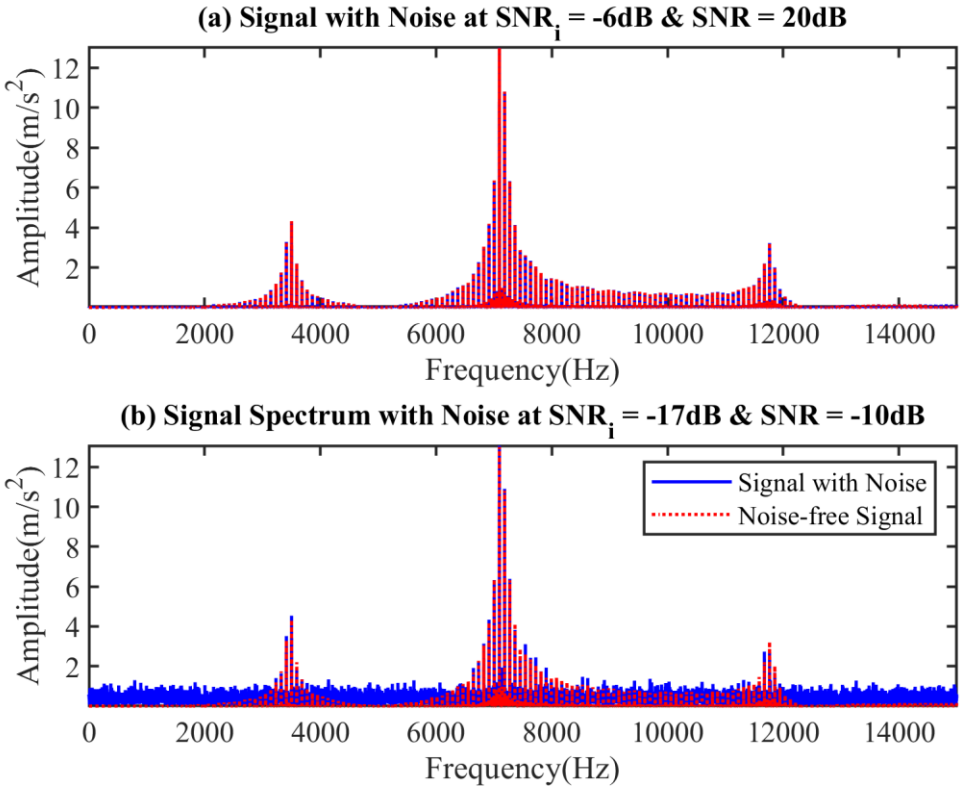


Figure 4-22 Spectra of the simulated outer race fault signals with non-Gaussian noise at: (a) SNR_i -6dB and SNR 20dB, (b) SNR_i -17dB and SNR -10dB

The simulated bearing signal at SNR_i -6dB and SNR 20dB is supposed to be the easiest case for selecting the frequency band because of a significantly high SNR. However, this signal seems to be a challenge for the approach of Kurtogram. It is shown in the Figure 4-23 that the Kurtogram denotes the best frequency band from 12.5kHz to 62.5kHz. The result indicates that one of the deficiencies of Kurtogram is that its powerless to deal with the irregular pulses. Because the signal does not contain any white Gaussian noise, nearly all the frequency bands allow the detection and diagnosis of the bearing defects. Consequently, the frequency band recommended by the Kurtogram is not meaningful in obtaining the fault signatures. The demodulation analysis is described in depth and discussed in the next section. Figure 4-24 shows that the proposed system identification method, SSI, can accurately identify the natural frequencies and furthermore, the Normalised Rates of identified three modes are approaching to one, which demonstrates that the extracted modal information has a very high SNR. The identified natural frequencies are the carrier frequencies in the bearing fault detection and diagnosis.

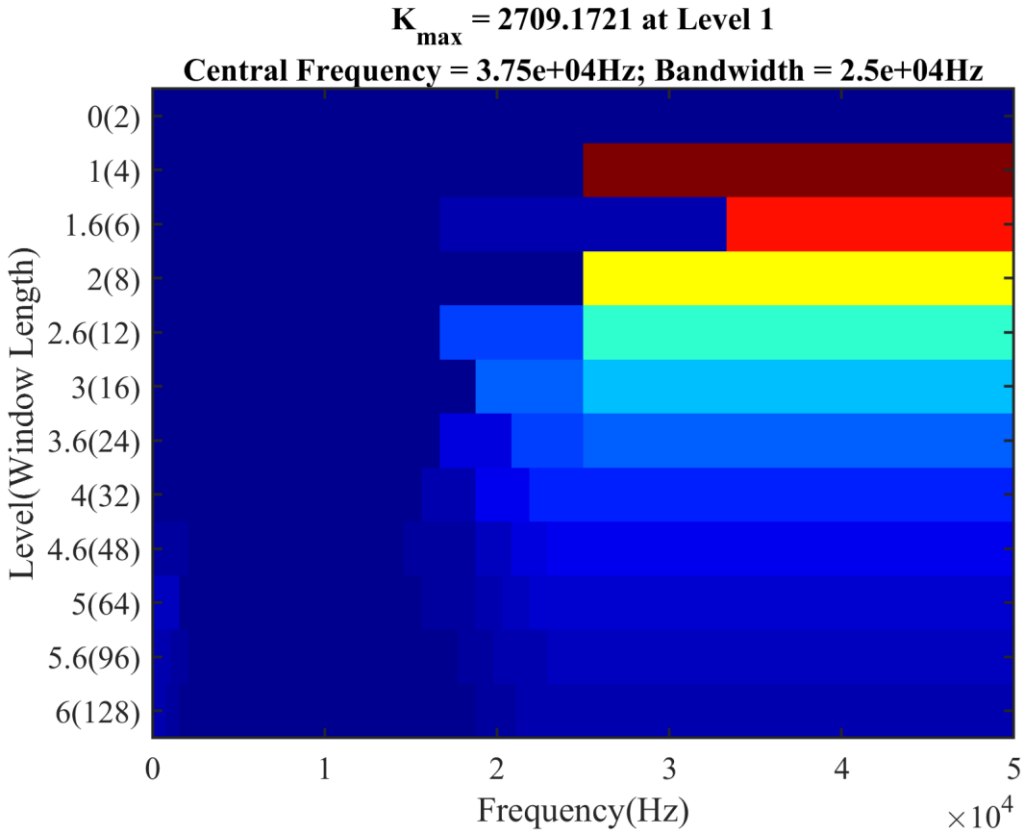


Figure 4-23 Kurtogram for the case of outer race faults at SNR_i -6dB and SNR 20dB

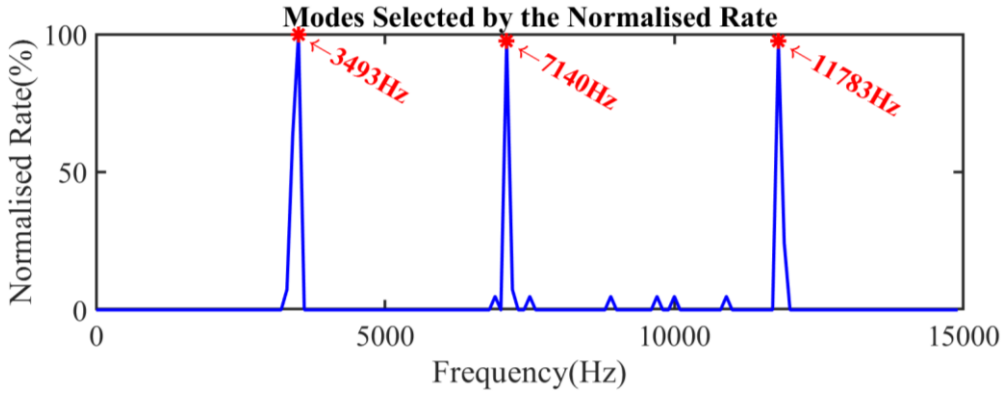


Figure 4-24 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR_i -6dB and SNR 20dB

Both Gaussian and non-Gaussian noise were added to the bearing signals in this case study and the results of the Kurtogram and SSI are depicted in the Figure 4-25 and Figure 4-26 respectively. The Kurtogram in the Figure 4-25 recommends the whole frequency band as the optimum candidate for the envelope analysis. The Figure 4-26 shows that the SSI based approach discloses all the natural frequencies of the bearing model with very high Normalised Rates, which are the carrier frequencies in the amplitude modulation signals.

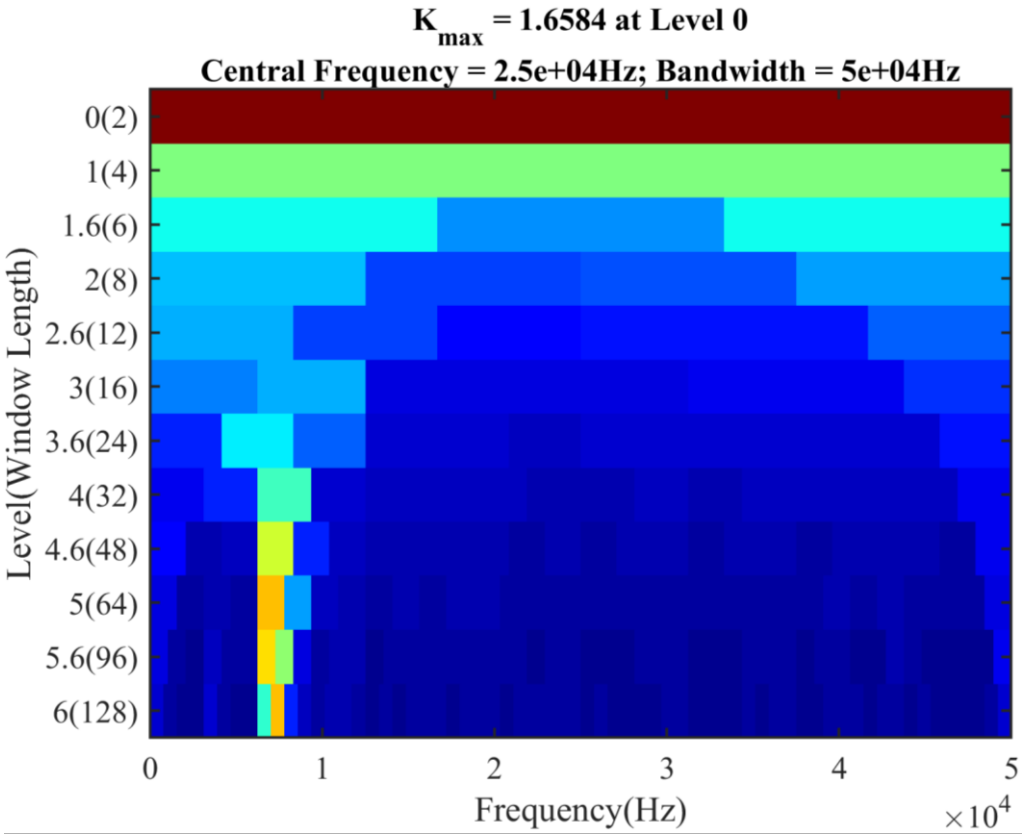


Figure 4-25 Kurtogram for the case of outer race faults at SNR_i -17dB and SNR -10dB

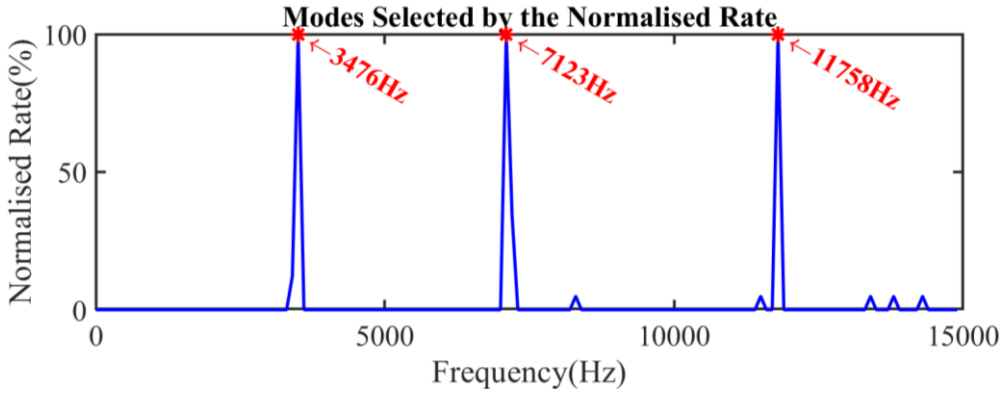


Figure 4-26 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR_i -17dB and SNR -10dB

4.4.2 Demodulation Results and Analysis

In the aforementioned section, several typical cases are simulated to address the realistic conditions of the rolling element bearings. These signals were used to examine the performance of the two approaches, Kurtogram and SSI, in determining the optimal frequency bands for the further demodulation analysis. In this part, the demodulation analysis including the developed EAAS and the benchmark method (conventional envelope, CE) is present. Since Darlow [10] developed the techniques of envelope analysis, this demodulation technique has been popular for approximately

40 years. The proposed method, EAAS, is discussed and compared based on the optimal frequency bands selected in the previous section. The benchmark method, the conventional envelope, is obtained by the ensemble average of the segments' spectra, which makes the fault detection and diagnosis more reliable. The detailed steps to obtain the CE results in this thesis are shown in the Figure 4-27. The ensemble average of the amplitude spectra can result in a more robust result than the directly calculated envelope spectrum. However this ensemble average does not suppress the background noise in the signals, and it cannot detect and diagnose the incipient faults when the SNR is very low.

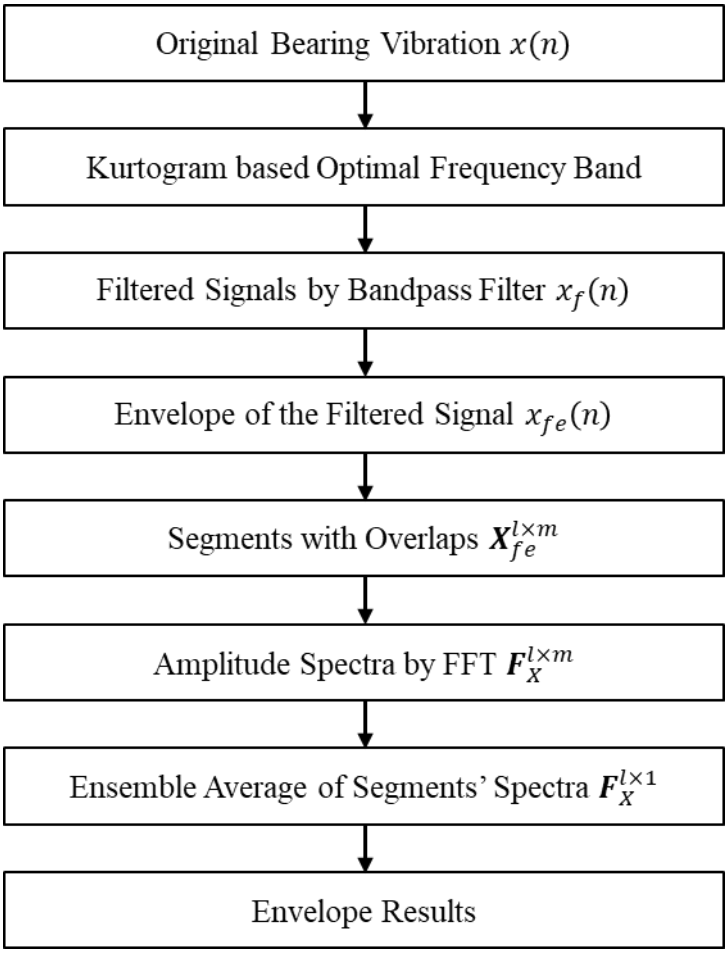


Figure 4-27 Flow chart of CE in this thesis

The demodulation results of EAAS and CE are shown and discussed below. The results of EAAS and CE of the outer race fault signal at SNR -10 dB are shown in the Figure 4-28 respectively. In the Figure 4-28 (a), EAAS denotes a substantial sparse spectrum comprised by the characteristic fault frequency and corresponding harmonics based on the central frequency selected by the SSI method. In order to compare the frequency bands by the Kurtogram and SSI, the CE based on the optimal frequency band by the SSI is also calculated and the spectrum of the SSI based CE is shown in Figure 4-28 (b), which gives a more convincing comparison of the EAAS and CE in the

demodulation analysis. The SSI based CE gives an accurate diagnosis of the bearing faults. As displayed in Figure 4-28 (c), the Kurtogram based CE gives a similar spectrum as the SSI based CE. The noise floor in the CE spectrum is obviously higher than that in the EAAS spectrum because the conventional envelope does not have the capability of noise reduction.

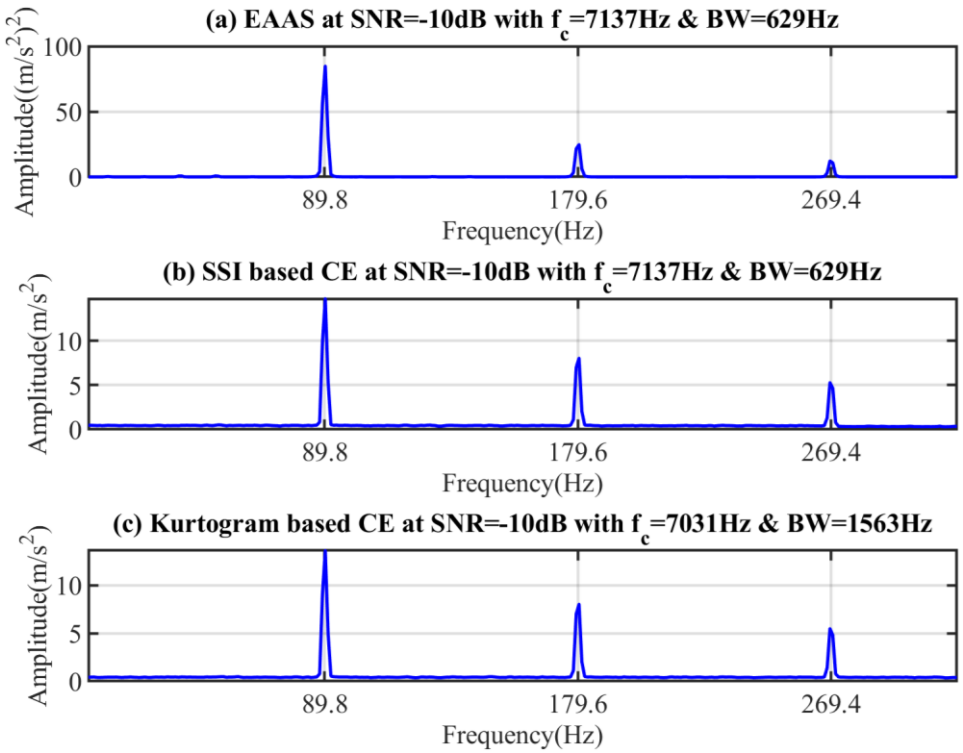


Figure 4-28 EAAS and CE of outer race faults at SNR -10dB

Figure 4-29 depicts the spectra of the EAAS, SSI based CE, and Kurtogram based CE. With the increase of the background noise, the EAAS method gives a remarkable representation of the fault information, which is shown in Figure 4-29 (a). In this case, both Kurtogram and SSI find the similar frequency band for demodulation analysis. As shown in Figure 4-29 (b) and (c), the noise floors in the spectra of two CE approaches are increasing due to the random noise. In addition, the SSI based CE in Figure 4-29 (b) displays a higher amplitude of the theoretical fault frequency than the Kurtogram based CE, which demonstrates that the frequency band by the SSI is better.

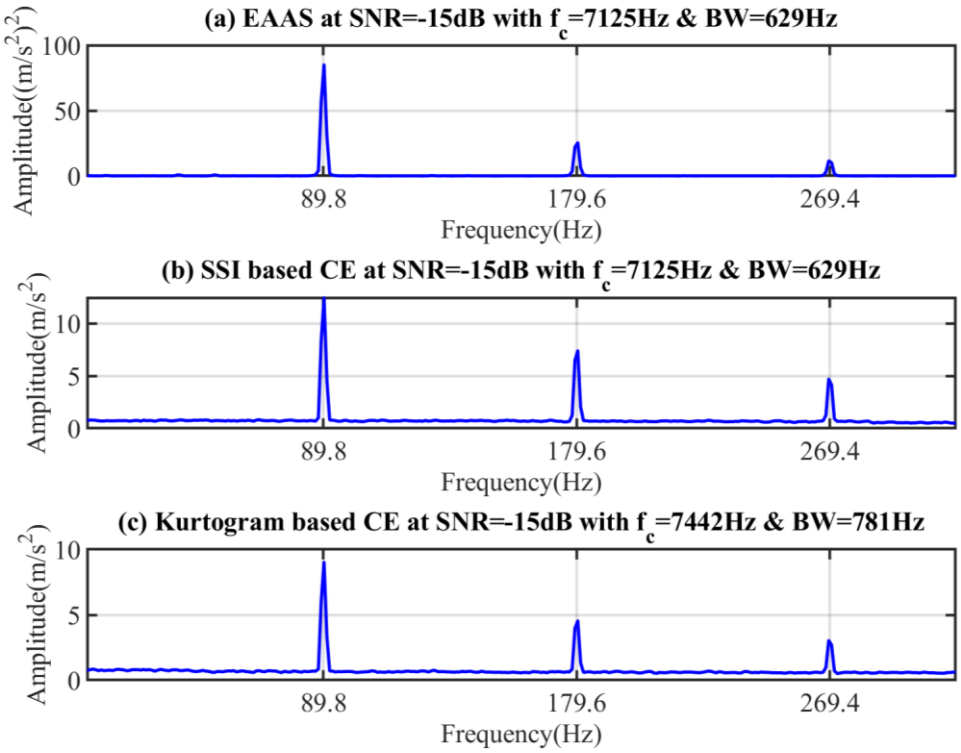


Figure 4-29 EAAS and CE of outer race faults at SNR -15dB

In the case of outer race signals at -20dB, the high-level noise has little influence on the proposed EAAS, which is shown in Figure 4-30 (a). The spectrum obtained by the EAAS shows the fault signatures clearly with a naught noise floor. Figure 4-30 (b) describes that the SSI based CE can accurately detect and diagnose the bearing faults. In this case, the Kurtogram is affected by the background noise. Based on the frequency band selected by Kurtogram, the CE can demodulate the characteristic fault frequencies. The Kurtogram chose a very wide frequency band from 0Hz to 25,000Hz, which includes all the natural frequencies of the bearing model. This then gives a chance for the Hilbert transform based envelope to detect and diagnose the bearing faults. However, a wide range frequency band based envelope spectrum also takes account of the negative influence of the noise, which does not utilise the advantage of the high SNR around the natural frequencies. If the signals are poorer than -20dB, the envelope results can be misleading which can be verified in the next -25dB case.

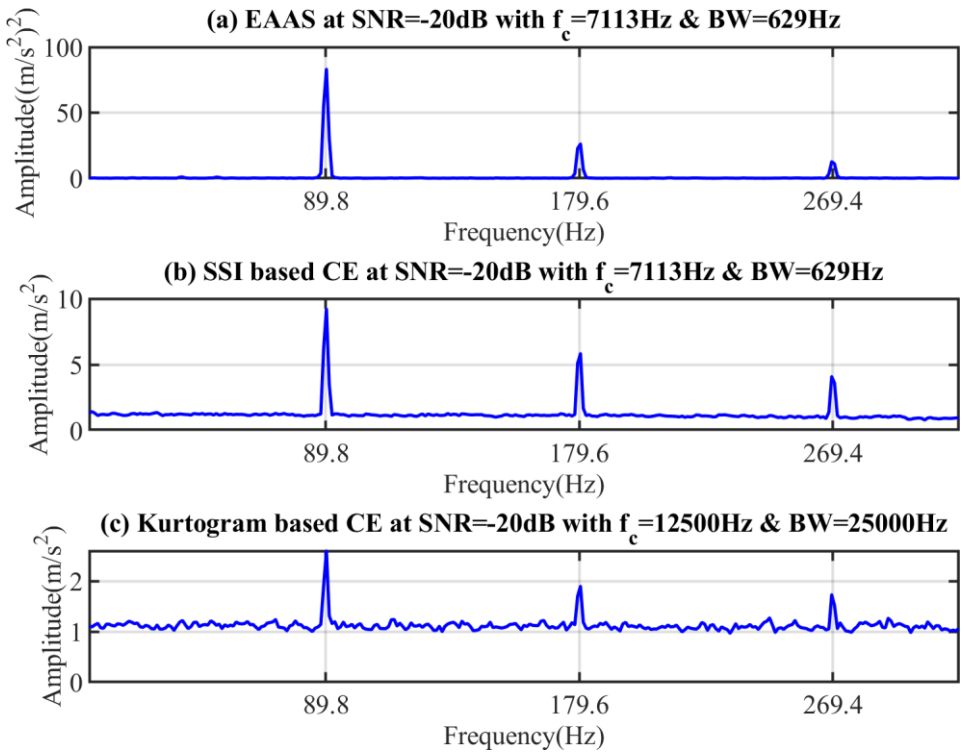


Figure 4-30 EAAS and CE of outer race faults at SNR -20dB

When the bearing signals are heavily contaminated into a level of the SNR at -25dB, the EAAS in the Figure 4-31 (a) shows three sparse components which are corresponding to the theoretical fault frequency and higher order harmonics. The Kurtogram based CE in the Figure 4-31 (c) fails to extract the fault information with a wide frequency band although the frequency band contains the carrier frequencies. In this circumstance, the monitoring results of the Kurtogram based CE are misleading. In contrast, the SSI based CE in the Figure 4-31 (b) gives an impressive statement that the outer race faults were accurately detected by a narrow frequency band around the natural frequency 7113Hz. The EAAS has a great capacity of noise reduction and hence, the characteristic fault frequency and the harmonics are obvious in the results.

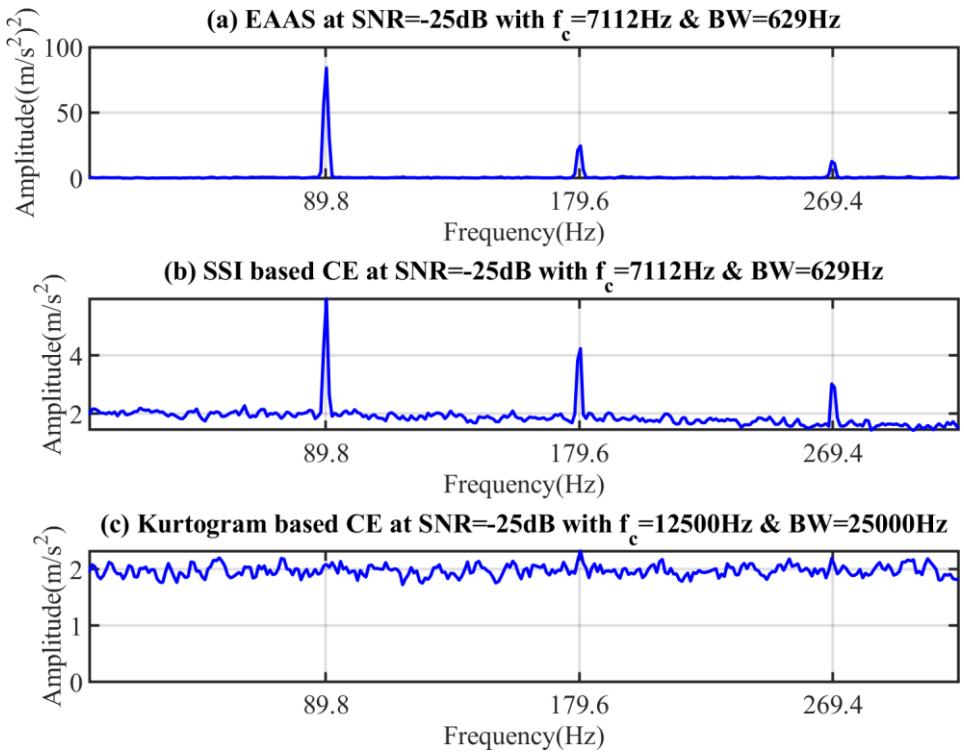


Figure 4-31 EAAS and CE of outer race faults at SNR -25dB

If the signal is more challenging, the methods without the ability of noise suppression are very difficult to use to extract the fault information behind the random noise. In this simulated case, the Signal to Noise Ratio becomes -30dB. As shown in Figure 4-32 (a), the EAAS upon the frequency band selected by the SSI approach displays a satisfied result and the spectrum clearly shows the characteristic frequency of the outer race defects and its harmonics. It is not beyond the expectation that the Kurtogram based CE in Figure 4-32 (c) is impossible to demodulate the fault signatures from such noisy signals because the frequency band is so large that too much noise is involved in the demodulation process. When the frequency band is optimal, the CE has the potential to find the fault features. As shown in Figure 4-32 (b), the frequency peaks obtained by the SSI based CE can be seen above the noise floor but the noise carpet is too high to submerge the fault signatures, therefore the incipient fault detection and diagnosis desired is outstanding capacity of noise reduction.

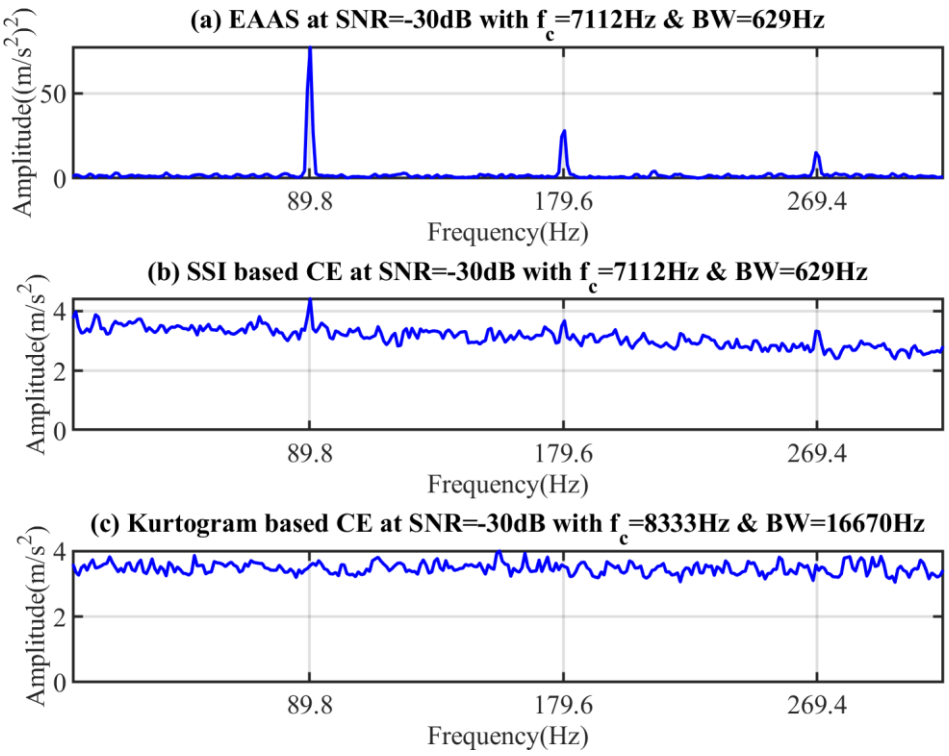


Figure 4-32 EAAS and CE of outer race faults at SNR -30dB

The worst situation in the outer race fault simulation study is that the SNR of the signal becomes -35dB, which leads to the useful fault information relating to the fault is totally submerged by the random noise in both time and frequency domain. In this circumstance, the demodulation methods, such as conventional envelope are no longer effective no matter how well the frequency bands are selected. From the results of EAAS in the Figure 4-33 (a), the proposed method is still operative owing to the reduction of the random noise. In contrast, the CE in the Figure 4-33 (b) and (c) fails to detect the bearing faults based on the frequency bands determined by either the Kurtogram or the SSI.

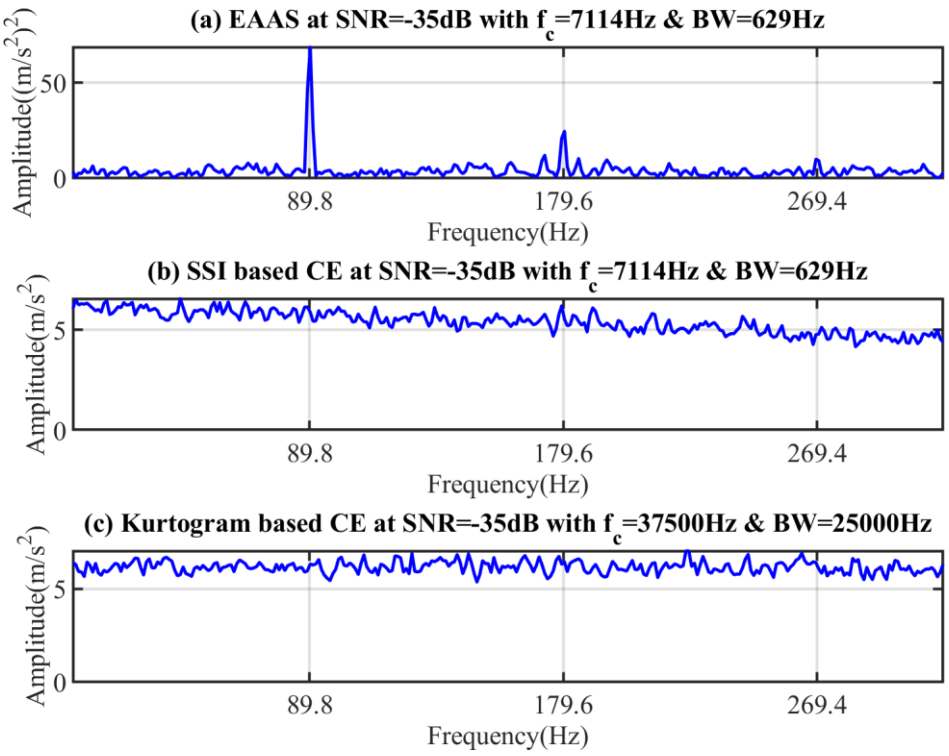


Figure 4-33 EAAS and CE of outer race faults at SNR -35dB

With the increase of the additive noise, the performance of CE based on the optimal frequency band is degraded gradually because the envelope cannot suppress random noise. It is evident that EAAS achieves the best results in the outer race fault detection and diagnosis.

4.4.3 Quantification of Outer Race Fault Signatures

To quantify the fault signatures extracted by the EAAS and CE, a strength indicator of bearing fault signatures is designed by following the concept of the Signal to Noise Ratio and the indicator is denoted as

$$SNR_{signature} = 10\,log_{10}\left(\frac{\frac{1}{N_h}\sum_{k=1}^{N_h} A_h(k)^2}{\frac{1}{N_s}\sum_{k=1}^{N_s} A_s(k)^2}\right) \qquad (4.68)$$

where, N_h and A_h are the length and amplitude of the fundamental characteristic fault frequency and its harmonics; N_s and A_s are the length and the amplitude of the remaining values in the spectrum. In this study, the limit of the spectrum used to calculate the strength of the fault features are from 10Hz to 314.3Hz (3.5 times of characteristic fault frequency). To decrease the drawback of the spectral leakage, a short frequency band with the width of 6Hz around the fault frequency and harmonics is adopted to include the energy of the fault frequency as much as possible. The frequency band of the fault features is determined by the width of the characteristic fault frequency in the spectrum.

The signature indicators in the Figure 4-34 show that the proposed method is more effective than the conventional envelope. The feature strength of EAAS decreases with the increase of noise levels but the $SNR_{signature}$ is above 14dB in all six simulated cases. The maximum $SNR_{signature}$ of the SSI based CE and Kurtogram based CE are 22dB and 11dB respectively. With the increase of the noise, the fault signatures are degraded gradually and in the last case, the fault indicator is approximately zero, which means the fault information is submerged by the random noise.

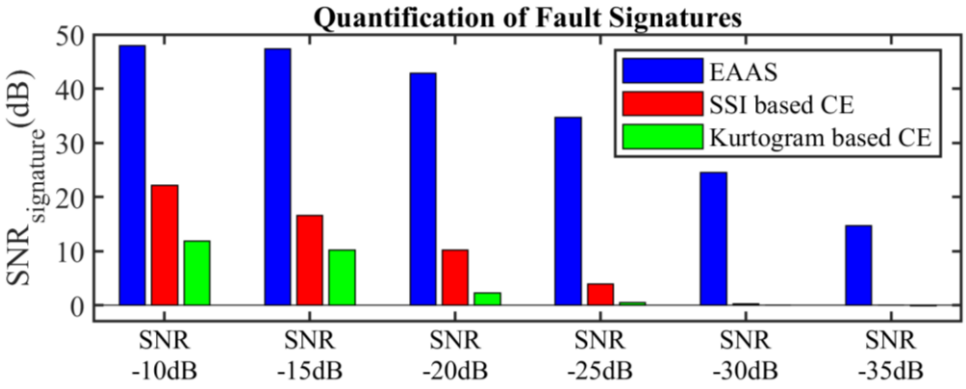


Figure 4-34 Quantification of outer race fault signatures

4.5 Demodulation Analysis of Inner Race Fault Signals

Similar to the simulation studies of the outer race faults, investigation of the inner race faults was carried out by following the same procedure, which are the selection of the optimal frequency bands and the demodulation analysis.

4.5.1 Determination of Optimal Frequency Bands

4.5.1.1 Effect of Gaussian Noise

The inner race faults are more complex than the outer race faults in nature because the location of the defects is rotating along with the inner race of the bearing. Consequently, the input forces of the excitation are varying with respect to the relative angle between the localised defects and the loading zone of bearings. The varying inputs themselves are a modulation signal and then the modulation forces result in the further modulation phenomenon by the transfer function of the bearing system. The inner race fault signatures are usually comprised of the theoretical fault frequency and the sidebands with the periodicity of the varying forces.

Six cases of the bearing inner race fault are simulated to be the SNR from -10dB to -35dB and the investigation is repeated to study the performance of the proposed methods. Figure 4-35 and Figure 4-36 display the simulated inner race fault signals in the time and frequency domain respectively. The SNRs of the signals decrease from -10dB to -35dB with an interval of -5dB.

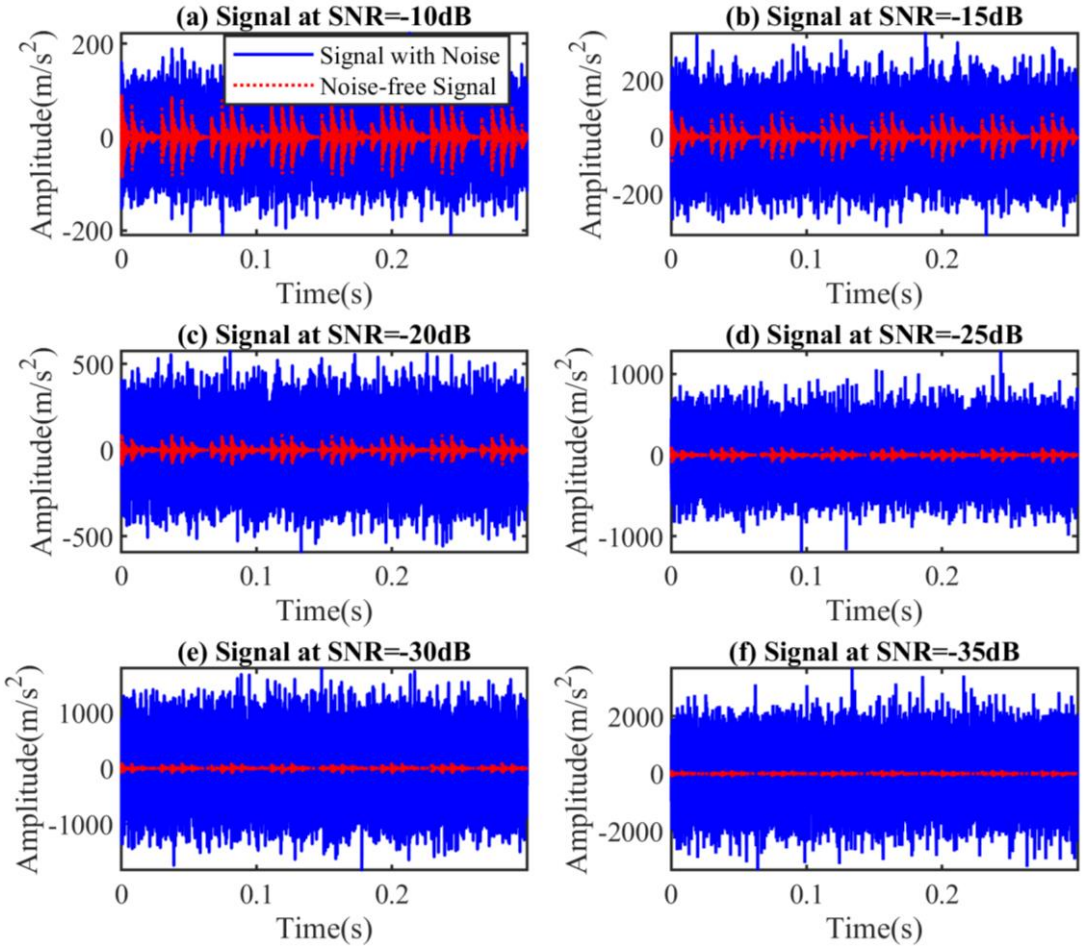


Figure 4-35 Time waveform of the simulated inner race fault signals at SNR of : (a) -10dB, (b) -15dB, (c) -20dB, (d) -25dB, (e) -30dB, (f) -35dB

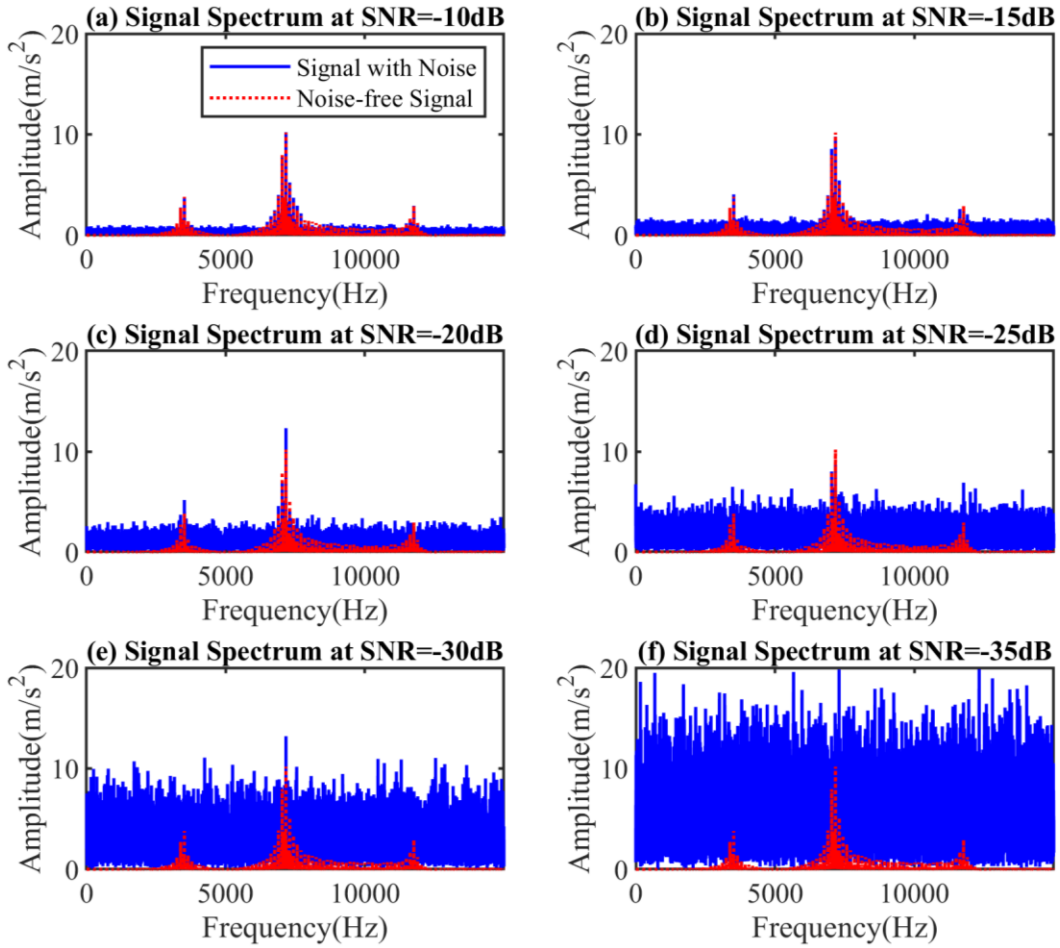


Figure 4-36 Spectra of the simulated inner race fault signals at SNR of : (a) -10dB, (b) -15dB, (c) -20dB, (d) -25dB, (e) -30dB, (f) -35dB

The inner race fault signals are more challenging compared to the outer race faults because the signals contain not only the characteristic fault frequencies but also the sidebands. The Kurtogram in Figure 4-37 depicts the most impulsive frequency band is centred at 7031Hz with a frequency bandwidth of 1563Hz, which is the theoretically optimal frequency band. This case is a clearly self evident in bearing detection and diagnosis. The noise added is white Gaussian noise and moreover the signal quality is very good. Therefore, the frequency band selected by the Kurtogram is the most appropriate. The SSI based frequency band selection can identify the modal parameters but the harmonics make the OMA method extract a fake mode in the stabilisation diagram of Figure 4-38. Based on the Normalised Ratio in Figure 4-39, the threshold for selecting modes is very high and it gives up the fake mode as well as a real mode, which is a deficiency of stochastic subspace identification. Several methods can improve this disadvantage, and the application of SSI in this scenario allows the existence of multiple modes because the fault signatures extracted from several candidate frequency bands can be easily compared to find the most effective.

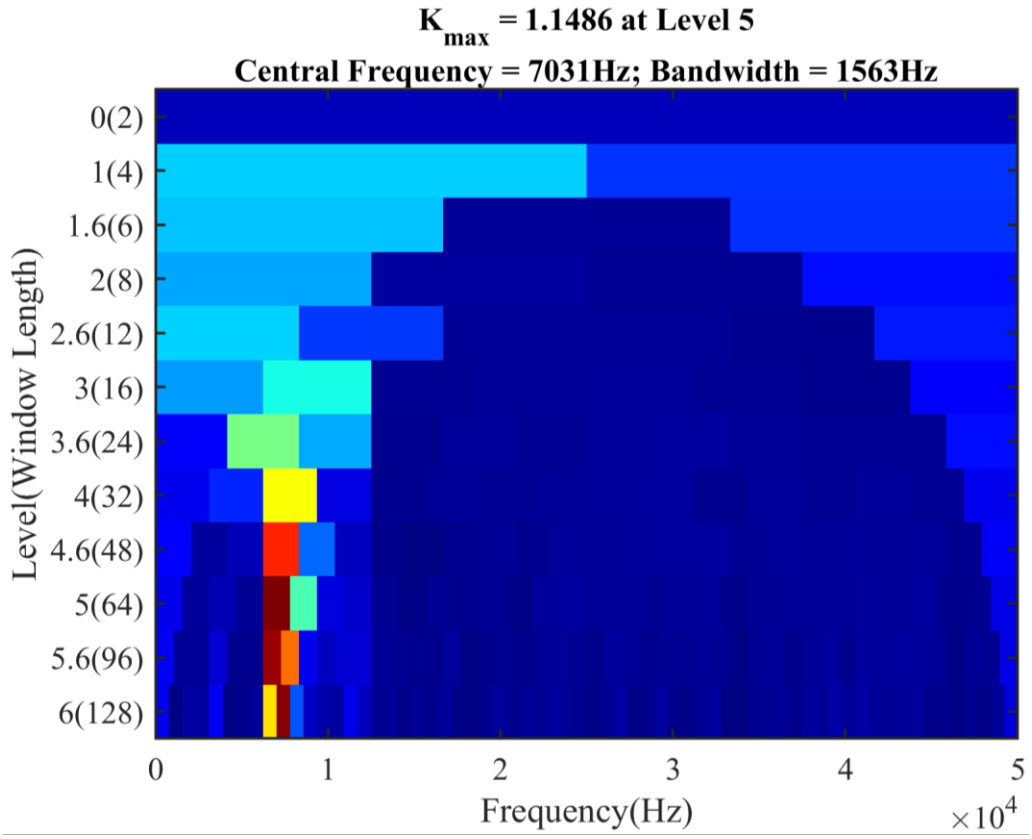


Figure 4-37 Kurtogram for the case of inner race faults at SNR -10dB

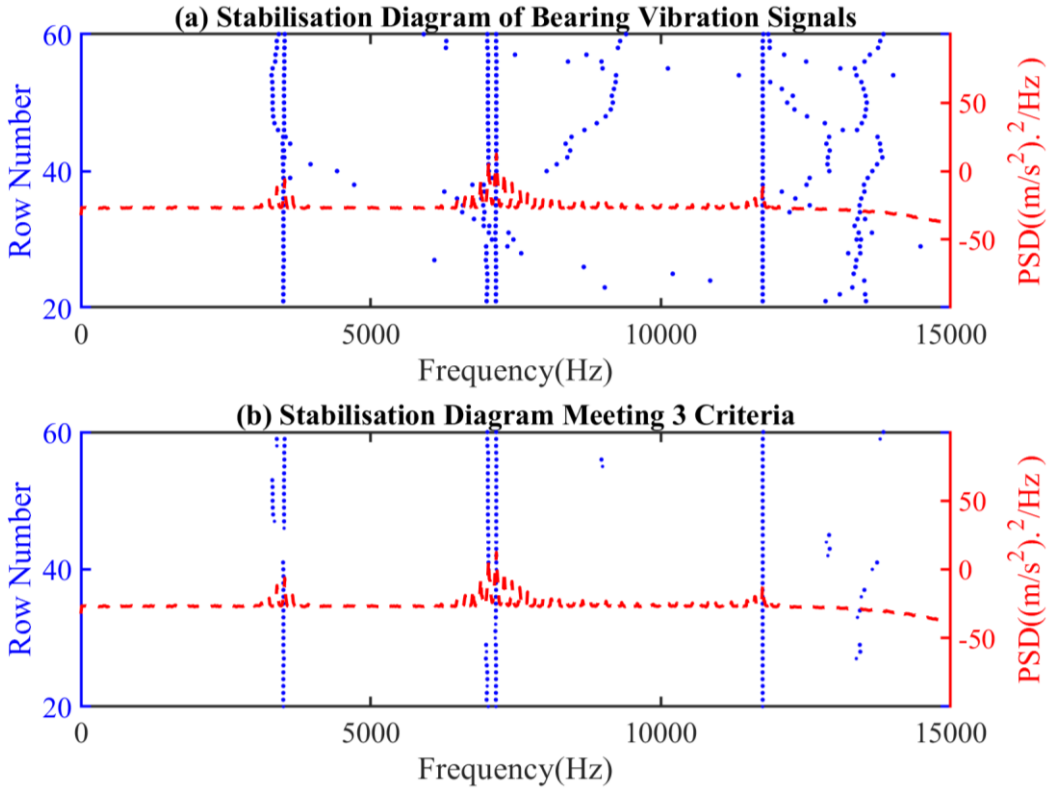


Figure 4-38 Stabilisation Diagram of SSI for the case of inner race faults at SNR -10dB: (a) all modes identified; (b) modes meeting three criteria

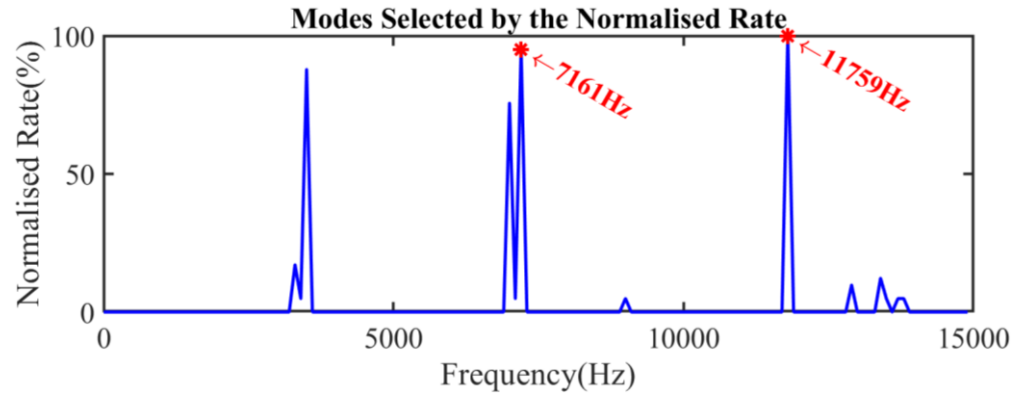


Figure 4-39 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at SNR -10dB

In the case of the inner race fault signal at SNR -15dB, both Kurtogram and SSI can identify the optimal central frequency, which is shown in Figure 4-40 and Figure 4-41 respectively. The frequency bands are close to each other, which are the second mode of the bearing system. This case is similar to the case study in the outer race fault diagnosis.

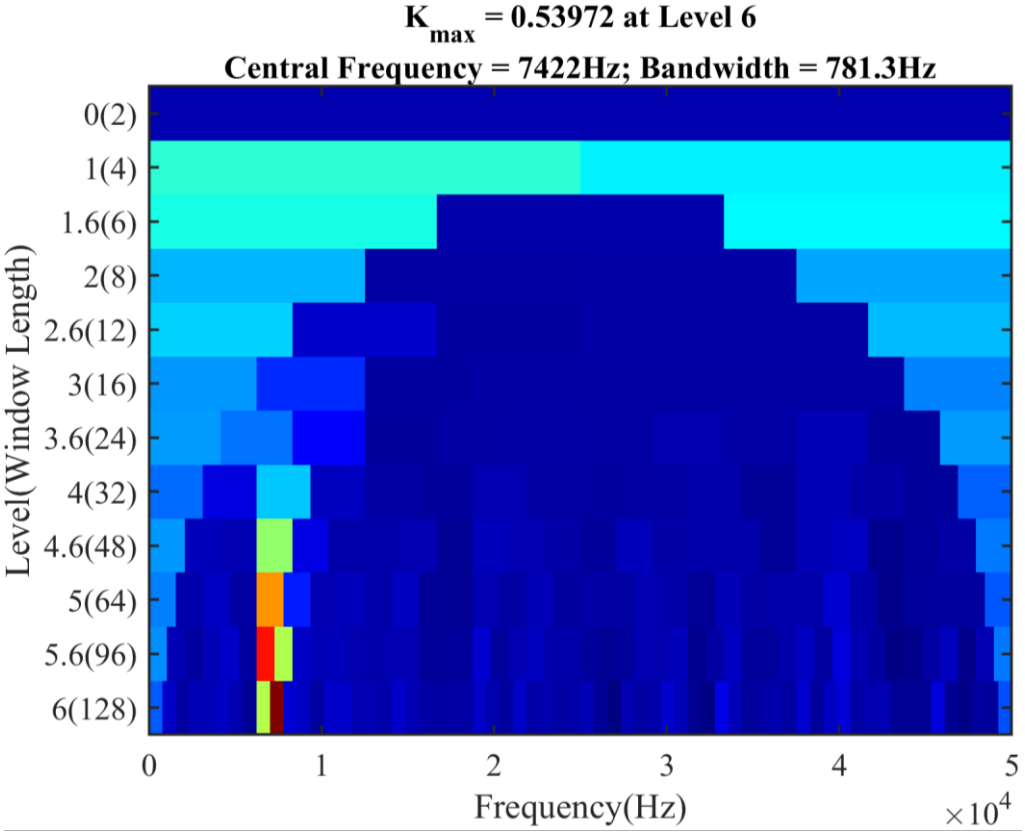


Figure 4-40 Kurtogram for the case of inner race faults at SNR -15dB

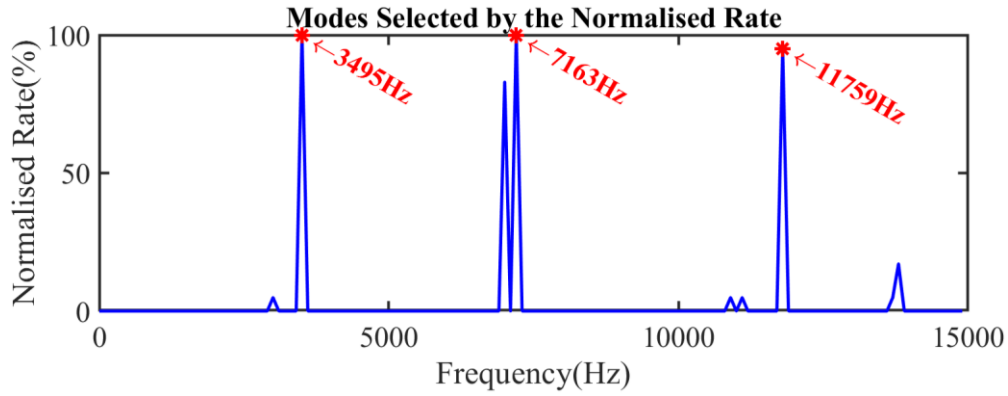


Figure 4-41 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -15dB

When bearing signals are more severely contaminated by the white Gaussian noise, the Kurtogram fails to select the optimal frequency band. As displayed in Figure 4-42, the frequency band picked by the Kurtogram is from 0Hz to 25,000Hz. It is different with the benchmark method in that the SSI in Figure 4-43 identifies the three carrier frequencies but only two modes are selected due to the noise interference.

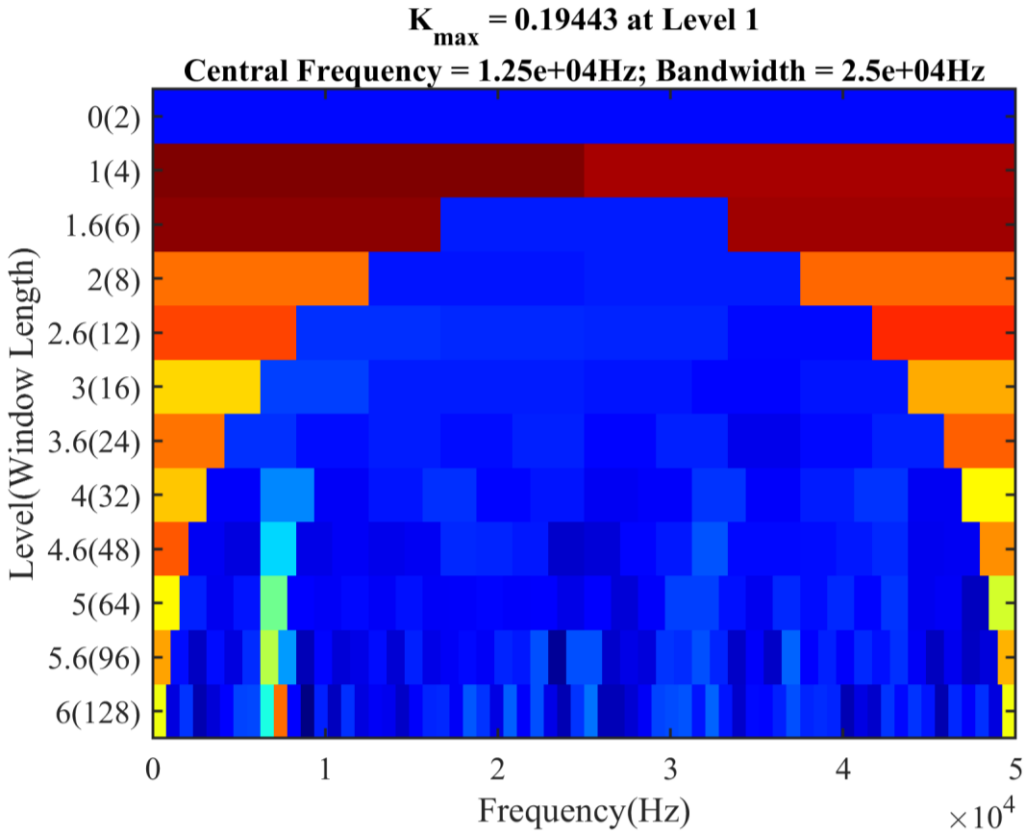


Figure 4-42 Kurtogram for the case of inner race faults at SNR -20dB

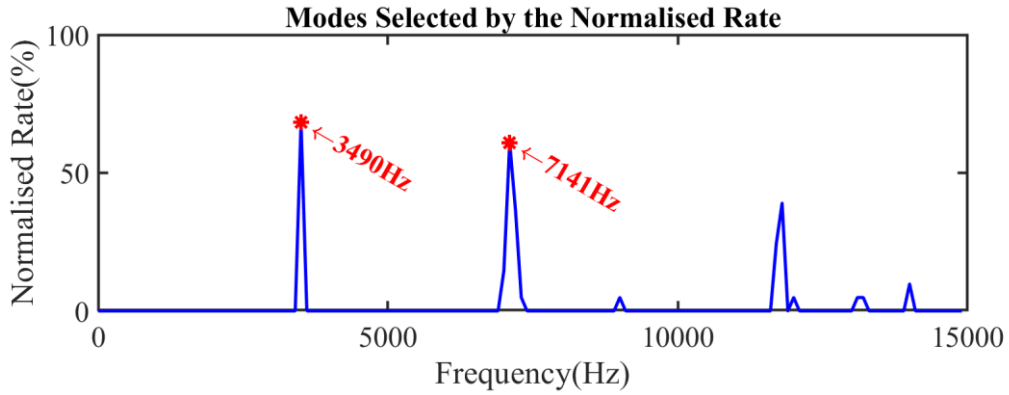


Figure 4-43 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -20dB

The simulated signals at SNR from -25 to -35dB are significantly difficult to handle by the Kurtogram because of the negative impact of the random noise, which are shown in the Figure 4-44 Figure 4-46 and Figure 4-48. As displayed in Figure 4-45 Figure 4-47 and Figure 4-49, the performance of the SSI approach also deteriorates but it is still able to obtain the carrier frequencies. The candidate central frequencies allow the effective and reliable fault detection and diagnosis.

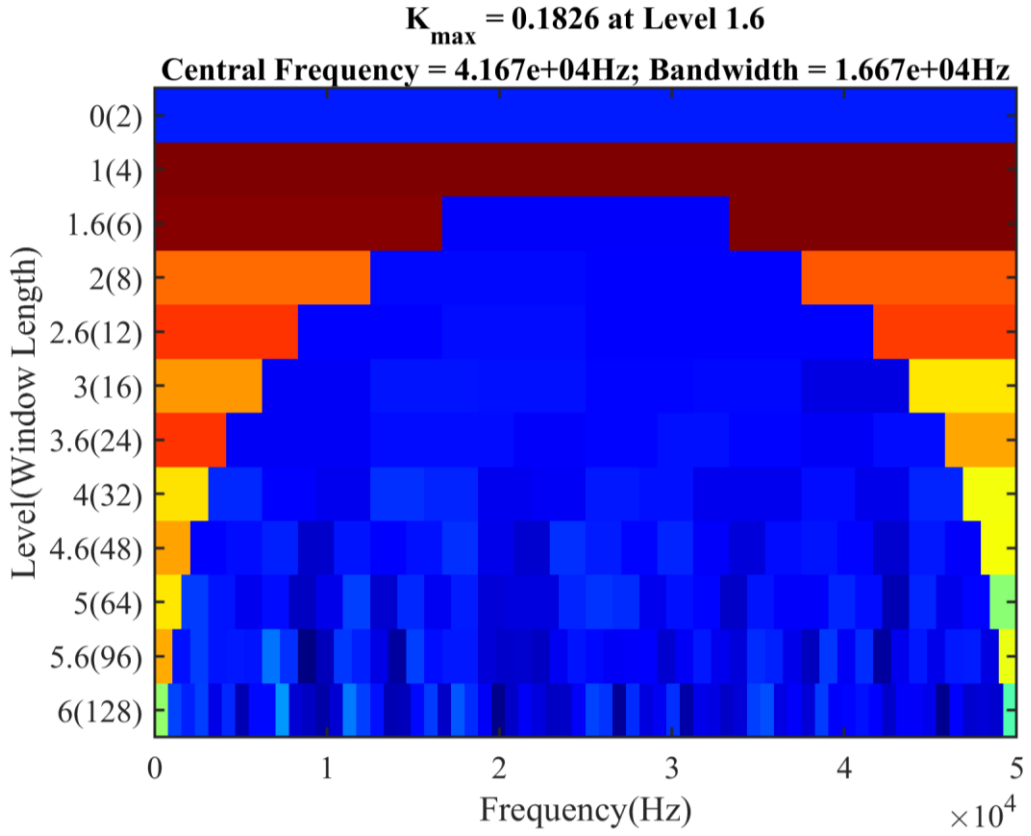


Figure 4-44 Kurtogram for the case of inner race faults at SNR -25dB

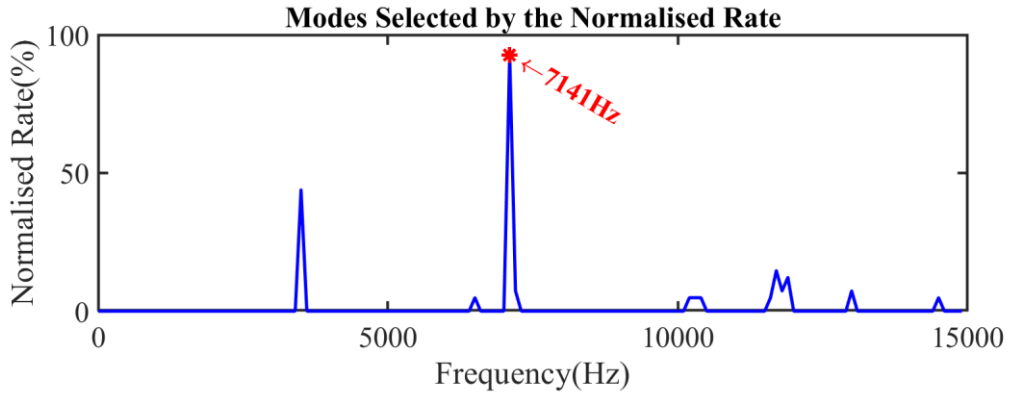


Figure 4-45 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -25dB

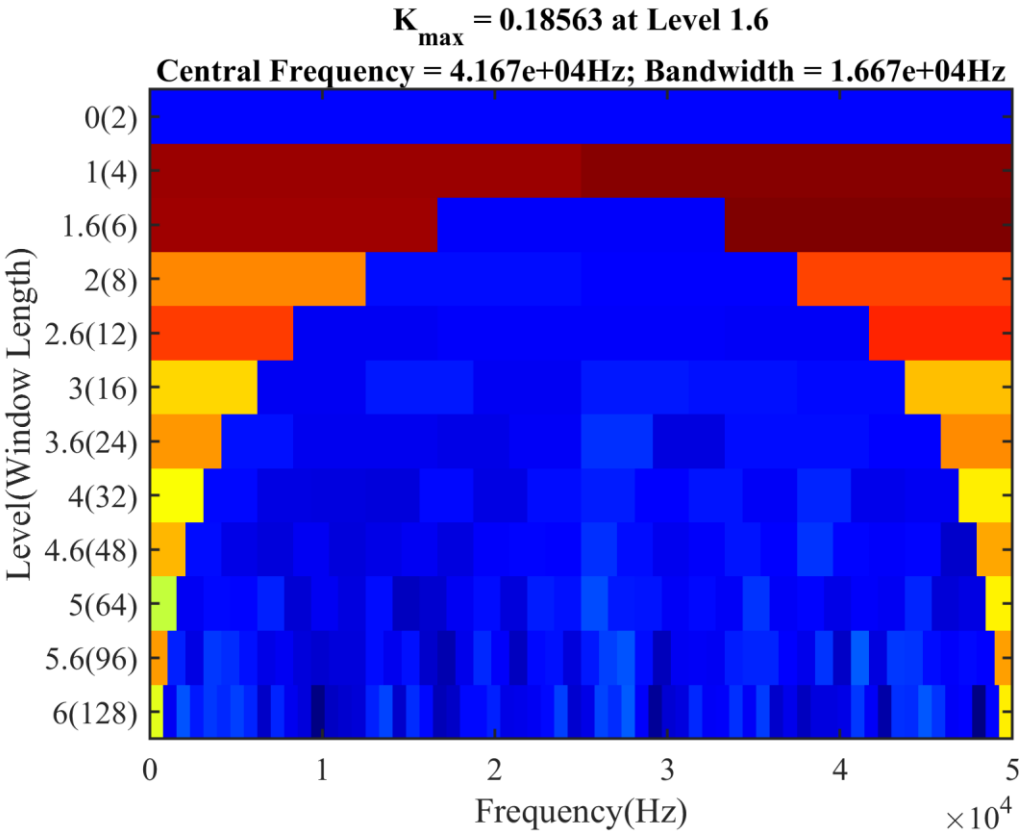


Figure 4-46 Kurtogram for the case of inner race faults at SNR -30dB

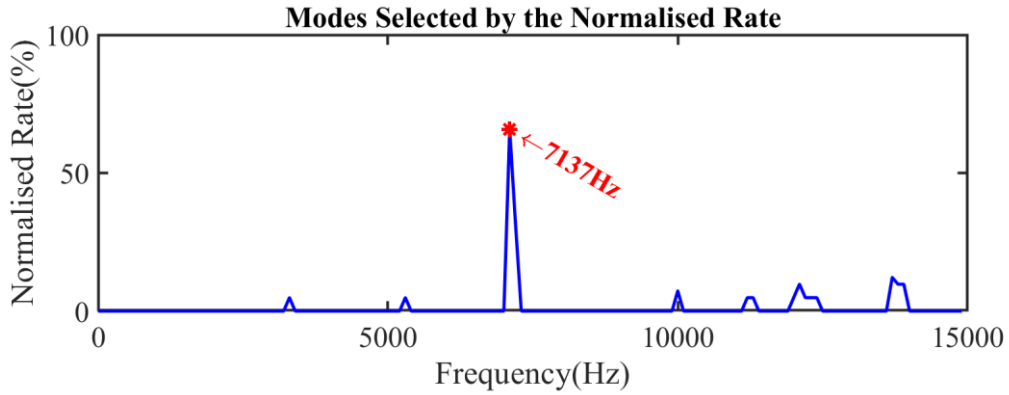


Figure 4-47 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -30dB

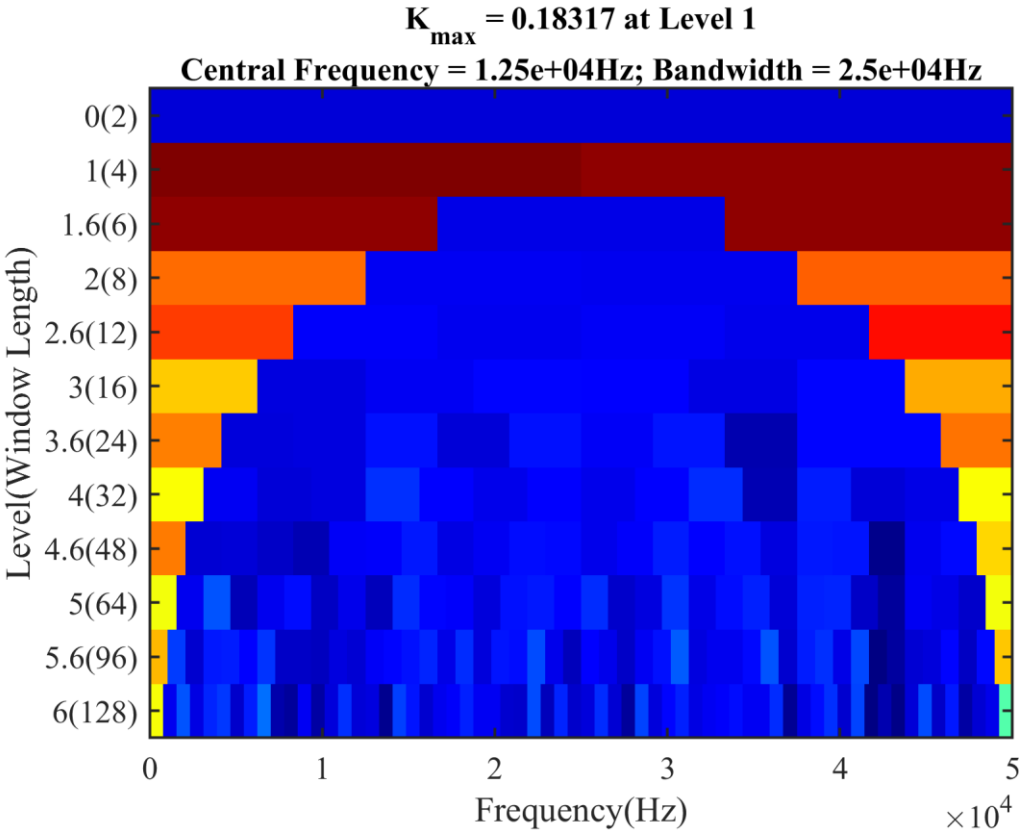


Figure 4-48 Kurtogram for the case of inner race faults at SNR -35dB

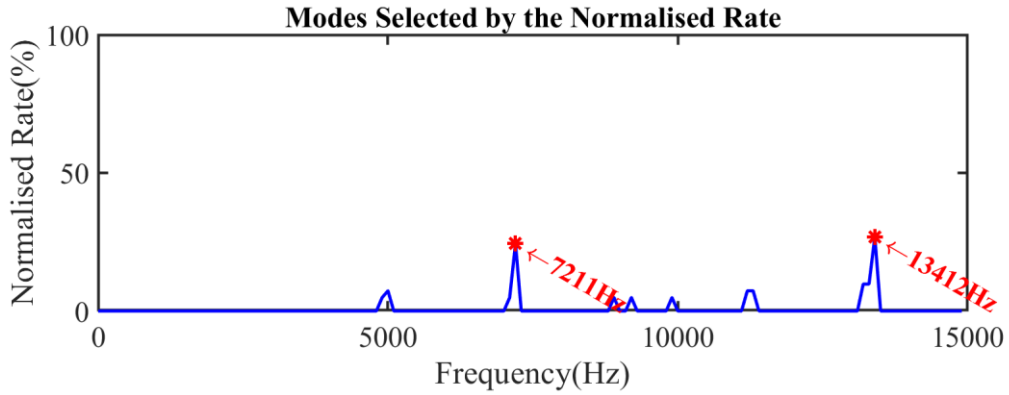


Figure 4-49 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at SNR -35dB

4.5.1.2 Effect of Non-Gaussian Noise

By following the same way in the outer race simulation, a series of random pulses are generated, of which the amplitude is two times of the maximum values of the noisy signals. Figure 4-50 and Figure 4-51 exhibit the temporal waveforms and spectrum of the bearing fault signals with non-Gaussian noise.

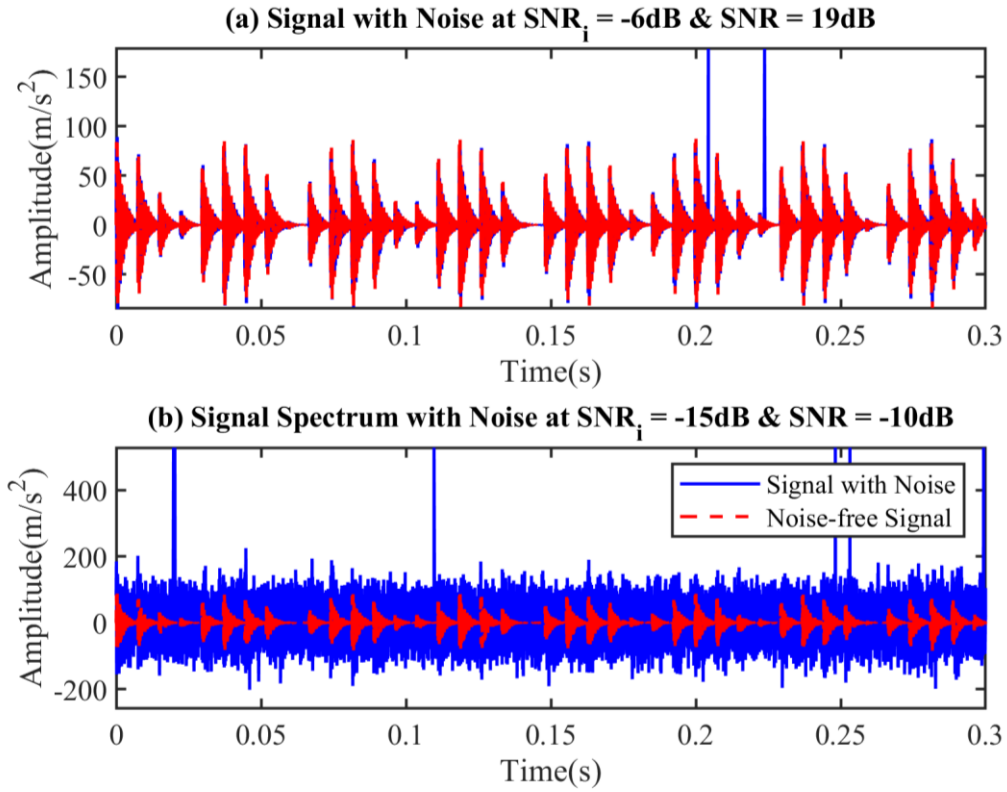


Figure 4-50 Time waveform of the simulated inner race fault signals with non-Gaussian noise at: (a) SNR_i -6dB and SNR 19dB, (b) SNR_i -15dB and SNR -10dB

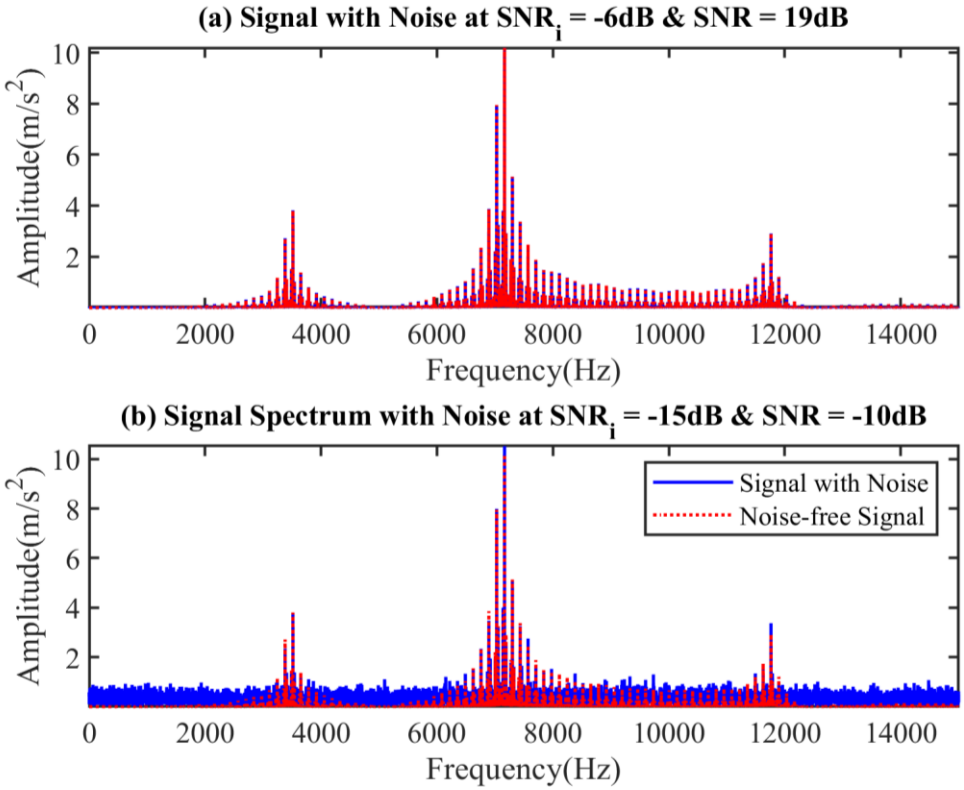


Figure 4-51 Spectra of the simulated outer race fault signals with non-Gaussian noise at: (a) SNR_i -6dB and SNR 19dB, (b) SNR_i -15dB and SNR -10dB

The determination of the optimal frequency bands by the Kurtogram and SSI approaches are shown in Figure 4-52 and Figure 4-53 respectively. Due to the unfavourable influence, the Kurtogram cannot find the right frequency band even though the SNR of the simulated signals is high. In contrast, the SSI extract the three modes of the bearing model. A fake mode around 7100Hz is also identified due to the influence of the harmonics.

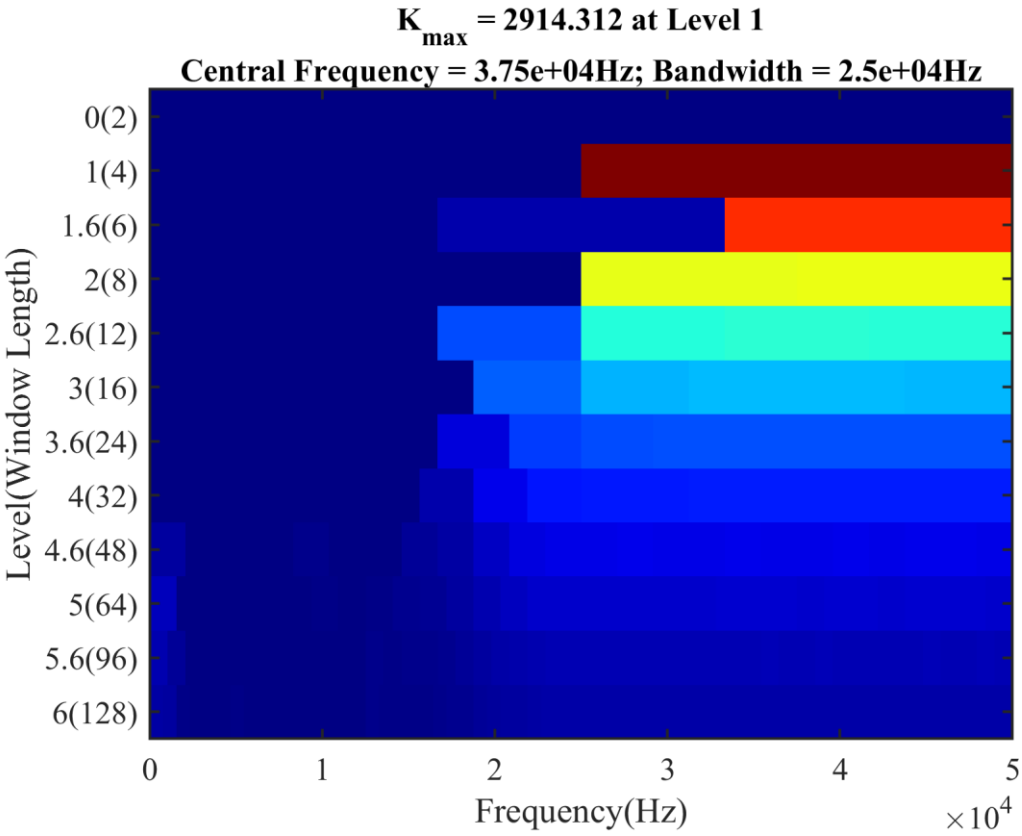


Figure 4-52 Kurtogram for the case of inner race faults at SNR_i -6dB and SNR 19dB

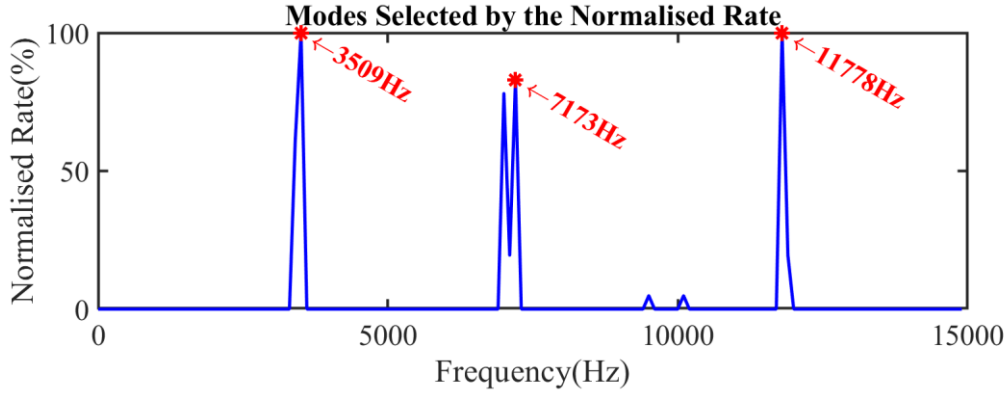


Figure 4-53 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at SNR_i -6dB and SNR 19dB

Figure 4-54 and Figure 4-55 shows the frequency band recommended by the Kurtogram and SSI. It is not beyond expectation that the Kurtogram is unable to find the proper frequency band with the increase of the Gaussian and non-Gaussian noise. The normalised rate in the SSI results shows that the four modes including three real and one fake are identified. The fake modes cause issues in the dynamic analysis of the systems and substantial quantities of research has been done to improve the accuracy of the identification. The fakes modes are not a big problem in the demodulation analysis because the method is used to select a frequency band for the demodulation.

The fake modes are mainly from the harmonics in the bearing signals and consequently the frequency is mainly within the frequency band. Moreover, these candidates of the frequency bands are easily compared, and the proposed indicator of the fault signatures is able to make the comparison and selection automatically.

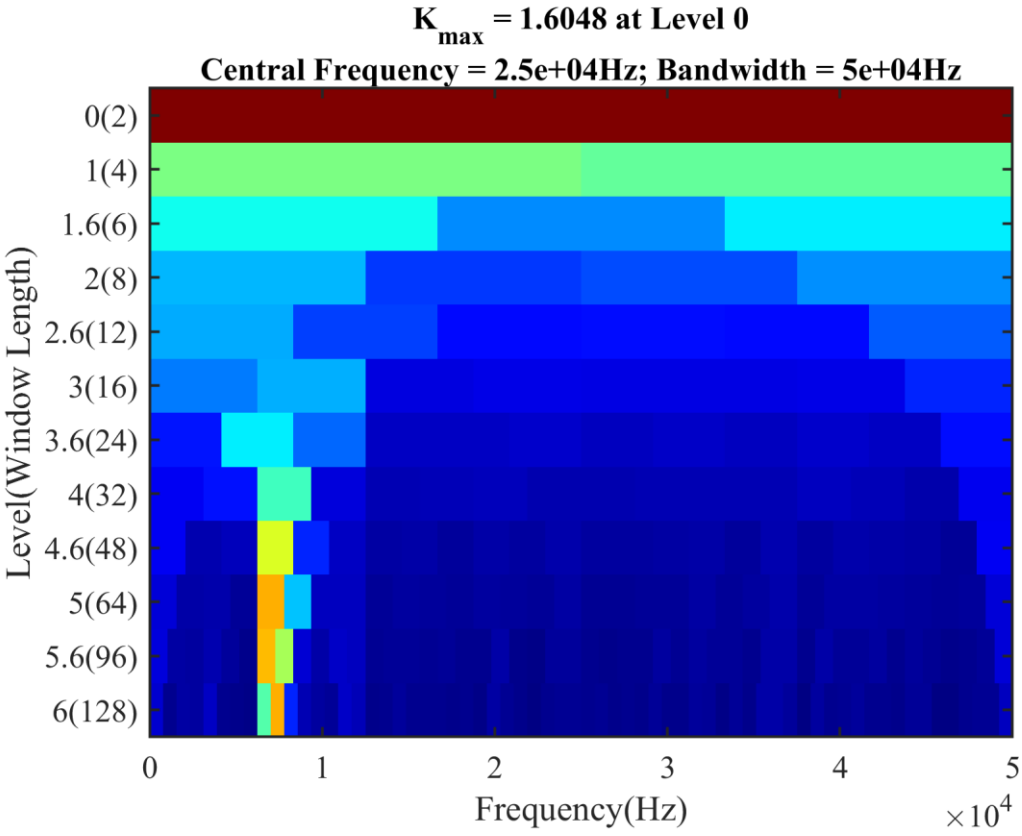


Figure 4-54 Kurtogram for the case of inner race faults at SNR_i -15dB and SNR -10dB

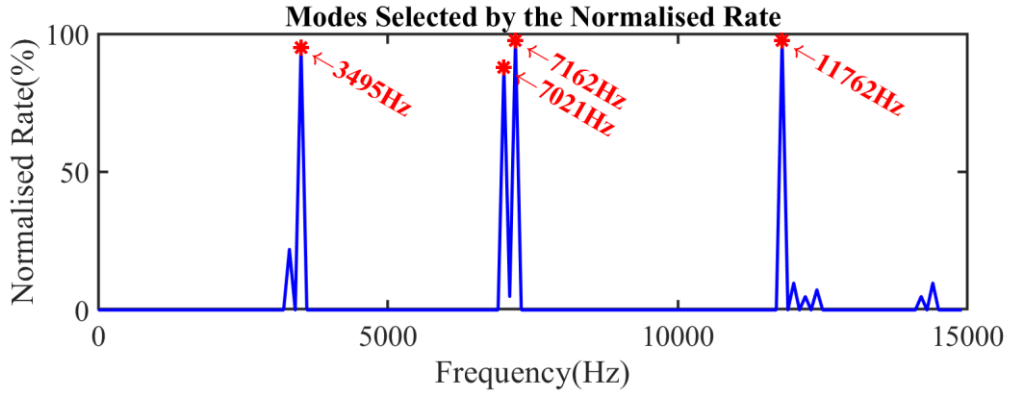


Figure 4-55 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at SNR_i -15dB and SNR -10dB

4.5.2 Demodulation Results and Analysis

Based on the frequency bands selected by the Kurtogram and SSI, the results of EAAS and CE are compared in the fault detection and diagnosis of inner race faults using six simulated cases. Figure 4-56 and Figure 4-57 show that the EAAS and CE based on the optimal frequency band can lead to the effective diagnosis owing to the high SNR of the signals.

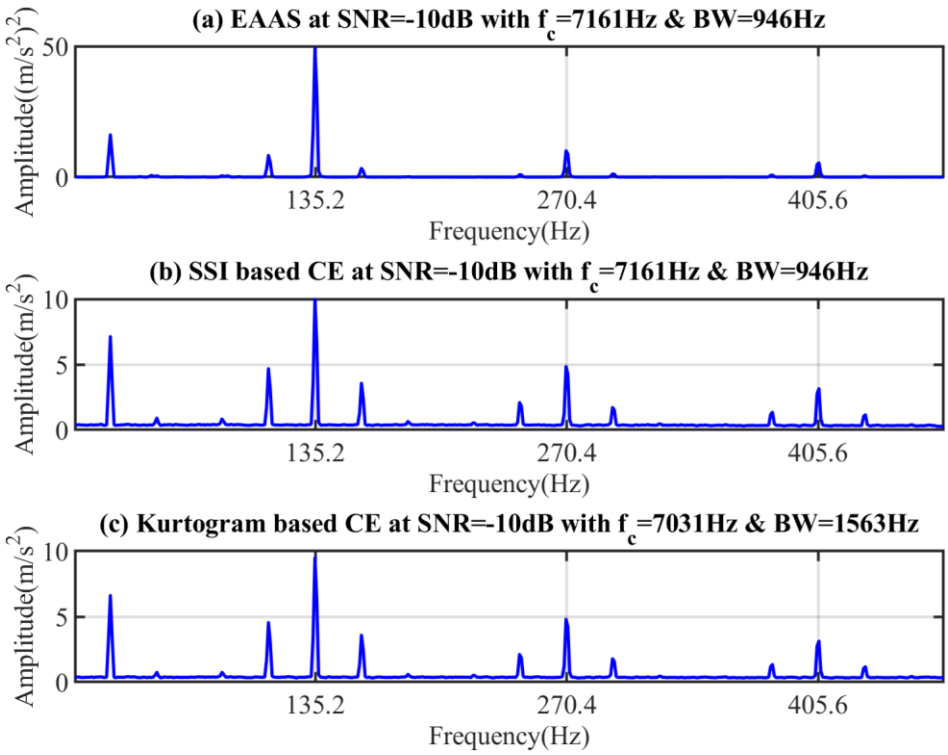


Figure 4-56 EAAS and CE of inner race faults at SNR -10dB

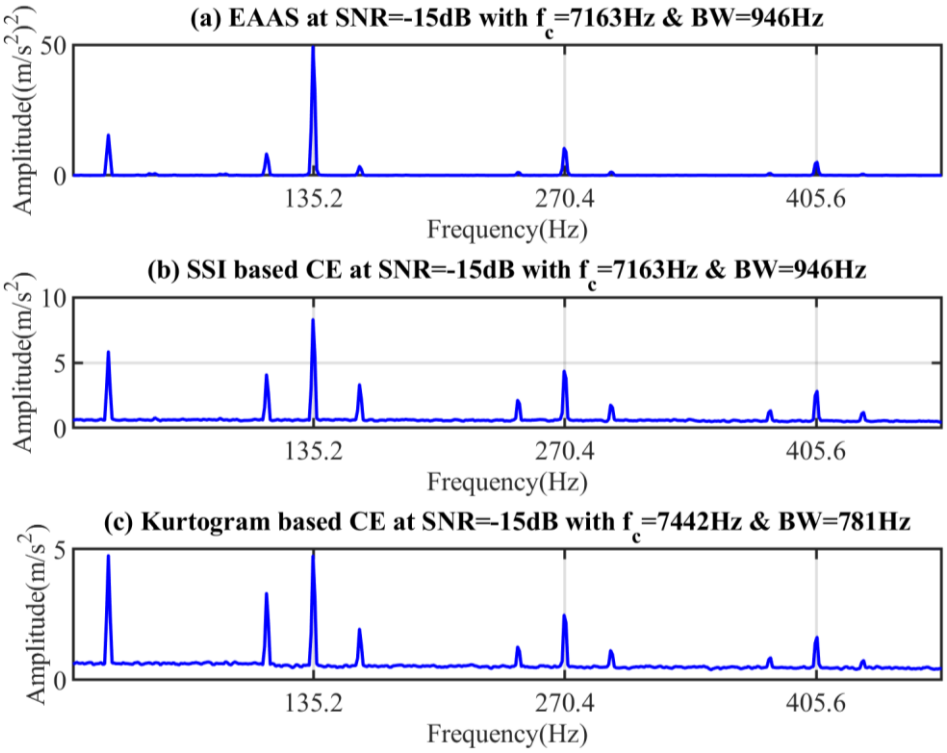


Figure 4-57 EAAS and CE of inner race faults at SNR -15dB

When the simulated inner race fault signal was contaminated by the random noise to the SNR of -20dB, the EAAS in Figure 4-58 (a) shows a significantly sparse spectrum, which only contains the shaft rotating frequency, the characteristic fault frequency, the fault frequency harmonics, and the sidebands. This EAAS result gives a very satisfied diagnosis with a nearly zero noise floor. The SSI based CE is displayed in Figure 4-58 (b) and the spectrum is rich in fault information. The undesired issue is that the noise floor is high due to the strong background noise. The worst result in this case is obtained by the Kurtogram based CE, which is displayed in Figure 4-58 (c). The noise floor is high and the amplitude of fault features is not pronounced compared to the noise floor. In addition, the sidebands of the second and third harmonic are not visible in the spectrum, which denotes the performance of the Kurtogram based CE is not satisfied in the inner race fault detection.

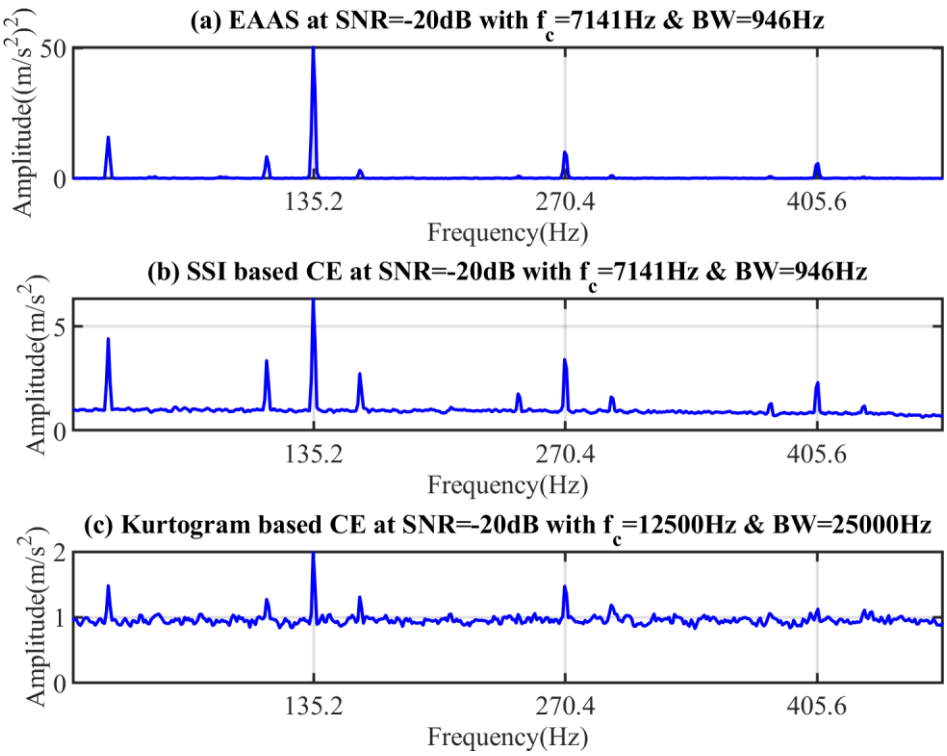


Figure 4-58 EAAS and CE of inner race faults at SNR -20dB

The vibration signals of the inner race fault are contaminated to the level of SNR -25dB. The demodulation results in Figure 4-59 denote that the EAAS is the most effective and efficient approach compared to the other two means. The fault signatures extracted by EAAS is of high quality, which has not only the high amplitude of fault frequency but also the extremely low noise floor. The SSI based CE with a proper frequency band in Figure 4-59 (b) can detect and diagnose the inner race faults whilst the Kurtogram based CE in Figure 4-59 (c) does not succeed in extraction of the defect features. Based on the results in Figure 4-59 (b) and (c), the main problem for this CM failure is the unsuitable frequency band, which results in a misleading result.

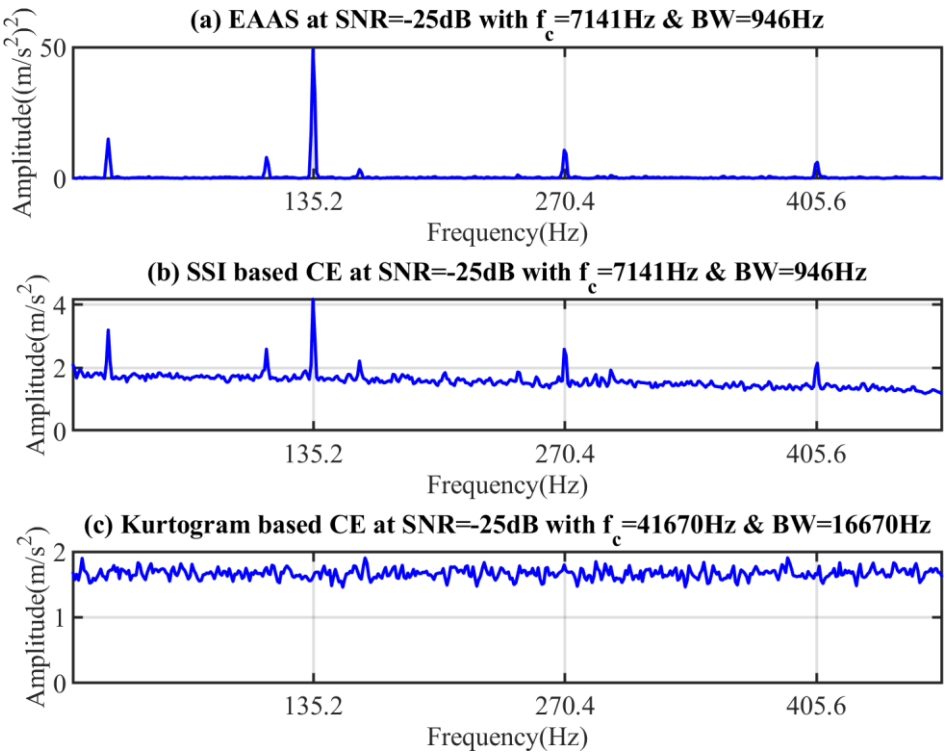


Figure 4-59 EAAS and CE of inner race faults at SNR -25dB

When the vibration signals are more severely distorted by the random noise, the capability of noise reduction becomes critical in the early fault detection and diagnosis of rolling bearings. The spectra in the Figure 4-60 and Figure 4-61 indicate that the CE cannot obtain the fault information from such a noisy signal. No matter how accurate the frequency bands are selected, the approaches without noise suppression are unable to tackle the challenging task of fault detection and diagnosis in a very early stage. To fulfil the early fault detection and diagnosis, the proposed EAAS is very capable to suppress the random noise and enhance the periodic fault features. Therefore, with the frequency band from SSI, the EAAS successfully detects the inner race defects of the rolling bearing.

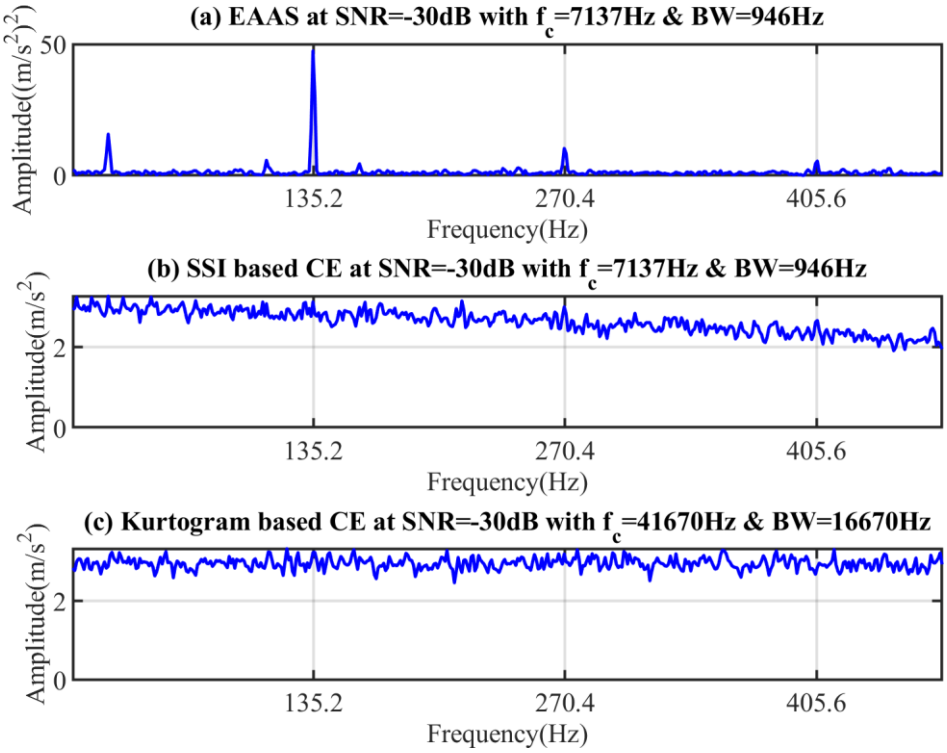


Figure 4-60 EAAS and CE of inner race faults at SNR -30dB

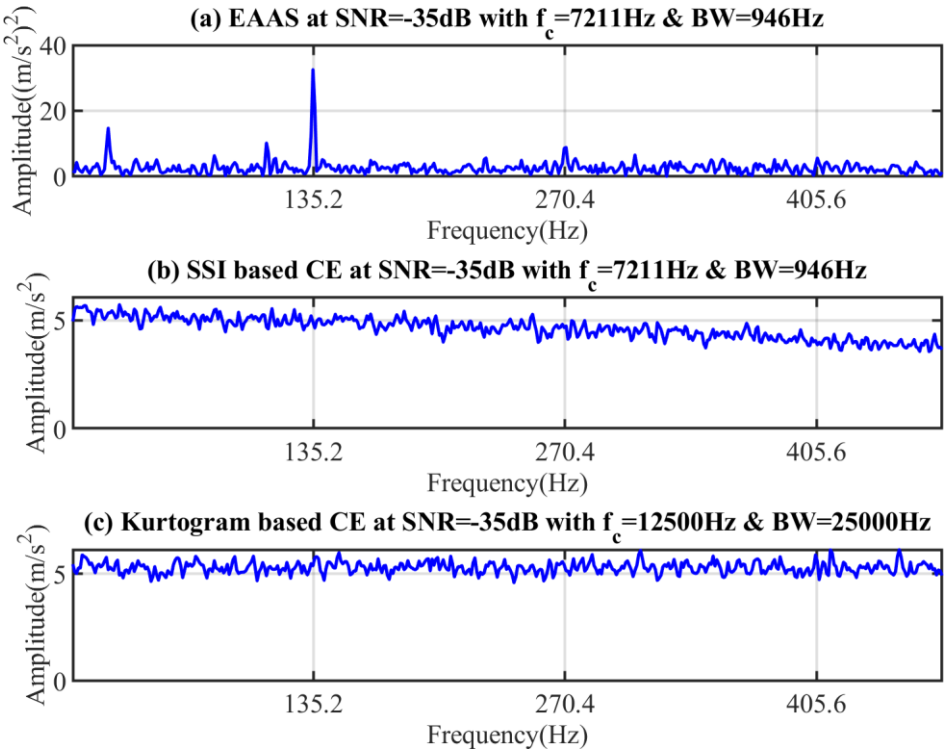


Figure 4-61 EAAS and CE of inner race faults at SNR -35dB

4.5.3 Quantification of Inner Race Fault Signatures

The demodulation results show the EAAS has the best performance in bearing fault detection and diagnosis. The strength of the fault signatures is quantified by the indicator $SNR_{signature}$ proposed in Equation (4.68). In comparison with the outer race faults, the inner race faults have the sidebands

due to the excitation variation. The sidebands around the fault frequency are highly correlated with the fault features and consequently, the sidebands are also taken in account in A_h. The limit of the spectrum used to calculate the strength of the fault features are from 10Hz to 473.2Hz (3.5 times of characteristic fault frequency). To decrease the drawback of the spectral leakage, a short frequency band with the width of 6Hz is adopted to include the energy of the fault frequency as much as possible. The frequency band of the fault features is determined by the width of the characteristic fault frequency in the spectrum.

Figure 4-62 shows that the signature indicators $SNR_{signature}$ of EAAS is more effective than the conventional envelope. With the increase of the noise, the fault signatures are degraded gradually and in the last cases, the signature indicator of CE is approximately zero, which means the fault information is submerged by the random noise.

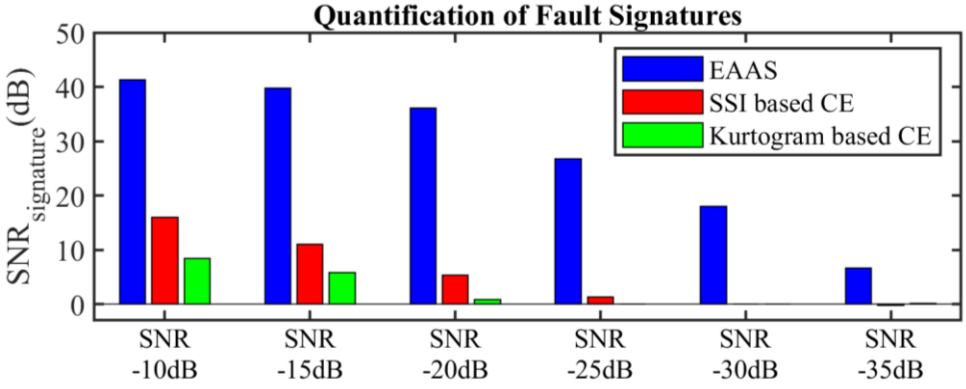


Figure 4-62 Quantification of inner race fault signatures

4.6 Summary

This chapter investigates the periodic input based modulation signals from the rolling element bearings. To simulate the dynamic responses under the deterministic excitations, a state space model of a rolling bearing with three degree of freedom was developed to serve the simulation studies in this thesis. A series of periodic impacts was created to excite the linear bearing model for simulating the vibration responses. The prior step of the demodulation analysis is usually to determine the optimal frequency bands. The Kurtogram is a widely used approach for selecting the optimal frequency bands, which looks for the frequency bands containing the most impulsive behaviour. The resonant modulation is actually the resonance of the system and consequently, an operational modal analysis method, SSI, is employed to select the optimal central frequency for the demodulation analysis. Benchmarked by the Kurtogram, the SSI based optimal frequency bands are around the natural frequencies of the bearing system, which plays the role of the carrier frequencies in the resonant modulation. The SSI approach is more robust to both Gaussian and

non-Gaussian noise. A novel method, named ensemble average of autocorrelation signals (EAAS), was developed to suppress the random noise for effective and efficient fault detection and diagnosis at a very early stage. Both outer race fault signals and inner race fault signals at six levels of SNR from -10dB to -35dB demonstrate that the EAAS is a more reliable and more accurate approach for identifying the fault signatures from severely contaminated vibration signals.

Chapter 5 Experimental Study of Ball Bearing Fault Detection and Diagnosis

In order to verify the aforementioned methods, experiments on ball bearings are carried out. The vibration from ball bearings are considered to be dominated by the deterministic components. In this chapter, the bearing experiments are introduced from the aspects of the test rig, data acquisition system, and test procedure. These vibration measurements from six fault cases are used to verify the performance of the proposed EAAS in characterising the periodic impacts based resonant modulation. The periodic fault signatures extracted by the EAAS are benchmarked by the Kurtogram based envelope analysis.

5.1 Test Facilities

5.1.1 Mechanical and Electric Systems

Bearing data sets are acquired from a test rig shown in Figure 5-1, which mainly consists of an induction motor and a DC generator. The motor and generator are connected by a flexible coupling, allowing small misalignment. The 4kW induction motor is a four-pole machine and its rated speed is 1420rpm. The load applied on the motor is supplied by the DC generator. The motor is connected to the DC generator by a flexible jaw coupling that allows a certain degree of misalignment. The essential control unit can control the speeds and power outputs and it also balances the torques by adjusting the magnetic fields. The power consumed in the tests is tiny because the power flow is a close loop due to the advanced control unit. The core of the control unit is the Variable Frequency Drive (VFD) which alternates the input frequency or voltage to control the motor speeds.

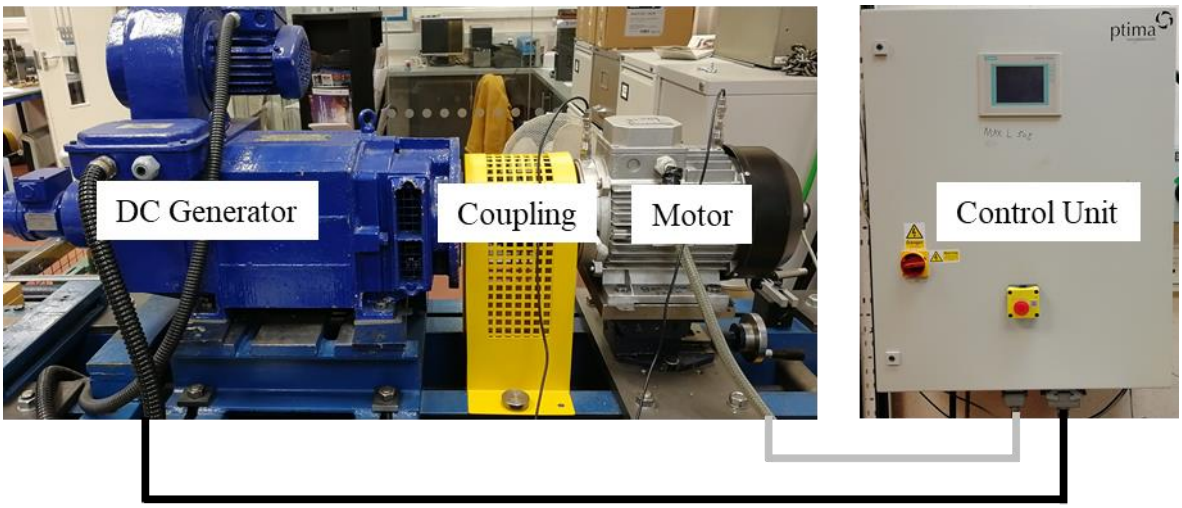


Figure 5-1 Layout of the motor bearing test rig

The key specifications of the test rig equipment are listed in Table 5-1.

Table 5-1 Equipment of motor bearing test rig

Equipment	Model	Key Specification
VFD	650V	0-240Hz, 15kW
Induction Motor	112MA/4	1420rpm, 4kw
DC Generator	CBH5025	1500rpm, 4kw

5.1.2 Data Acquisition System

Apart from the electromechanical system, the data acquisition system is another important type of test facility. The vibration transducers used in this experiment are accelerometers because the modulation usually occurs at thousands of hertz and the acceleration based vibration is more effective at high frequency measurements. As shown in Figure 5-2, two accelerometers are

installed at both vertical and horizontal position on the motor shell around the target bearing. To acquire the vibration data sets, a high speed analogue-to-digital converter (ADC) is essential to obtain the digital signals for further processing. In the meantime, a commercial encoder is connected to the free end of the motor by the plastic tube as an adapter. The encoder signals are used to calculate the rotating speed of the motor and then calculate the theoretical characteristic frequencies of the ball bearing faults. A K type thermocouple is also implemented to measure the internal temperature by drilling a hole into the internal room. The temperature measurements can monitor the health condition of motors and allow to align the working conditions of motors at different test cases. Two accelerometers, an encoder and a thermocouple are connected to the data acquisition device by the 50ohm BNC cables.

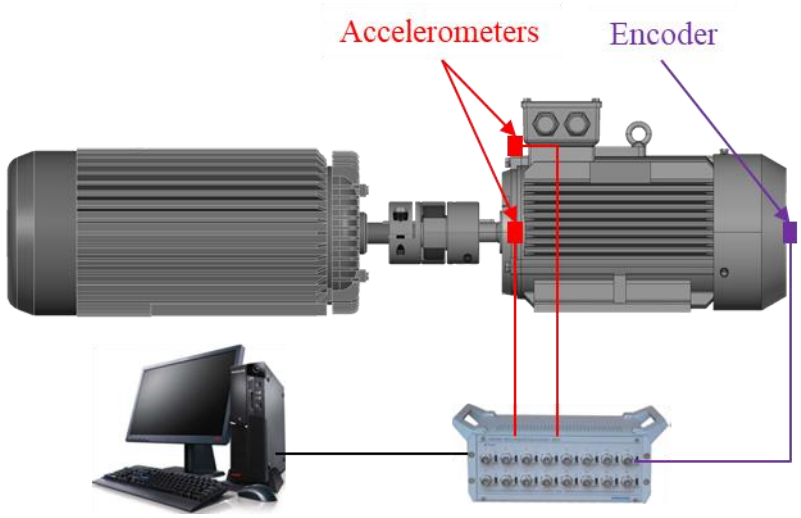


Figure 5-2 Schematic diagram of the motor bearing test rig

The apparatus used in the test rig is listed in Table 5-2 and the key specifications of these devices are also attached in the table.

Table 5-2 Data acquisition device used in motor bearing tests

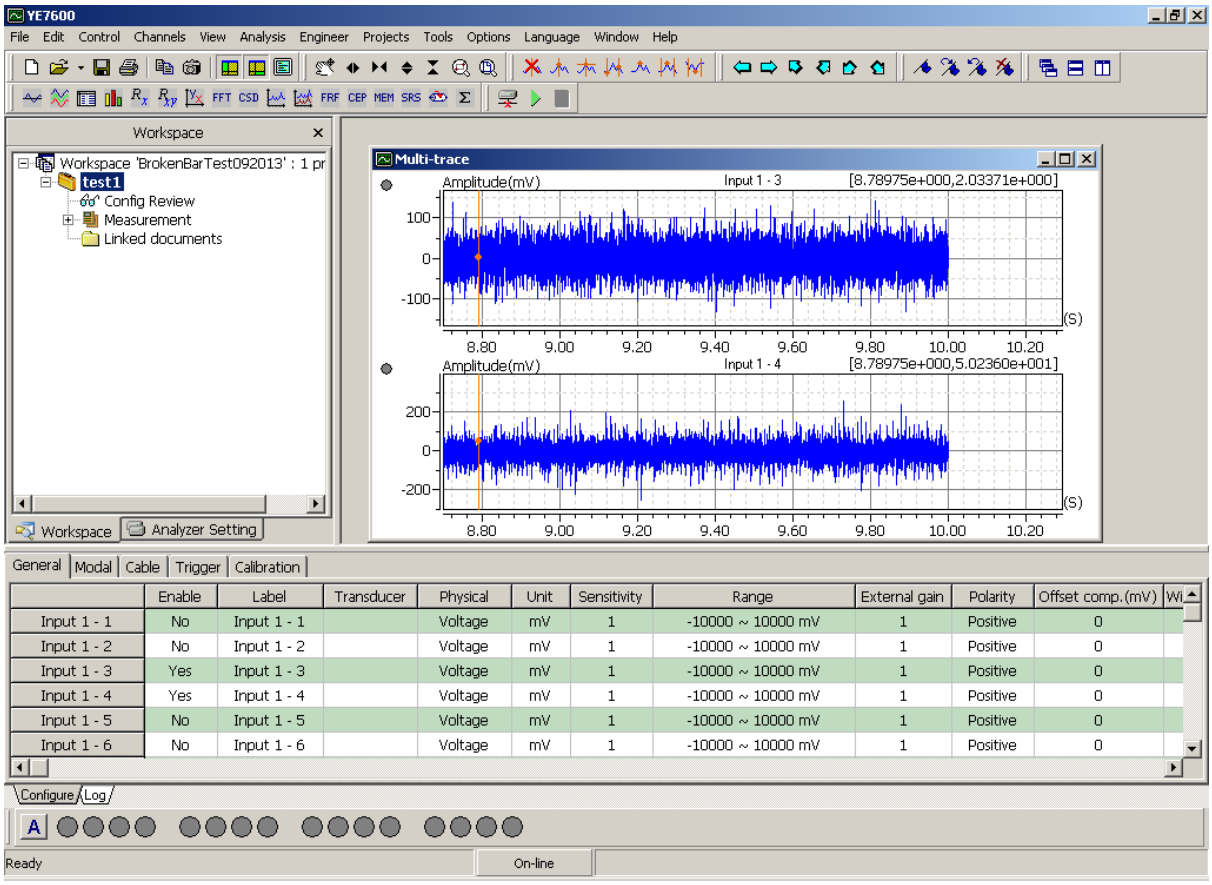
Equipment	**Model**	**Key Specifications**
Data Acquisition Device	YE6232B	16 channels, 96kHz
Accelerometer	CA-YD-185	5mv/(m/s^2), 0.5-5000Hz
Encoder	260-3724	100ppr, 6000rpm
Thermocouple	K type	0 to 200°C

The data acquisition device is a 16-channel parallel sampling hardware and it can collect the data samples through multiple channels by using a sampling rate of up to 96kHz per channel. The detailed specifications of the data acquisition device are shown in Table 5-3.

Table 5-3 Specifications of the data acquisition device

Parameters	**Specifications**
Model	YE6232B
Channels	16
A/D bits	24bit
Input Mode	V/IEPE
IEPE Power Supply	4mA/+24VDC
Signal Input Range	≤±10VP
Signal Frequency Range	DC-30kHZ(-3dB±1dB)
Gain	×1, ×10, ×100
Filter	Independent Anti-filtering
Accuracy	±0.5%
Sample Rate (Max.)	96kHz/CH, Parallel
Interface	USB 2.0
Trigger Modes	Signal Trigger
Dimensions (mm)	236W×88H×277D
Weight	4 kg
Power Supply	AC: 220V 50Hz; 110V 60Hz

The data acquisition software for driving the hardware is YE7600, which allows users to control the sampling frequency and sample duration on demand. The graphical user interface of the software is displayed in Figure 5-3.



	Enable	Label	Transducer	Physical	Unit	Sensitivity	Range	External gain	Polarity	Offset comp.(mV)
Input 1 - 1	No	Input 1 - 1		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0
Input 1 - 2	No	Input 1 - 2		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0
Input 1 - 3	Yes	Input 1 - 3		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0
Input 1 - 4	Yes	Input 1 - 4		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0
Input 1 - 5	No	Input 1 - 5		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0
Input 1 - 6	No	Input 1 - 6		Voltage	mV	1	-10000 ~ 10000 mV	1	Positive	0

Figure 5-3 Graphical user interface of the data acquisition software

5.2 Test Procedure

The reason to implement experiments on motors is that bearing failure is a common fault in induction motors which play an important role in power sources. The failure of motors can cause the substantial decrease of productivity and safety. The condition monitoring of induction motors is highly desired in industrial appliances. The ball bearings are most frequently used in induction motors and the bearings in the test are deep groove ball bearings. The bearings are installed at both sides of the motor rotors, which can reduce the friction, handle the axial loads and guide the rotary movement. There are a pair of ball bearings at the drive and non-drive end and in the tests, only the bearings at the drive end is the target for monitoring, which is shown in Figure 5-4.

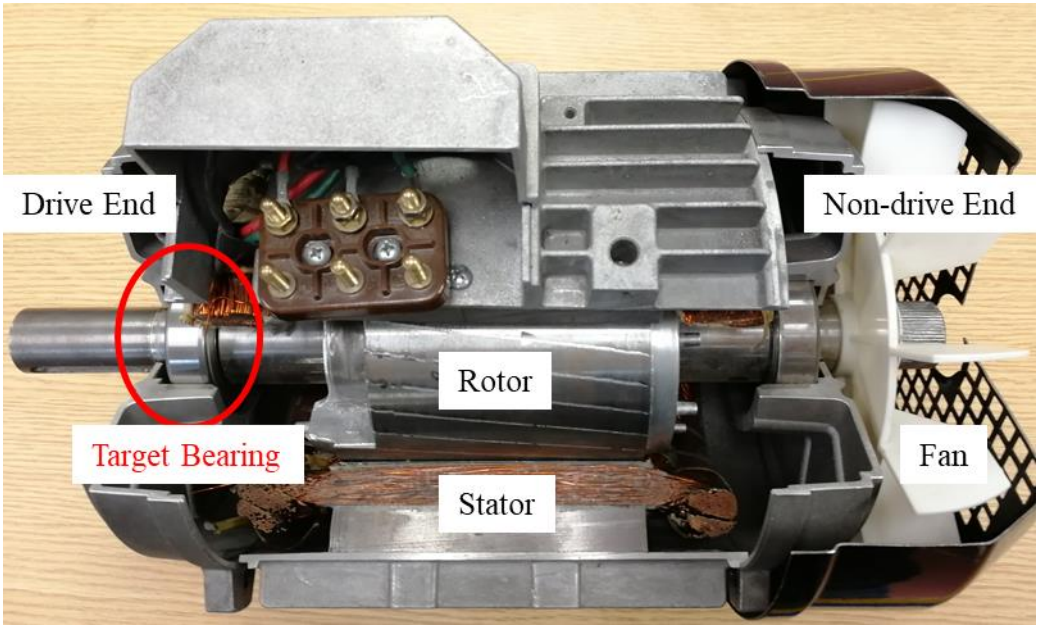


Figure 5-4 Target ball bearing in the motor

As shown in the Figure 5-4, the faulty bearing circled by the red ellipse is located at the drive end of the motor. The specifications of the bearing are shown in Table 5-4.

Table 5-4 Specification of the bearing

Parameters	**Specifications**
Model	6206
Number of Balls	9
Ball Diameter	9.53 mm
Pitch Circle Diameter	46.4 mm
Contact Angle	0

Two types of bearing faults are simulated in this experiment, they are outer race faults and inner race faults. The faults of bearings are seeded by making a pitting on the race ways. According to the pitting size, the faults are classified into three type: large (about $4mm \times 2mm$), medium (about $2mm \times 1mm$) and small (about $1mm \times 0.2mm$). The defects are manually seeded and the depth of the defects are very small (less than $0.05mm$). The outer race faults are displayed in Figure 5-5 and the photographs of the inner race defects are shown in Figure 5-6.

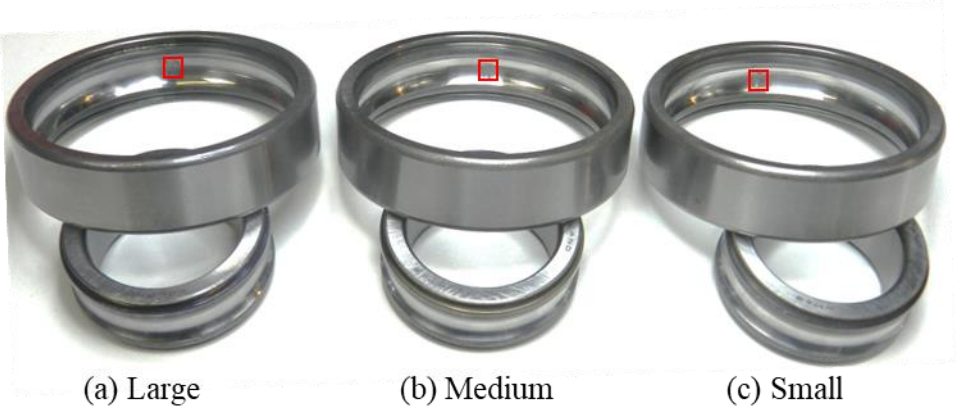


Figure 5-5 Outer race faults

Figure 5-6 Inner race faults

After seeding faults on the bearings, tests of every fault case were carried out to obtain the vibration signals. The tests are scheduled to proceed in full speed (1500rpm) versus five loads (0, 25%, 50%, 75% and 100%) for every fault bearing. The 100% load is the full power output of the induction motor, which generates about the torque of 25Nm. In each test, the thermocouple is used to monitor the motor temperature after the warmup stage. If the temperature goes over 40°C, the designed speed and load can be reached by the control system to start data logging. The threshold of temperature can reduce the influence from the motor performance parameters and lead to reliable data sets for further analysis.

The main concern in the process of data acquisition is to meet the Nyquist sampling theorem. Theoretically a sampling rate of twice interested frequency components can preserve the relative information but in practice a sampling frequency of 10 times higher than the highest frequency is empirically recommended for more satisfied analysis results. Therefore, the sampling rate in all cases is configured to be 96kHz and the time duration for each data is 30 seconds. The sample length is based on the demand of the propose methods for achieving a decent averaging time. A longer time duration can achieve better results, but it consumes considerable storage space, which

is opposite to the CM aim of cost-effective means. Besides, the test loop in every case was repeated three times to decrease the measurement errors and ensure the reliable data sets.

As the speed in the whole test is configured to be nearly the same at 1500rpm, the theoretically characteristic fault frequencies can be calculated according to Equation (4.1) to (4.4) with bearing parameters. The theoretically characteristic fault frequencies of the ball bearing are listed in Table 5-5.

Table 5-5 Theoretically characteristic fault frequencies of the ball bearing

Characteristic Fault Frequencies	**Frequency Value**
Outer Race Frequency	89.2Hz
Inner Race Frequency	135.3Hz
Ball Frequency	58.1Hz
Cage Frequency	9.9Hz

Three dynamic signals obtained from the test rig are shown in Figure 5-7. The temperature signals are not displayed here because the temperature data is a DC component, which usually vary slowly and can be represented by the mean values. The structure of the time domain signals is regular because of the periodicity of the rotating machine. The Figure 5-7 (a) describes the pulse signals from the encoder in every revolution, which can be utilized to calculate the rotating speed. Additionally, the encoder signals are significantly important in the time synchronous averaging (TSA) or order tracking methods. The Figure 5-7 (b) and (c) display the vertical and horizontal vibration obtained at the 100% load of large outer race faults. The vibration at two positions is different due to the different transmission paths and the orthogonal directions of the dynamic responses.

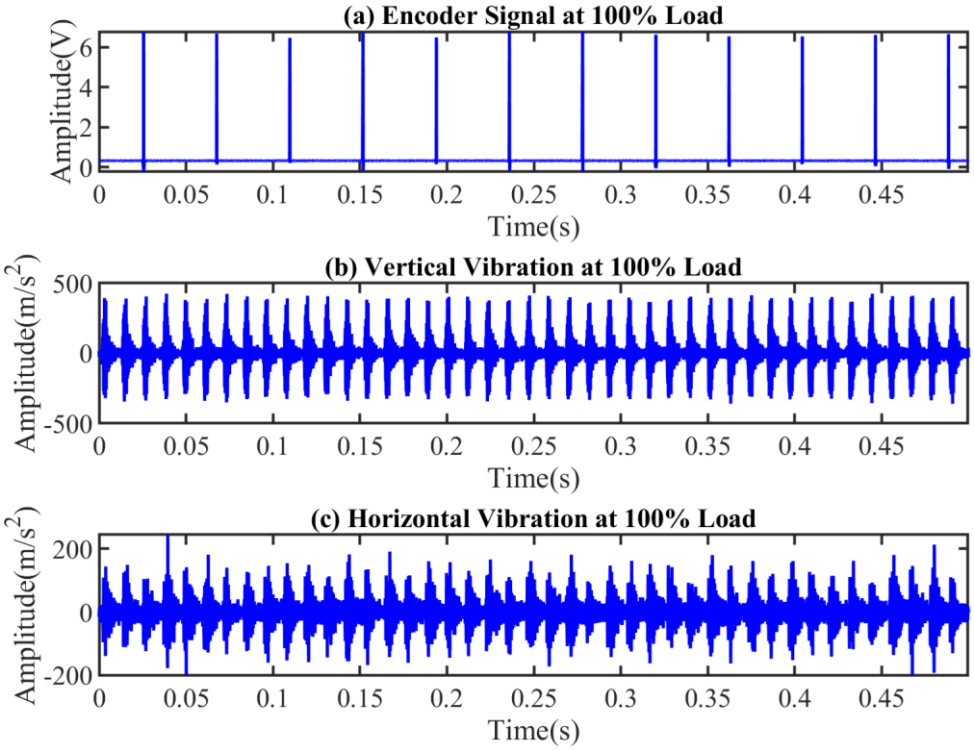


Figure 5-7 Time domain signal

5.3 Signal Analysis and Discussion

5.3.1 Time Domain Analysis

Figure 5-8 illustrates the temporal vibration signals of the bearings with outer race faults from the vertical accelerometer. The vibration signals of the cases with large, medium and small faults are displayed in Figure 5-8 (a), (b) and (c) respectively. It is clear that the periodic peaks are induced by the bump when balls pass the defect on the outer race. The periodic events represent the characteristic fault result from the local defect of the bearing. The time interval between impulses, especially for the cases of large outer race faults, is 0.011s, which is actually the fault characteristic period of the outer race defects. In addition, the amplitude of the vibration decreases with the decrease of fault size.

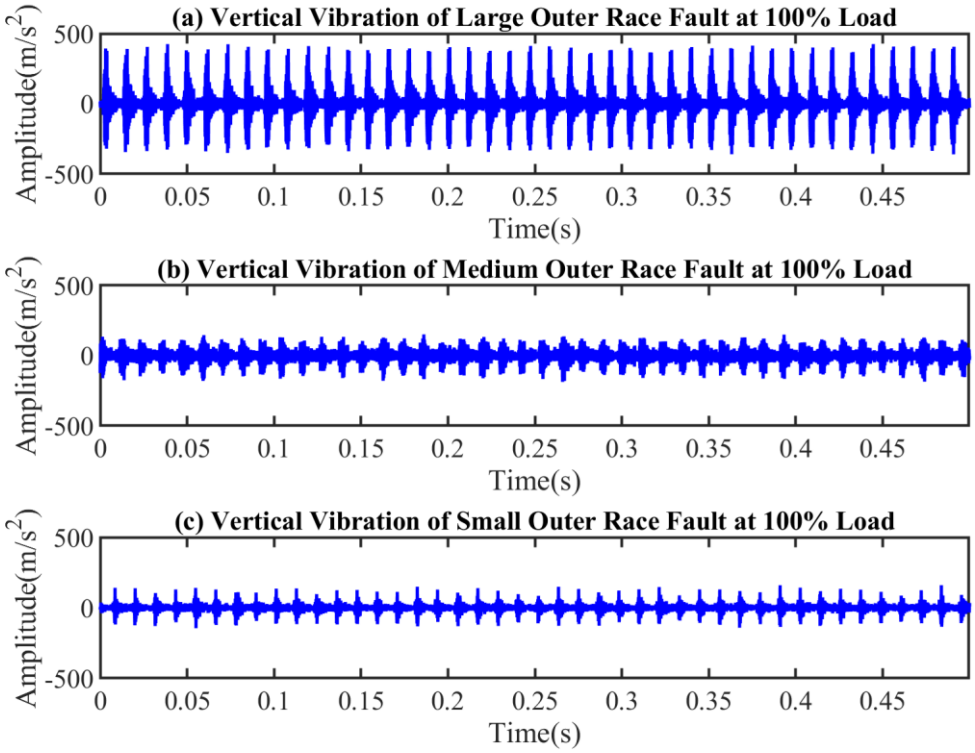


Figure 5-8 Time waveform of the vibration induced by the outer race faults

The vibration signals acquired in the cases of inner race faults are depicted in the Figure 5-9. The inner race faults of the ball bearing are also caught in the time domain analysis.

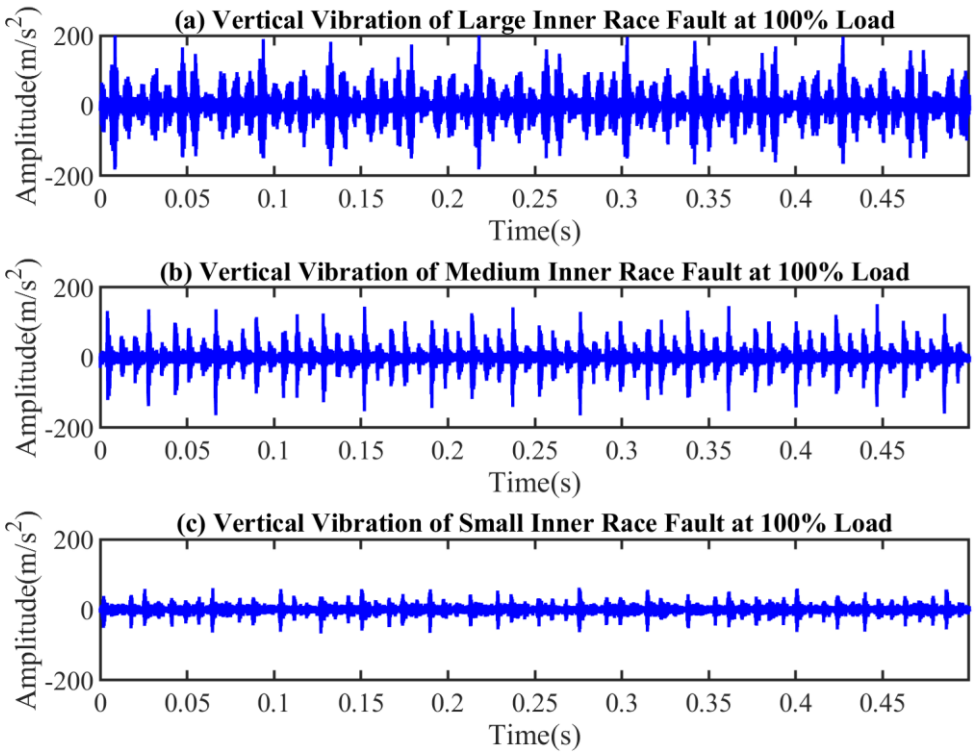


Figure 5-9 Time waveform of the vibration induced by the inner race faults

Both outer race and inner race faults can be identified from the time domain signals but the method is not easy to implement in practice. The signal acquired from the test rig is usually much better than the one from the actual working machines because the running condition of the test rig including speed and load is more stable, and furthermore, the noise from other systems and ambient environments is less. To summarise, the time domain analysis is able to detect the bearing faults but practically it is not reliable especially when the impulses are submerged by the noise.

5.3.2 Determination of Optimal Frequency Bands

The vibration signals from the large outer race faults are analysed by the Kurtogram in Figure 5-10 and the optimal frequency band recommended by the Kurtogram is centred at 33.5kHz with a frequency bandwidth of 1000Hz, which locates at the level 5.6.

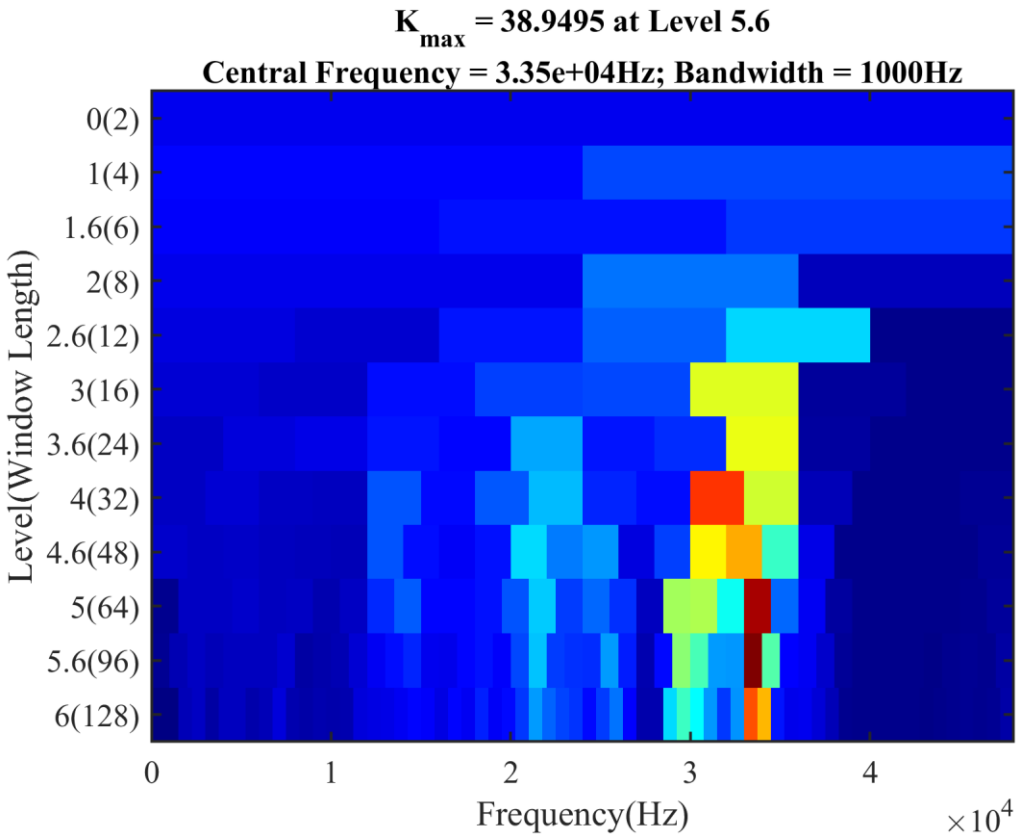


Figure 5-10 Kurtogram for the case of large outer race fault at 100% load

The Stabilisation Diagram of SSI in identifying the system modes at the large outer race faults are shown in Figure 5-11. It can be seen that the modes identified are located at the peak in the power spectrum, which is identical to the empirical knowledge.

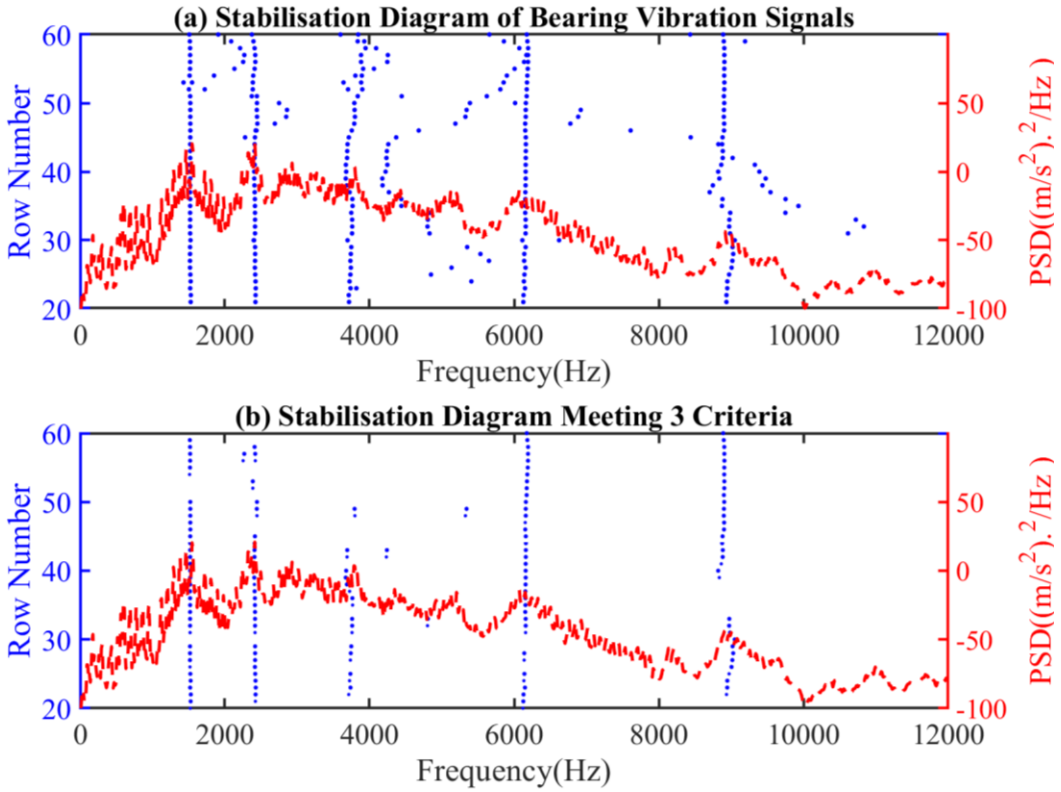


Figure 5-11 Stabilisation Diagram of SSI for the case of large outer race fault at 100% load

Calculated from the Stabilisation Diagram, the resonant frequencies of the bearing system with high normalised rate are shown in Figure 5-12 and the central frequencies over the threshold are 1530Hz and 2383Hz. The method here is not for the purpose of modal analysis and hence the analysis does not focus on the identification of the full modal properties. The pronounced modes identified by the SSI are more likely to be the carrier frequencies because the modal responses are manifest in the vibration signals and are easily extracted.

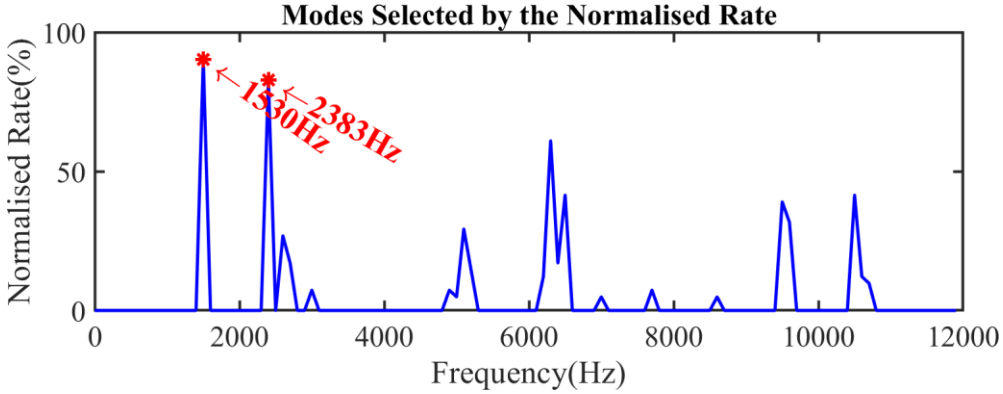


Figure 5-12 Normalised Rate of stable modes in the Stabilisation Diagram for the case of large outer race fault at 100% load

In the medium fault case, the most impulsive frequency band is from 0Hz to 24kHz denoted by the Kurtogram at level 1 in Figure 5-13. The system identification method, SSI, selects five central frequency candidates, which are 1898Hz, 2523Hz, 5011Hz, 6496Hz and 10882Hz in Figure 5-14.

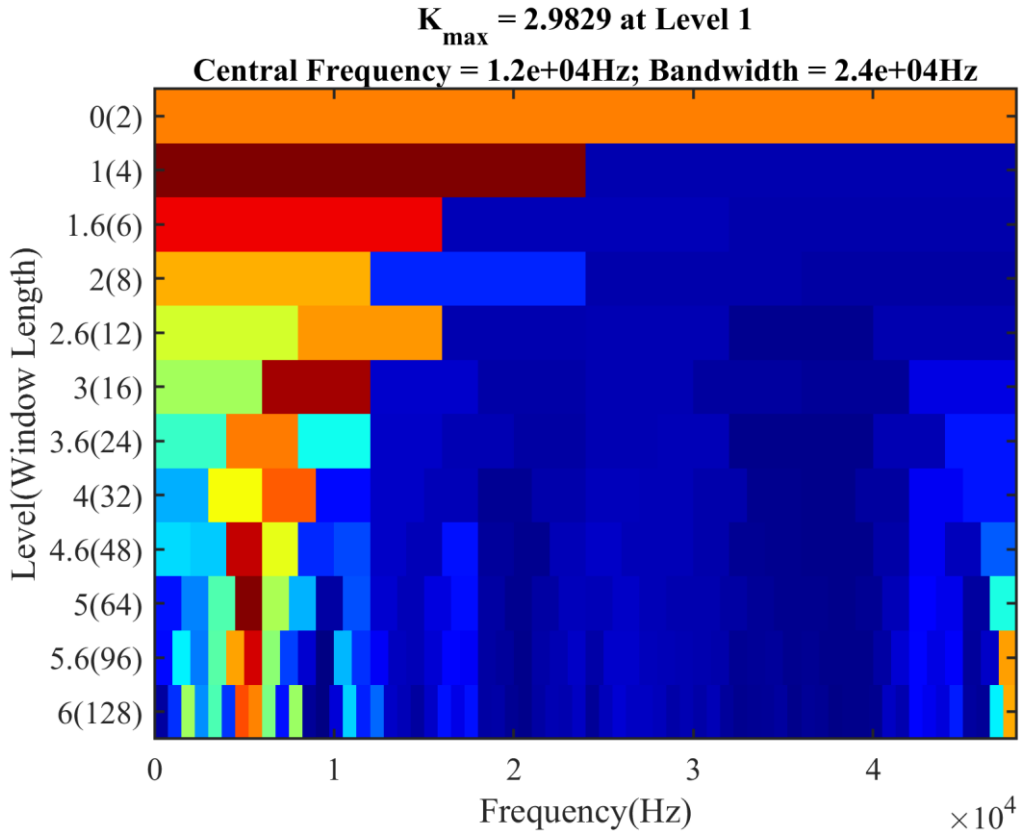


Figure 5-13 Kurtogram for the case of medium outer race fault at 100% load

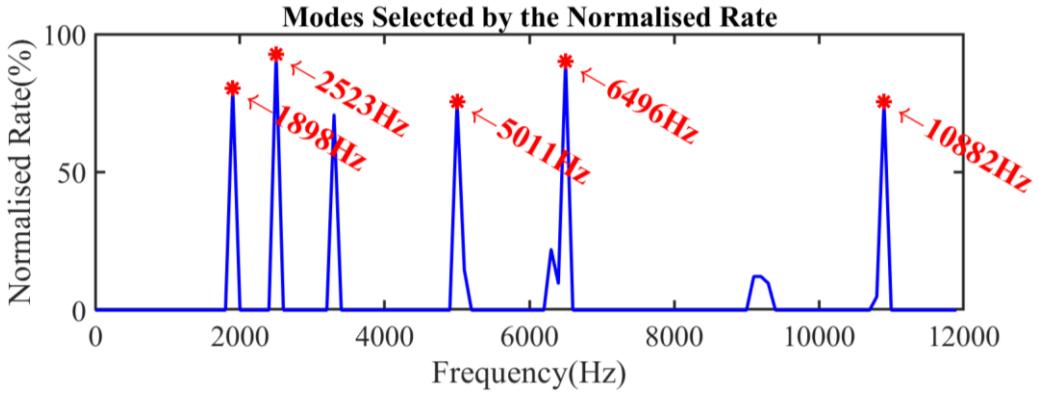


Figure 5-14 Normalised Rate of stable modes in the Stabilisation Diagram for the case of medium outer race fault at 100% load

For the small fault case, the Kurtogram shows in Figure 5-15 that the optimal frequency band is centred at 29.5kHz with a bandwidth of 1000Hz. The bearings used in these tests are the same model, and the materials and dimensions of the bearings are similar. The modal parameters are slightly different due to the manufacture errors and installation but the modal properties of the bearing system are considered to be not significantly different. Considering not all modes can be excited in various working conditions, the candidates for the potential optimal central frequency can be different but should be one of the bearing modes. The results from the SSI method are more reasonable because the candidates of the central frequencies overlap with each other. In the case

of small outer race faults, the potential carrier frequencies are 1479Hz and 6061Hz, which are shown in Figure 5-16.

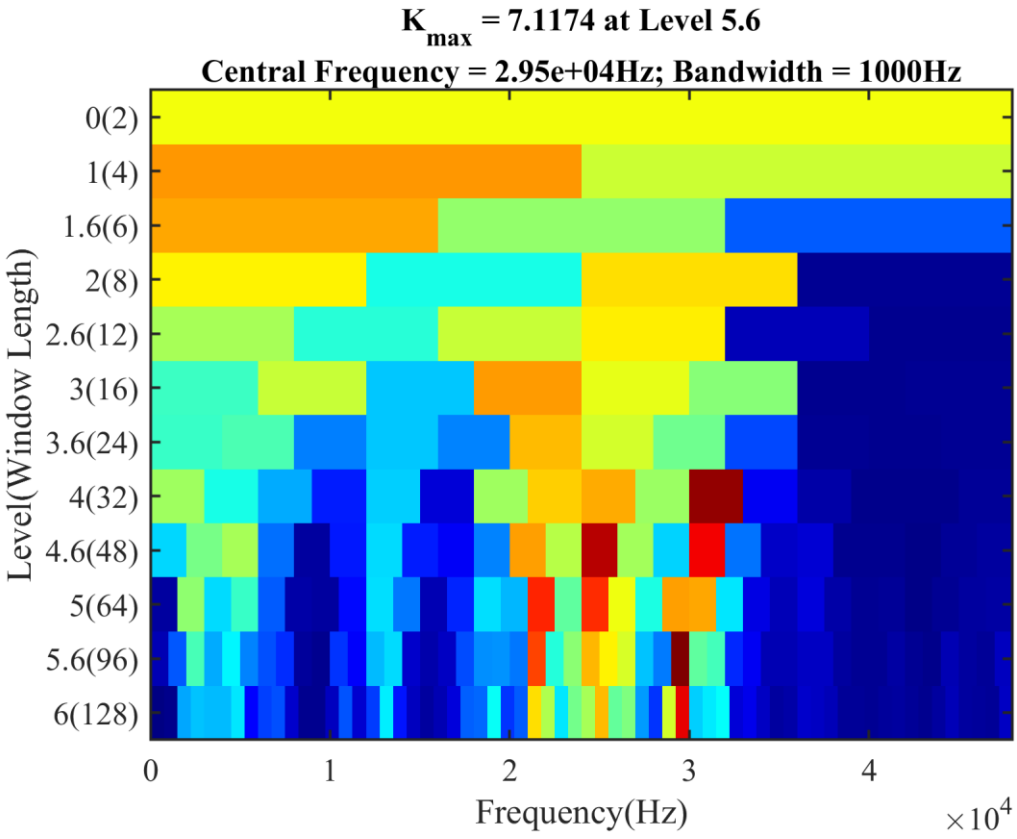


Figure 5-15 Kurtogram for the case of small outer race fault at 100% load

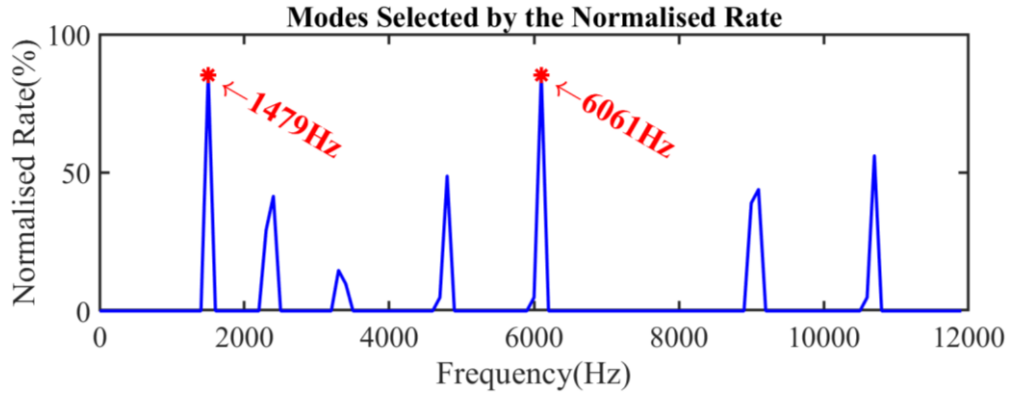


Figure 5-16 Normalised Rate of stable modes in the Stabilisation Diagram for the case of small outer race fault at 100% load

For the case of the large bearing inner race fault, the vibration signal is processed by the Kurtogram and the optimal frequency band selected in Figure 5-17 is centred at 35.5kHz with a frequency bandwidth of 1000Hz. Figure 5-18 displays that the SSI identified the natural frequencies of 2424Hz, 3402Hz, 4763Hz, 6387Hz and 9582Hz.

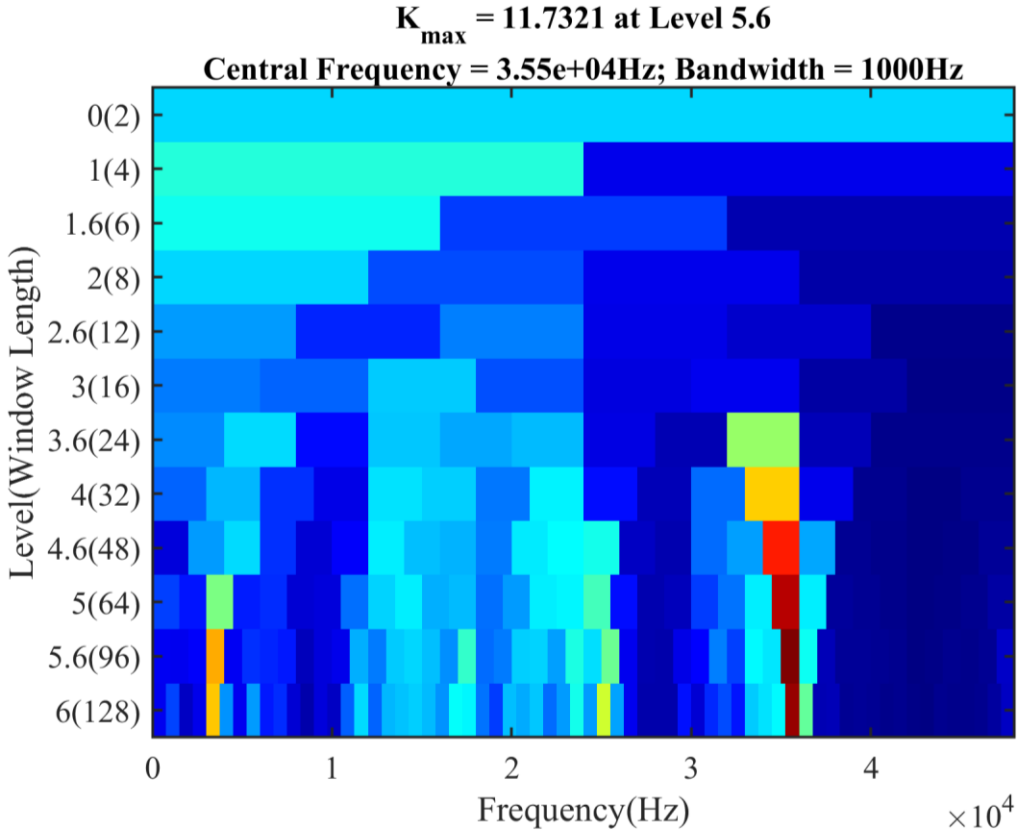


Figure 5-17 Kurtogram for the case of large inner race fault at 100% load

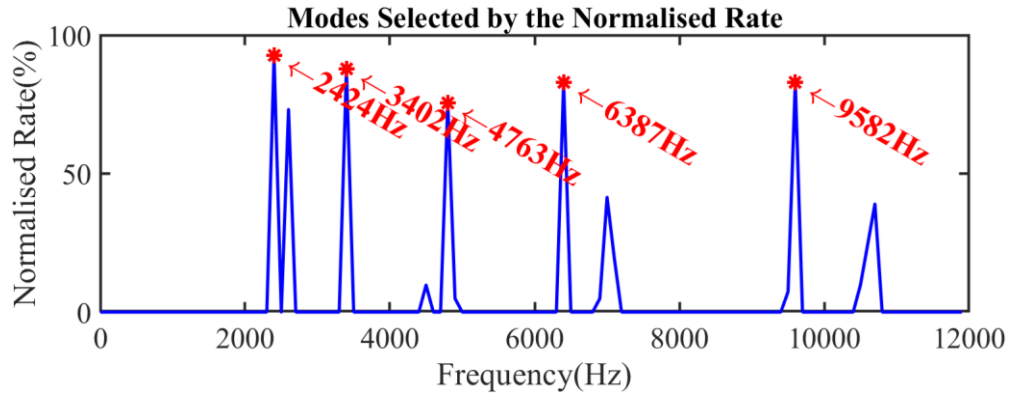


Figure 5-18 Normalised Rate of stable modes in the Stabilisation Diagram for the case of large inner race fault at 100% load

The medium inner race fault generates weak impulsive responses and in the Figure 5-19, the Kurtogram displays the frequency band from 21kHz to 24kHz. The normalised rates in Figure 5-20 show that the optimal central frequencies identified by the SSI are 3324Hz, 5412Hz and 6117Hz.

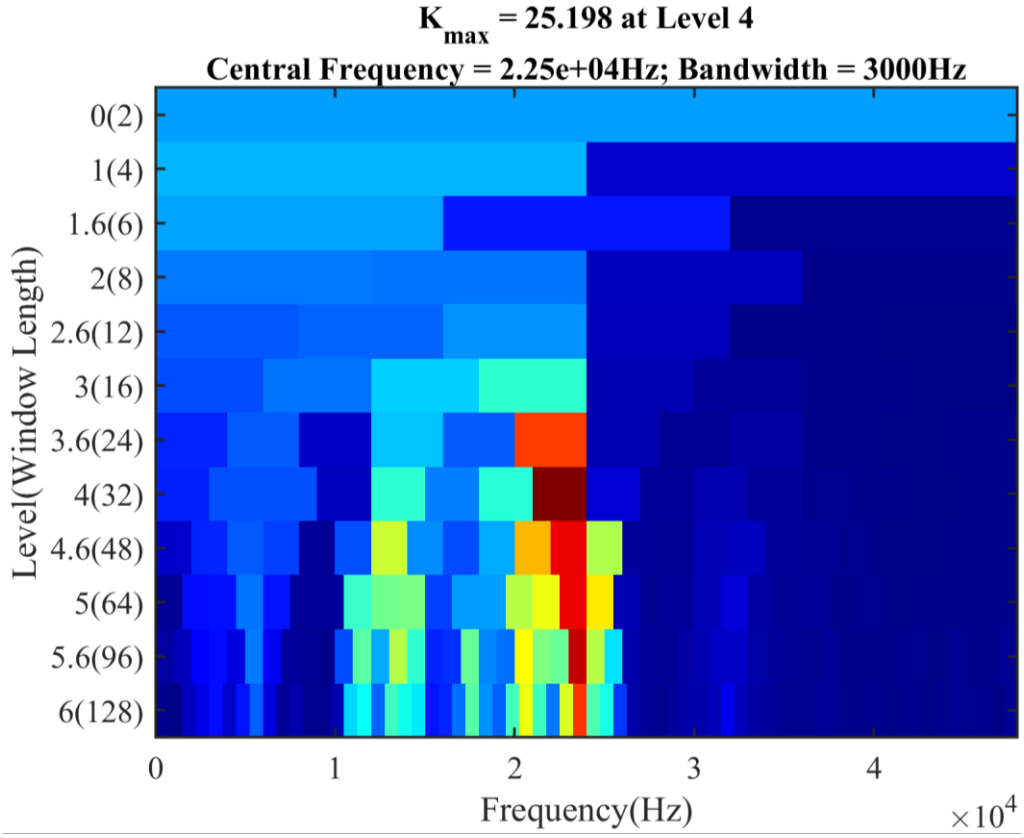


Figure 5-19 Kurtogram for the case of medium inner race fault at 100% load

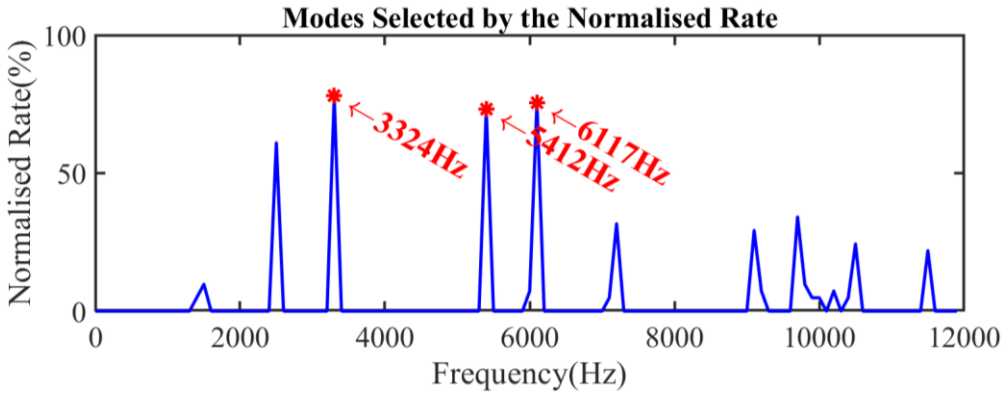


Figure 5-20 Normalised Rate of stable modes in the Stabilisation Diagram for the case of medium inner race fault at 100% load

As depicted in the Figure 5-21, the Kurtogram in the small inner race fault case shows that the optimal frequency band is from 0Hz to 16kHz. The frequency band is too wide and some undesired noise would be involved in the demodulation analysis. Figure 5-22 shows that the SSI recommends the optimal centre frequency at 2998Hz.

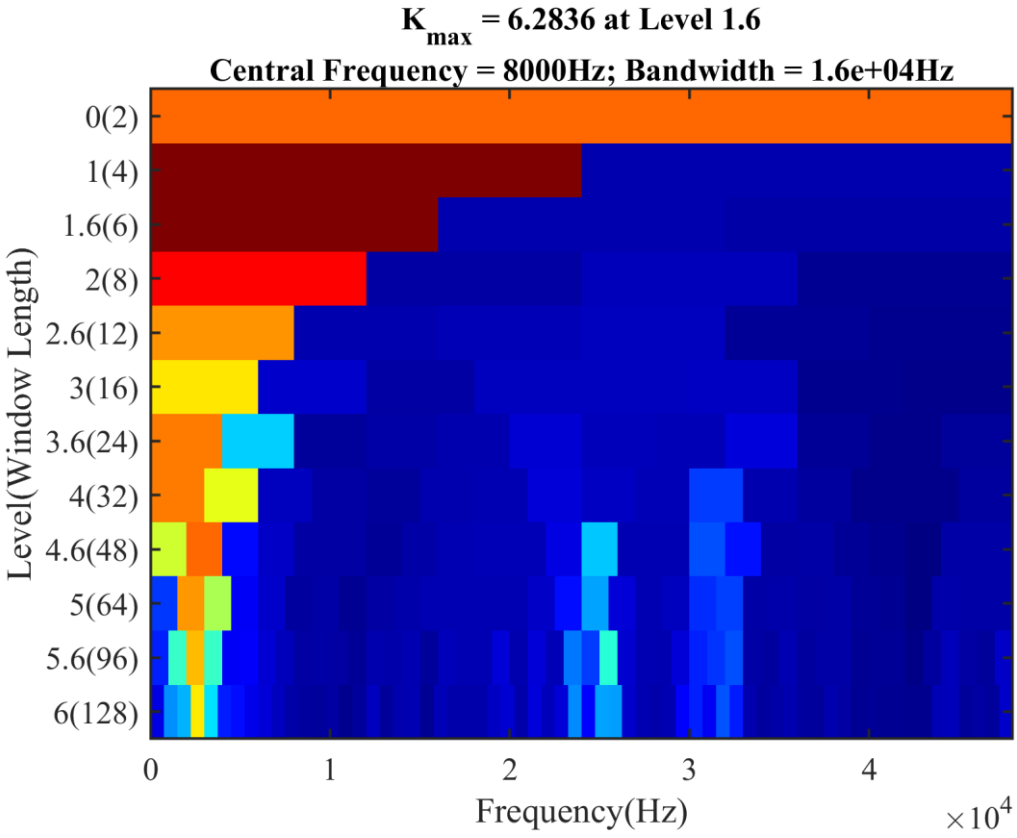


Figure 5-21 Kurtogram for the case of small inner race fault at 100% load

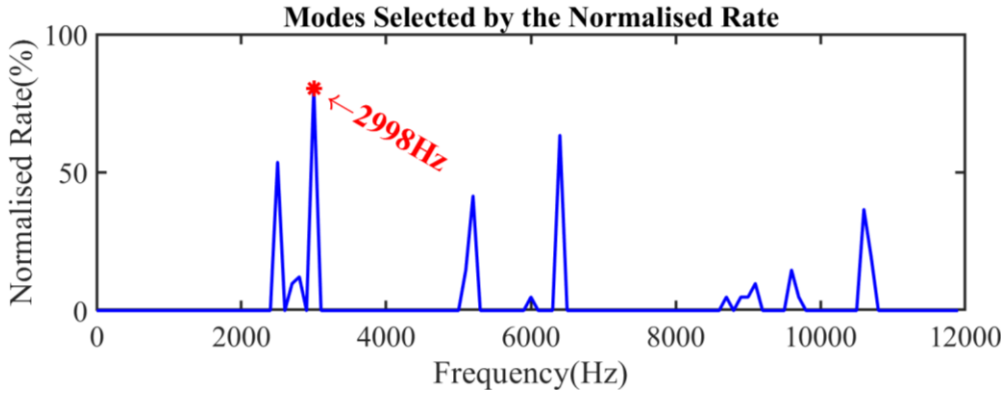


Figure 5-22 Normalised Rate of stable modes in the Stabilisation Diagram for the case of medium inner race fault at 100% load

Stochastic subspace identification can extract the resonant frequencies of rolling element bearings with different faults. The stabilization diagrams demonstrate that the identification results vary with the fault size and fault position. Hence, it is logically considered that the faults have an influence on the dynamics of the bearing systems. Furthermore, the dimension and material of the same model bearings are not identical. The identification methods need to be further studied especially the three conditions for stabilization diagram. Usually the natural frequencies identified by the SSI are multiple and these frequencies, especially with high normalised rates, are potentially the optimal carrier of the bearing fault signatures. One of the extracted natural frequencies can

lead to the optimal diagnosis results, which can be achieved by the trial and error studies using the proposed method. The optimal frequency bands extracted by the Kurtogram and SSI are summarised in the Table 5-6.

Table 5-6 Optimal Frequency Bands

Faults	Kurtogram		SSI	
	Central Frequency	Frequency Bandwidth	Central Frequency	Frequency Bandwidth
Large Outer Race Faults	33500Hz	1000Hz	1530Hz	602Hz
Medium Outer Race Faults	12000Hz	24000Hz	2523Hz	609Hz
Small Outer Race Faults	29500Hz	1000Hz	1479Hz	602Hz
Large Inner Race Faults	35500Hz	1000Hz	2424Hz	903Hz
Medium Inner Race Faults	22500Hz	3000Hz	3324Hz	900Hz
Small Inner Race Faults	8000Hz	16000Hz	2998Hz	895Hz

5.3.3 Demodulation Results and Analysis

The optimal central frequencies and frequency bandwidths are obtained by Kurtogram and SSI approaches. The demodulation analysis based on the frequency bands is introduced in this section. Not all loads are present in this section. Only the graphical results of vibration signals at 100% load conditions are discussed here.

Figure 5-23 shows the EAAS and CE results based on the optimal frequency bands selected by the SSI and Kurtogram approaches. The EAAS in Figure 5-23 (a) denotes the characteristic fault frequency of the outer race faults. Furthermore, the characteristic fault frequency of the bearing carrier can be seen. It is reasonable that the cage is not perfect. The faults of bearings are manually seeded and during the dismantling and assembling of bearings components, the cage has been damaged. The Kurtogram based CE in Figure 5-23 (c) shows that this method fails to find the faults that have occurred. The SSI based CE in Figure 5-23 (b) gives a fair demonstration of the outer race faults. Due to the leak of noise suppression, the SSI based CE has a high noise floor compared to EAAS. Consequently, the cage fault features are submerged by the random noise, leading to an ambiguous diagnosis conclusion.

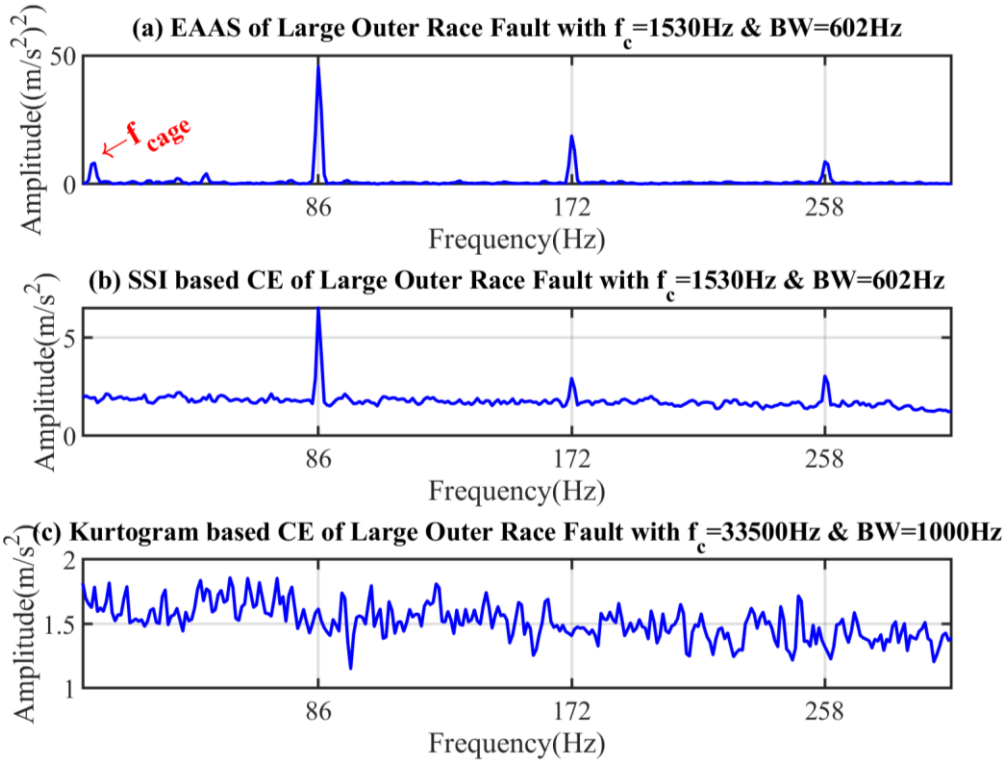


Figure 5-23 EAAS and CE for the case of large outer race fault at 100% load

For the medium outer race defect, the EAAS gives a promising representation of the bearing failure in Figure 5-24 (a). As in the aforementioned explanations, the cage frequency in the spectrum of EAAS denotes its imperfection. The multiple harmonics of the fundamental frequency and the sidebands around the outer race characteristic fault frequency are obvious to reveal the improper working conditions of the bearing carrier. The Figure 5-24 (b) and (c) show that both Kurtogram based CE and SSI based CE are able to detect the medium outer race faults because the frequency bands selected are inclusive of the carrier frequencies of the bearings. The Kurtogram based CE is valid by using a wide frequency band because the modulated fault features are above the noise carpet, which is the major benefit of the demodulation analysis. The amplitude of the characteristic fault frequency and the harmonics show that the fault signatures extracted by the SSI based CE are stronger than the Kurtogram base CE. The results will be quantitively compared by the signature strength ratio in the following section.

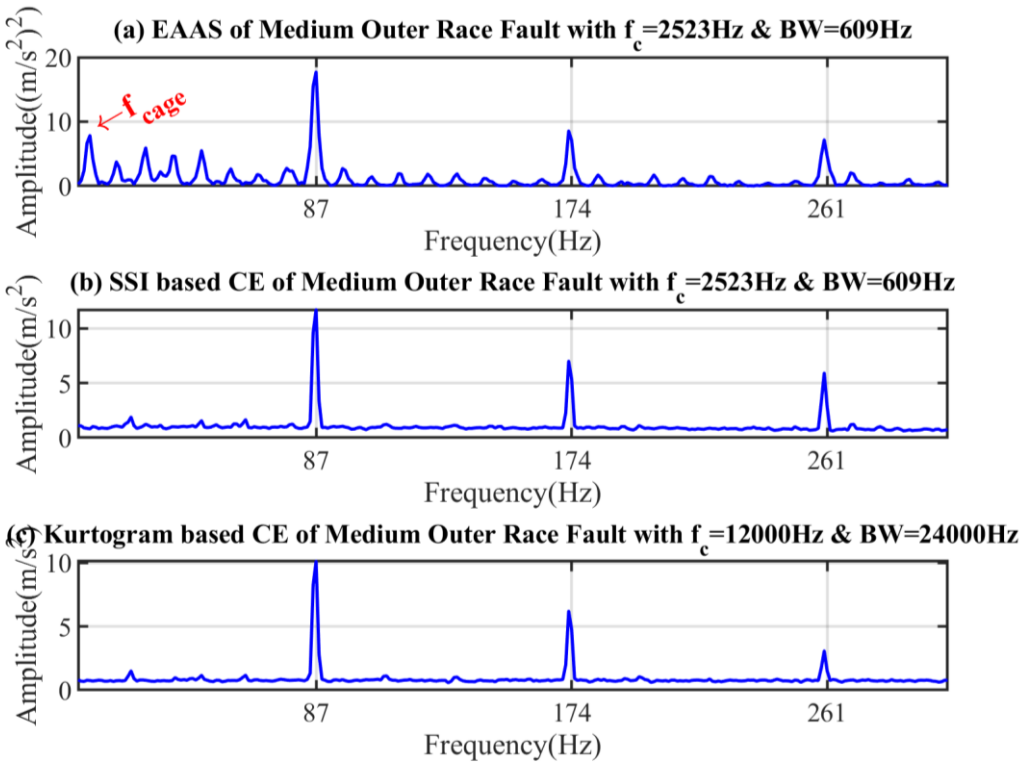


Figure 5-24 EAAS and CE for the case of medium outer race fault at 100% load

In the case study of a small outer race fault, the effectiveness of the aforementioned methods is expected to be thoroughly evaluated and compared. However, the vibration signals are of high SNR owing to the optimal laboratory environments and simple transmission path. Therefore, the small faults can be detected by the methods of EAAS in Figure 5-25 (a) and SSI based CE in Figure 5-25 (b). As shown in in Figure 5-25 (c), the Kurtogram based CE fails to diagnose the outer race failure because the frequency band selected by the Kurtogram does not contain the natural frequencies of the bearing.

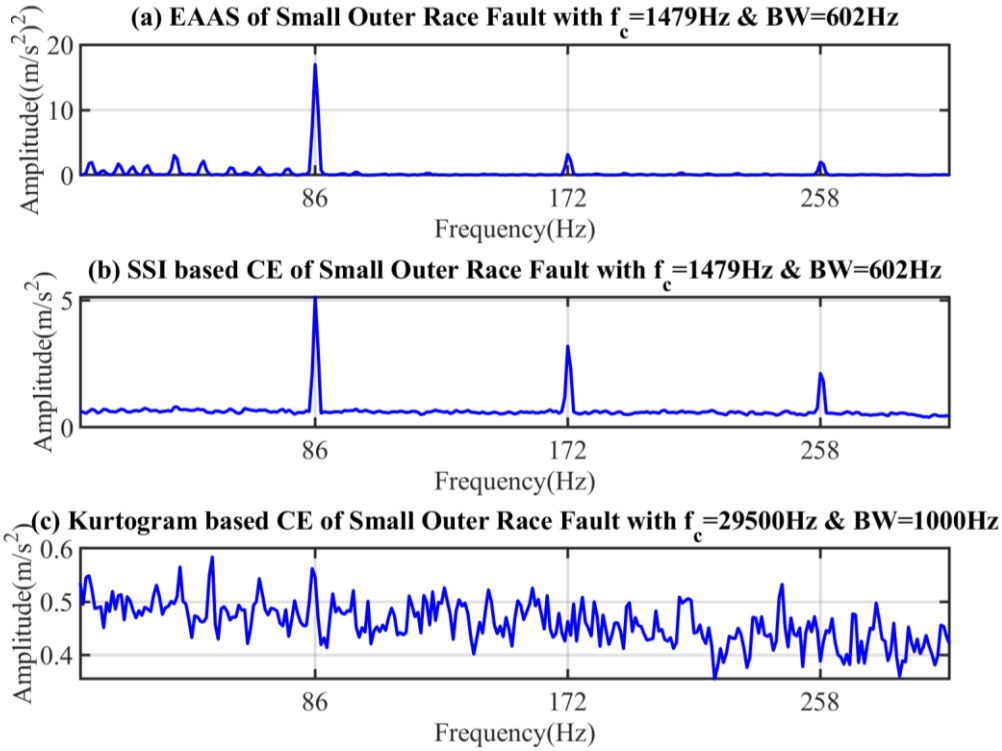


Figure 5-25 EAAS and CE for the case of small outer race fault at 100% load

The faults on the outer rings are more easily detected than the defects on the inner raceways. The periodically varying forces make the modulation phenomenon more complicated, resulting in the appearance of sidebands. The demodulation result of EAAS is displayed in Figure 5-26 (a) and the characteristic frequency of the inner race faults are manifest. Additionally, the rotating frequency of the shaft and second order harmonic is significantly high due to the rotor imbalance and the misalignment. The SSI based CE in Figure 5-26 (b) shows an identical conclusion with the EAAS. Unfortunately the Kurtogram based CE in Figure 5-26 (c) cannot demodulate the fault information due to an improper frequency band.

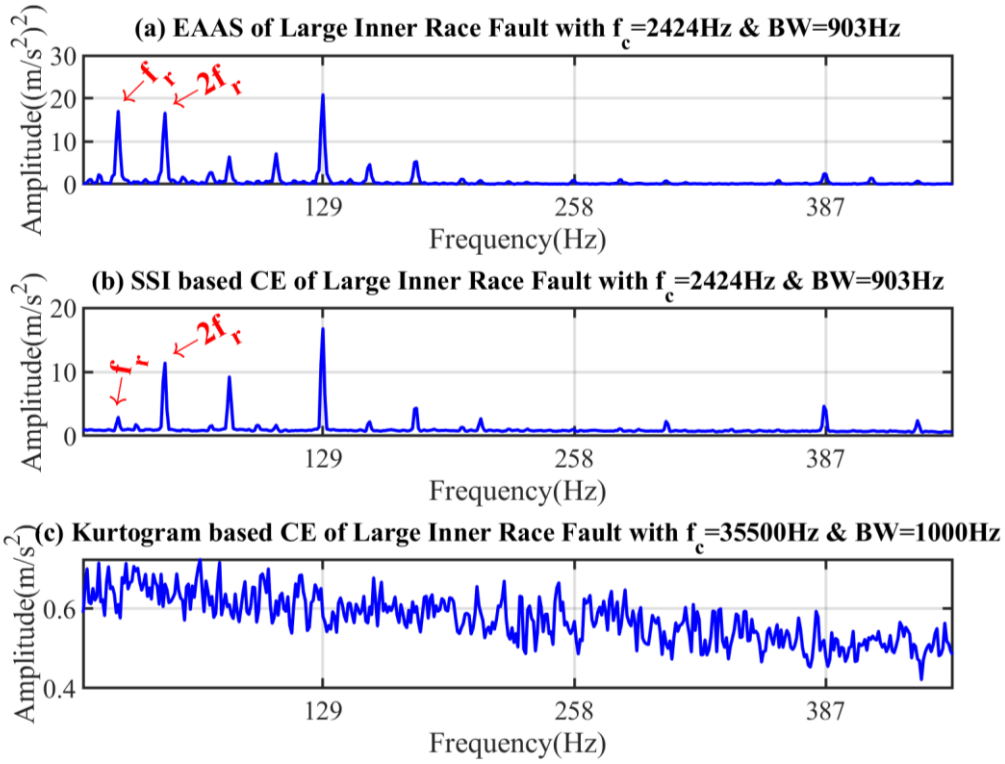


Figure 5-26 EAAS and CE for the case of large inner race fault at 100% load

The next case is based on the medium severity of the inner race defects. The demodulation results from the system resonance are depicted in Figure 5-27 and the most informative spectrum is obtained by the EAAS. The resolution of the spectrum has been defined in order to accurately locate the components relating to the dynamics of the bearing elements. Figure 5-27 (a) shows that besides the fault signatures of the inner race, the sidebands around the fault frequency are also visible in the EAAS spectrum. The sidebands are mainly the second harmonic of the shaft rotating frequency and the modulated frequency from the shaft rotating frequency and the cage fault frequency. The second harmonic of the shaft rotating frequency is significantly obvious because the rolling element running in and out the loaded zone of the bearing can be roughly considered as a symmetric variation of the impact force. Consequently, the variation period is half of the shaft rotating period, leading to a substantially high peak in the spectrum. The SSI based CE in Figure 5-27 (b) denotes the fault features of the inner race ways and the sidebands are double of the shaft rotating frequency. The fault information of the cage is not shown in the SSI based CE spectrum. It is expected that the Kurtogram based CE in Figure 5-27 (c) cannot detect the inner race faults by using an inappropriate frequency band.

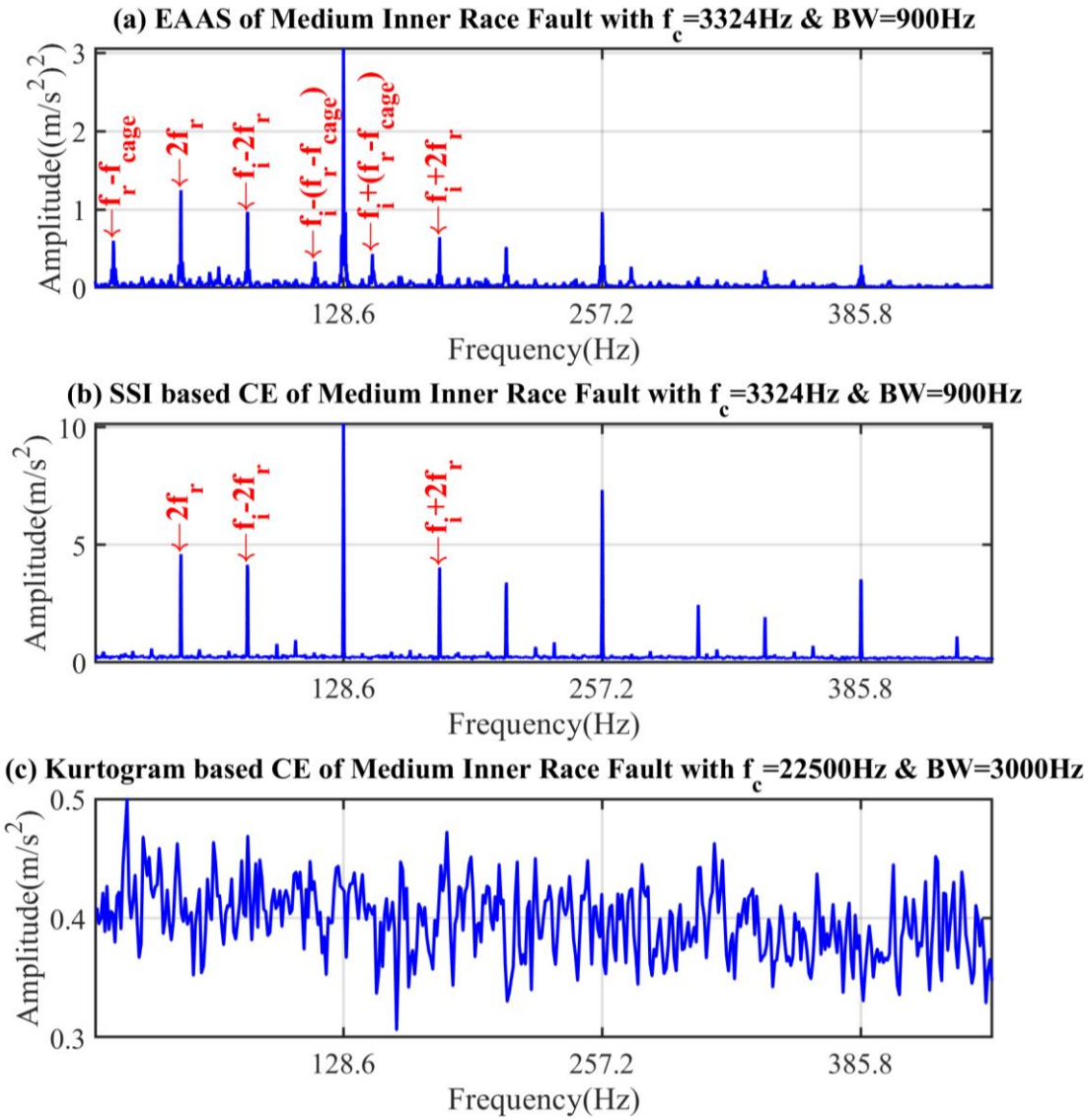


Figure 5-27 EAAS and CE for the case of medium inner race fault at 100% load

The last case in this chapter is the small fault on the inner race. As shown in Figure 5-28 (a), the EAAS approach can obtain an informative representation in the frequency domain. The characteristic fault frequency of the inner race defects has a large amplitude in the EAAS spectrum. Furthermore, the other components including the shaft rotating frequency and its harmonics, the modulated cage fault frequency are also pronounced in the frequency domain. As displayed in Figure 5-28 (b) and (c), both the SSI based CE and SSI based CE can identify the inner race faults but the fault signatures relating to the cage deformation are not obvious in the CE spectrum.

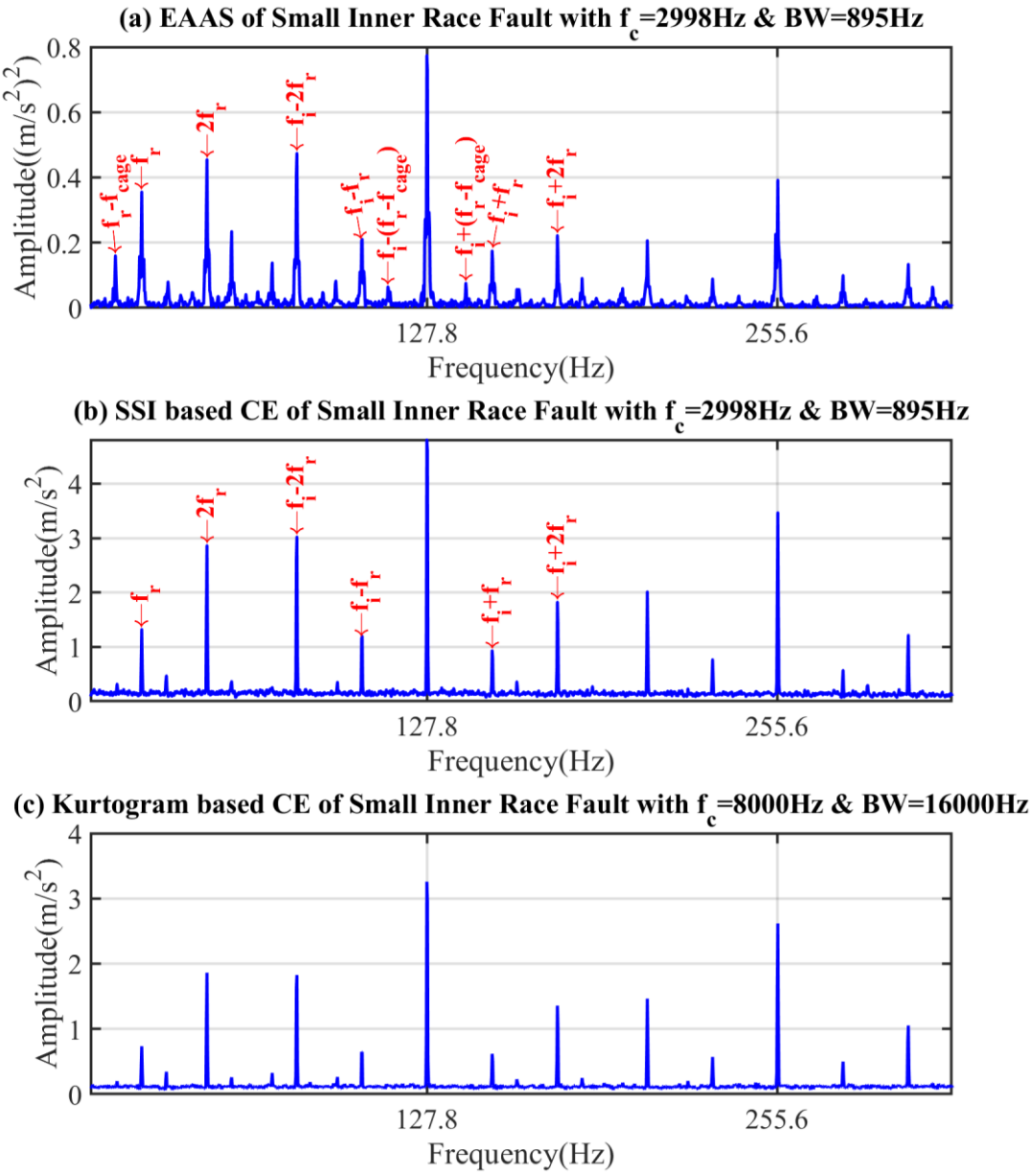


Figure 5-28 EAAS and CE for the case of small inner race fault at 100% load

Unlike the amplitude of the characteristic fault frequency for the outer race fault cases, the fault signatures extracted for the inner race fault cases are significantly smaller. The reason behind this is the effect of phase noise. Ideally the bearings are expected to roll along the raceways, but the rolling elements running in and out of the loaded zones have a small amount of slippage. Consequently, the phase noise destroys the periodicity of the signals, resulting in the nonperiodic measurements. The EAAS is a specific method for characterising the deterministic signals and any phase noise can distort the autocorrelation functions based on Equation (4.46).

The results of the fault signatures also denote that the incipient defects on inner raceways are more vulnerable to the random slippage. The presence of the phase noise enervates the capability of the EAAS and hence the fault signature strength is lower than the conventional envelope analysis.

Another important point in this analysis is the fault severity and among the six cases, the small inner race fault has actually the minimum defect, which can be verified by the amplitude of the fault features identified by the CE approach. The attenuation of the performance of EAAS denotes that the hypothesis of pure rotation in ball bearings is violated. The negative effect can be omitted in the cases of large bearing defects, but it will be a critical factor when the fault is small. The methods that are not robust to phase noise are unable to achieve the incipient fault detection and diagnosis. The results in this section also denote that the slippage in ball bearings is quite limited, which will be demonstrated in the next simulation study.

5.3.4 Quantification of Bearing Fault Signatures

The demodulation results in the last section have shown the outstanding performance of the proposed EAAS. To quantify the signature strength of the bearing faults, the $SNR_{signature}$ of three demodulation methods are depicted from Figure 5-29 to Figure 5-34.

The signature strength of the EAAS and conventional envelope for the large outer race fault signals is shown in Figure 5-29. The strength of the fault features extracted by EAAS is much higher than the conventional envelope analysis because the EAAS is capable of noise reduction and envelope analysis has no ability of noise suppression. The SSI based CE has a more obvious fault feature than the Kurtogram based CE because the SSI can select the proper carrier frequency.

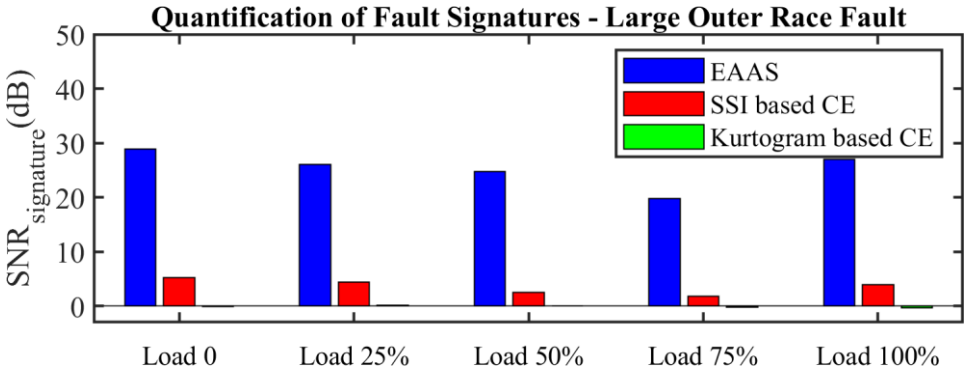


Figure 5-29 Quantification of fault signatures for large outer race defects

The medium fault on the outer raceway can be detected by the EAAS and CE. The $SNR_{signature}$ of the EAAS in Figure 5-30 is not as high as the other case because the EAAS in Figure 5-24 detects dozens of frequency components related to the cage fault. The equation for calculating the $SNR_{signature}$ is not exclusive to the cage fault information in the spectrum of EAAS and consequently attenuates the strength of the fault features.

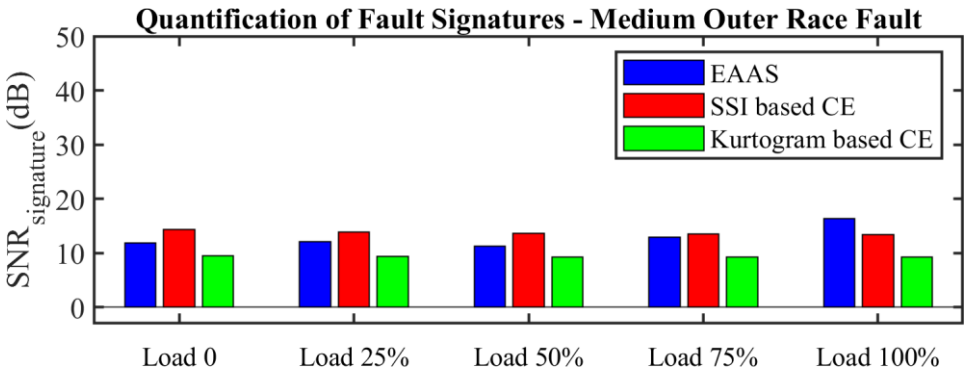


Figure 5-30 Quantification of fault signatures for medium outer race defects

As shown in Figure 5-31, the strength of the small outer race fault is within the expectation that the EAAS is the most promising approach for ball bearing fault detection and diagnosis.

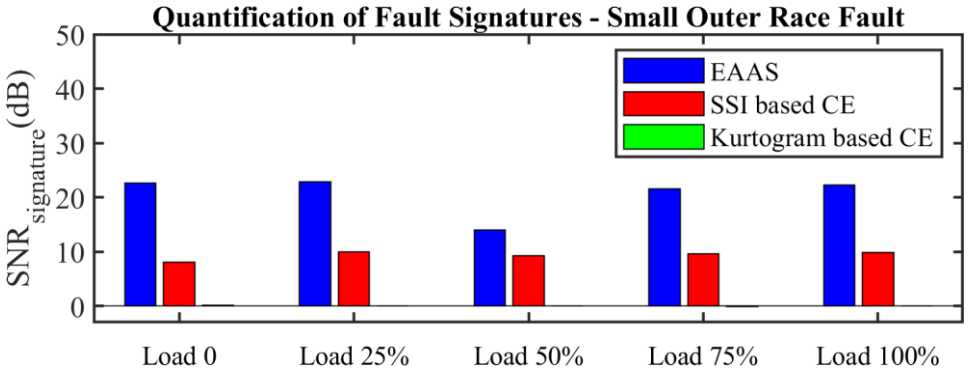


Figure 5-31 Quantification of fault signatures for small outer race defects

The fault signatures for the inner race faults of ball bearings are shown in Figure 5-32 to Figure 5-34. The results of the inner race faults denote that the EAAS performs better than the conventional envelope on the noise reduction. The test rig in the laboratory usually works in a better environment than the industrial scenario. Therefore the SSI based CE can effectively diagnose the bearing faults based on an appropriate frequency band.

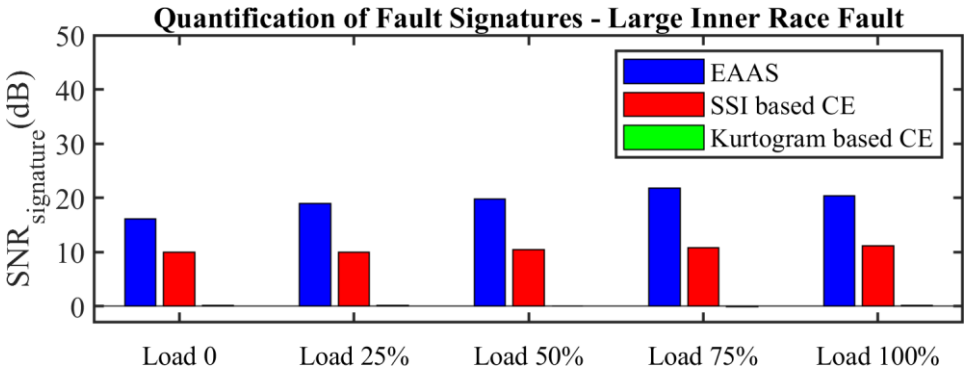


Figure 5-32 Quantification of fault signatures for large inner race defects

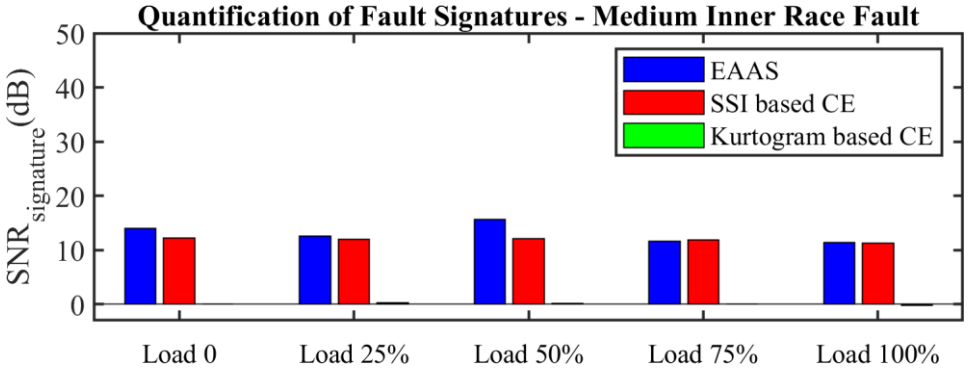


Figure 5-33 Quantification of fault signatures for medium inner race defects

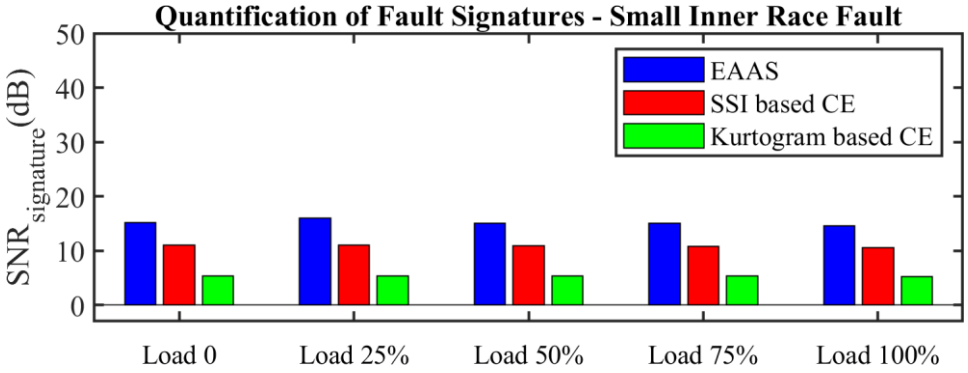


Figure 5-34 Quantification of fault signatures for small inner race defects

5.4 Summary

The experimental studies on ball bearings were carried out to verify the proposed EAAS and SSI methods for characterising the resonant modulation vibration signals. Six faulty cases of outer and inner race defects were carried to obtain the vibration signals at five different loads. Then the system identification approach was employed to analyse these vibration signals for selecting the optimal frequency bands. Based on the frequency bands recommended by the SSI and Kurtogram, the filtered signals around the natural frequencies can be demodulated by the EAAS and CE methods. The demodulated fault signatures show that EAAS is the most effective and efficient for detecting and diagnosing the ball bearing faults among the three approaches. The $SNR_{signature}$ of the fault gives quantified measurements of the fault features to the noise carpet in the spectrum. The experimental studies have verified that the proposed SSI and EAAS are effective for demodulating the periodic impacts induced resonant modulation signals.

Chapter 6 Simulation Study of Tapered Roller Bearing Fault Detection and Diagnosis

This chapter focuses on investigating resonant modulation induced by the aperiodic impacts through the simulation studies. The state space model of the rolling element bearings introduced in Chapter 4 is employed by amending the inputs to be a series of nonstationary impacts. The cyclostationary responses are part of the nonstationary signals and the deterministic approach, EAAS, is not effective under this condition. Then, two novel signal processing methods, Ensemble Average of Autocorrelated Envelopes (EAAE) and Phase Linearisation based Modulation Signal Bispectrum (PL-MSB), are developed to detect and diagnose the bearing faults at a very early stage. The different levels of phase noise are simulated to examine the EAAE and PL-MSB which are benchmarked by the conventional envelope analysis.

6.1 Random Slippage in Bearing Modelling

Rolling element bearings are designed to purely rotate during the service and the rolling behaviour is an effective means to minimise friction between raceways and rolling elements. However, the bearings inevitably suffer from the relative slippage between the bearing rings and rolling elements in practice because the working conditions of bearings are undesired, for instance load variations, localised deformation, random impacts, and lubrication failure. These reasons result in the distortion of the pure rolling and lead to the undesirable slippage. According to reference [110], approximately 1% to 2% slippage happens in the rolling element bearing. A further study [37] shows that the bearing vibration signals consist of phase noises. The slippage also leads to the differences between the theoretically characteristic fault frequencies with the experimental one [22]. Additionally, the slippage can become more severe if the radial clearances of bearings turn out to be larger with the increase of service time. At this stage, bearings are more likely to start localised fatigue defects. This small random slippage results in variations between the periods of impulses and consequently the impacts from the localised defects are not periodic. The input of the bearing system becomes nonstationary and these nonstationary impacts are classified into the cyclostationary signals due to the periodic variation of the statistical properties.

Due to the random slippage (phase noise), the EAAS cannot guarantee the best results in all cases because this stationary signal processing method cannot handle the nonstationary signals. The EAAS is sensitive to the phase noise, leading to the significant decrease of the fault signatures. The random slippage in ball bearings due to the incipient fault is tiny and consequently, the EAAS is still effective to detect the faults occurred. However, some bearings, for instance tapered roller bearings, are more vulnerable to phase noise due to the installation errors. In such circumstances, the early faults are substantially challenging in being identified by the conventional methods. The capable methods are expected to characterise nonstationary signals and suppress the strong background noise effectively.

It is simple to generate the modulation signal based on equation, however, the vibration including fault information is a phase-lock signal and it cannot indicate the slippage of bearings in practical working conditions. To simulate the random slippage induced inputs for the bearing model, the dynamics of the slippage is investigated. It has been found that the pitting induced impacts are not periodic anymore, which becomes a Markov process. The impacts occurred currently are only determined by its most recent impacts. Therefore, the successive impacts are simulated as a Markov chain, which is expressed as follows.

$$T_f^s(k) = T_f^s(k-1) + T_f + \Delta t(k) \qquad (6.1)$$

where, $T_f^s(k)$ is the moment of the k_{th} impact; T_f is the cycle of impacts or the fault cycle; and $\Delta t(k)$ is the random slippage at k_{th} impact. Consequently, the impact forces with the random slippage can be denoted as

$$f(t) = \begin{cases} \sum_{k=1}^{+\infty} A_f \delta\left(t - T_f^s(k)\right), & outer\ race\ faults \\ \sum_{k=1}^{+\infty} A_f \left|cos\left(2\pi \frac{f_r}{2} t + \varphi_f\right)\right| \delta\left(t - T_f^s(k)\right), & inner\ race\ faults \end{cases} \quad (6.2)$$

The bearing model developed in Chapter 4 is still used here and the inputs of the model are changed into the nonstationary forces in Equation (6.2). $\Delta t(k)$ is an import variable to show the degree of the slippage between the rolling elements and race ways. According to the different levels of slippage, the bearing vibration with localised faults can be simulated more practically. To more precisely control the slippage level, the random phase errors are generated to obey to the uniform distribution. In total, four cases are simulated to investigate the nonstationary responses of the bearings. As this research aims to diagnose the bearing faults at an early stage, the SNR of the simulated signals are set up to a very low value. Combing the effects of white Gaussian noise and the phase noise, the developed fault detector can be evaluated and verified in different conditions.

The periodicity of the outer race faults is shown in Figure 6-1 and it can be seen that the variation of the fault periods increases with more slippage.

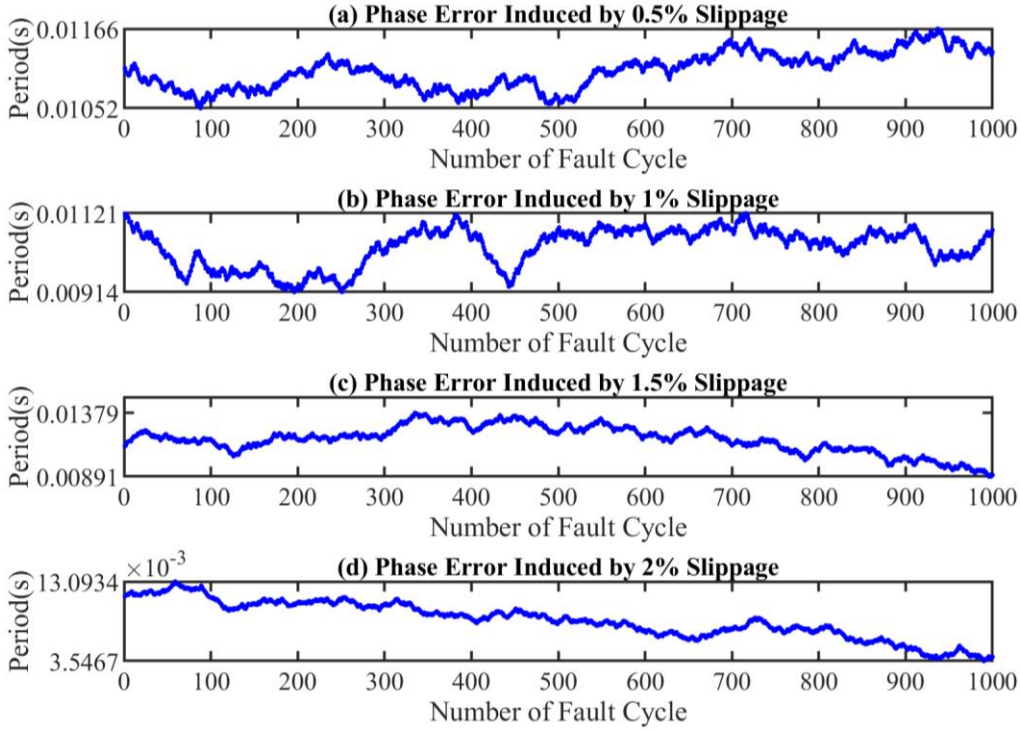


Figure 6-1 Phase error induced by slippage

The instantaneous characteristic fault frequency of the outer race fault is displayed in Figure 6-2. The variation of the fault frequency varies from 0.89Hz to 3.59Hz with respect to the slippage of 0.5% to 2%. The randomness of the bearing faults is simulated to fit to a uniform distribution, which allows a more accurate control of the slippage level. The variation of the instantaneous fault frequency manifestly obeys the uniform distribution.

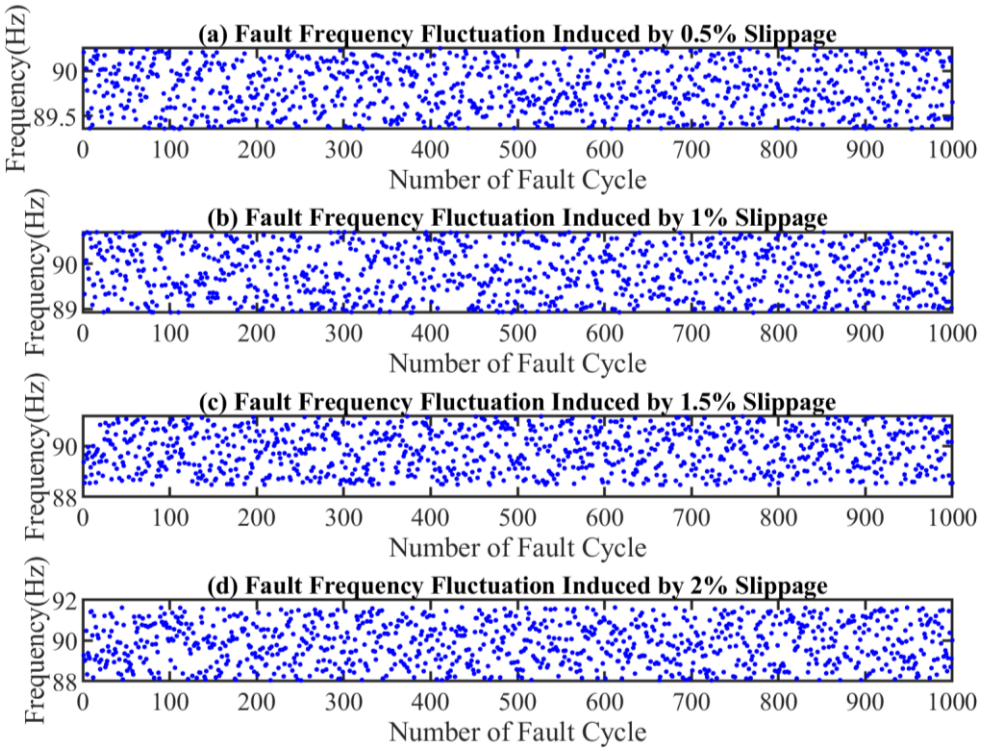


Figure 6-2 Fault frequency fluctuation induced by slippage

The variation of the impact timing makes the periodic signals to cyclostationary signals. The fluctuation of the fault cycles makes conventional fault detectors ineffective.

6.2 Method of Ensemble Average of Autocorrelated Envelopes (EAAE)

6.2.1 Ensemble Average of Autocorrelated Envelopes (EAAE)

Encouraged by the significant noise reduction, the EAAE is improved from the EAAS for tolerating the phase noise whilst suppressing the random noise. In some types of bearings, the slippage at the early fault stage are also large enough to destroy the periodicity of the fault information, which makes fault detection more difficult [110]. The input of the system becomes from the periodic forces into aperiodic impacts and hence the output of a linear system is difficult to characterise by stationary signal processing methods. To address these problems, the Ensemble Average of Autocorrelated Envelopes (EAAE) is derived based on autocorrelation functions of the envelope of the raw vibration signals. The detailed steps to implement EAAE is shown in Figure 6-3.

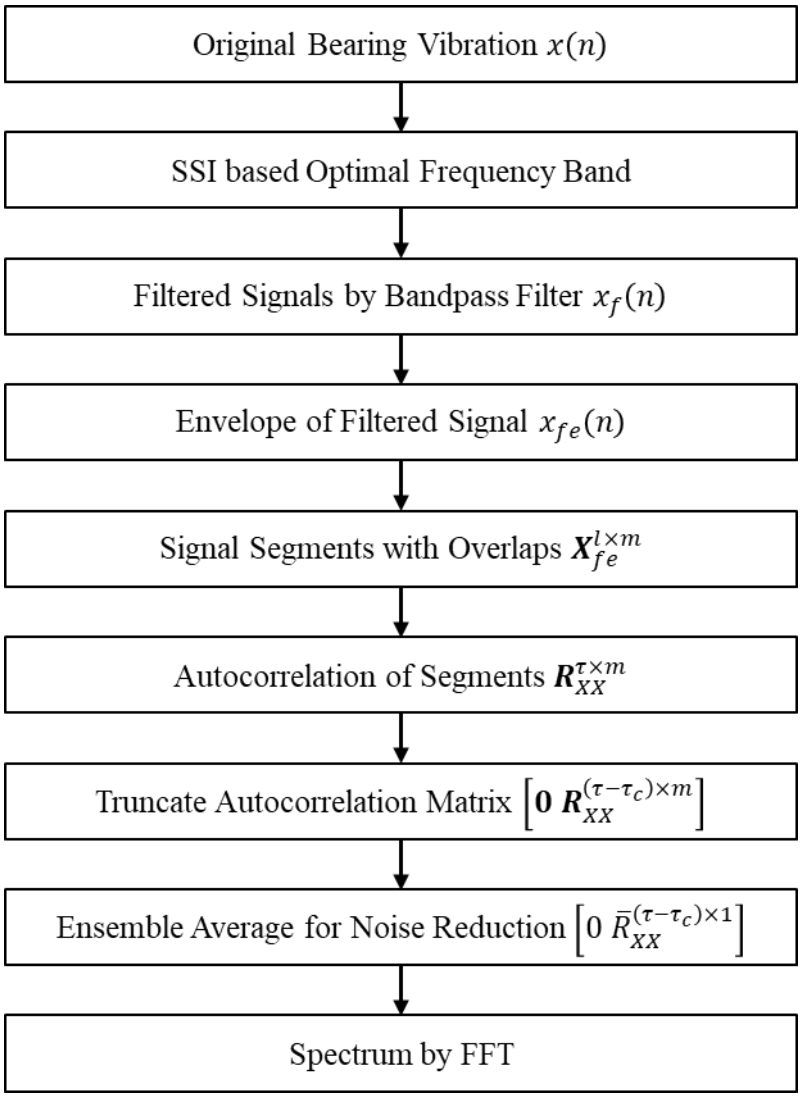


Figure 6-3 Flow chart of EAAE

In detail, the steps of EAAE are described as follows:

Step 1. The original vibration signal $x(n)$ with a length of N is processed by the Stochastic Subspace Identification to find the natural frequencies of the bearing system, which are used as the optimal central frequency in the following steps.

Step 2. The filtered signal $x_f(n)$ is obtained by implanting a Fourier bandpass filter based on the selected central frequency and the frequency band. The Fourier bandpass filter is achieved by only retaining the Fourier coefficients within the range of the bandpass filter. The other coefficients in the two-sided spectrum are assigned to be zero. The inverse Fourier transform of the modified Fourier coefficients can lead to the filtered signal $x_f(n)$.

Step 3. The envelope signals $x_f^e(n)$ are then obtained from the analytic signal of the filtered signal $x_f(n)$. The analytic signal can be calculated using Hilbert transform.

Step 4. The envelope vector is re-represented to be a buffer matrix that comprises of short segments with overlaps. The reorganisation of the signals is shown in Equation (4.58). The filtered signal with a length of N is rearranged to $m+1$ short segments with a length of l and an overlap of $l-k+1$. If the last segment is less than l, the segment is zero padded to l.

$$\boldsymbol{X}_{fe}^{l\times m}=\begin{bmatrix} x_f^e(1) & x_f^e(k) & \cdots & x_f^e\left(k+(m-2)(k-1)\right) & \vdots \\ \vdots & \vdots & \cdots & \vdots & x_f^e(N) \\ x_f^e(k) & x_f^e\left(k+(k-1)\right) & \cdots & x_f^e\left(k+(m-1)(k-1)\right) & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_f^e(l) & x_f^e\left(l+(k-1)\right) & \cdots & x_f^e\left(l+(m-1)(k-1)\right) & 0 \end{bmatrix} \quad (6.3)$$

The segment length l is determined by the demand of frequency resolutions in the diagnostic results. In this thesis, the resolution of 1Hz is used to distinguish the characteristic fault frequencies and the length l is double of the sampling frequency, which will be explained in the next step. The overlap ratio $(l-k+1)/l$ is usually recommended to be 0.5 to 0.7 by experience, which increases the average times in the ensemble average operation. A too high overlap ratio does not contribute to the optimisation of the final results, which is not meaningful in the practical signal analysis.

Step 5. The segments are used to calculate the unbiased autocorrelation signals, which converts the time domain signal into the lag domain with a long lag to ensure a high frequency resolution for fault detection. The lag length directly determines the final frequency resolution. In this thesis, the resolution of 1Hz is used and the lag length is the same as the sampling frequency to ensure a high frequency resolution for fault detection. Moreover, to increase the accuracy of the autocorrelation, the length l of the signal segments is chosen to be double that of autocorrelation lag τ.

$$\boldsymbol{X}_{fe}^{l\times m}\Rightarrow\begin{bmatrix} R_{xx}^1(1) & R_{xx}^2(1) & \cdots & R_{xx}^m(1) \\ R_{xx}^1(2) & R_{xx}^2(2) & \cdots & R_{xx}^m(2) \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^1(\tau) & R_{xx}^2(\tau) & \cdots & R_{xx}^m(\tau) \end{bmatrix}=\boldsymbol{R}_{XX}^{\tau\times m} \quad (6.4)$$

Step 6. The beginnings of the autocorrelated envelope signals contain a greater amount of noise due to the limited signal lag. To eliminate the influence of extreme values at the beginnings, the autocorrelated envelopes of each segment only retain the lags which are greater than τ_c and the others are assigned to be zero. The τ_c is empirically selected to be about one tenth of the lag τ.

$$\boldsymbol{R}_{XX}^{\tau\times m} \Rightarrow \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^{1}(\tau_c+1) & R_{xx}^{2}(\tau_c+1) & \cdots & R_{xx}^{m}(\tau_c+1) \\ R_{xx}^{1}(\tau_c+2) & R_{xx}^{2}(\tau_c+2) & \cdots & R_{xx}^{m}(\tau_c+2) \\ \vdots & \vdots & \cdots & \vdots \\ R_{xx}^{1}(\tau) & R_{xx}^{2}(\tau) & \cdots & R_{xx}^{m}(\tau) \end{bmatrix} \quad (6.5)$$

Step 7. The ensemble average of the autocorrelated envelope segments can be calculated as follows.

$$\bar{\boldsymbol{R}}_{XX}^{\tau\times 1} = E[\boldsymbol{R}_{XX}^{\tau\times m}] \quad (6.6)$$

Step 8. The spectrum is obtained by the Fourier transform, which makes the fault detection and diagnosis easily confirmed by the sparse representation in the frequency domain.

6.2.2 Similarity and Difference between EAAS and EAAE

The similarity and difference of EAAS and EAAE are displayed in Figure 6-4. The EAAS is designed to process the stationary signals, which can significantly suppress the random noise in the periodic signals. The EAAS is more powerful to extract the periodic information in the stationary signals because the amplitude and phase information are identical in different cycles whilst the EAAE is good at processing the cyclostationary signals, especially for the bearing vibration signals. The transient responses of the rolling elements passing the defects are the core information in detecting and diagnosing bearing faults. The outlines of the transient responses are the envelopes extracted despite the noise. Considering the transient responses are not periodic, the phase noise results in the ineffectiveness of the autocorrelation. In contrast, the envelope can align the outlines to a certain degree and make the phase to be periodic despite the nonstationary timing of the transient responses, which allows the autocorrelation to retain the periodic parts and remove the nonstationary parts. The EAAE exploits the merits of the envelope analysis to characterise the nonstationary (especially cyclostationary) signals, which can greatly increase the signal to noise ratio of the desired information. For practical applications, the difference of EAAS and EAAE is the sequence of the envelope and autocorrelation, however these two methods are under different theoretical background. The segments lead to the autocorrelation matrix $[R_{xx}(t_1,\tau), R_{xx}(t_2,\tau), \cdots R_{xx}(t_n,\tau)]$, which can increase the robustness of the method to nonstationary components in the signals.

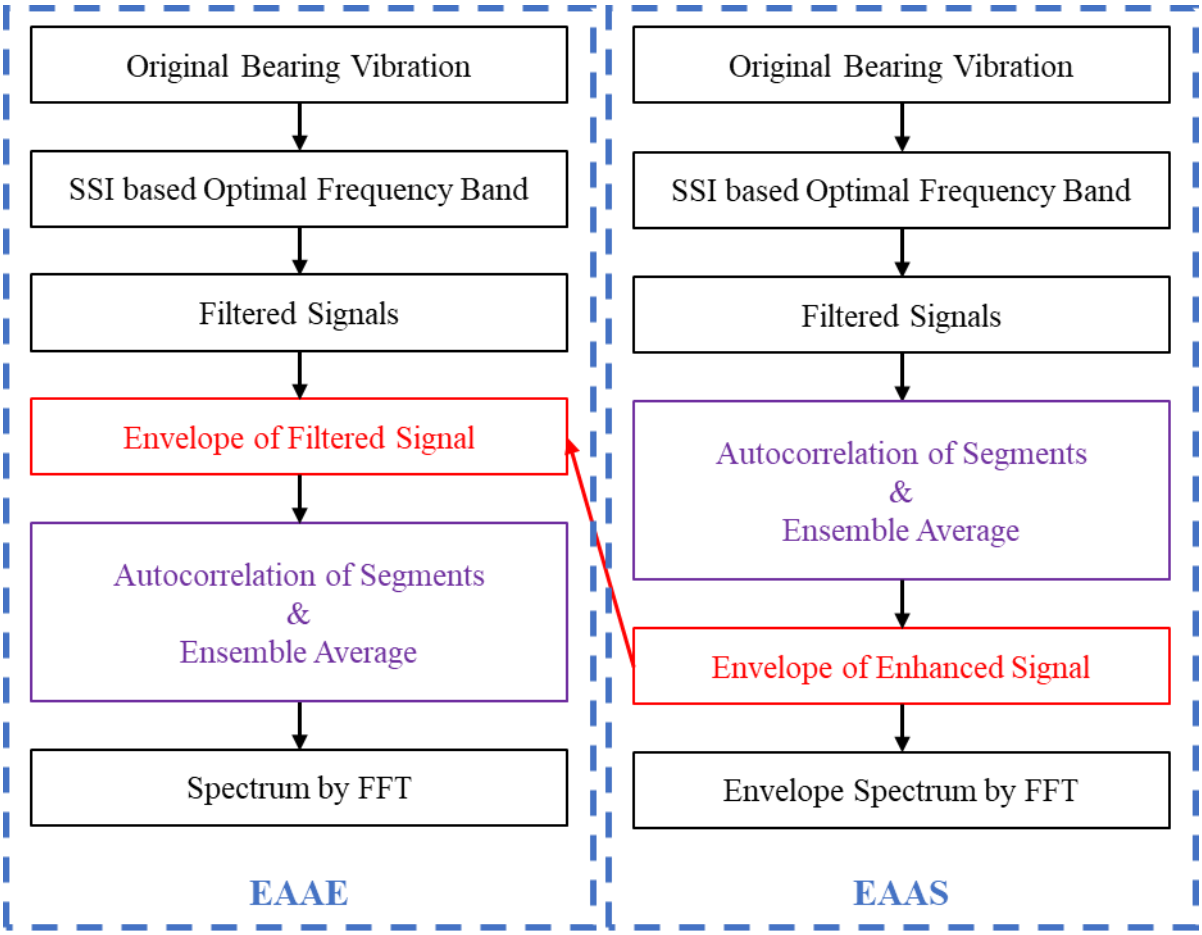


Figure 6-4 Similarity and difference between EAAS and EAAE

6.3 Method of Phase Linearisation based Modulation Signal Bispectrum (PL-MSB)

6.3.1 Modulation Signal Bispectrum (MSB)

The method of Modulation Signal Bispectrum (MSB) originates from high order spectrum analysis. A conventional second order spectrum analysis method, bispectrum (CB), is defined as follows [29]

$$B_{CB}(f_1, f_2) = E\{X(f_1)X(f_2)X^*(f_1 + f_2)\} \quad (6.7)$$

in which, $X(f) = \int x(t)e^{-i2\pi ft}dt$ denotes the Fourier transform of a temporal signal $x(t)$; f_1 and f_2 are two frequency components from the spectrum $X(f)$; $X^*(f)$ is the complex conjugate of $X(f)$; and $E\{\ \}$ denotes the ensemble average. The conventional bispectrum is a complex third order measurement and it contains both the magnitude and phase of the interacting components. If the frequency components of f_1 and f_2 are nonlinearly coupled, the additional component $f_1 + f_2$ would be generated. The total phase in the bispectrum phrase $X(f_1)X(f_2)X^*(f_1 + f_2)$ is denoted as

$$\varphi_{CB}(f_1, f_2) = \varphi(f_1) + \varphi(f_2) - \varphi(f_1 + f_2) \quad (6.8)$$

The phase information of the coupled components would be constant, which shows good performance in the demodulation analysis of the quadratic coupled frequency components f_1 and f_2. If the components f_1 and f_2 are independent components or noise, the total phase $\varphi_{CB}(f_1, f_2)$ will be random within $[-\pi, \pi]$ and the averaging of the bispectrum phrase $X(f_1)X(f_2)X^*(f_1 + f_2)$ will be close to zero. With ensemble averaging in the Equation (6.7), the coupled components are enhanced owing to the same phase information. The other independent components and random noise will tend towards zero due to random phase.

The MSB is an improved version of the conventional bispectrum and it considers the lower and upper sideband at the same time, which allows accurate description of modulation signals. The MSB method is defined as

$$B_{MS}(f_1, f_2) = E\{X(f_1 + f_2)X(f_1 - f_2)X^*(f_1)X^*(f_1)\} \quad (6.9)$$

where, $X(f)$ denotes the Fourier transform of the temporal signal $x(t)$; f_1 and f_2 are two frequency components in the spectrum $X(f)$; $X^*(f)$ is the complex conjugate of $X(f)$; and $E\{\ \}$ denotes the operation of ensemble average. For a periodic modulation signal, the phase of the MSB phrase $X(f_1 + f_2)X(f_1 - f_2)X^*(f_1)X^*(f_1)$ can be calculated as

$$\phi(f_1, f_2) = \phi(f_1 + f_2) + \phi(f_1 - f_2) - \phi(f_1) - \phi(f_1) \quad (6.10)$$

If the components f_1 and f_2 are coupled by modulation (AM or PM), the subsequent phase can be expanded as

$$\begin{cases} \phi(f_1 + f_2) = \phi(f_1) + \phi(f_2) \\ \phi(f_1 - f_2) = \phi(f_1) - \phi(f_2) \end{cases} \quad (6.11)$$

Consequently, the total phase of the MSB phrase will be constant. The constant phase of the coupled components allows the ensemble average of the MSB phrase from different signal segments. The averaging can enhance modulation characteristics and suppress both the independent components and random noise. However, the potential lower sideband in modulation signals is not inclusive in the CB. The MSB method is more effective in characterising modulation signals because it considers the upper and lower sidebands from a modulation signal simultaneously. The MSB approach has been investigated in references [34], [122]–[124] and the MSB is able to extract the modulating components in the vibration, acoustic, and motor current signals for diagnosing machine faults. It has been demonstrated that the MSB method is highly effective in extracting incipient fault signatures. The CB is not adequate in the analysis of the modulation signals [36]. Sidebands around the carrier frequency component are usually much smaller than the carrier signal. The estimation of the sidebands by the FFT spectrum is usually not

reliable because the background noise is strong and camouflages these small but important components. The implementation of the MSB method is based on the conventional Fourier transform, which is not an efficient way to tackle cyclostationary signals. For this reason, the MSB is extended by using a phase linearisation method to improve the performance in characterising the cyclostationary signals. The improvement can increase the strength of the deterministic components, allowing the PL-MSB to process the cyclostationary signals effectively. The developed PL-MSB approach aims to process the cyclostationary signals from rolling element bearings. The cyclostationary signals are generated by the random slippage between bearing components. The random slippage in the rolling element bearings can reach up to 2% [110] and the slippage makes the theoretically periodic impacts into a pseudo-periodic series, terminologically named as cyclostationary signals.

6.3.2 Phase Linearisation based Modulation Signal Bispectrum (PL-MSB)

Many methods have been developed to characterise time varying vibration signals. One of the most popular is Time Synchronous Averaging (TSA). TSA is extensively used in vibroacoustic signal analysis due to the merits of noise reduction. The averaging can lead to high Signal to Noise Ratio (SNR), and it can achieve successful fault diagnosis of rotating machines [120], [125], [126]. The implementation of the TSA approach relies on reference signals from a tachometer. The shaft revolution is usually the working cycle of machines, for instance gears and engines, and therefore, the dynamic responses of the rotating machine are repeatable in revolutions. Consequently, the vibroacoustic signals from different shaft revolutions can be averaged. The deterministic components can be enhanced for the purpose of accurate diagnoses. However tachometer signals are not always available, and hence 'tacho-less' TSA techniques have been developed to overcome this deficiency [127]–[130]. The tacho-less approaches rely on the accurate estimation of the instantaneous frequency/phase [125], [126], [131]. With the accurate frequency/phase information of the target components, the raw vibroacoustic signals can be segmented in the time domain and then averaged in the angular domain. However, the TSA method in bearing fault diagnosis is not applicable because the bearing fault frequencies is usually not the integer of the shaft frequency. Moreover, the phase noise in bearing vibration signals is transient, which is difficult to tackle in the TSA method. Therefore, the phase linearisation is developed here to minimise the influence of the phase noise.

A monocomponent signal is described by one single "ridge" in the time and frequency domain, and the ridge represents the energy concentrated region in the time-frequency representation. Such a monocomponent signal $s(t)$ can be expressed in a complex form as

$$s(t) = Re\{a(t)e^{i\varphi(t)}\} \quad (6.12)$$

where, $a(t)$ and $\phi(t)$ are the instantaneous amplitude and instantaneous phase respectively; $Re\{\ \}$ represents the real part of the complex number. For a periodic signal, the instantaneous phase $\varphi(t)$ is a linearly incremental function of time, which can be expressed as

$$\varphi(t) = \omega t + \varphi(0) \quad (6.13)$$

where, $\varphi(0)$ is the initial phase of this periodic signal and ω is the angular frequency of the periodic signal. Periodic signals are an exception and nonstationary signals are generic in nature. The periodicity in vibration signals is distorted by the unstable working conditions, which are due to the internal and external factors. If the instantaneous phase is distorted by the random slippage, and the monocomponent signal $s(t)$ can be expressed as

$$s(t) = Re\{a(t)e^{i[\varphi(t)+n(t)]}\} \quad (6.14)$$

in which, $n(t)$ denotes the phase noise from random slippage and speed oscillation. Then, the distorted instantaneous phase $\phi(t)$ yields to

$$\phi(t) = \varphi(t) + n(t) \quad (6.15)$$

Therefore, the instantaneous phase of the distorted periodic signal is

$$\phi(t) = \omega t + \varphi(0) + n(t) \quad (6.16)$$

The random phase error $n(t)$ is an independent variate, which is not correlated with the linear increment ωt and the initial phase $\varphi(0)$. The dominate component in $\phi(t)$ is the deterministic component $\omega t + \varphi(0)$. Consequently, a linear regression model is a good way to explain the variation of the instantaneous phase. The phase noise $n(t)$ becomes the random deviations in the linear representation. The linear regression model removes the random errors and retains the linear part, which can be shown as

$$\hat{\phi}(t) = at + b \quad (6.17)$$

where, the coefficient a and the instance b can be obtained from the linear fitting model.

An amplitude modulation signal composed of multiple components can be described by the individual mono-component signals. In this manner the phase noise induced by random slippage within a bearing can be supressed and the strength of the deterministic components can be enhanced by the phase linearisation. In detection and diagnosis of incipient bearing faults, the raw signals are severely contaminated by strong background noise. The estimation of the instantaneous phase is easily influenced by the random noise. For the purpose of accurate estimation of the phase information, a linear prediction filter is used to suppress random noise [132], [133]. A linear prediction model can predict the deterministic random signals, including Gaussian white noise.

The residual signal then retains the bearing fault related components but with much less noise, and this in turn allows the more accurate estimation of instantaneous phase. After phase correction, the cyclostationary signal is modified to be approximately periodic one. The new phase relationship can be more realistically characterised by the MSB method to further suppress noise. The PL-MSB can be achieved by the following equation

$$\hat{B}_{MS}(f_c, f_x) = E\left\langle \hat{X}(f_c + f_x)\hat{X}(f_c - f_x)\hat{X}^*(f_c)\hat{X}^*(f_c) \right\rangle \quad (6.18)$$

where, $\hat{X}(f)$ is the Fourier transform of the reconstructed signal $\hat{x}(t) = x(t)e^{-in(t)}$. The steps of the proposed PL-MSB is shown in the Figure 6-5.

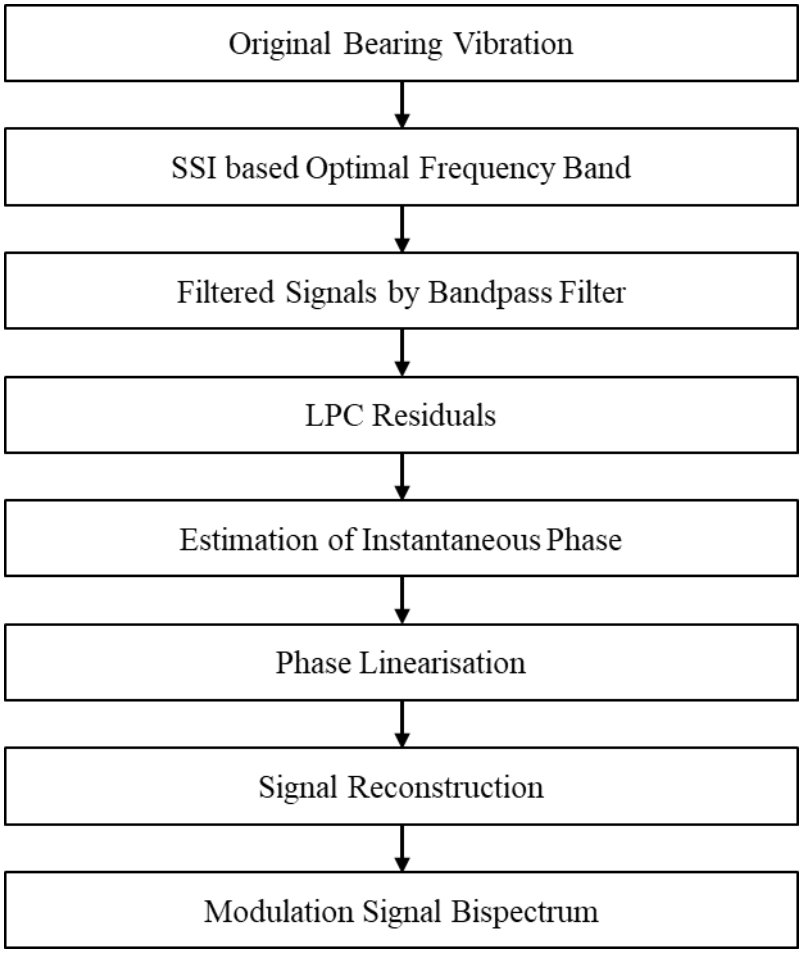


Figure 6-5 Flow chart of the PL-MSB method

The details of each step are introduced as follows.

Step 1: The original vibration signals $x(t)$ are pre-processed by the Stochastic Subspace Identification to identify the natural frequencies and then the natural frequencies can be evaluated to find the optimal frequency bands.

Step 2: The filtered signal is obtained by implanting a Fourier bandpass filter based on the selected central frequency and the frequency band. The Fourier bandpass filter is achieved by only

retaining the Fourier coefficients within the range of the bandpass filter. The other coefficients in the two-sided spectrum are assigned to be zero. The inverse Fourier transform of the modified Fourier coefficients can lead to the filtered signal $x_f(n)$.

Step 3: The filtered signal is processed by the linear prediction method for roughly supressing the noise. The deterministic random process, for example the Gaussian white noise, can be perfectly predicted by the linear prediction. The order of LPC is usually determined by the trial and error method. In this proposed method, the order of the linear prediction is selected as high as 200 to allow the residual signal to be obtained.

Step 4: The filtered residual signal is used to obtain the analytic signal by the Hilbert transform. The analytic signal is employed to estimate the instantaneous phase $\phi(t)$ of the residual signal.

Step 5: The instantaneous phase is linearised by a linear fitting and then generates a linearly incremental phase. The new phase array is used to reconstruct the new signal $\hat{x}(t)$ via the spline interpolation.

Step 6: The new signal phase $\hat{x}(t)$ is used as the input of the Modulation Signal Bispectrum (MSB) for the MSB magnitudes and coherence. The concise and interpretable result is extracted by the peak of the carrier frequency.

6.4 Demodulation Analysis of Cyclostationary Bearing Signals

The demodulation analysis of the bearing faults in this section focuses on evaluating the performance of the proposed methods in the incipient bearing fault detection and diagnosis. The main problem discussed is the influence of the random phase errors in the cyclostationary signals under the conditions of extremely low SNR (less than -30dB).

6.4.1 Demodulation Analysis of Outer Race Fault Signals

As explained at the beginning, the analysis of the severely contaminated bearing signals is presented in this section. The outline of the noisy signals is displayed to address the challenge of the vibration signals. Figure 6-6 shows the spectrum of the simulated cyclostationary bearing signals at different levels of random slippage under the SNR of -31dB. The spectrums denote that the phase noise significantly decreases the amplitude of the resonant areas in the frequency domain, which can be verified by the spectrums in the Figure 4-7. Due to the nonstationary input, the output of the bearing model is too difficult to characterise by the stationary signal processing approaches. Consequently the Fourier transform based spectrum shows that the spectral energy is smeared over frequency bins. The spectrums of noise-free signals in the Figure 4-7 is three times higher than the

amplitude in the Figure 6-6. The leakage of the spectrum represents three clusters in the frequency domain other than the sparse components in the periodic signals.

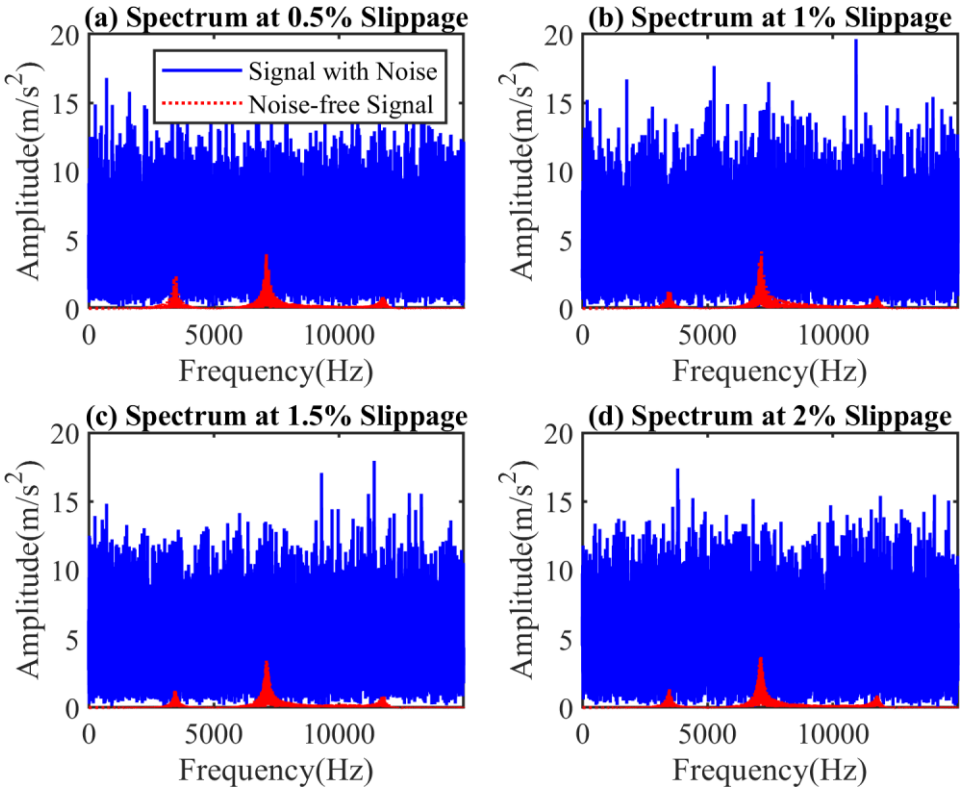


Figure 6-6 Spectra of cyclostationary outer race fault bearing signals at SNR -31dB and: (a) 0.5% slippage; (b) 1% slippage; (c) 1.5% slippage and (d) 2% slippage

The prior step in the proposed methods is to select an appropriate frequency band for the purpose of extracting fault signatures accurately. Therefore, the results of the Kurtogram and SSI are shown in the following contents. The recommended frequency bands by the Kurtogram and SSI are displayed in the Figure 6-7 to Figure 6-14. In the case of outer race faults at 0.5% slippage and SNR -31dB, the most impulsive frequency band is from 0Hz to 25kHz denoted by the Kurtogram at level 1 in Figure 6-7. The SSI selects the central frequency of 7120Hz in Figure 6-8.

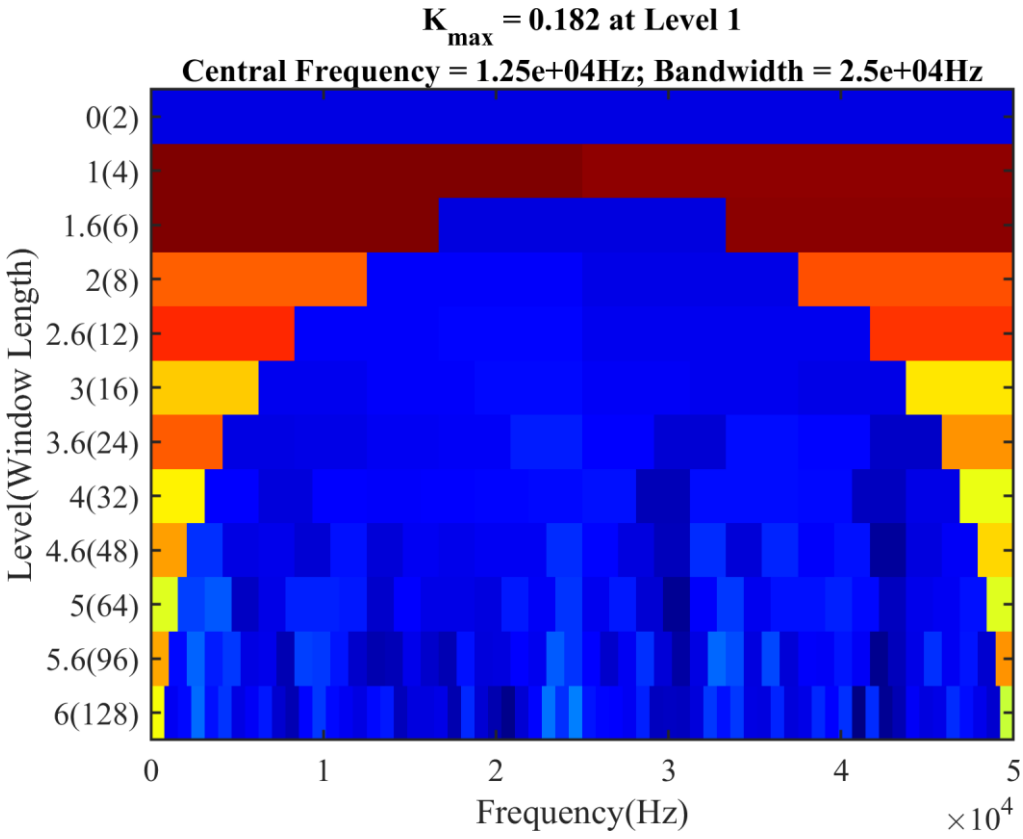


Figure 6-7 Kurtogram for the case of outer race faults at 0.5% slippage and SNR -31dB

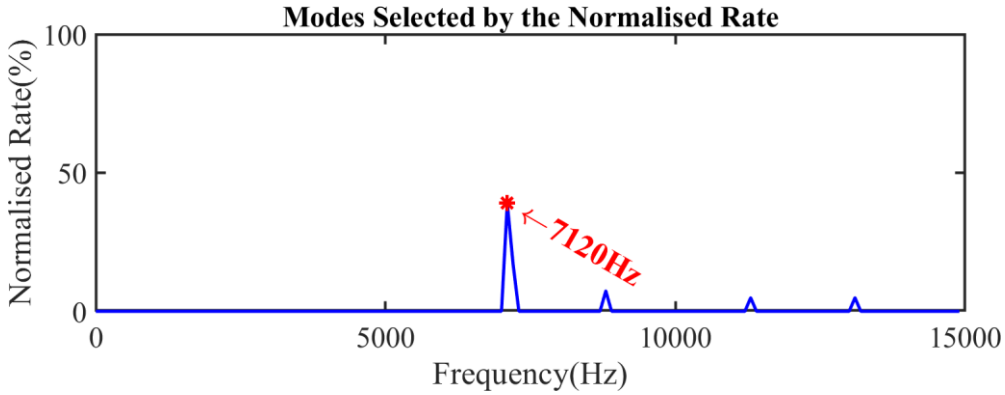


Figure 6-8 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at 0.5% slippage and SNR -31dB

In the case of outer race faults at 1% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 41.67kHz with a frequency bandwidth of 16.67kHz at level 1.6 in Figure 6-9. The SSI selects the central frequency of 7080Hz in Figure 6-10.

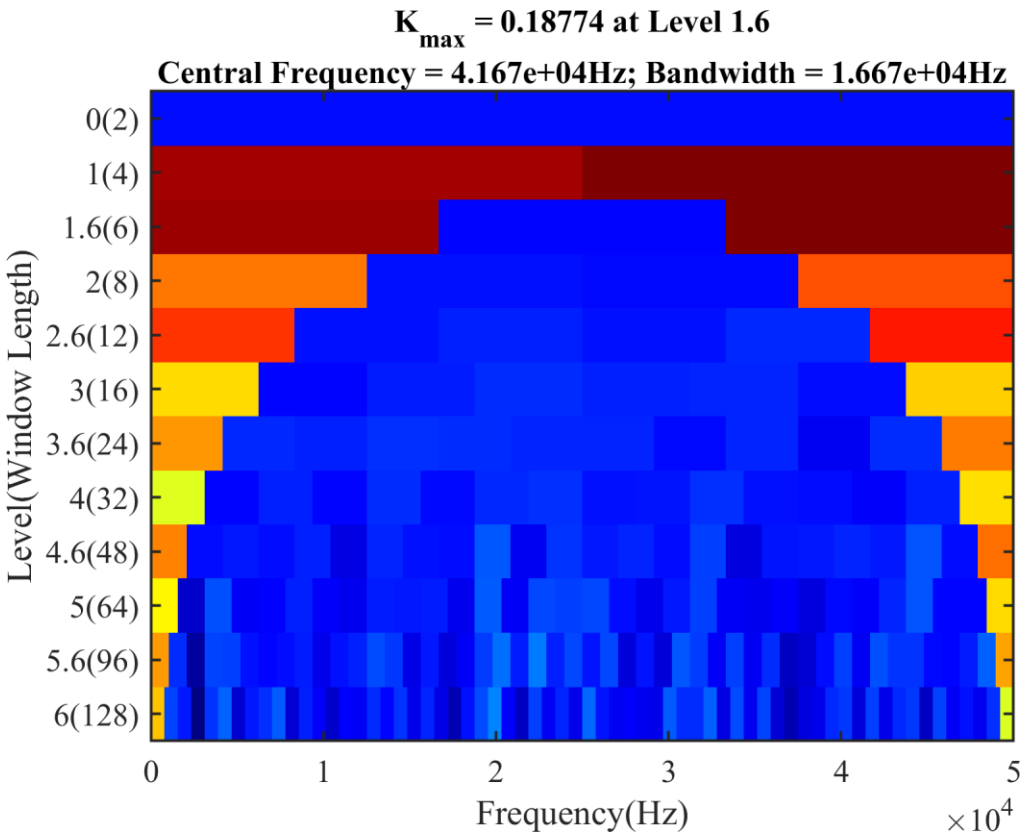


Figure 6-9 Kurtogram for the case of outer race faults at 1% slippage and SNR -31dB

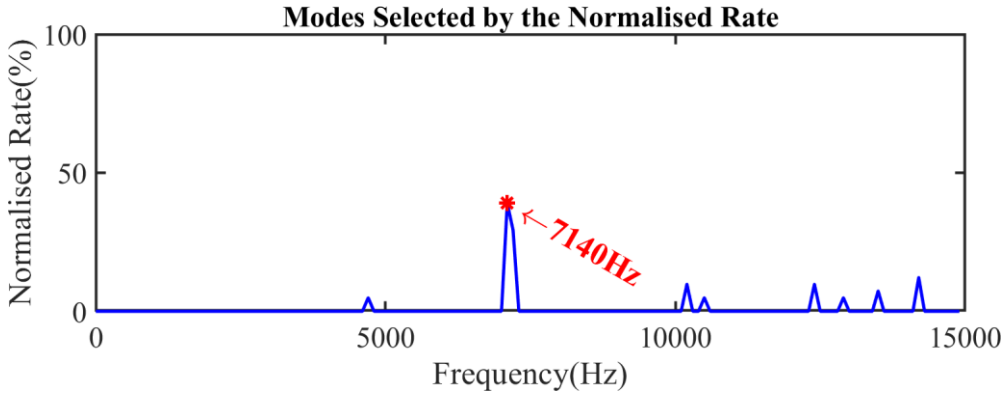


Figure 6-10 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at 1% slippage and SNR -31dB

In the case of outer race faults at 1.5% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 41.67kHz with a frequency bandwidth of 16.67kHz at level 1 in Figure 6-11. The SSI selects the central frequency of 7114Hz in Figure 6-12.

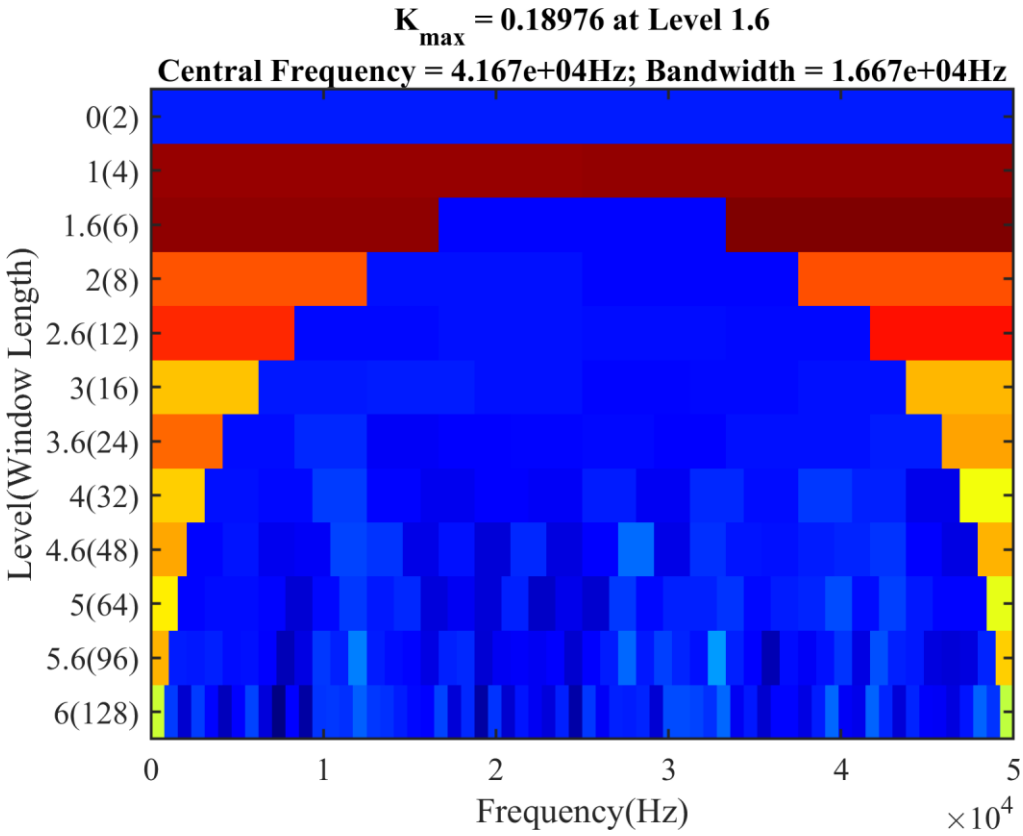


Figure 6-11 Kurtogram for the case of outer race faults at 1.5% slippage and SNR -31dB

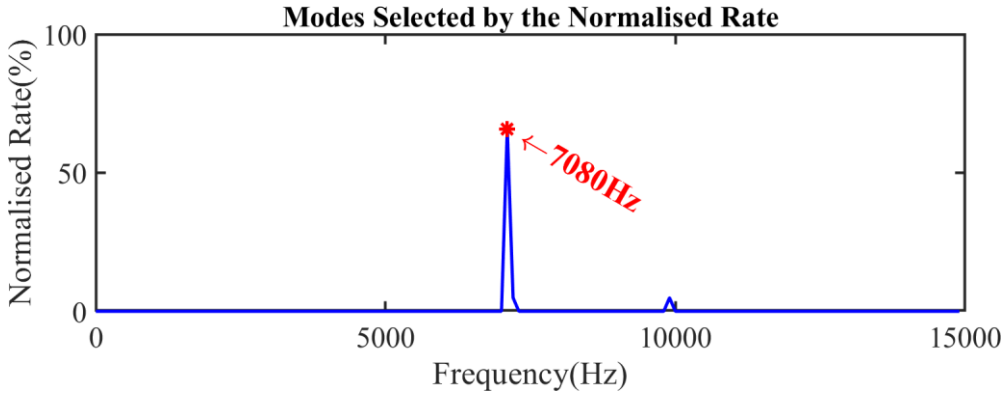


Figure 6-12 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at 1.5% slippage and SNR -31dB

In the case of outer race faults at 2% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 41.67kHz with a frequency bandwidth of 16.67kHz at level 1 in Figure 6-13. The SSI selects the central frequency of 7140Hz in Figure 6-14.

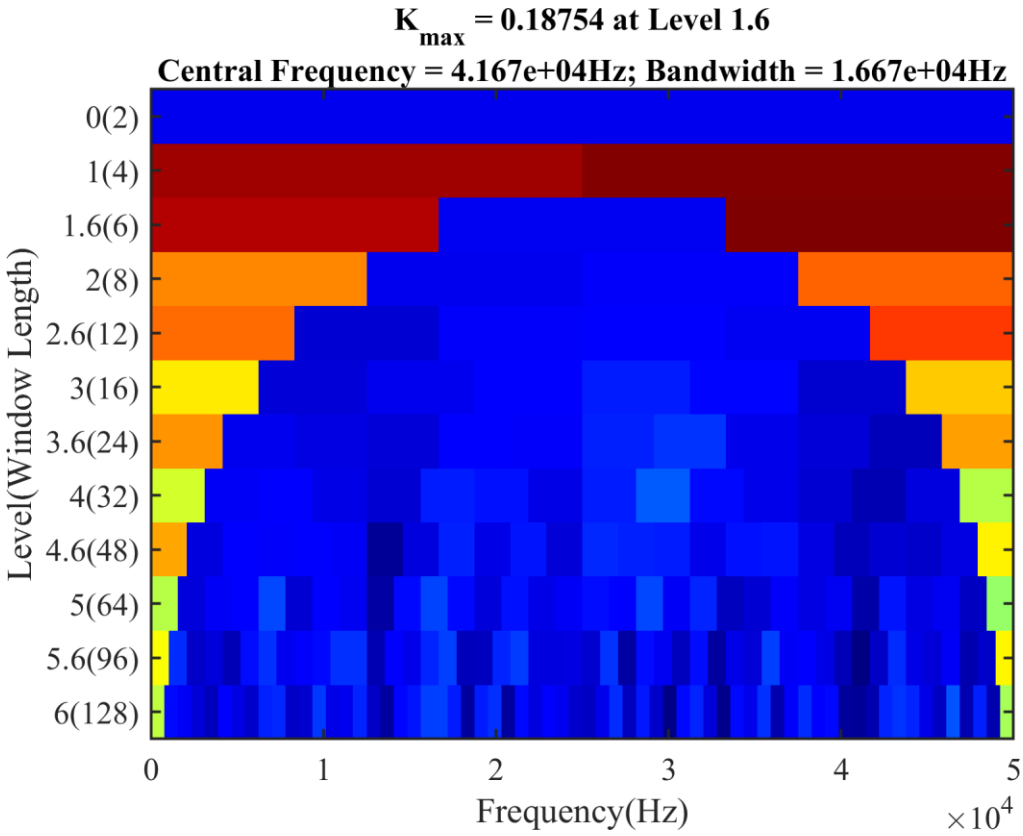


Figure 6-13 Kurtogram for the case of outer race faults at 2% slippage and SNR -31dB

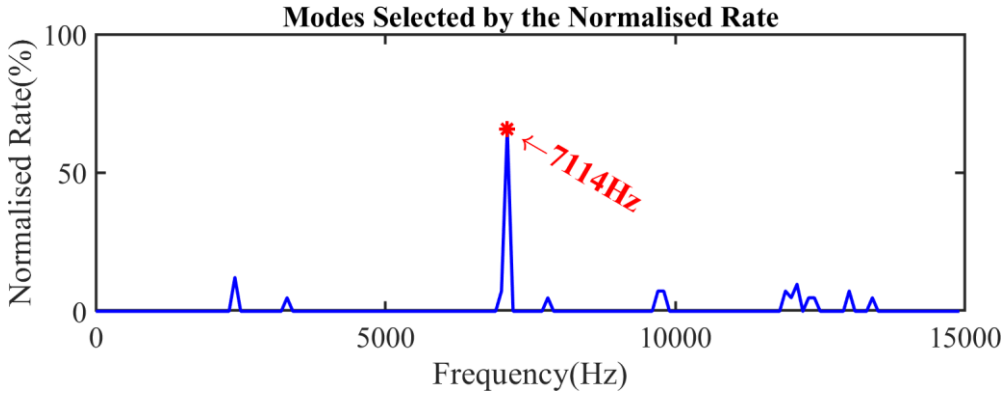


Figure 6-14 Normalised Rate of stable modes in the Stabilisation Diagram for the case of outer race faults at 2% slippage and SNR -31dB

The optimal central frequencies in the four slippage cases are shown in Table 6-1. Compared to the Kurtogram, the SSI method gives more reasonable determination of the optimal frequency bands for further demodulation analysis.

Table 6-1 Optimal frequency bands of cyclostationary outer race fault signals

Cases of Outer Race Faults	Kurtogram		SSI	
	Central Frequency	Frequency Bandwidth	Central Frequency	Frequency Bandwidth
0.5% Slippage	12500Hz	25000Hz	7120Hz	629Hz
1% Slippage	41666Hz	16666Hz	7140Hz	629Hz
1.5% Slippage	41666Hz	16666Hz	7080Hz	629Hz
2% Slippage	41666Hz	16666Hz	7114Hz	629Hz

Due to the large quantities of white noise, the frequency bands selected by the Kurtogram is misleading, which is a very wide frequency band covering all the natural frequencies of the bearing model. The wide frequency band inevitably involves a large amount of noise and consequently a wide frequency band is unable to give the promising fault diagnosis results especially for the incipient fault detection. The SSI method does not focus on finding the most impulsive frequency band but extracting the modal parameters from the vibration signals. The critical point in the OMA is that the system modes are excited by the external inputs, which means the impacts from the rolling elements passing the localised faults excite the identified natural frequencies. Therefore the identified modes from the bearing faulty signals are carrier frequencies in the modulation signals. The natural frequencies accurately identified by the SSI approach are the desired candidates in the demodulation analysis. However, the SSI is unable to be independent of background noise. If the SNR of the signals is too low, this can reach to the limit of the method. It can be seen that not all bearing modes can be extracted from the simulated vibration signals because the input for the bearing model is not stationary and furthermore the noise totally submerges the informative components.

Based on the central frequencies selected by the Kurtogram and SSI methods, the demodulation analysis of the proposed methods and the benchmarked methods was carried out. As the cyclostationary characteristics of the faulty bearing vibration signals were addressed here, the simulation studies are mainly to compare the capability of the five demodulation methods under different levels of random slippages. It has to be mentioned here that the aim of these methods is to detect and diagnose the bearing faults at an early stage. Therefore, a large amount of white Gaussian noise up to -31dB is employed to contaminate the bearing fault signals. The first cyclostationary case is that the bearing components occur a slippage of 0.5%, which means that the instantaneous outer race fault frequency varies from 89.35Hz to 90.25Hz in the simulated

signal. The variation of the instantaneous fault frequency and its harmonics are different compare to most publications [134]–[136]. Most of the research focuses on the vibration signals with slowly varying speeds whilst the cyclostationary bearing vibration signals are extremely transient responses. The conventional approaches for slow time varying signals are not applicable here.

As shown in Figure 6-15, the envelope spectrum of the phase linearised signal cannot reveal the outer race bearing faults at SNR -31dB and 2% slippage. Therefore, the phase linearisation is not sufficient for the early fault detection and diagnosis of bearings, which tells the necessity of the further noise reduction method. Consequently, the PL-MSB is developed to achieve the early fault diagnosis of rolling bearings.

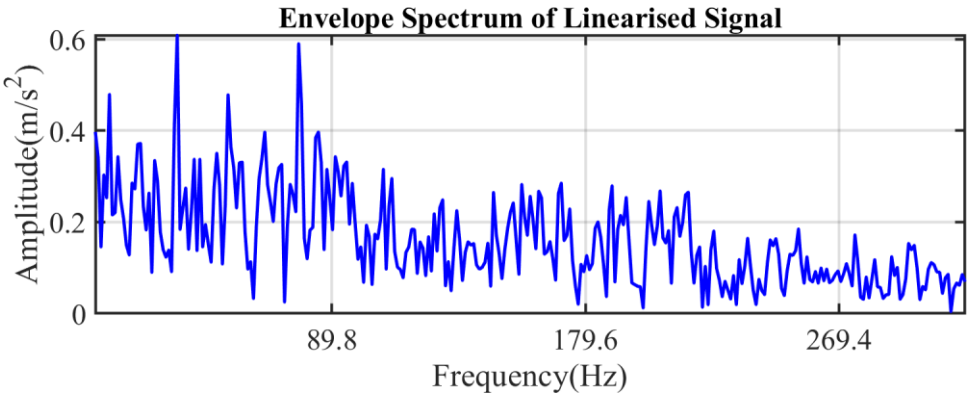


Figure 6-15 Envelope spectrum of the phase linearised signal for the case of outer race fault at SNR -31dB and 2% slippage

The results of the five methods, including three novel ones and two conventional ones, are depicted in the Figure 6-16. The Figure 6-16 (a) shows the proposed EAAE for the bearing signals with random slippages. The theoretical fault frequency of outer race defects is visible in the spectrum. The second harmonic of the fault theoretical frequency is not so pronounced due to the large quantities of amplitude and phase noise. The Figure 6-16 (b) displays the slice of the PL-MSB magnitudes which represents the concise two dimensional results for an easily interpretable diagram. The fault signature of the outer race failure is most obvious in these five methods. The PL-MSB based sparse spectrum denotes that the noise of both amplitude and phase has been significantly suppressed. However, due to the poor quality of the signals, the higher order harmonics of the fault theoretical frequency is not shown in the frequency domain. The three remaining methods, including EAAS and two CE methods, are displayed in Figure 6-16 (c), (d) and (e). These methods fail to detect the outer race faults because either the EAAS is not designed for the cyclostationary signals or the CE methods are unable to reduce the noise in the signals.

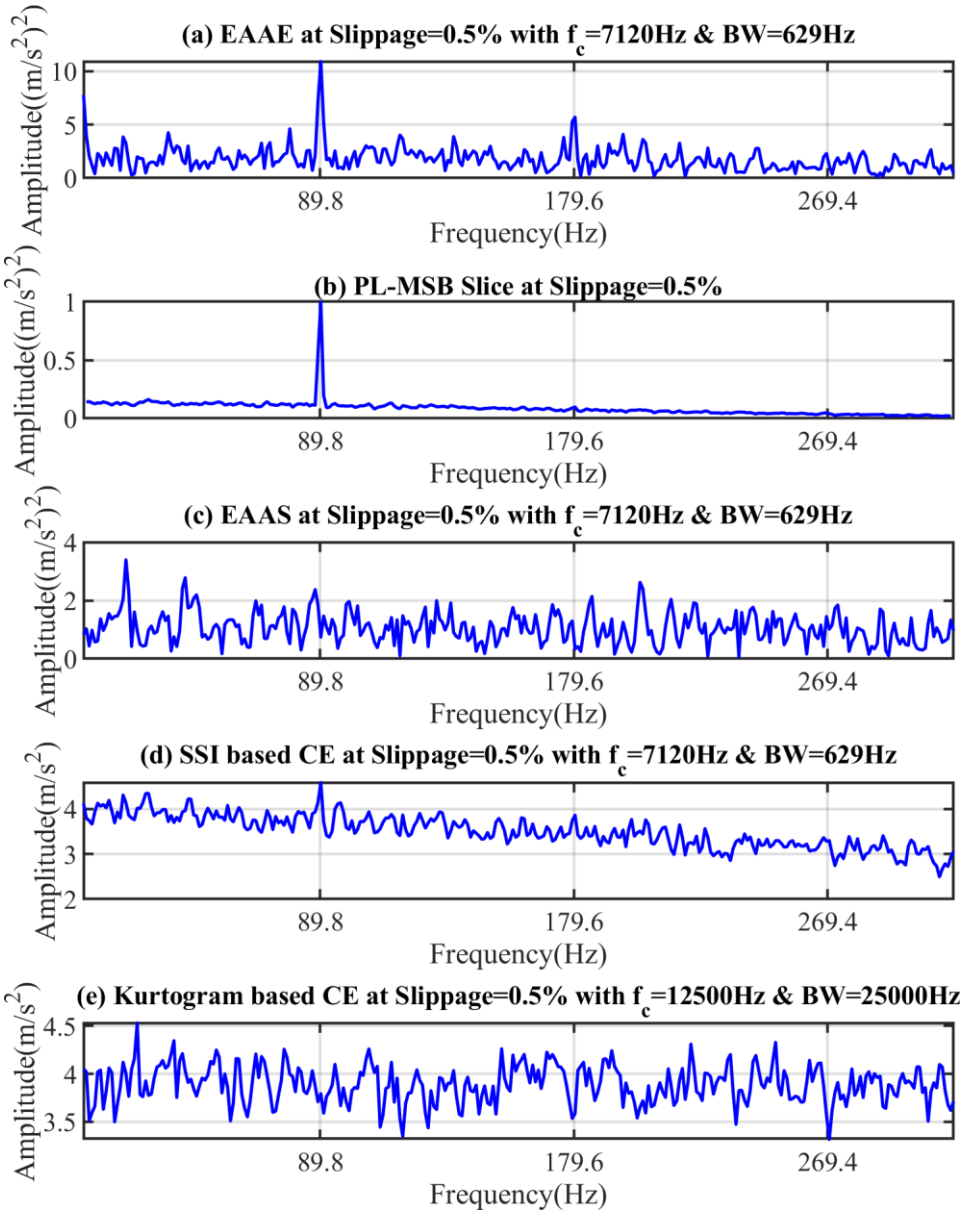


Figure 6-16 EAAE, PL-MSB, EAAS and CE for the case of outer race fault at SNR -31dB and 0.5% slippage

The increasing slippage case of the bearing with 1% phase noise is investigated to compare the effectiveness of the proposed methods. In the Figure 6-17 (a) and (b), both EAAE and PL-MSB are successful in obtaining the fault features from the severely contaminated noise. The capability of these two methods can be demonstrated sufficiently. The effectiveness of these two methods will be compared quantitatively based on the signature strength in Equation (4.68). Similarly to the case of 0.5% slippage, the EAAS is unable to extract the second order statistic parameters with periodicity. The definition of autocorrelation in Equation (4.46) is specific for the stationary signals, which exceeds the application scope when processing nonstationary signals. The SSI based CE in Figure 6-17 (d) can obtain the fault information but the noise carpet in the spectrum is very high, leading to the weak fault signatures. Additionally, due to the uncertainty of the added

noise, the SSI based CE is not always effective, which can be verified in the case of 0.5% slippage. The conventional envelope with the optimal frequency band has difficulty in identifying the faults and therefore the Kurtogram based CE is not beyond the expectation based on a significantly wide frequency band.

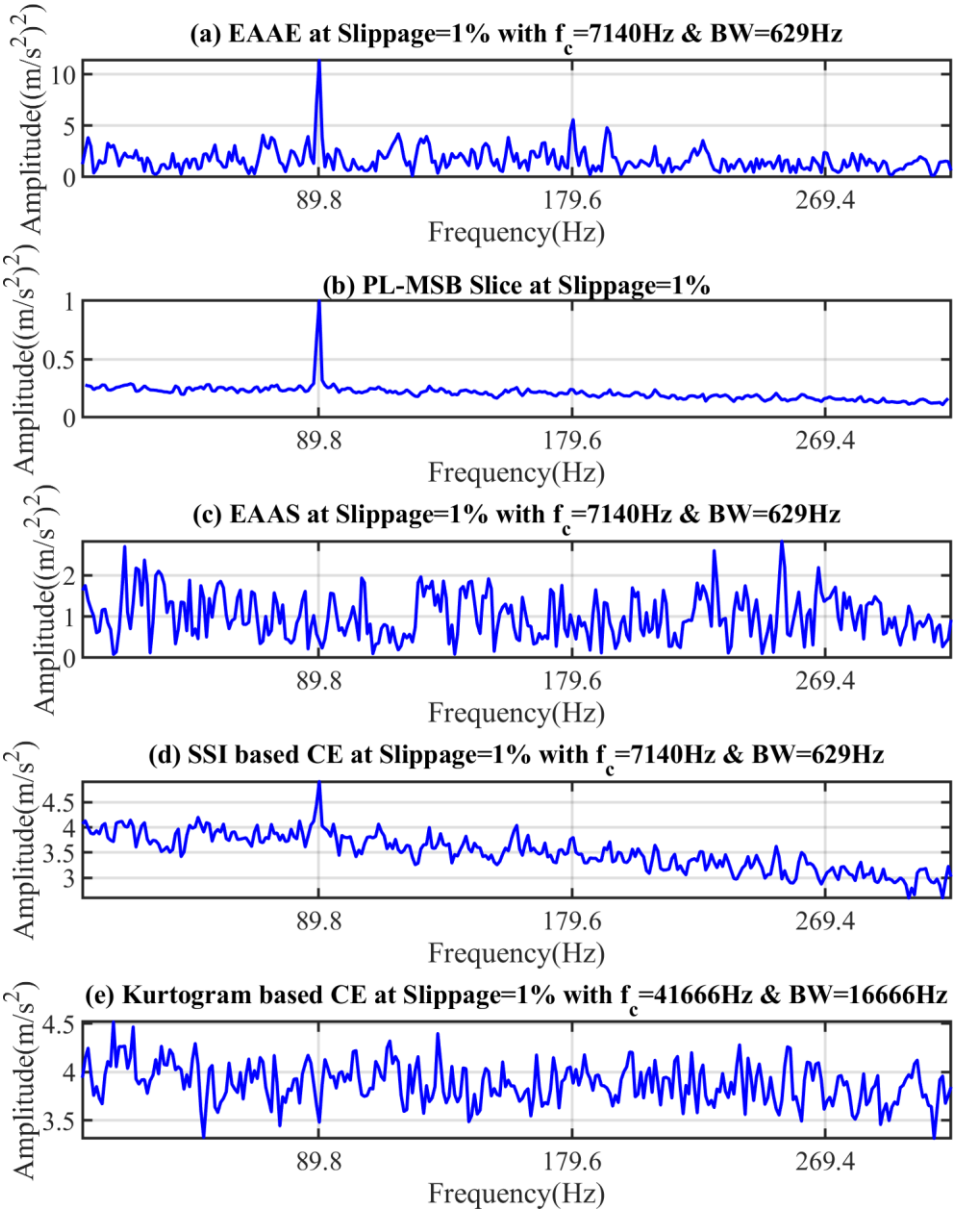


Figure 6-17 EAAE, PL-MSB, EAAS and CE for the case of outer race fault at SNR -31dB and 1% slippage

The demodulated signals in the frequency domain for the case of 1.5% slippage are similar to that for the case of 1% slippage.

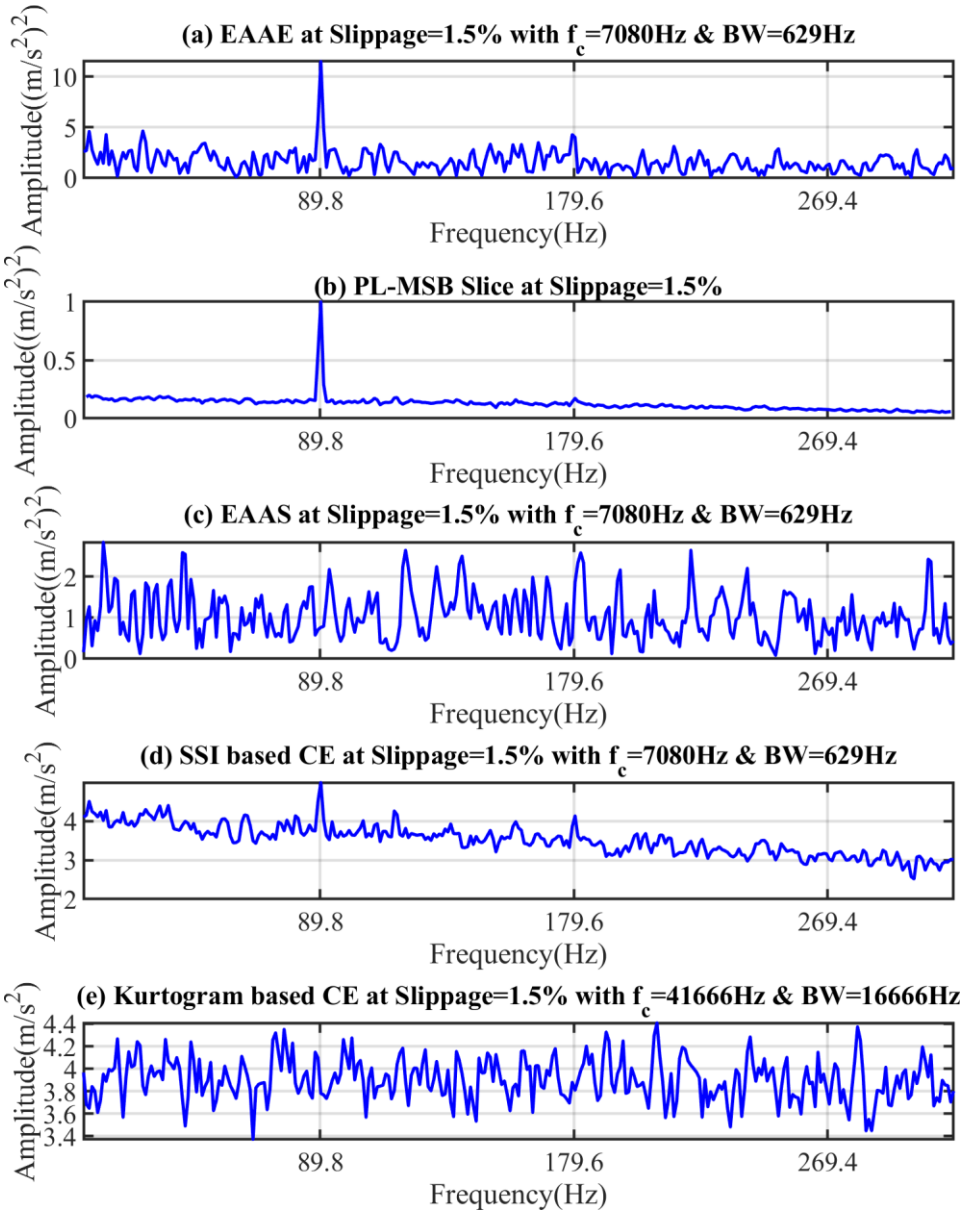


Figure 6-18 EAAE, PL-MSB, EAAS and CE for the case of outer race fault at SNR -31dB and 1.5% slippage

For the harshest case with 2% slippage between the rolling elements and raceways, the instantaneous frequency of the outer race faults varies from 88.0Hz to 91.6Hz. The difference of the fault frequency can reach up to 3.6Hz. The EAAE in Figure 6-19 (a) still works in the simulated signals but the strength of the fault indicators is not satisfied. It is shown in Figure 6-19 (b) that the PL-MSB is sufficient to give a reliable diagnosis result. However, the EAAS, SSI based CE and Kurtogram based CE are unable to extract fault features, which are shown in the Figure 6-19 (c), (d) and (e).

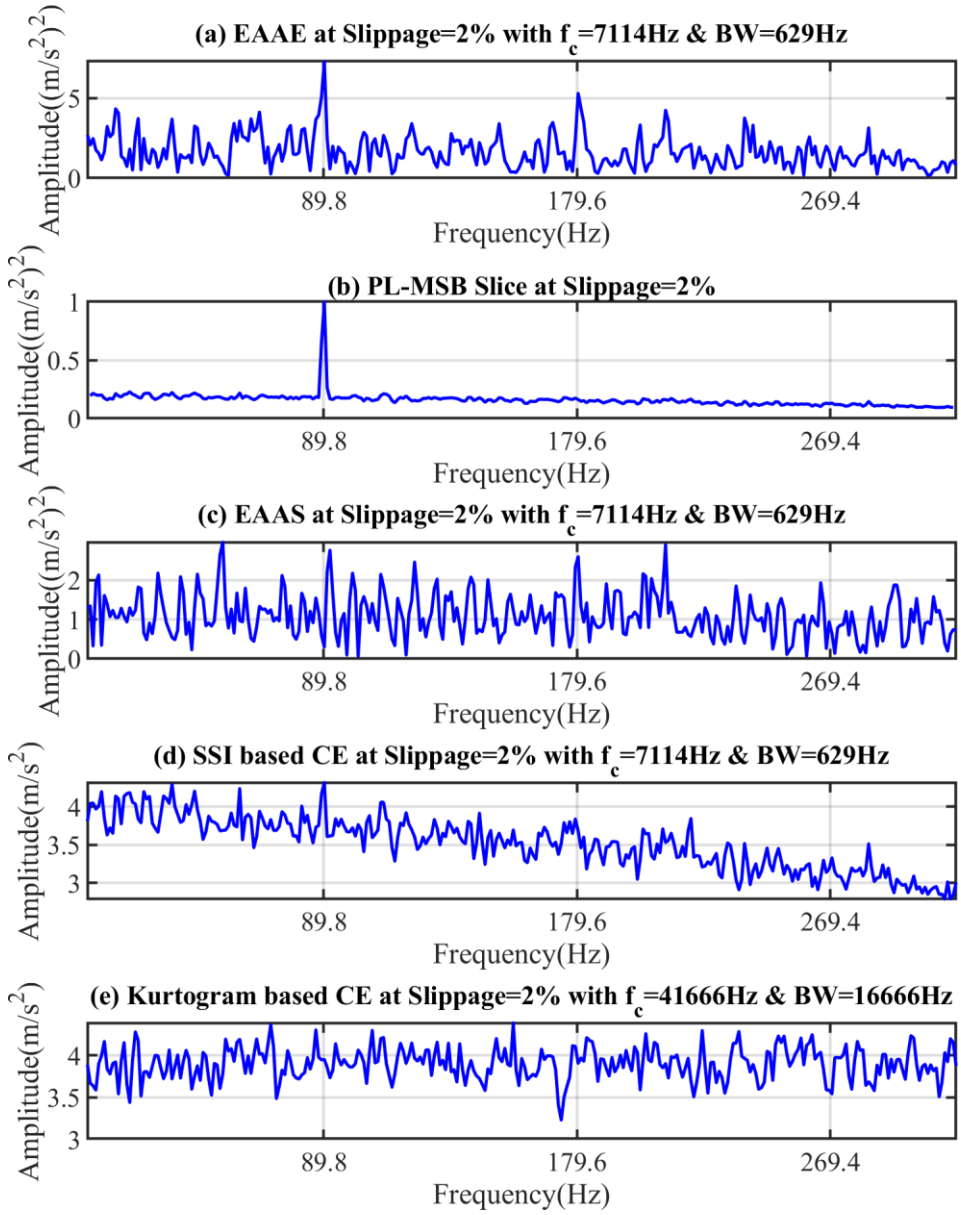


Figure 6-19 EAAE, PL-MSB, EAAS and CE for the case of outer race fault at SNR -31dB and 2% slippage

6.4.2 Demodulation Analysis of Inner Race Fault Signals

The simulation studies on the inner race faults follow the route arranged in the outer race faults. Figure 6-20 shows the spectrum of the simulated cyclostationary bearing signals at different levels of random slippage under the SNR of -31dB. The spectrums denote that the phase noise significantly decreases the amplitude of the resonant areas in the frequency domain, which manifests comparing to Figure 4-36. Due to the nonstationary input, the output of the bearing model is too difficult to characterise by the stationary signal processing approaches. Consequently the Fourier transform based spectrum shows that the spectral energy is smeared over frequency

bins. The leakage of the spectrum represents three clusters in the frequency domain other than the sparse components in the periodic signals.

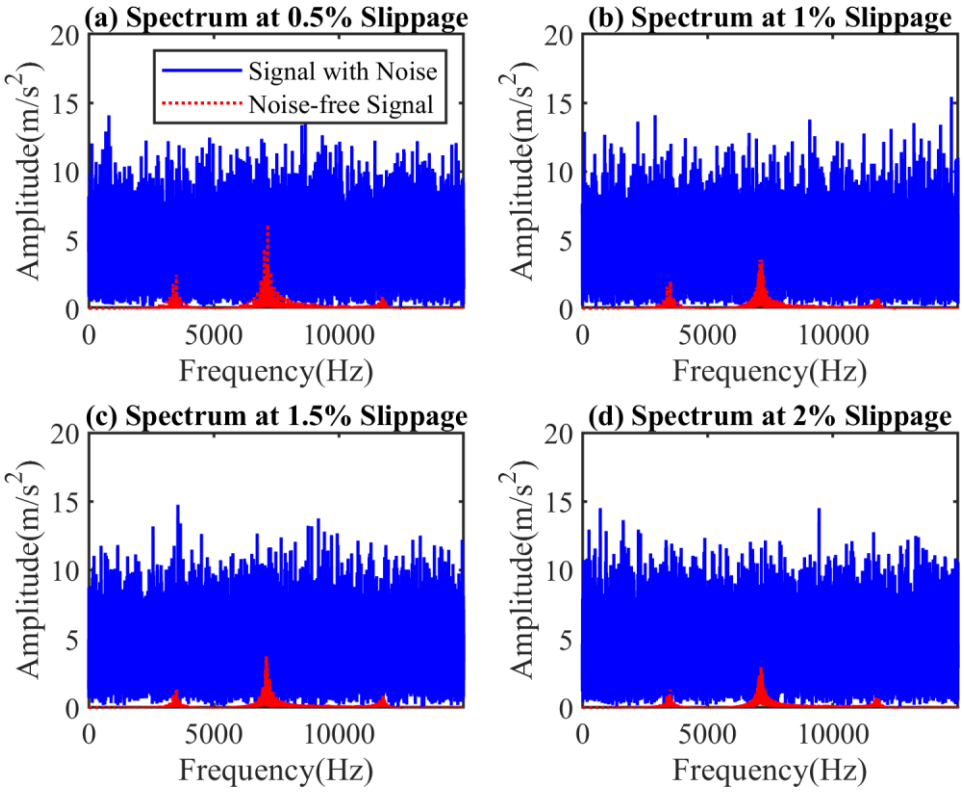


Figure 6-20 Spectra of cyclostationary inner race fault bearing signals at SNR -31dB and: (a) 0.5% slippage; (b) 1% slippage; (c) 1.5% slippage and (d) 2% slippage

The optimal frequency bands in these simulated inner race signals are shown in the following figures. The Figure 6-21, Figure 6-23, Figure 6-25, and Figure 6-27 denote the frequency bands by the Kurtogram which may contain the most impulsive behaviour. To summarise, the frequency bands recommended by the Kurtogram are substantially wide, which is not theoretically good in the early fault detection and diagnosis. The frequency bands selected by the operational modal analysis, SSI, are displayed in Figure 6-22, Figure 6-24, Figure 6-26, and Figure 6-28. In four slippage cases, only the second order mode of the bearing model at 7117Hz is identified from the cyclostationary signals.

In the case of inner race faults at 0.5% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 8333Hz with a frequency bandwidth of 16.67kHz at level 1.6 in Figure 6-21. The SSI selects the central frequency of 7111Hz in Figure 6-22.

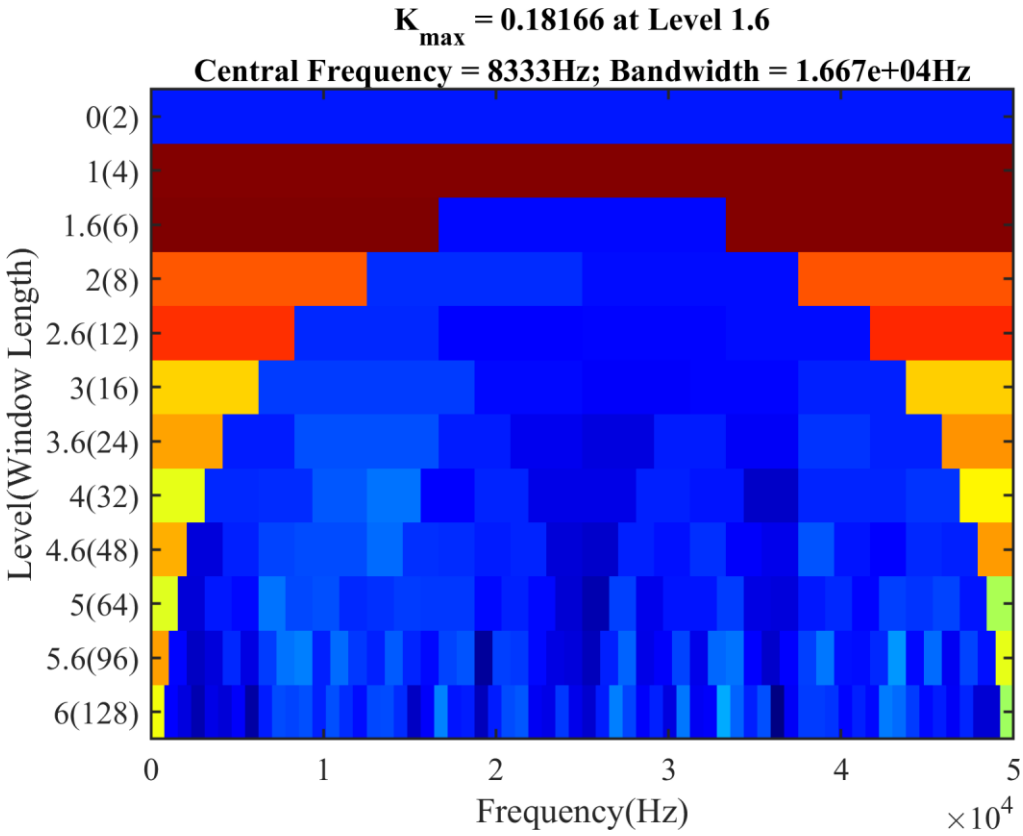


Figure 6-21 Kurtogram for the case of inner race faults at 0.5% slippage and SNR -31dB

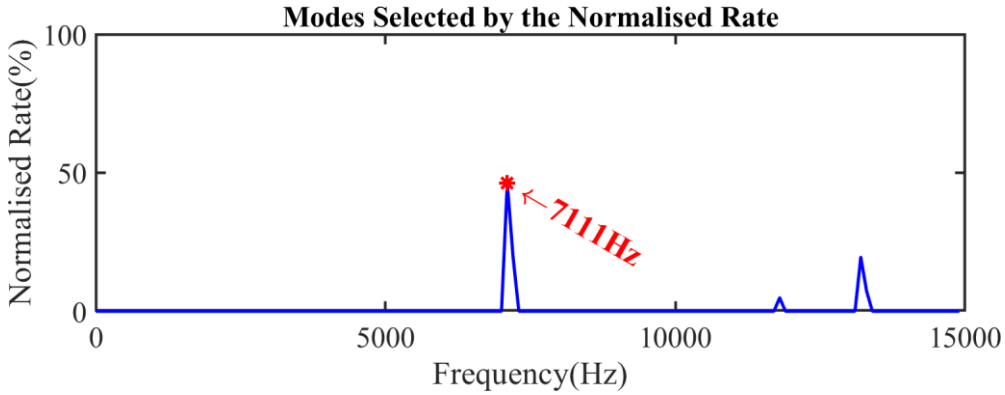


Figure 6-22 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at 0.5% slippage and SNR -31dB

In the case of inner race faults at 1% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 8333Hz with a frequency bandwidth of 16.67kHz at level 1.6 in Figure 6-23. The SSI selects the central frequency of 7111Hz in Figure 6-24.

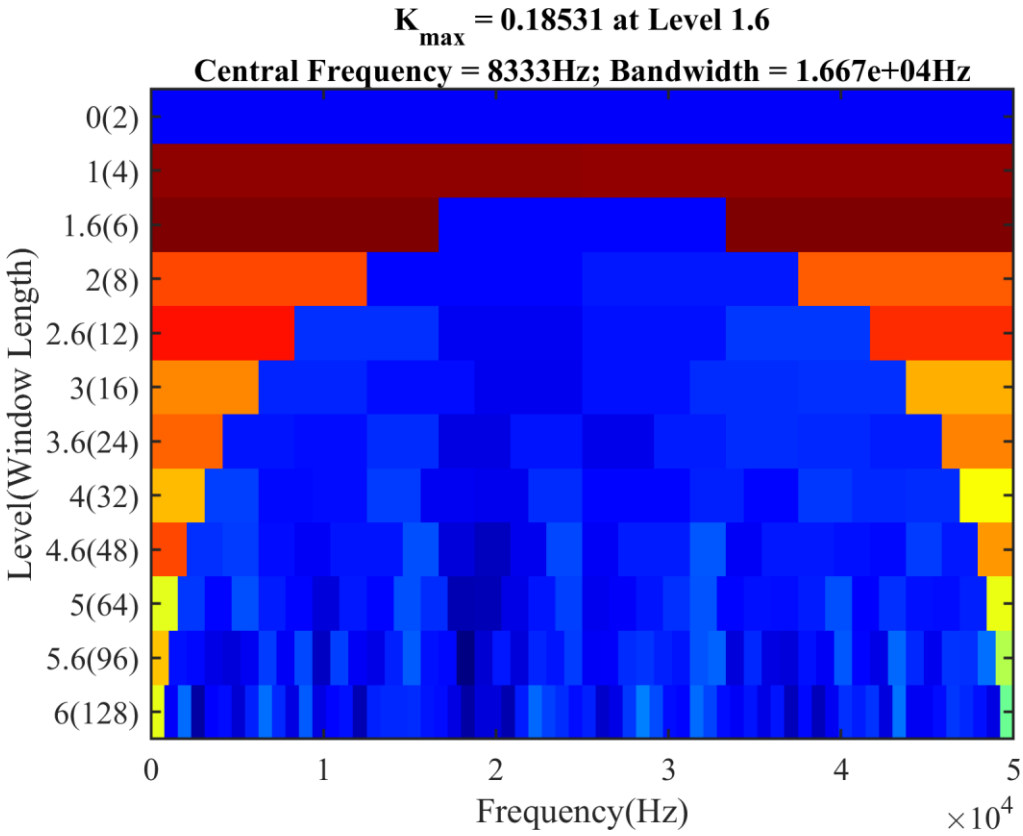


Figure 6-23 Kurtogram for the case of inner race faults at 1% slippage and SNR -31dB

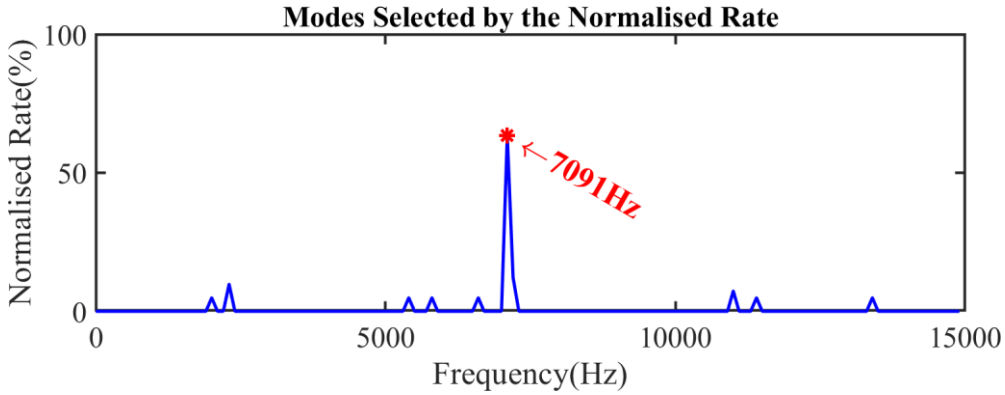


Figure 6-24 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at 1% slippage and SNR -31dB

In the case of inner race faults at 1.5% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 37.5kHz with a frequency bandwidth of 25kHz at level 1 in Figure 6-25. The SSI selects the central frequency of 7099Hz in Figure 6-26.

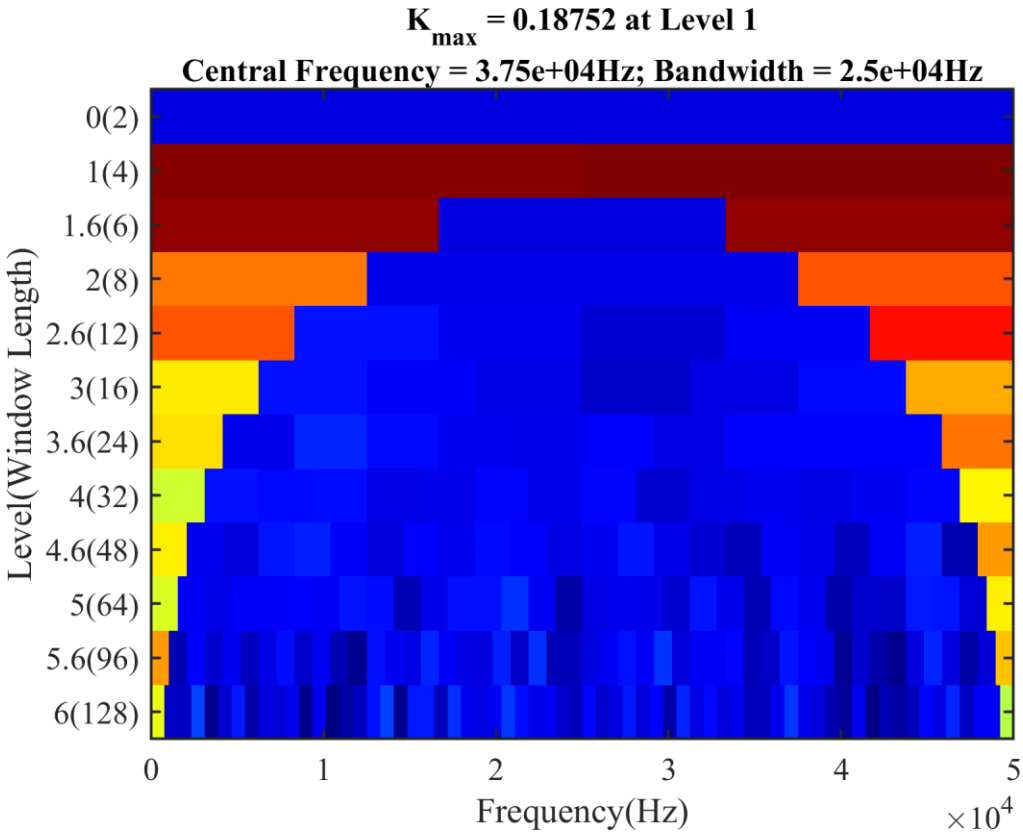


Figure 6-25 Kurtogram for the case of inner race faults at 1.5% slippage and SNR -31dB

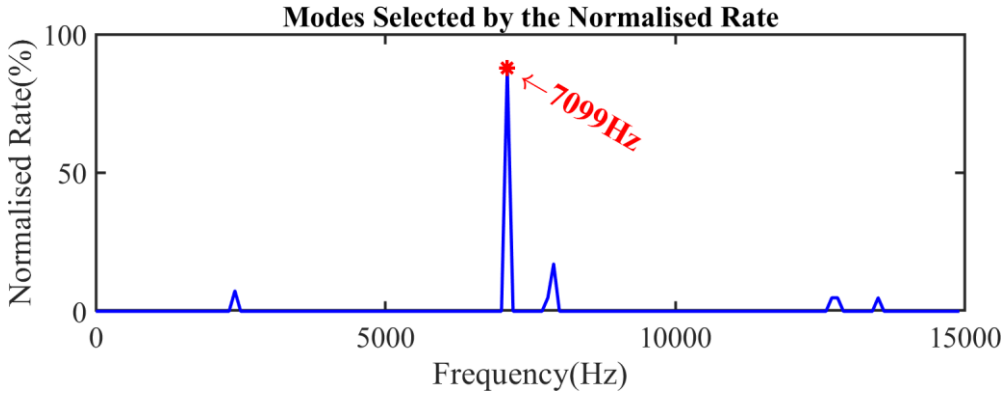


Figure 6-26 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at 1.5% slippage and SNR -31dB

In the case of inner race faults at 1.5% slippage and SNR -31dB, the most impulsive frequency band by the Kurtogram is centred at 12.5kHz with a frequency bandwidth of 25kHz at level 1 in Figure 6-27. The SSI selects the central frequency of 7180Hz in Figure 6-28.

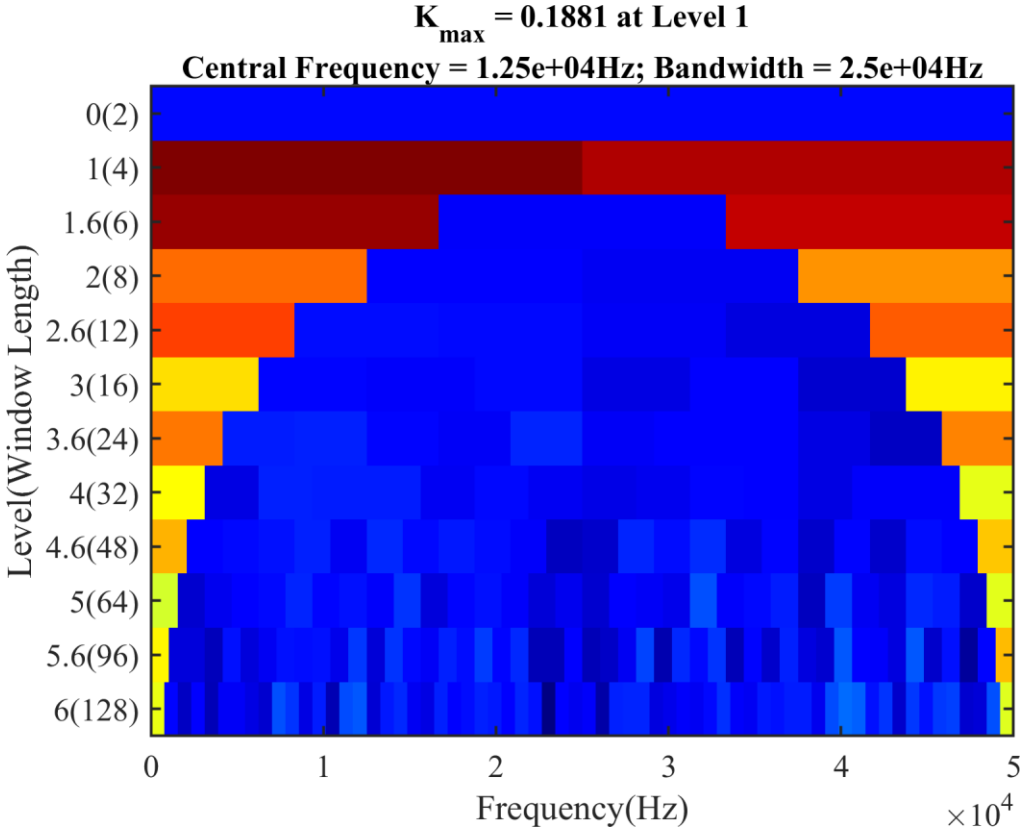


Figure 6-27 Kurtogram for the case of inner race faults at 2% slippage and SNR -31dB

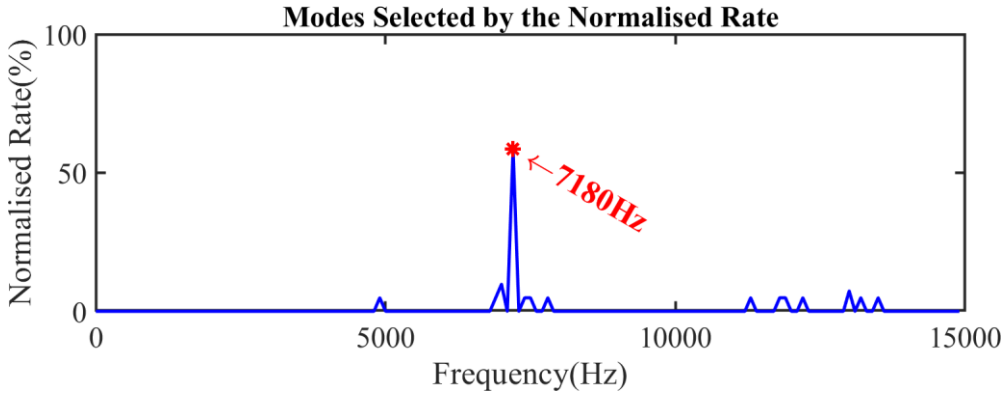


Figure 6-28 Normalised Rate of stable modes in the Stabilisation Diagram for the case of inner race faults at 2% slippage and SNR -31dB

The optimal central frequencies in the four slippage cases of the inner race faults are summarised in Table 6-1. Compared to the Kurtogram, the SSI method gives more reasonable determination of the optimal frequency bands for further demodulation analysis.

Table 6-2 Optimal frequency bands of cyclostationary inner race fault signals

Cases of Inner Race Faults	Kurtogram		SSI	
	Central Frequency	Frequency Bandwidth	Central Frequency	Frequency Bandwidth
0.5% Slippage	8333Hz	16666Hz	7111Hz	946Hz
1% Slippage	8333Hz	16666Hz	7091Hz	946Hz
1.5% Slippage	37500Hz	25000Hz	7099Hz	946Hz
2% Slippage	12500Hz	25000Hz	7180Hz	946Hz

The inner race faults are often a more challenging task due to the long transmission path and varying excitation forces. The impacts of rolling elements passing the localised defect are not only amplitude varying but also direction varying. The accelerometers on the bearing housing can only obtain the vibration on a single direction and the vibration projecting on the accelerometer axis cannot capture all the dynamic responses relating to the bearing detects, and hence, gives more barriers to detect and diagnose the bearing failure. Figure 6-29 depicts the demodulation results in the simulated case of 0.5% slippage. It is obvious that only EAAE and PL-MSB can find the fault theoretical frequency of the inner race faults from the extremely low SNR signals. However, the high order harmonics is very difficult to detect and the EAAE shows some potential for high order harmonics but with very low amplitude. The other three methods are not effective under these conditions.

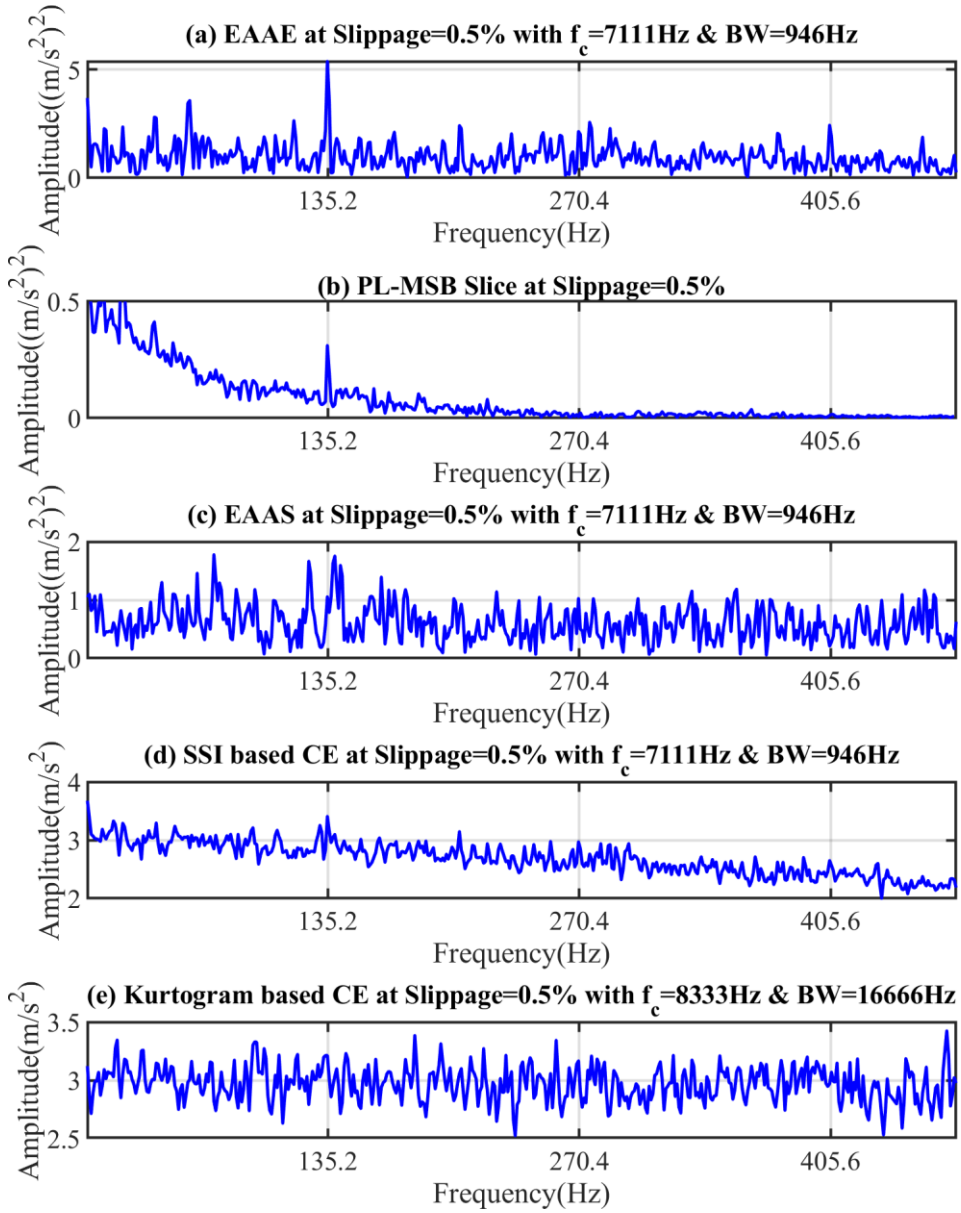


Figure 6-29 EAAE, PL-MSB, EAAS and CE for the case of inner race fault at SNR -30dB and 0.5% slippage

The slippage of the bearings goes into the next level. The demodulation results of the five methods are denoted in Figure 6-30. The performance of these methods is similar to the case of 0.5% slippage.

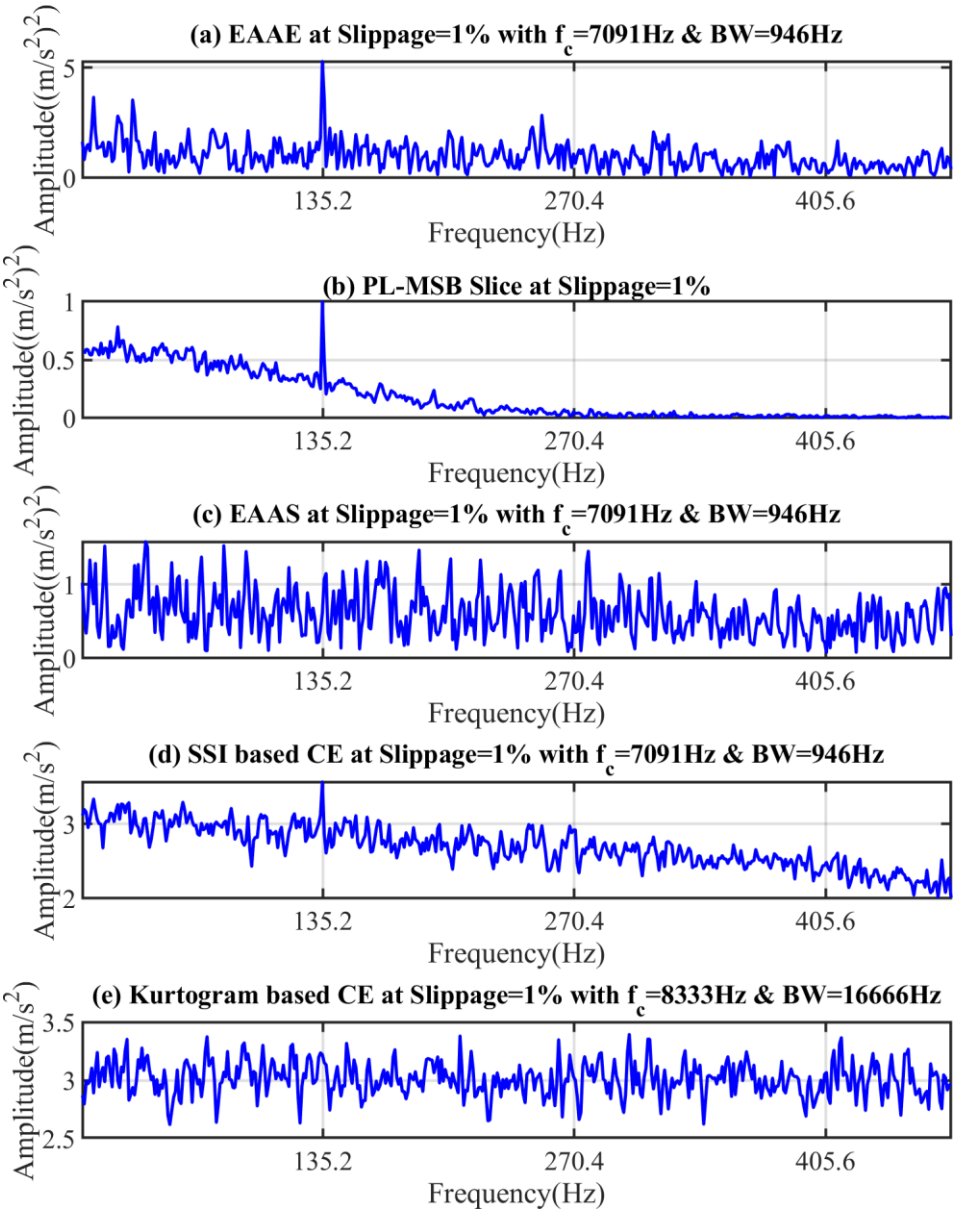


Figure 6-30 EAAE, PL-MSB, EAAS and CE for the case of inner race fault at SNR -30dB and 1% slippage

For the inner race fault signals with 1.5% slippage, the EAAE and PL-MSB give the accurate diagnosis of the inner race faults. To be different, the SSI based CE gives an accurate fault indicator about the inner race fault. However, based on the Monte Carlo tests, this method cannot give an accurate indication.

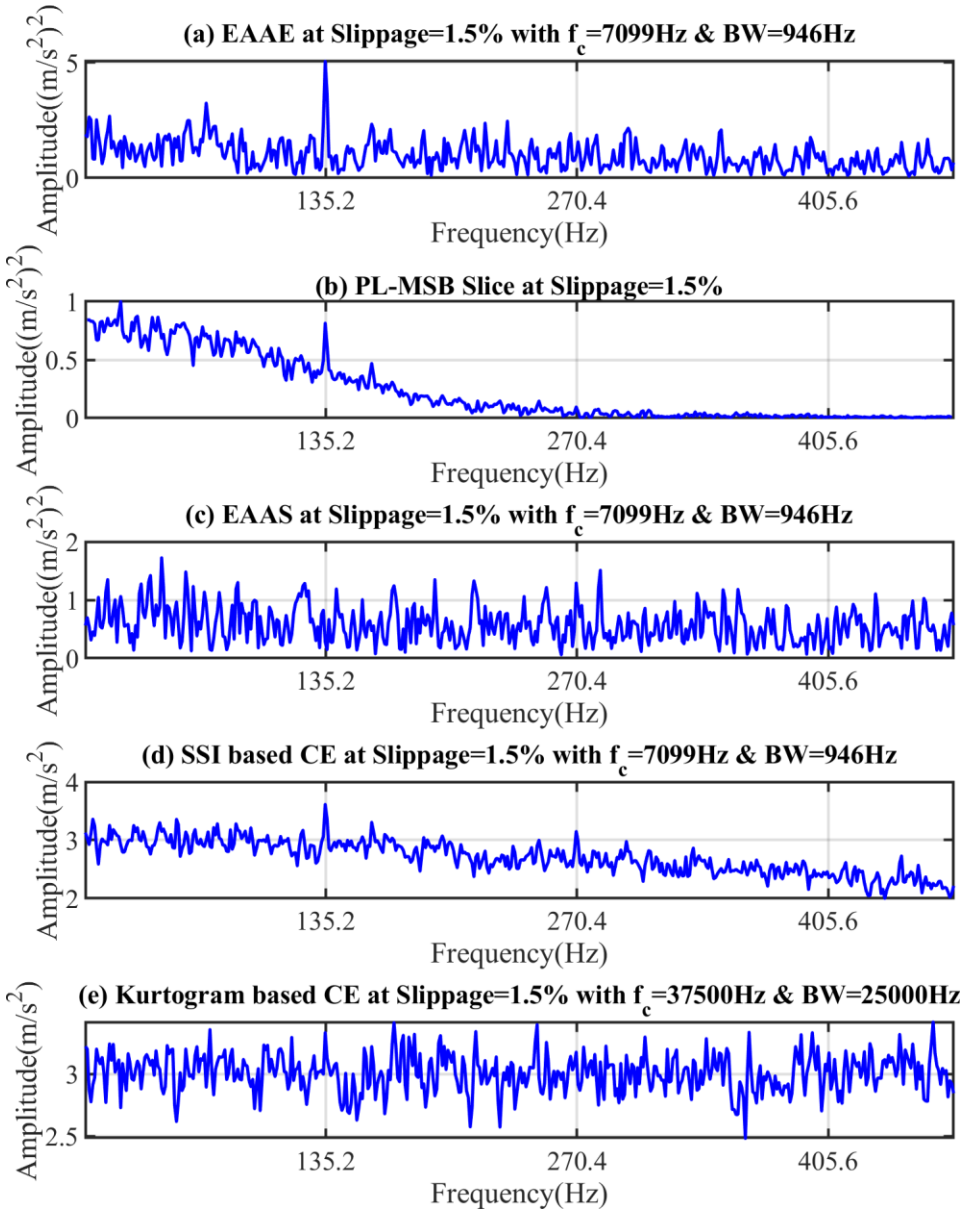


Figure 6-31 EAAE, PL-MSB, EAAS and CE for the case of inner race fault at SNR -30dB and 1.5% slippage

The Figure 6-32 illustrates the analysis results of the bearing signals under the worst condition in the simulation studies. Based on the spectra in the Figure 6-32, the EAAE gives the best representation of the fault signatures. The fundamental fault frequency as well as the second order harmonic is significantly obvious in the EAAE spectrum, which outperforms the other methods. The PL-MSB in Figure 6-32 (b) also allows the effective detection and diagnosis of the inner race defects. The other methods fail to diagnose the inner race faults.

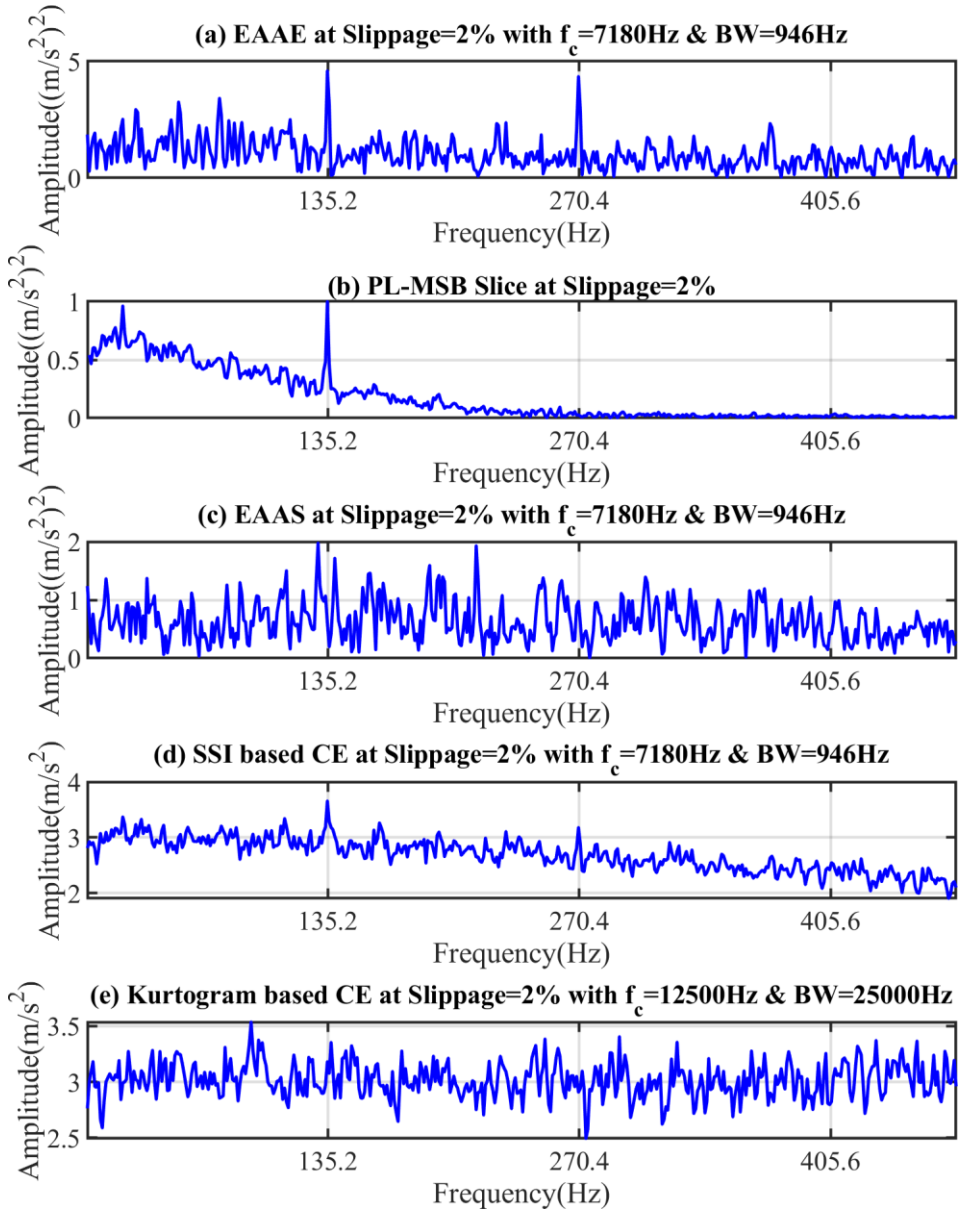


Figure 6-32 EAAE, PL-MSB, EAAS and CE for the case of inner race fault at SNR -30dB and 2% slippage

6.4.3 Quantification of Bearing Fault Signatures

The spectrums of the five demodulation approaches at different levels of slippage show that the novel methods, EAAE and PL-MSB, are much more effective and efficient than the conventional methods. The strength of the fault signatures were calculated by using Equation (4.68) and the quantified results of the outer race and inner race are shown in Figure 6-33 and Figure 6-34. The EAAE and PL-MSB can generate stronger and more reliable fault signatures than the conventional approaches. Additionally, the method EAAS is effective in processing the periodic impacts based resonant modulation signals but it is not reliable in the aperiodic impacts based resonant modulation signals. The conventional envelope analysis has difficulty in finding the fault

information from the noisy signals. Between the five demodulation methods, the PL-MSB is the most powerful approach for extracting the most effective fault signatures.

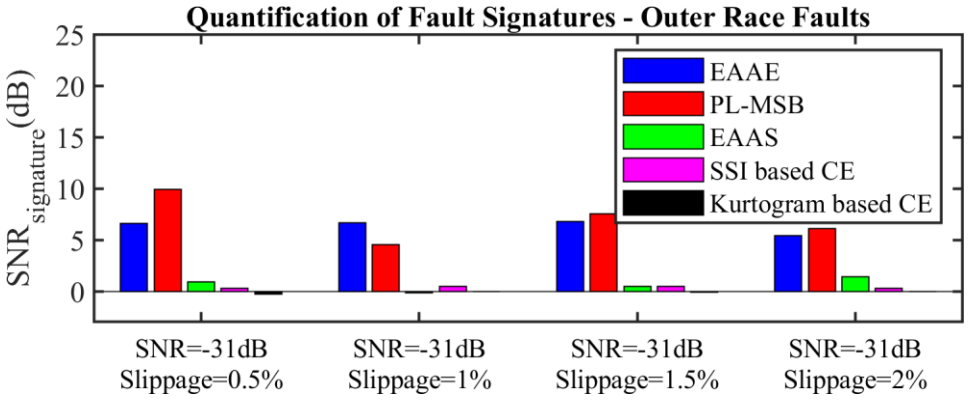


Figure 6-33 Quantification of fault signatures for tapered bearing outer race defects

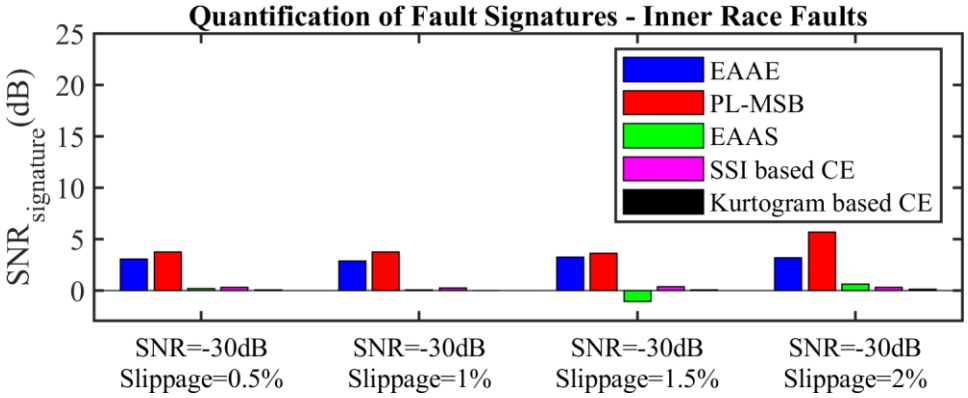


Figure 6-34 Quantification of fault signatures for tapered bearing inner race defects

6.5 Summary

In this chapter, the responses of the bearing under the aperiodic impulsive excitations are investigated for the effective demodulation of the bearing fault signals. To simulate the slippage between bearing elements, a series of random impacts that satisfied the Markov process was simulated to drive the linear bearing model. The cyclostationary responses under these excitations are significantly different to the periodic impacts in Chapter 4. The nonstationary inputs lead to the cyclostationary responses and the resonant modulation enlarges the periodicity of the second order statistics. To tackle the challenging vibration signals, two novel methods, EAAE and PL-MSB, were developed to handle the phase noise and suppress the temporal noise. The EAAE used the Hilbert transform based envelope analysis to tolerate the slippage in the cyclostationary responses and it has been verified that the method is effective for the tapered bearing vibration signals. The PL-MSB reduces the phase noise by using the linear fitting of the phase variation along the timestamp and, after linearisation, the coupled components can be enhanced by the ensemble averaging. The PL-MSB demonstrates the method is effective for the aperiodic

impulsive excitations based resonant modulation. Both simulation and experimental studies demonstrate that the proposed methods can detect the bearing defects from the cyclostationary responses by demodulating from the limited band in resonant zones.

Chapter 7 Experimental Study of Tapered Roller Bearing Fault Detection and Diagnosis

Based on the simulation studies upon the cyclostationary signals in Chapter 6, the experimental investigation on the tapered roller bearings were introduced in this chapter. The experiments aim to verify the findings in the simulation studies. Two novel methods, Ensemble Average of Autocorrelated Envelopes (EAAE) and Phase Linearisation based Modulation Signal Bispectrum (PL-MSB), are verified based on the vibration signals obtained from the tapered bearing test rig.

7.1 Test Facilities

7.1.1 Mechanical and Electric Systems

The experimental studies upon tapered roller bearings were conducted to verify the proposed fault detection approaches. The vibration signals are captured from an in-house bearing test system. The slippage rates of bearings are well controllable, which is an easy way to investigate the proposed methods. However, in experimental studies, it is actually very difficult or nearly impossible to carry out the bearing tests under the working conditions of 2% slippage and less than -30dB SNR. Compared to the widely used ball bearings, the tapered roller bearings are more susceptible to random slippage because the internal clearance is determined by manual control. The inappropriate clearance has a critical impact on the slippage, making it difficult to measure the slippage level in practice. The slippage level can be roughly controlled by adjusting the clearance in the bearings. The more clearance between bearing components, the more slippage when the tapered bearings occur.

Figure 7-1 shows the photograph of the tapered bearing test rig. The test system includes a tapered bearing, a deep groove ball bearing and an induction motor. The tapered bearing is the target for experiments and the ball bearing is to support the long shaft. The main shaft and the inductor motor are connected by a flexible coupling.

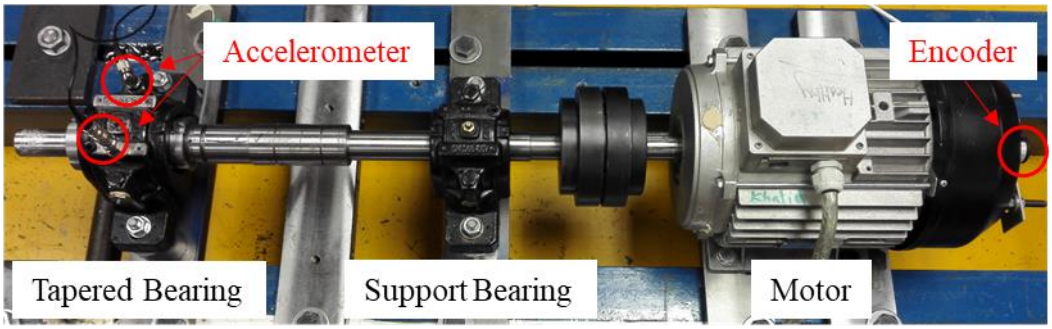


Figure 7-1 Photograph of the tapered bearing test rig

Figure 7-2 shows that the internal clearance of the tapered bearing can be adjusted by the adjustable nut allowing it to move axially upon the threads on the shaft. The outer race of the bearing is fixed by the housing and the inner race is installed on the shaft. The bearing clearance is mapped with the gap between the adjustable nut and the flange on the main shaft. The positioning of the adjustable nut can control the internal clearance of the tapered bearing, which also controls the preload on the bearing. The internal clearance of the tapered bearings can directly influence the contact stiffness and determines the rotating condition of rollers. The reference gap Δd can be measured accurately by a gauge block tool box, which allows the precise control of the bearing clearance.

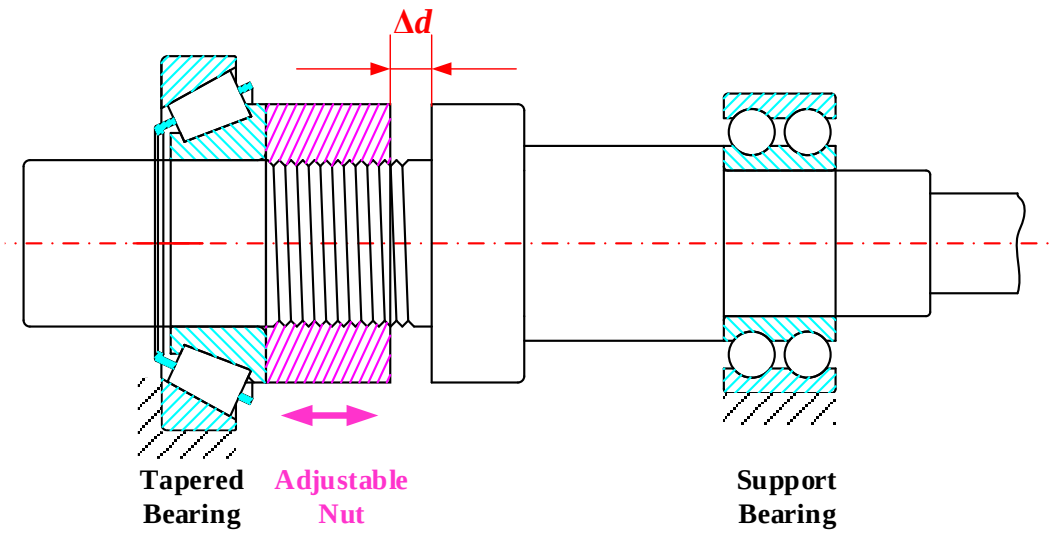


Figure 7-2 Schematic diagram of the tapered bearing test rig

7.1.2 Data Acquisition System

Two charge type accelerometers were installed on the tapered bearing housing to obtain the vibration signals of the bearing under different clearance. The charge type accelerometers are usually highly sensitive when comparing to the Integrated Electronics Piezo-Electric (IEPE) accelerometers. The output from charge type accelerometers has to be amplified by the charge amplifier, which can then be measured by the data acquisition device. The lubrication condition has a critical impact on the dynamic responses of the bearings and therefore a thermocouple was installed on the housing to monitor the temperature of the bearing during the running stage. Additionally, a commercial encoder was attached onto the fan end of the motor by a rubber tube, which is used to measure the rotational speed of the whole rig. The data acquisition card for collecting the outputs of four transducers is National Instrument PCI-6221, a multifunction input/output device with the resolution of 16bit and an overall sampling rate of 250kS/s.

The apparatus used in this experimental study is listed in Table 7-1.

Table 7-1 Specifications of the data acquisition system

Item	Manufacturer	Model	Key Specifications
Date Acquisition Card	National Instrument	NI 6221	16bit, 250kS/s overall
Charge Amplifier	Brüel & Kjær	B&K 2635	0.2Hz-100kHz with noise < 5×10^{-3}pC
Accelerometers	Sinocera	CA-YD-104T	3.640pC/ms^{-2} in 0.5Hz-7000Hz
Thermocouple	RS	K type	0-100°C
Encoder	Hengstler	RI32	100ppr, max 6000rpm

To meet the requirement of the data collection, a data acquisition software based on the LabWindows/CVI was developed to have a user-friendly graphical user interface (GUI) for online monitoring of temperature, vibration levels, and rotating speeds. The developed data collection software can ensure the similar running conditions of the bearing when logging the vibration signals, which eliminate the other undesired influence for the purpose of accurate comparison of the dynamic responses under different levels of slippage. All outputs of sensors were sampled simultaneously for 30 seconds using the data acquisition card at a sample rate of 50kHz.

7.2 Test Procedure

The test bearing used is a Timken type 31308 taper roller bearing. It has a bore diameter of 40mm, an outer diameter of 90mm and a width of 25.25mm. The specifications of the bearing are shown in Table 7-2.

Table 7-2 Specification of the tapered bearing

Parameters	**Specifications**
Model	31308
Number of Rollers	15
Bore	40mm
Cup Outer Diameter	90mm
Cone Width	23mm
Cup Width	17mm
Bearing Width	25.25mm
Bearing Weight	0.7kg

The faults simulated in this chapter are the outer race defects. Two different size of localised defects were seeded on the outer race of the bearings. Figure 7-3 shows the photographs of the faulty bearings. As shown in Figure 7-3 (a), a line scratch with 4mm length, 0.3mm width and 0.05mm depth was created near the edge of the outer race. The Figure 7-3 (b) displays the small outer race fault with a line scratch of 2mm length, 0.1mm width and 0.05mm depth. Often the bearing faults start with localised pitting, the line scratch with a very small width was used to simulate a small early stage fatigue defect on the outer race. Furthermore, more than 1mm scratch is out of the rolling path of the rollers because the fault location is too close to the edge of the outer ring.

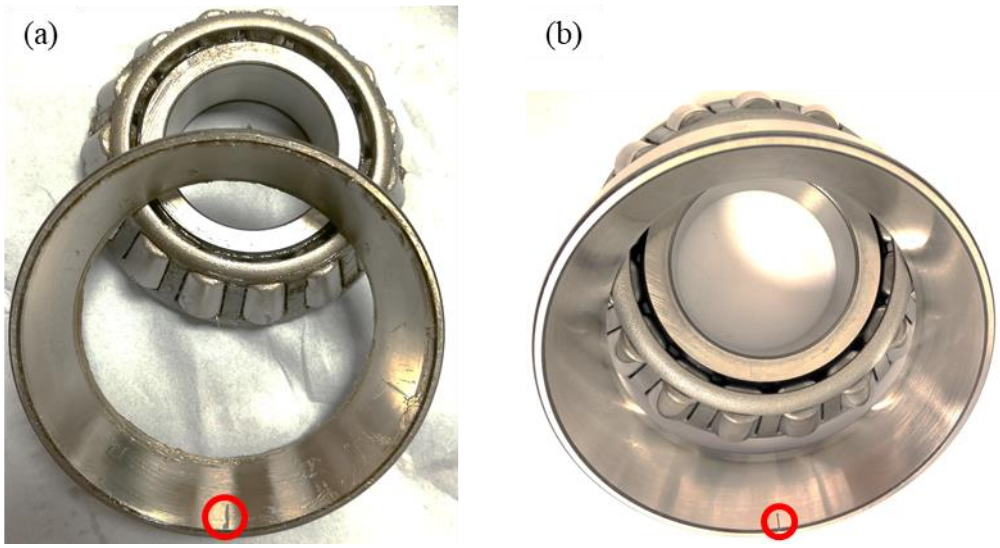


Figure 7-3 Photograph of the outer race faults: (a) 4mm defect; and (b) 2mm defect

The faulty bearings were tested with five successive levels of internal clearances to simulate different degrees of wear. As shown in Figure 7-2 the control of internal clearances of the testing tapered bearing is realised by turning an adjustable nut in forward or backward directions. In this experiment, the internal clearance of the bearings is represented by the gap between the nut and the shaft flange. In this experiment, the internal clearance of the faulty bearings is designed into five successive levels. The gaps between the nut and the shaft flange are configured to have five levels based on the axial movement of the adjustable nut. These gap levels correspond to wear degrees of bearings, which are detailed in the Table 7-3.

Table 7-3 Test cases

Cases	**Clearance Levels**	**Speed**	**Gap Δd**
Large Outer Race Fault	+60um	100% (1500rpm)	14.77mm
	+40um		14.79mm
	+20um		14.81mm
	0um		14.83mm
	-20um		14.85mm
Small Outer Race Fault	+60um		14.79mm
	+40um		14.81mm
	+20um		14.83mm
	0um		14.85mm
	-20um		14.87mm

The real values of the gaps are not identical because the test rig cannot be guaranteed to be identical after the dismantling and re-assembly. Therefore, the absolute gap values are different for two cases. The 0 gap of the tapered bearings are determined by the finger tight position. By taking the

position of the 0 gap as a reference, the other positions are achieved by moving the adjustable nut along the shaft direction. For every gap level, the test system runs for about 40 minutes to finish the warm-up stage until the temperature indicated by the thermocouple goes up to 40°C. after warm-up, the motor is powered by the VFD to run at 1500rpm for 5 minutes and from the third minute, the signals are logged for the duration of 30 seconds at the sampling frequency of 50kHz.

7.3 Demodulation Analysis of Cyclostationary Vibration Signals

7.3.1 Determination of Optimal Frequency Bands

The vibration signals obtained from the tapered bearing test rig are analysed following the basic route of signal processing. Figure 7-4 shows the temporal vibration waveforms from two accelerometers on the bearing housing. The vibration signals demonstrated were from the large outer race fault at the gap zero, which means the dynamics responses come from the zero internal clearance of the bearings. Both vertical and horizontal vibrations display the periodic impulses that are induced by the rollers passing the outer raceway.

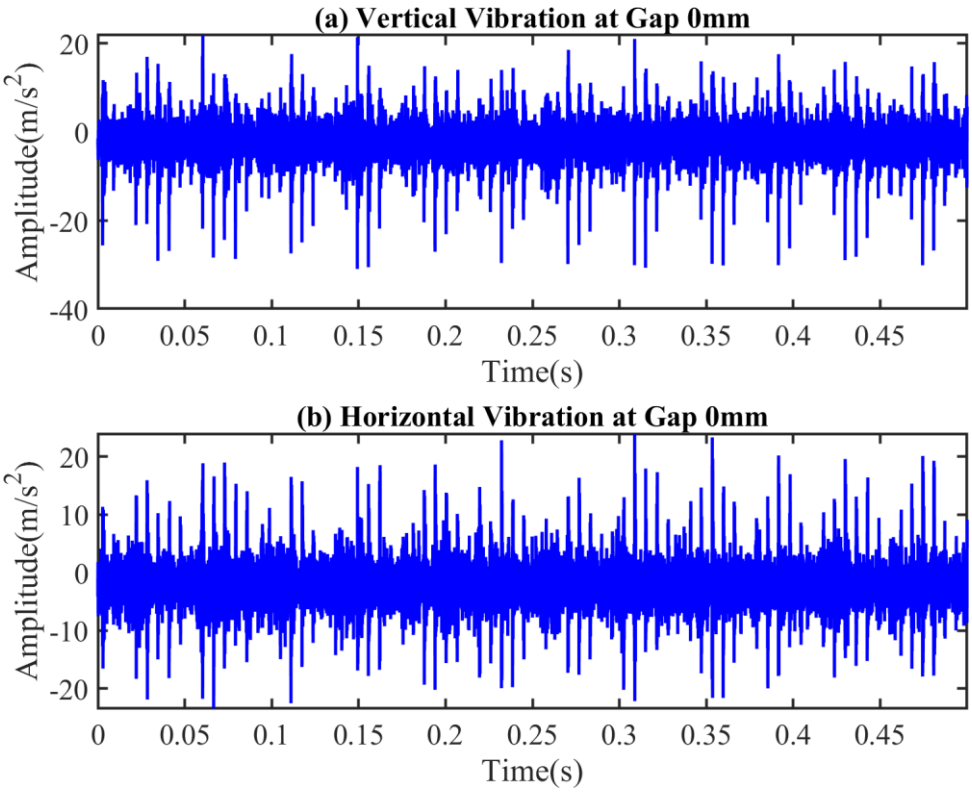


Figure 7-4 Raw vibration signals in the time domain: (a) vertical; and (b) horizontal

Several common statistic parameters are calculated and show the general characteristics of the bearing vibration. The Root Mean Square (RMS) values of the raw signals are shown in Figure 7-5. The RMS values are often used to present the violence of the vibration responses. Vibration

in both directions shows that the vibration energy is kept at the same level under different bearing clearances. Therefore the RMS is not robust in monitoring the wear degree of the bearings.

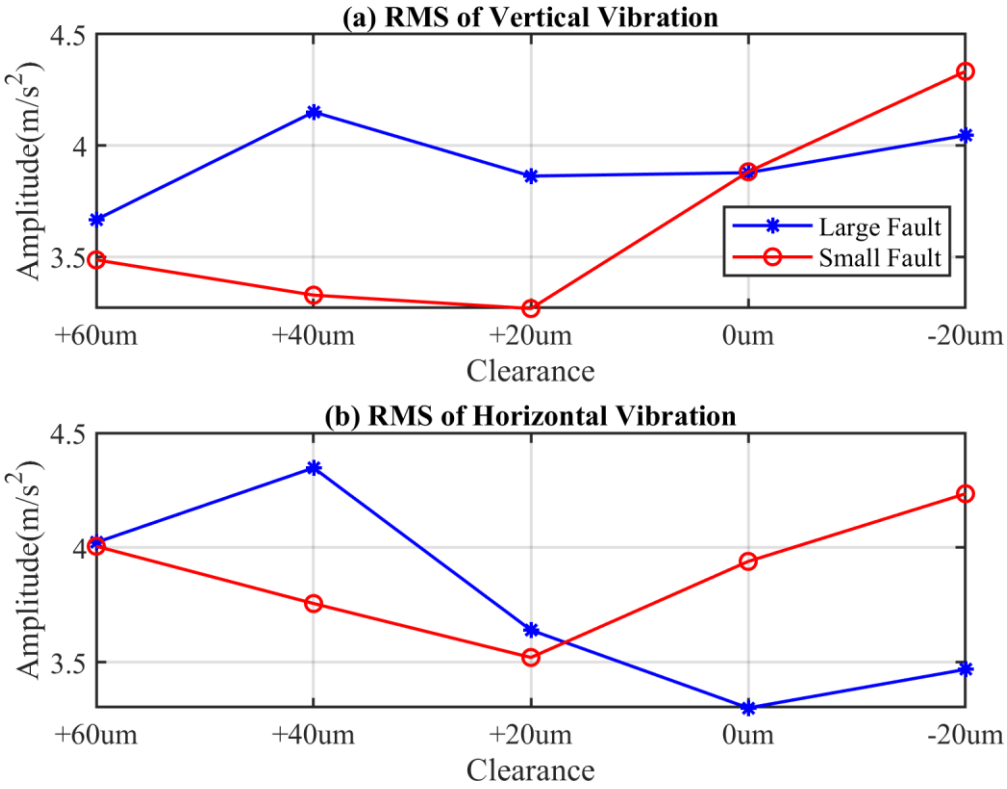


Figure 7-5 RMS of raw vibration signals: (a) vertical; and (b) horizontal

Kurtosis is another commonly used parameters to measure the impulsiveness of a random signal in essence. However kurtosis takes into account both periodic and aperiodic impulses and hence it can be easily interfered with. Although the kurtosis has its drawbacks, it is still a very good indicator in the modulation signal analysis. Figure 7-6 depicts the kurtosis values of each samples under five gap levels. The kurtosis vertical vibration signals show that a severe fault of the bearing can lead to more impulsive signals.

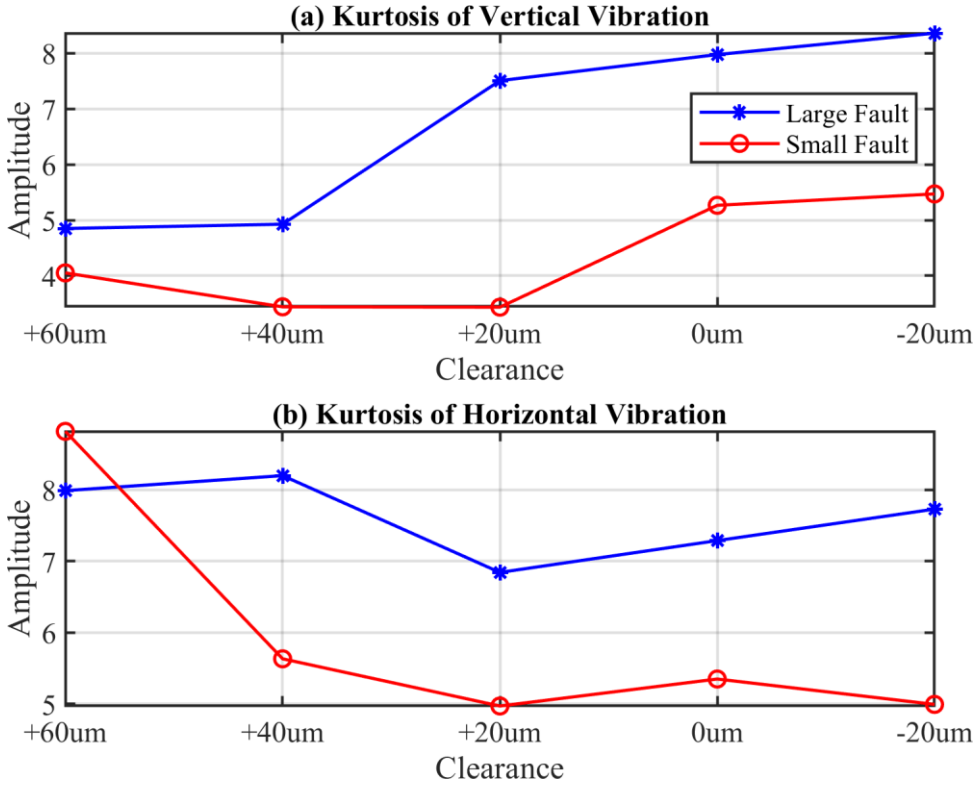


Figure 7-6 Kurtosis of raw vibration signals: (a) vertical; and (b) horizontal

The Gini index is a statistical parameter in the time domain for measuring the inequality of data sets. The index is in the range from 0 to 1. Zero means the distribution of the data is perfectly equal whilst 1 means the absolute inequality. The Gini index is originally an indicator in the economics and it has been extended into the vibration analysis. The Gini index is also sensitive to the impulsive responses in the time domain. The Figure 7-7 displays the trend of the Gini index computing from the vibration signals under five clearances. The vertical vibration shows that the more severer the fault of the bearing, this results in a larger Gini index therefore could be a good indicator for the fault severity. However the horizontal vibration does not show a meaningful relationship because the horizontal vibration contains not only the radial vibration but also the axial vibration.

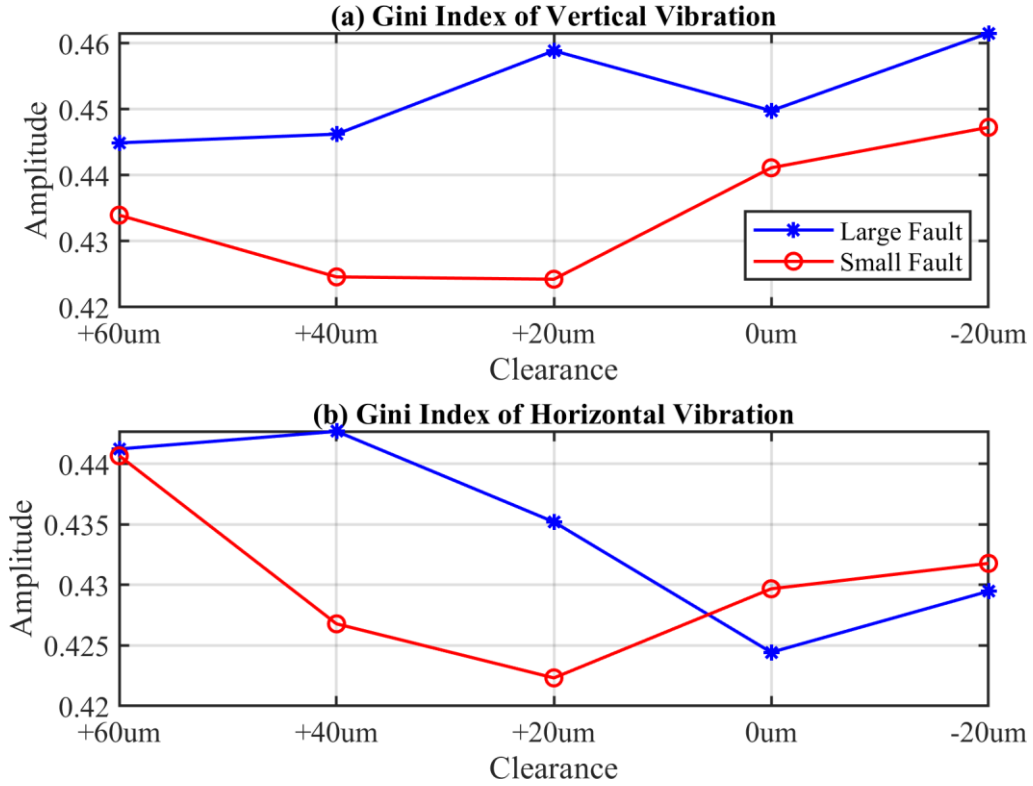


Figure 7-7 Gini index of raw vibration signals: (a) vertical; and (b) horizontal

The spectrum analysis of the vibration signals is a widely used technique in condition monitoring. The Fourier transform based frequency domain analysis is always a promising technique to show an informative result in another perspective. Figure 7-8 displays the power spectrum of the vibration signal at both directions. The power spectrum is estimated by the Welch's method using a window of 25,000 and an overlap of 60%. The solid line and dashed line represent the power spectrum of the vertical and horizontal vibration signals. The vibration shows a great distinction in the low and high frequency range. The vertical vibration has a higher energy level in almost frequency bands for two. The frequency band from 3000Hz to 7000Hz empirically contains several modes, which are the potential carrier frequencies for the demodulation analysis.

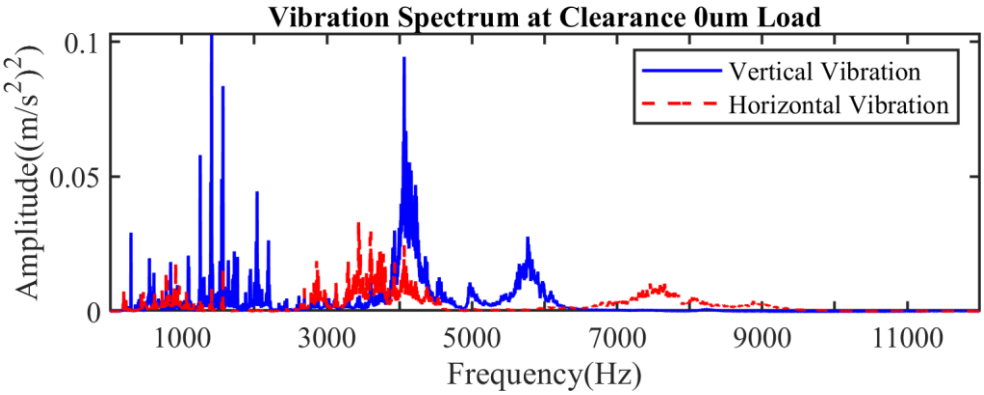


Figure 7-8 Power spectrum of vibration signals

At the beginning, the optimal frequency bands are obtained by Kurtogram and SSI methods. The internal clearances of the bearings have a significant influence on the rotating condition and contact stiffness of the rolling elements, and consequently have a direct impact on the dynamic responses. If the system inputs are not a wide frequency band excitation, the modes of the bearings cannot be excited and only the natural frequencies matched within the input range can be the potential carrier frequency. Therefore, these two methods are employed to analyse the vibration signals obtained at various clearances. Figure 7-9 to Figure 7-28 show the recommended frequency bands for the demodulation analysis.

In the case of the large outer race fault at clearance -20um, the most impulsive frequency band by the Kurtogram is centred at 14.84kHz with a frequency bandwidth of 1563Hz at level 4 in Figure 7-9. The SSI in Figure 7-10 selects three candidates of the central frequencies, which are 1441Hz, 4082Hz and 5797Hz.

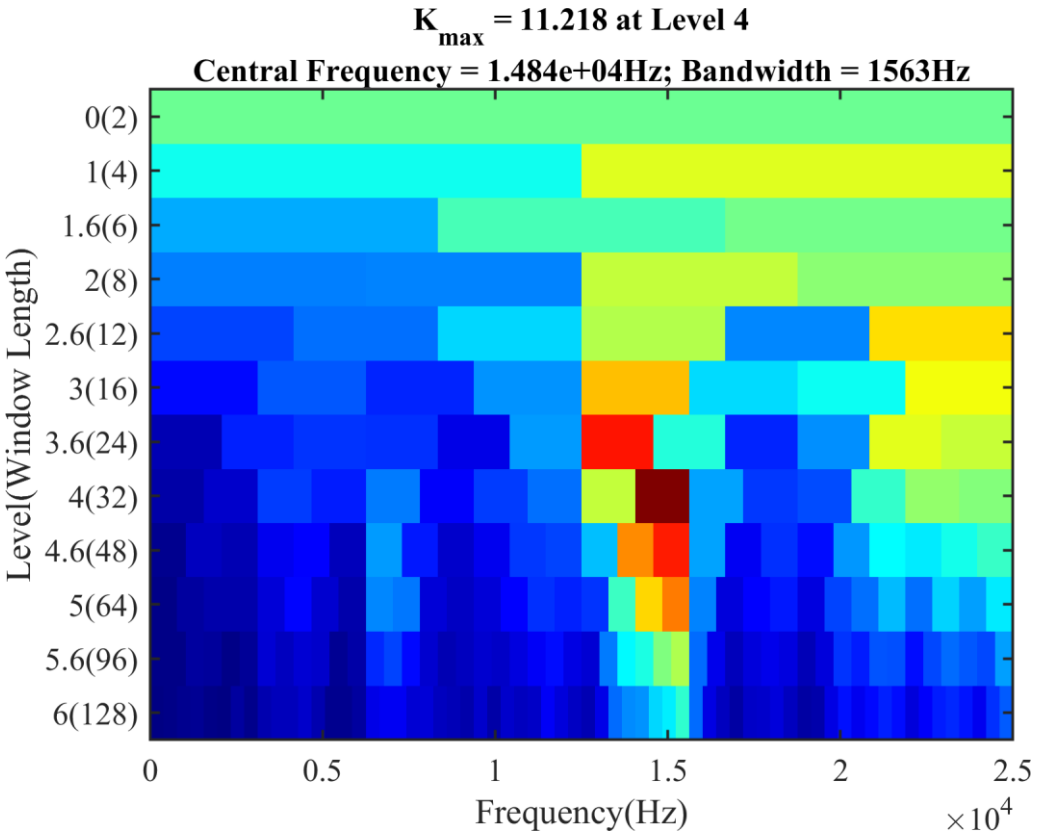


Figure 7-9 Kurtogram for large outer race fault at clearance -20um

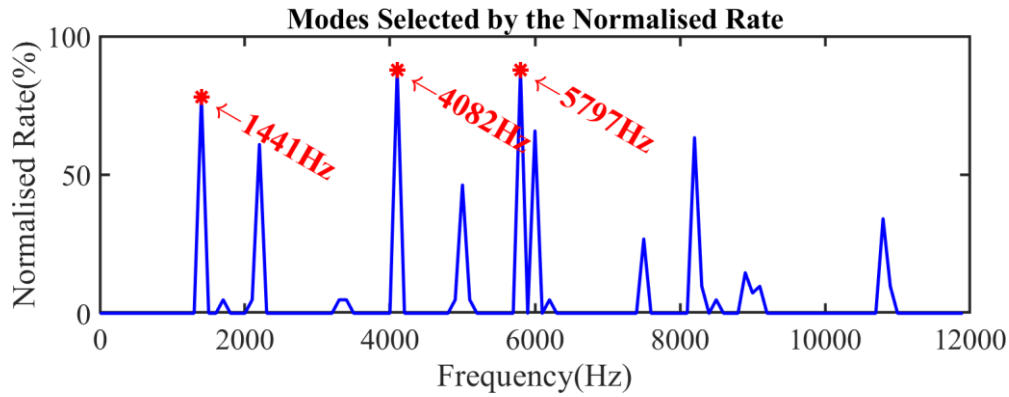


Figure 7-10 Normalised Rate of stable modes in the Stabilisation Diagram for large outer race fault at clearance -20um

In the case of the large outer race fault at clearance 0um, the most impulsive frequency band by the Kurtogram is centred at 13.54kHz with a frequency bandwidth of 2083Hz at level 3.6 in Figure 7-11. The SSI in Figure 7-12 selects three candidates of the central frequencies, which are 4079Hz, 5793Hz and 8237Hz.

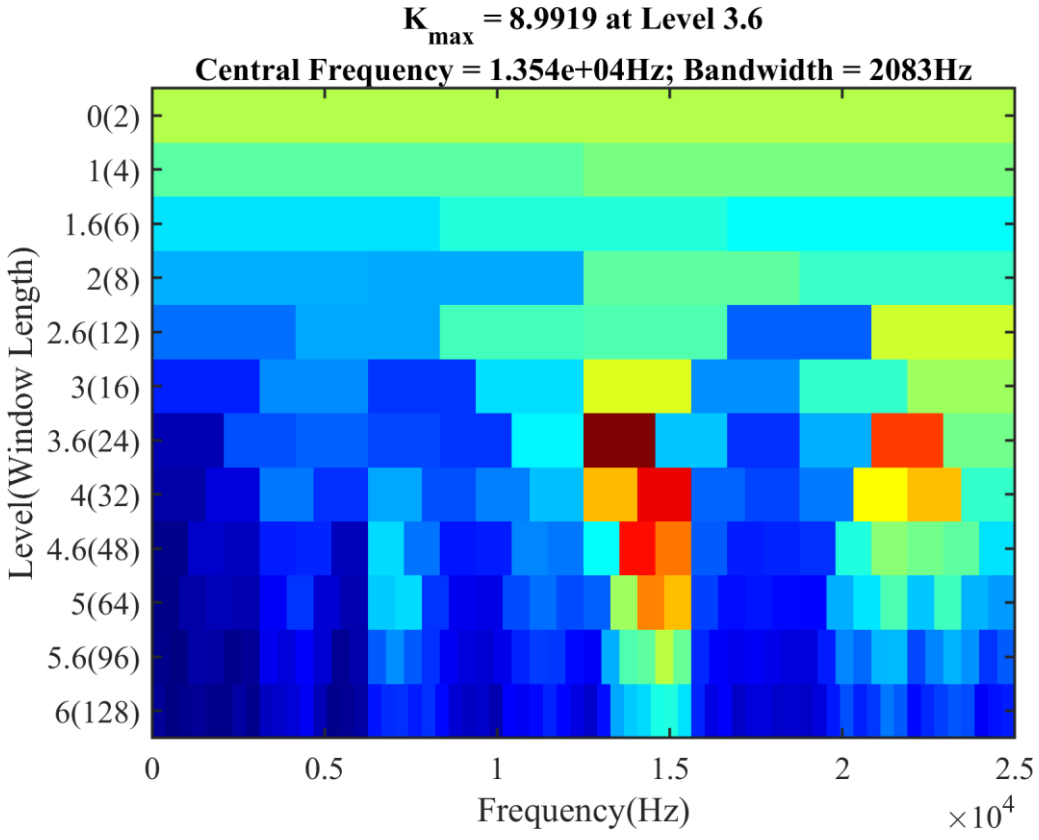


Figure 7-11 Kurtogram for large outer race fault at clearance 0um

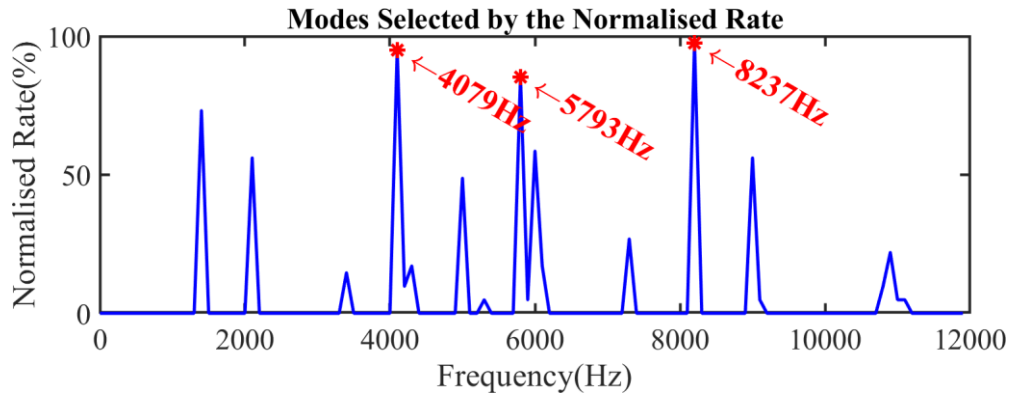


Figure 7-12 Normalised Rate of stable modes in the Stabilisation Diagram for large outer race fault at clearance 0um

In the case of the large outer race fault at clearance +20um, the most impulsive frequency band by the Kurtogram is centred at 21.87kHz with a frequency bandwidth of 2083Hz at level 3.6 in Figure 7-13. The SSI in Figure 7-14 selects two candidates of the central frequencies, which are 4083Hz and 10982Hz.

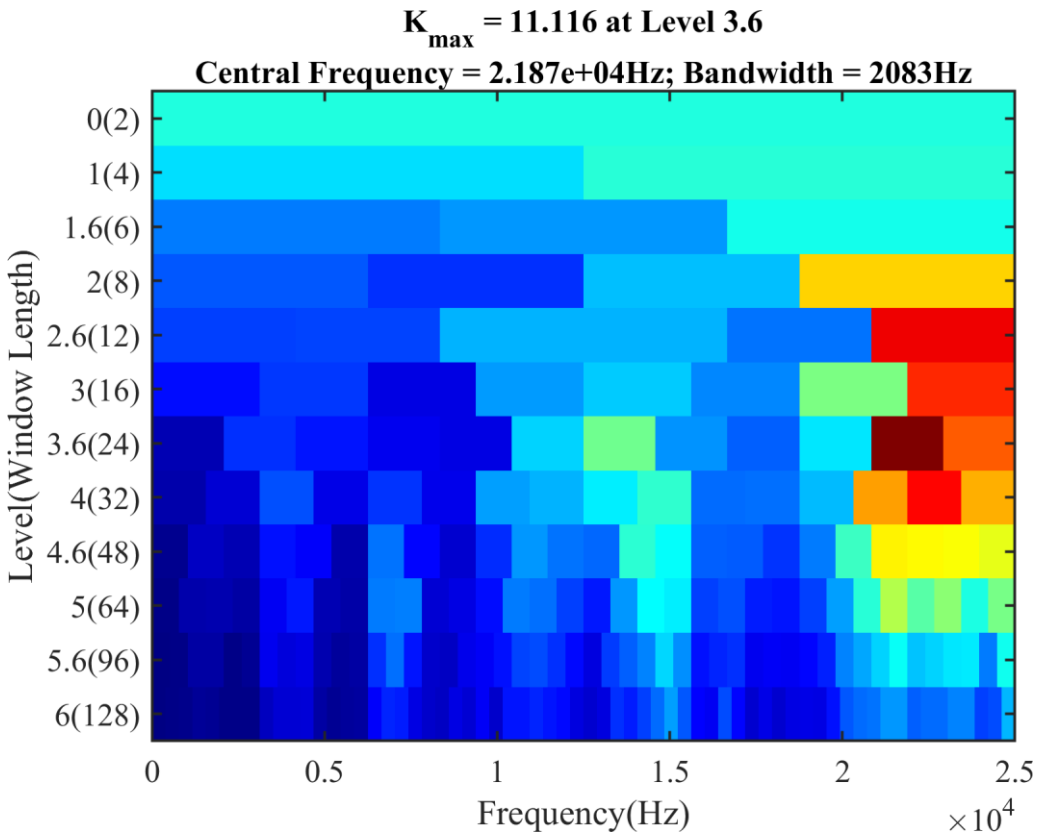


Figure 7-13 Kurtogram for large outer race fault at clearance +20um

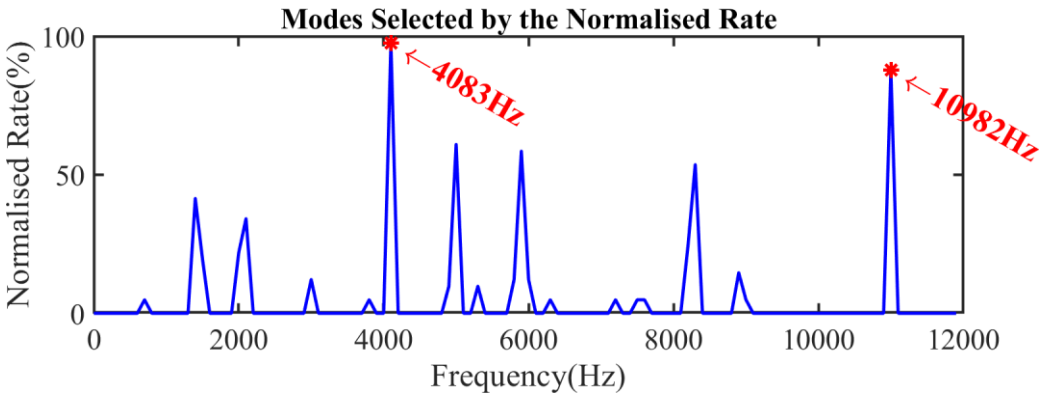


Figure 7-14 Normalised Rate of stable modes in the Stabilisation Diagram for large outer race fault at clearance +20um

In the case of the large outer race fault at clearance +40um, the most impulsive frequency band by the Kurtogram is centred at 22.66kHz with a frequency bandwidth of 1563Hz at level 4 in Figure 7-15. The SSI in Figure 7-16 selects three candidates of the central frequencies, which are 4073Hz, 4985Hz and 8241Hz.

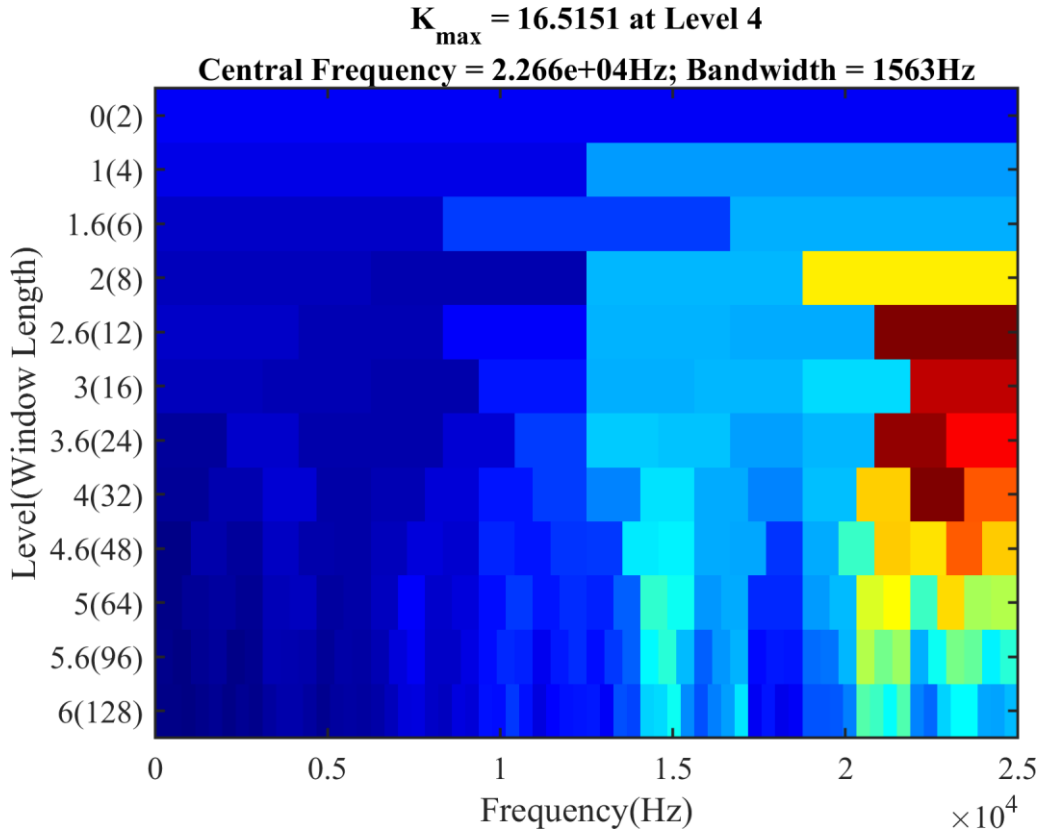


Figure 7-15 Kurtogram for large outer race fault at clearance +40um

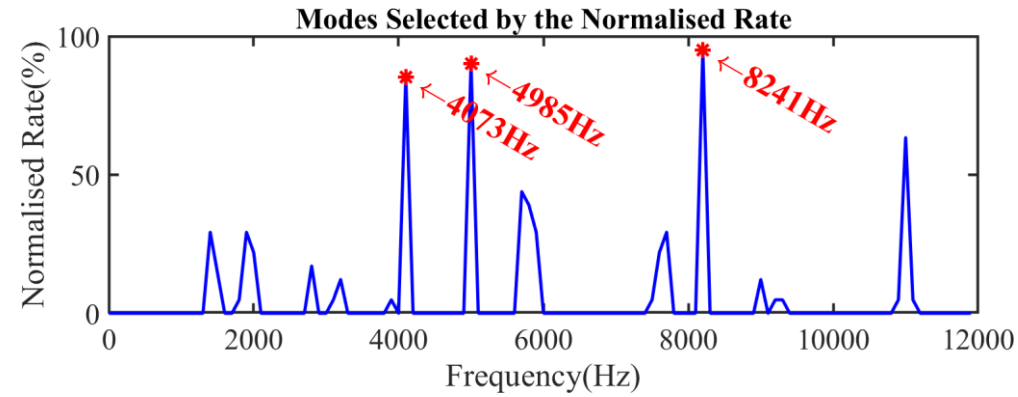


Figure 7-16 Normalised Rate of stable modes in the Stabilisation Diagram for large outer race fault at clearance +40um

In the case of the large outer race fault at clearance +60um, the most impulsive frequency band by the Kurtogram is centred at 22.92kHz with a frequency bandwidth of 4167Hz at level 2.6 in Figure 7-17. The SSI in Figure 7-18 selects two candidates of the central frequencies, which are 4075Hz and 8236Hz.

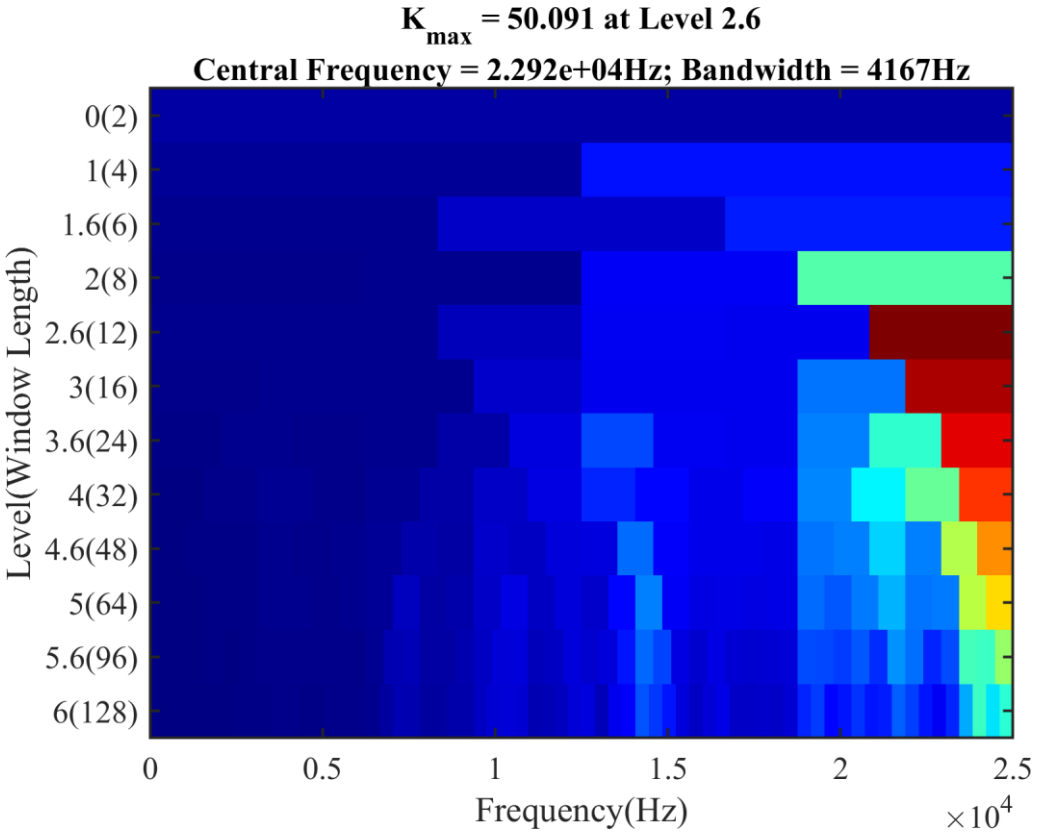


Figure 7-17 Kurtogram for large outer race fault at clearance +60um

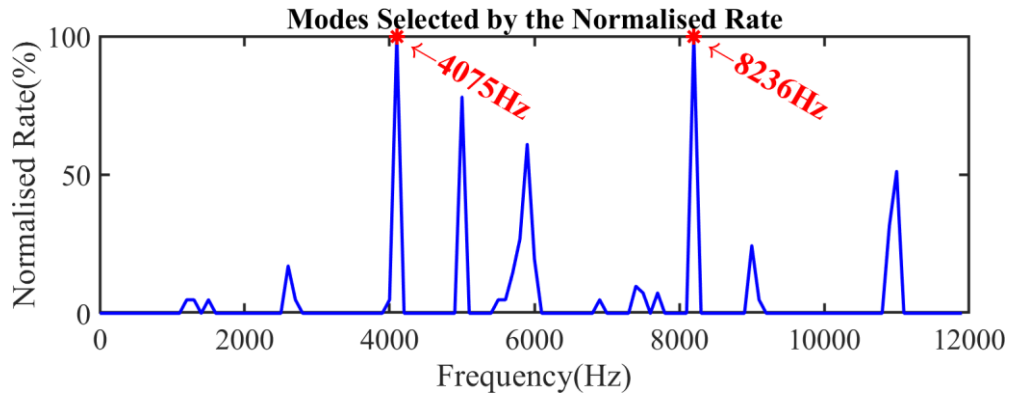


Figure 7-18 Normalised Rate of stable modes in the Stabilisation Diagram for large outer race fault at clearance +60um

In the case of the small outer race fault at clearance -20um, the most impulsive frequency band by the Kurtogram is centred at 22.27kHz with a frequency bandwidth of 781.3Hz at level 5 in Figure 7-19. The SSI in Figure 7-20 selects two candidates of the central frequencies, which are 5782Hz and 8035Hz.

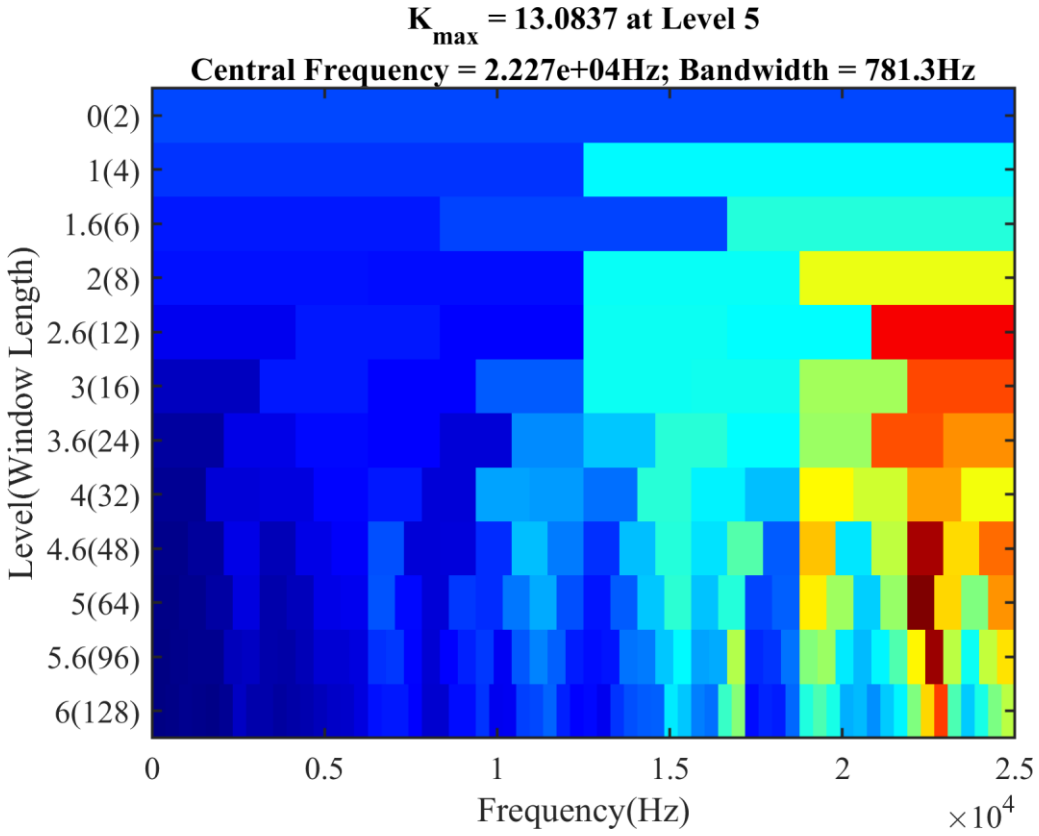


Figure 7-19 Kurtogram for small outer race fault at clearance -20um

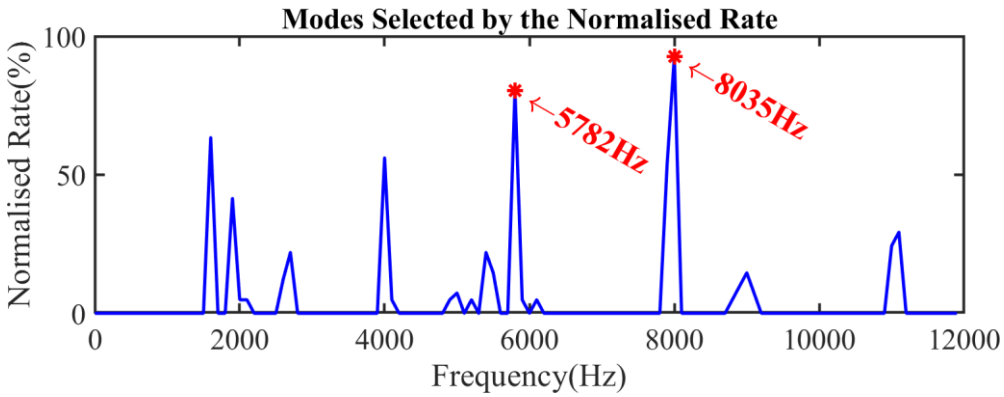


Figure 7-20 Normalised Rate of stable modes in the Stabilisation Diagram for small outer race fault at clearance -20um

In the case of the small outer race fault at clearance 0um, the most impulsive frequency band by the Kurtogram is centred at 22.66kHz with a frequency bandwidth of 520.8Hz at level 5.6 in Figure 7-21. The SSI in Figure 7-22 selects three candidates of the central frequencies, which are 3988Hz, 8001Hz and 9013Hz.

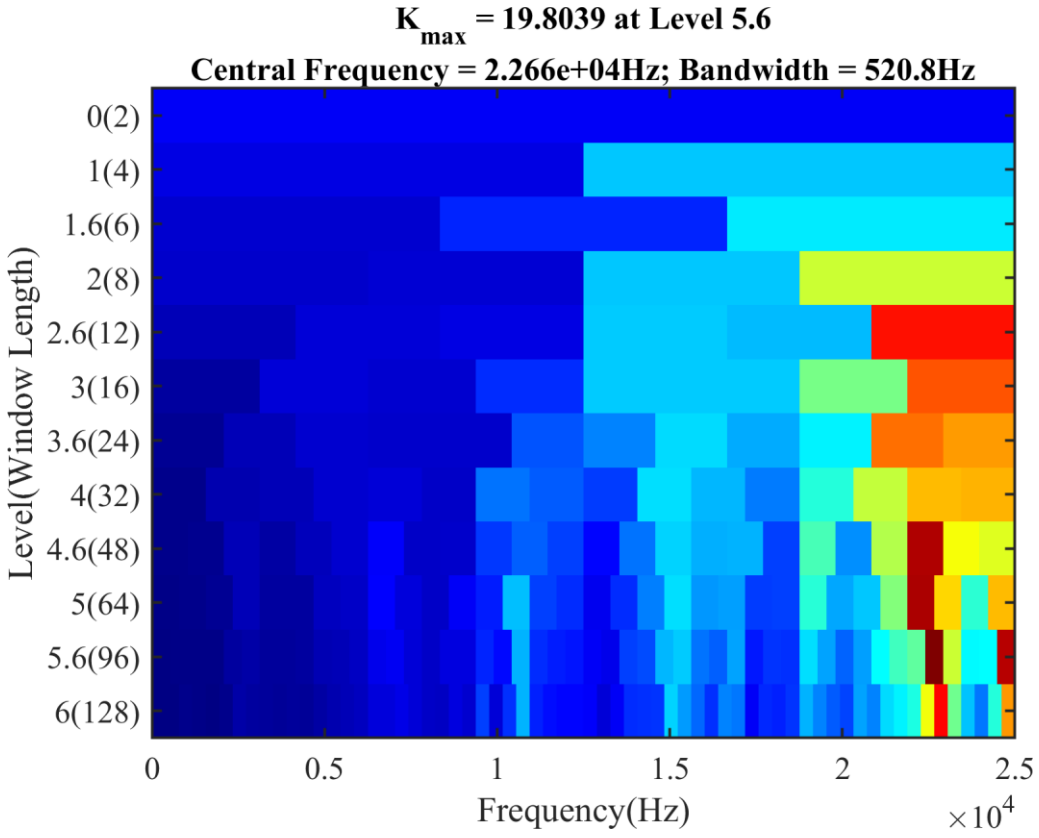


Figure 7-21 Kurtogram for small outer race fault at clearance 0um

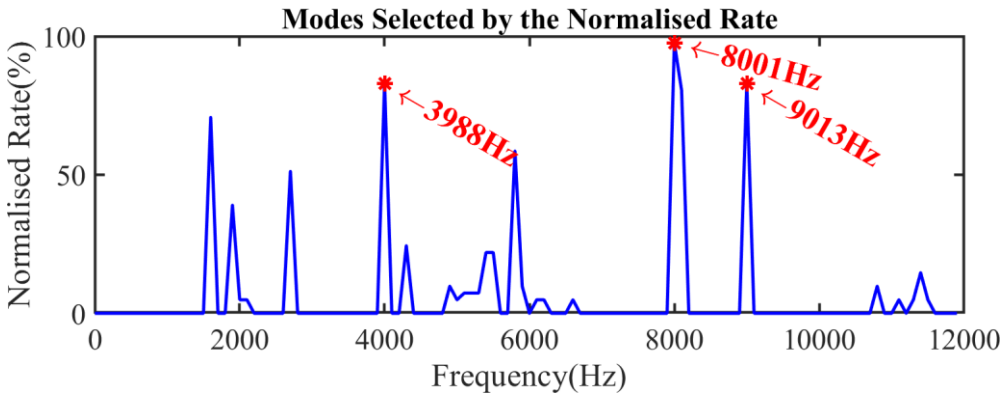


Figure 7-22 Normalised Rate of stable modes in the Stabilisation Diagram for small outer race fault at clearance 0um

In the case of the small outer race fault at clearance +20um, the most impulsive frequency band by the Kurtogram is centred at 11.98kHz with a frequency bandwidth of 1042Hz at level 4.6 in Figure 7-23. The SSI in Figure 7-24 selects three candidates of the central frequencies, which are 5795Hz, 8225Hz and 10985Hz.

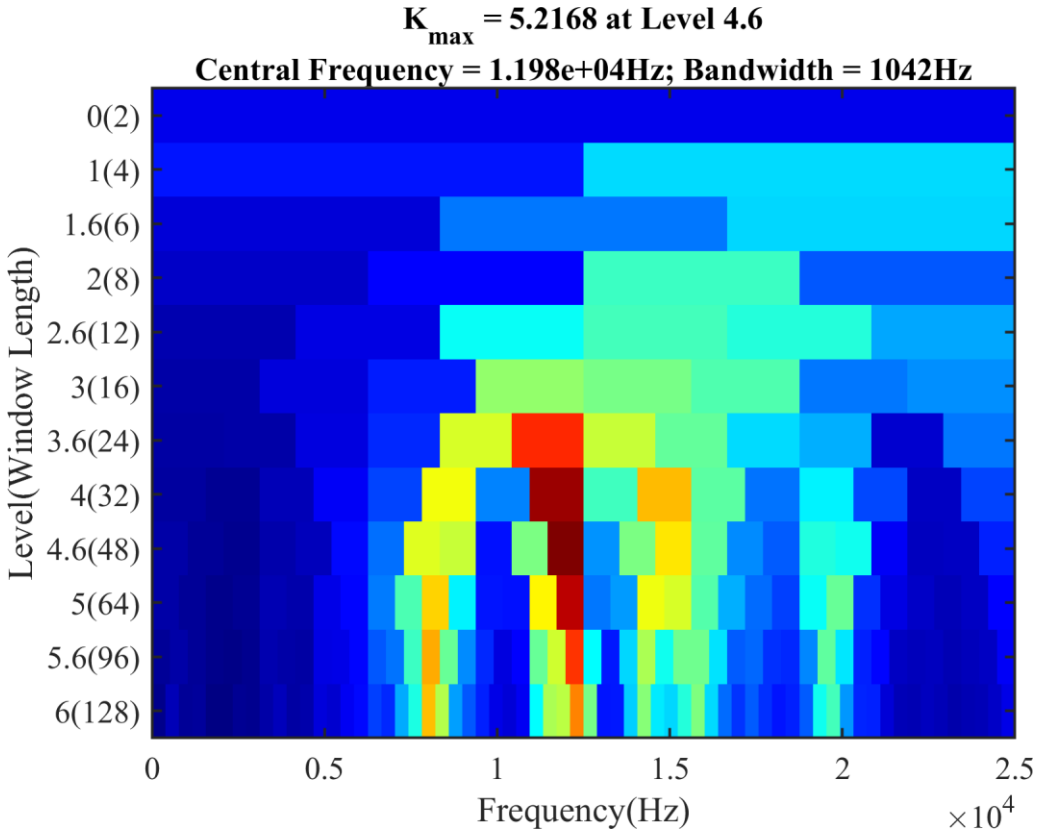


Figure 7-23 Kurtogram for small outer race fault at clearance +20um

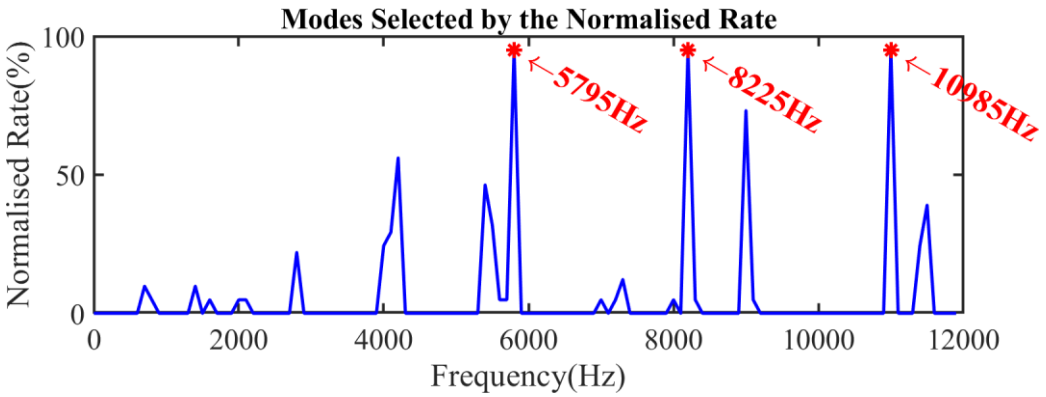


Figure 7-24 Normalised Rate of stable modes in the Stabilisation Diagram for small outer race fault at clearance +20um

In the case of the small outer race fault at clearance +40um, the most impulsive frequency band by the Kurtogram is centred at 11.98kHz with a frequency bandwidth of 1042Hz at level 4.6 in Figure 7-25. The SSI in Figure 7-26 selects five candidates of the central frequencies, which are 4176Hz, 5410Hz, 5813Hz, 8228Hz and 10984Hz.

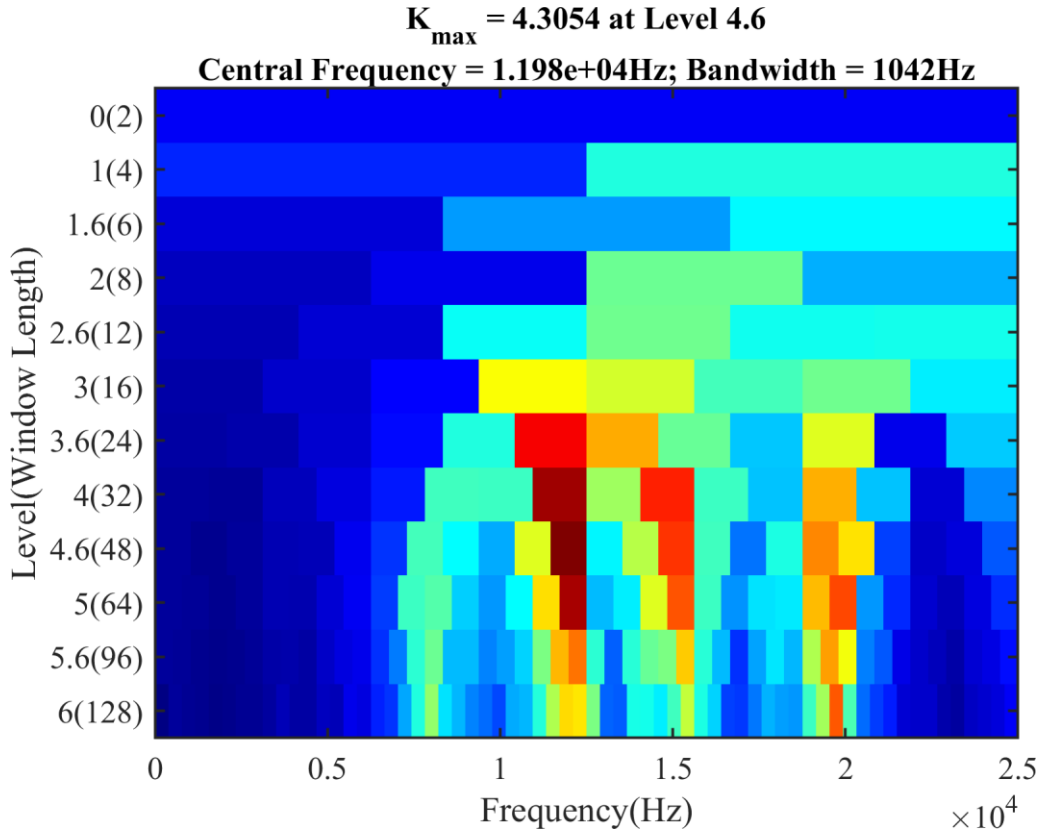


Figure 7-25 Kurtogram for small outer race fault at clearance +40um

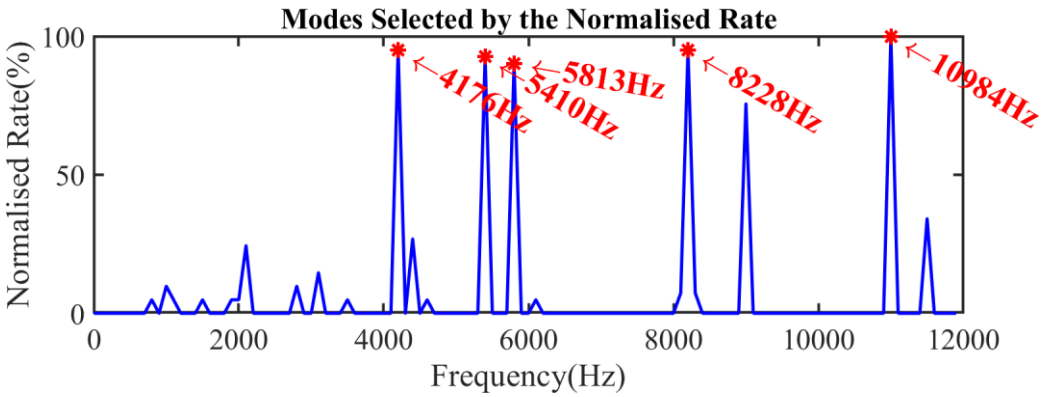


Figure 7-26 Normalised Rate of stable modes in the Stabilisation Diagram for small outer race fault at clearance +40um

In the case of the small outer race fault at clearance +60um, the most impulsive frequency band by the Kurtogram is centred at 21.87kHz with a frequency bandwidth of 2083Hz at level 3.6 in Figure 7-27. The SSI in Figure 7-28 selects four candidates of the central frequencies, which are 5394Hz, 5783Hz, 8235Hz and 10989Hz.

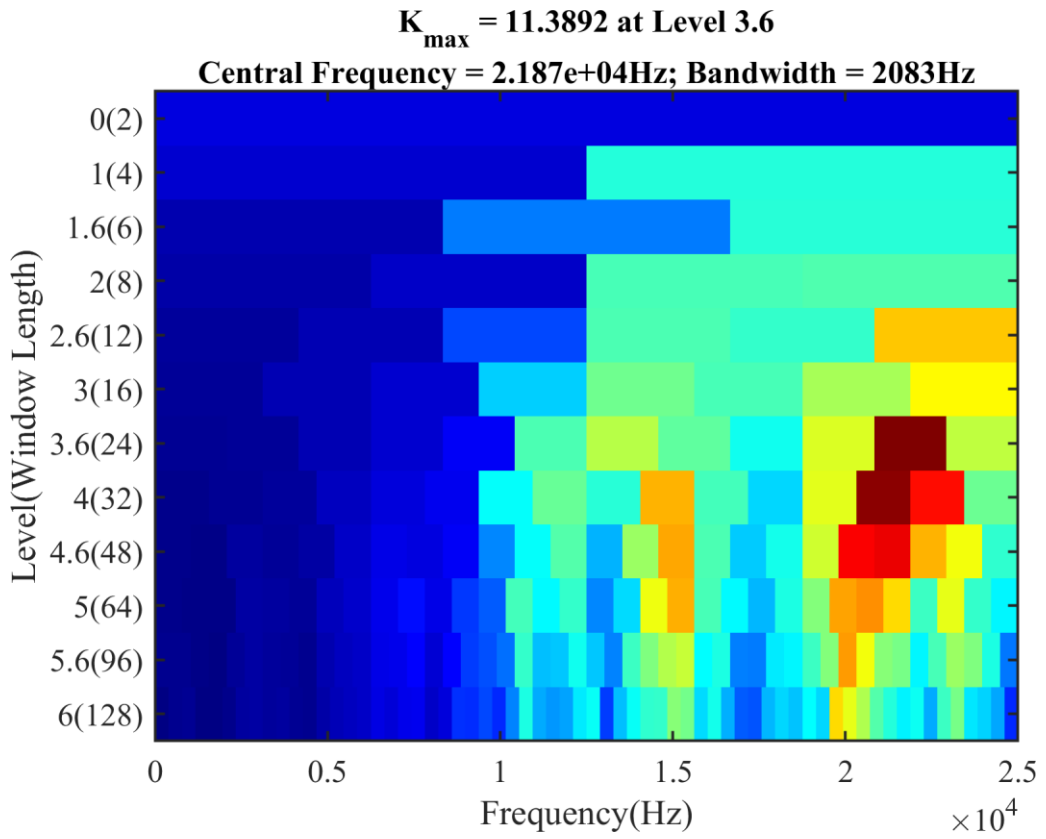


Figure 7-27 Kurtogram for small outer race fault at clearance +60um

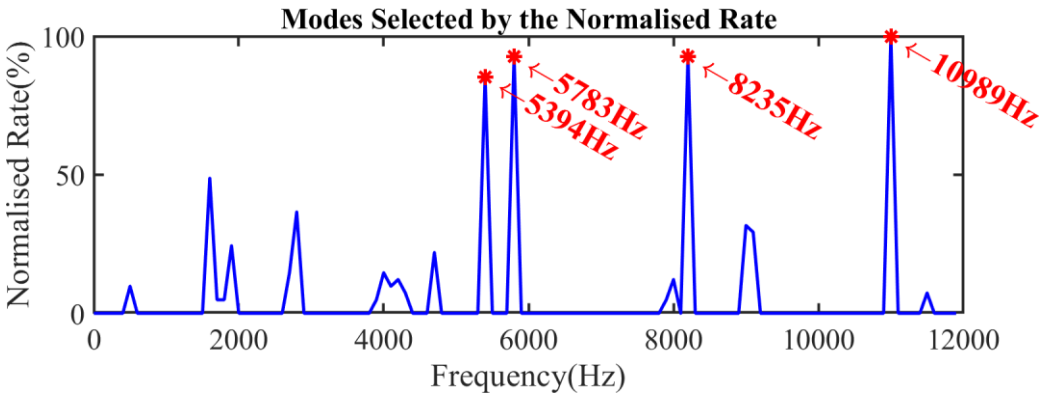


Figure 7-28 Normalised Rate of stable modes in the Stabilisation Diagram for small outer race fault at clearance +60um

A summary of the optimal frequency bands of the vibration signals are listed in the Table 7-4.

Table 7-4 Optimal Frequency Bands

Faults	**Clearance**	**Kurtogram**		**SSI**	
		Central Frequency	Frequency Bandwidth	Central Frequency	Frequency Bandwidth
Large Outer Race Faults	-20um	14844Hz	1563Hz	4082Hz	1099Hz
	0um	13542Hz	2083Hz	4079Hz	1099Hz
	+20um	21875Hz	2083Hz	4083Hz	1099Hz
	+40um	22656Hz	1563Hz	4073Hz	1099Hz
	+60um	22917Hz	4167Hz	4075Hz	1099Hz
Small Outer Race Faults	-20um	22266Hz	781Hz	5782Hz	1092Hz
	0um	22656Hz	521Hz	3988Hz	1099Hz
	+20um	11979Hz	1042Hz	5795Hz	1099Hz
	+40um	11979Hz	1042Hz	5813Hz	1099Hz
	+60um	21875Hz	2083Hz	5783Hz	1099Hz

7.3.2 Demodulation Results and Analysis

Based on the optimal frequency bands selected in the last section, the demodulation results of the EAAE, PL-MSB, EAAS and CE are compared to show the capability of fault detection and diagnosis. The discussion of the fault conditions starts from the most severe which means the vibration signals were collected from the faulty bearing with the large defect at the minimum clearance. The fault signatures are supposed to be most obvious and Figure 7-29 shows the demodulation results of the large outer race defect at a solid contact between rolling elements and raceways. All five methods can successfully extract the fault frequency of the rollers passing the localised defect. The EEAS in the Figure 7-29 (c) shows a fair result only because the tapered bearings are susceptible to the random slippage. The vibration signals of the tapered bearings are

nonstationary and are difficult to characterise by the stationary signal processing method EAAS. The other four methods generate a promising result for denoting the outer race fault. The sidebands around the fault frequency are the shaft rotating frequency. Usually the sidebands are not visible in the outer race fault because the location of defects does not rotate with the rotors and the impacts of the collisions between rollers and faults are nearly constant. In this case, the sidebands around the fault frequency and harmonics are obvious, which is induced by the imbalance of the shaft.

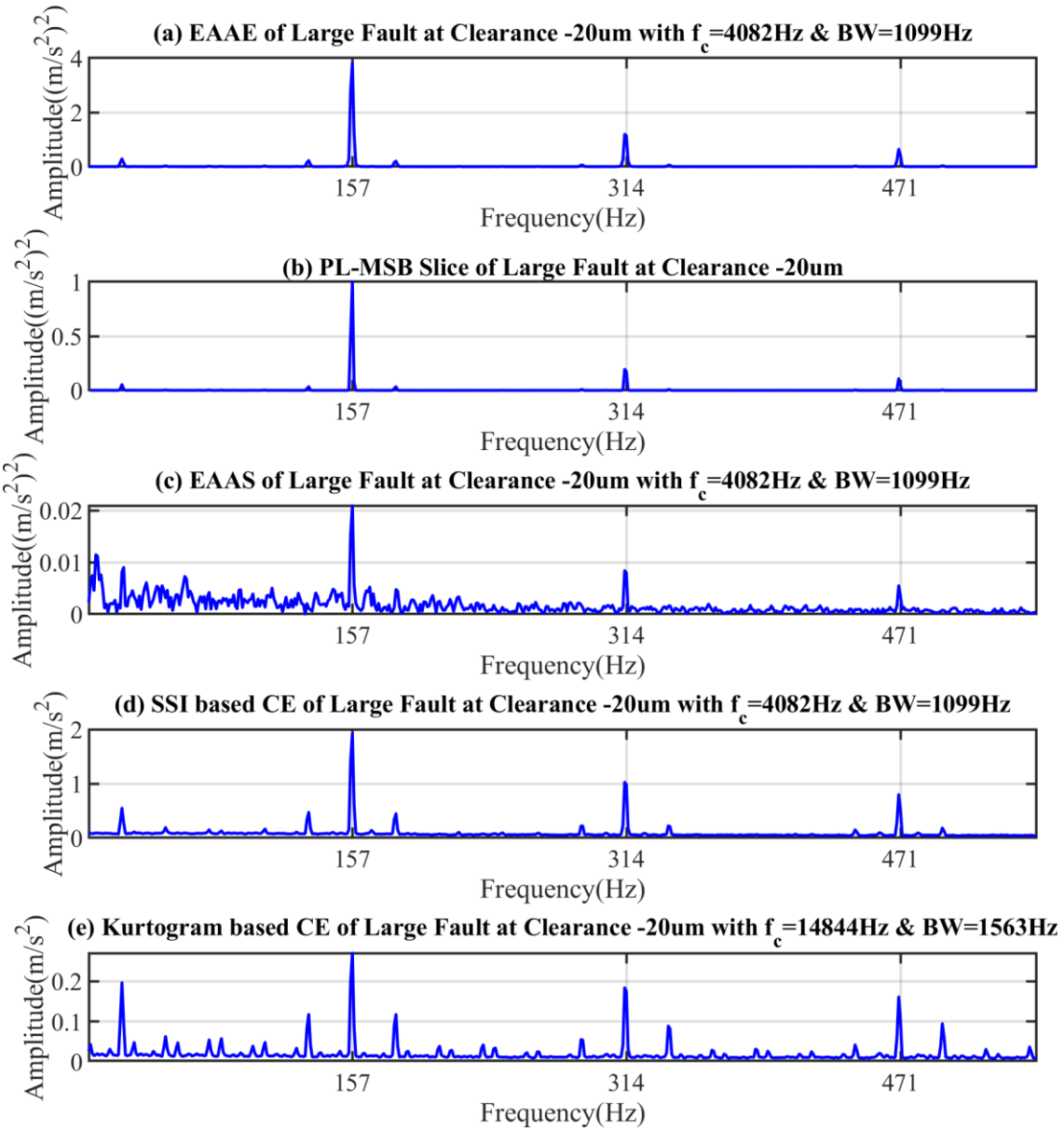


Figure 7-29 EAAE, PL-MSB, EAAS and CE for large outer race fault at clearance -20um

The dynamic responses of the bearing under zero clearance are similar to that under -20um clearance. The spectra in Figure 7-30 denote the diagnostic information extracted is similar to that at clearance -20um in spite of the amplitude of the fault signatures being lower because of the smaller contact stiffness.

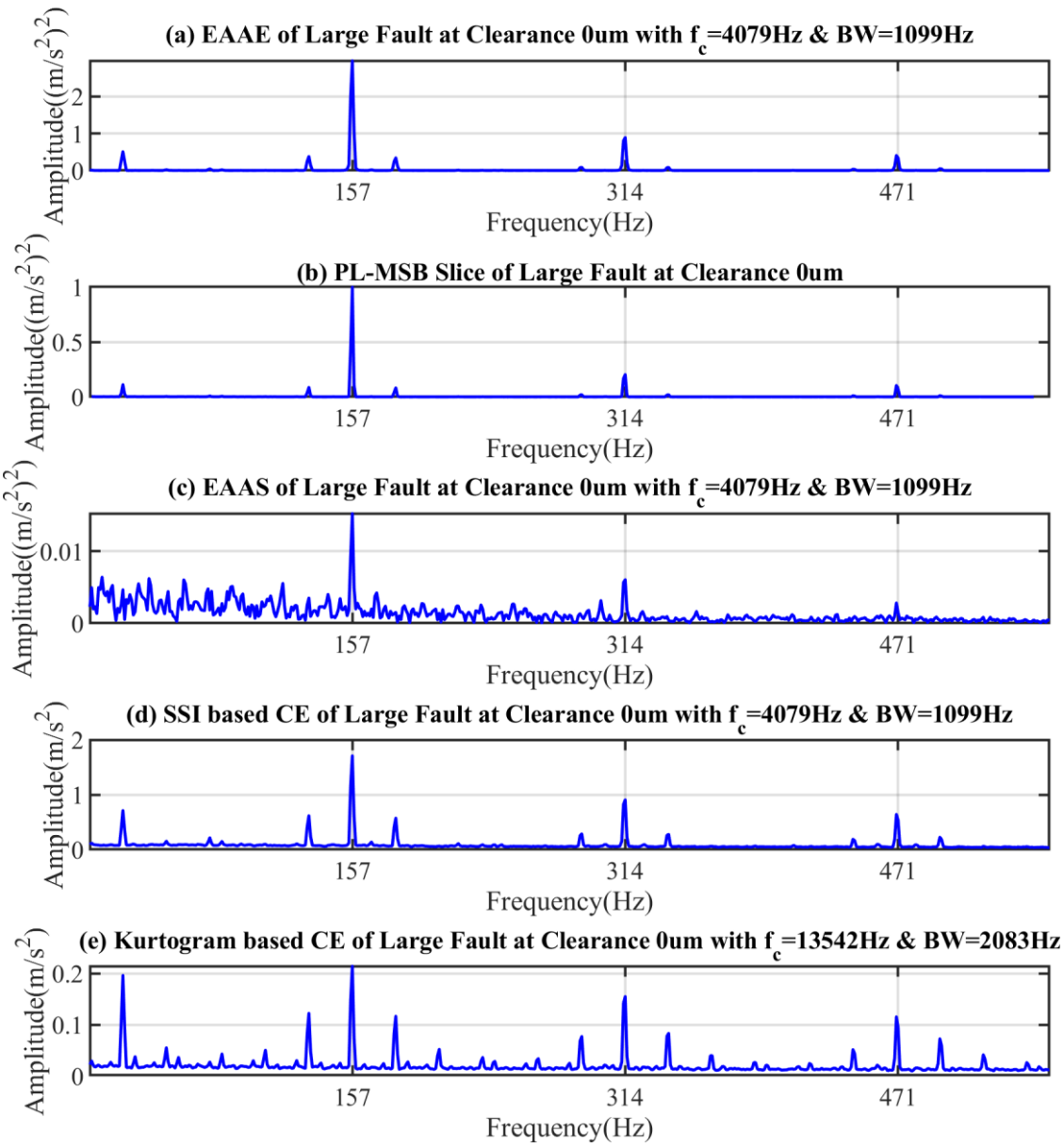


Figure 7-30 EAAE, PL-MSB, EAAS and CE for large outer race fault at clearance 0um

When the internal clearance of the bearings increases to +20um, the contacts between the rollers and the large outer race fault are still decent. Hence, the fault features of the outer race defects are very pronounced in the demodulation results. Based on the frequency bands selected by the Kurtogram and SSI methods, the demodulation analysis can obtain the fault information with high confidence.

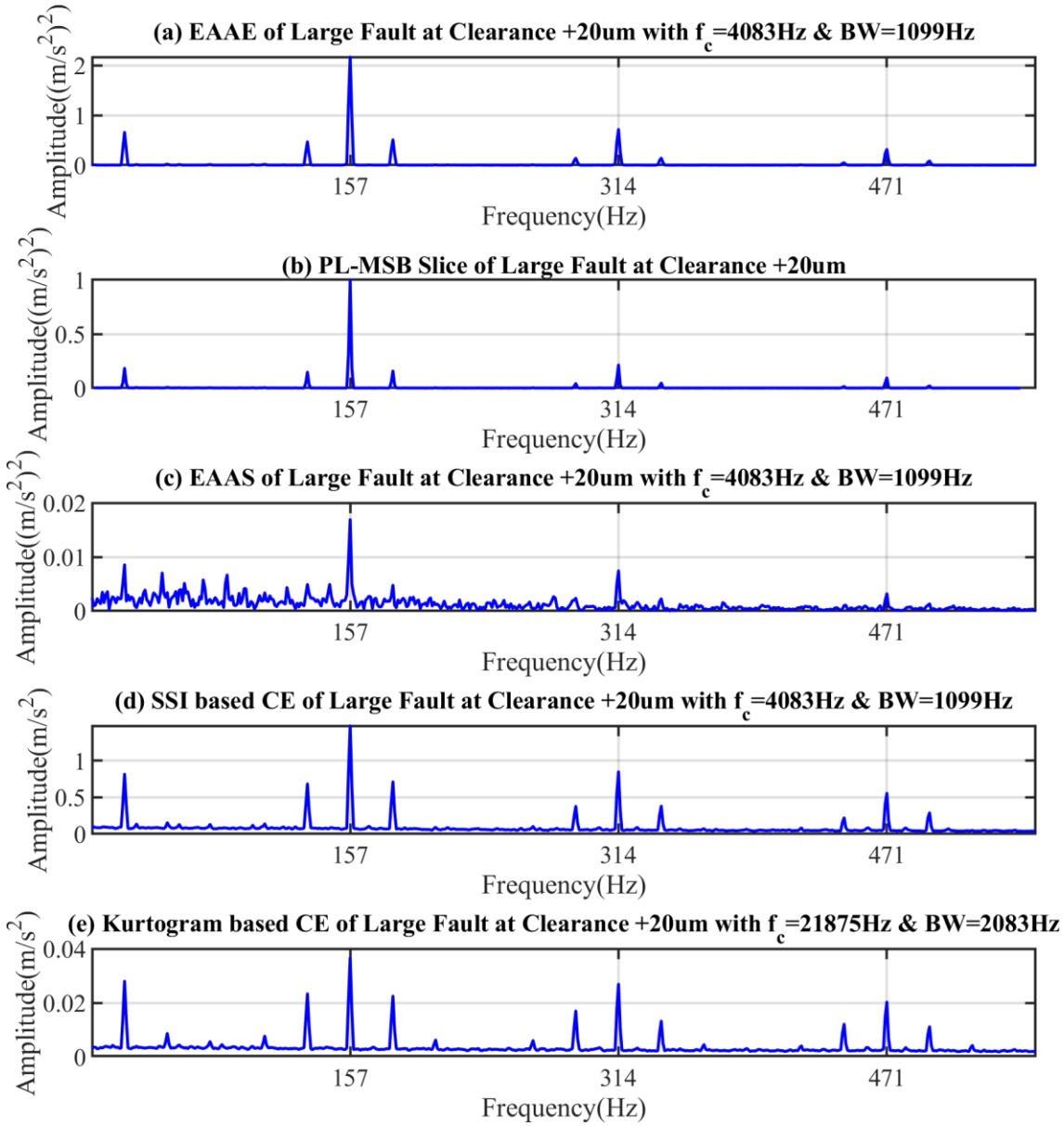


Figure 7-31 EAAE, PL-MSB, EAAS and CE for large outer race fault at clearance +20um

The internal clearance increases to +40um by rotating the adjustable nut backwards to make the gap between the nut and shaft flange as 14.79mm. With the increase of the clearance, the slippages between bearing elements rise, which can be verified by the result of EAAS in Figure 7-32 (c). Due to the increase of the nonstationarity, the EAAS cannot extract the fault signatures due to the increase of the phase noise. The other four methods are effective because the envelope can allow a certain degree of random phase noise in the analysis. With the increase of the clearance, the amplitude of the fault frequency is lower than 1 in the EAAE, which also denotes the more nonstationary characteristics in the vibration signals.

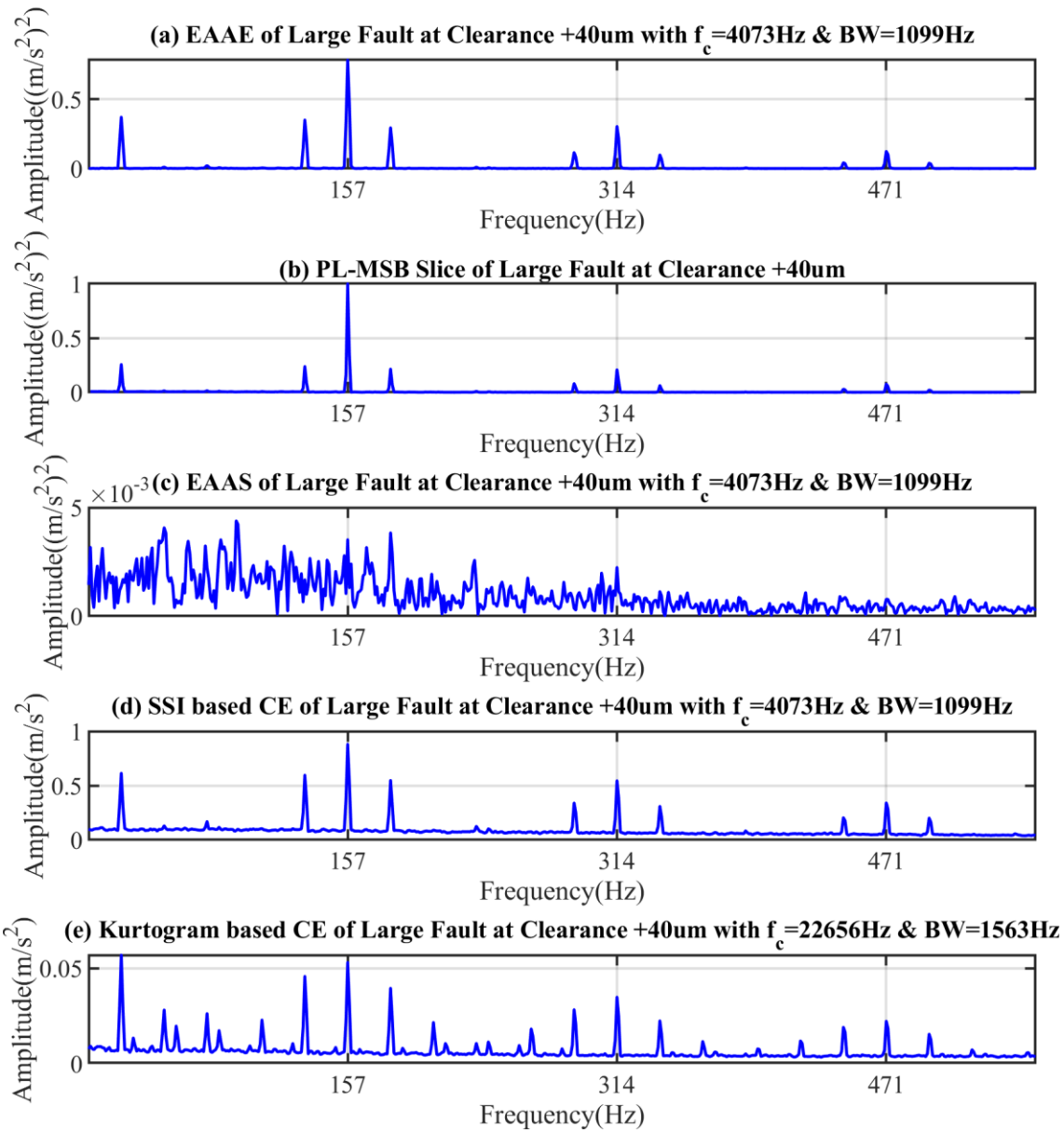


Figure 7-32 EAAE, PL-MSB, EAAS and CE for large outer race fault at clearance +40um

Among the test conditions of the large outer race defect, the internal clearance of +60um is the worst scenario in the vibration based fault detection. With this large clearance, the dynamics of the bearing components becomes extremely complicated and the responses due to the nonlinear system and nonstationary inputs increase the barriers for the effective fault detection and diagnosis especially in a very early stage. The results in Figure 7-33 show that only the EAAS is unable to find the bearing faults due to the capability of the approach limited within the range of stationary signals. The developed methods, EAAE and PL-MSB, represent a very sparse spectra which is informative with the fault frequency and the shaft rotating frequency. The fault frequency is much lower than that of the small clearance and the noise carpet in the conventional envelope at two frequency bands is higher than the approaches with the ability of noise suppression.

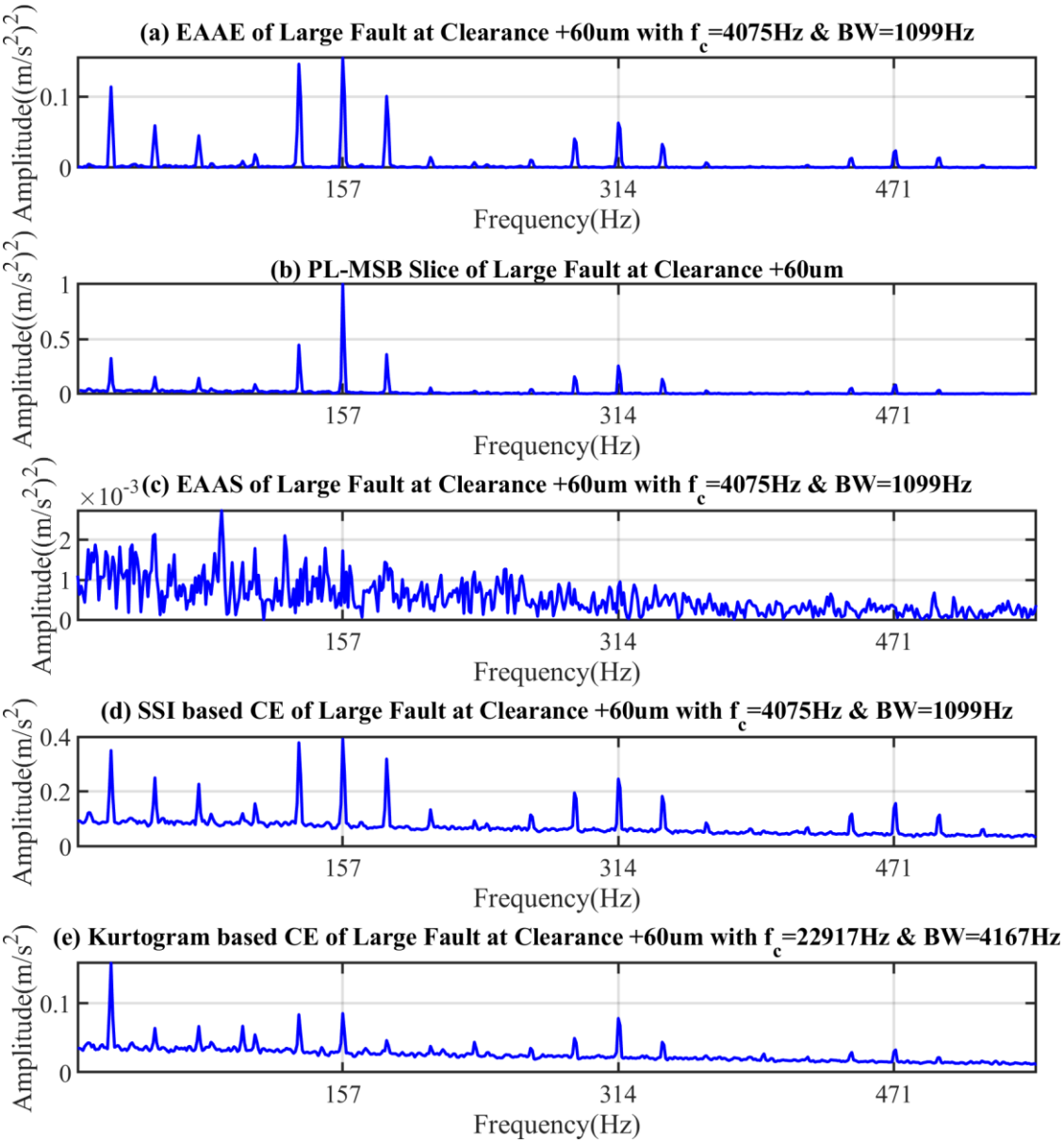


Figure 7-33 EAAE, PL-MSB, EAAS and CE for large outer race fault at clearance +60um

The five scenarios of large outer race faults show that the conventional demodulation methods can reveal the fault information in the vibration signals. The bandpass filter itself has a noise reduction effect and it can reduce the negative impacts on the diagnostic results. For the large fault cases, the background noise cannot shield the fault features, consequently the methods used in the section except for EAAS are effective enough to indicate the fault occurrence on the outer ring.

If the defects on the raceways are incipient, the methods without noise reduction are not capable for finding the tiny failure of bearings. The experiments of the small defect on the tapered roller bearings were carried out to examine the methods on incipient fault detection and diagnosis. The first scenario in the small outer race fault is the internal clearance -20um, which means the installed bearing has a preload without other external forces. The contact between bearing rollers and rings is quite solid and consequently the impacts induced by the localised defect are supposed to be the

strongest. The Figure 7-34 shows the diagnostic results obtained by the five methods. The results obtained in EAAE, PL-MSB and CE show that the fault signatures are much smaller than that in the large fault signals. The main component is the shaft rotating frequency, which dominates the demodulated signals.

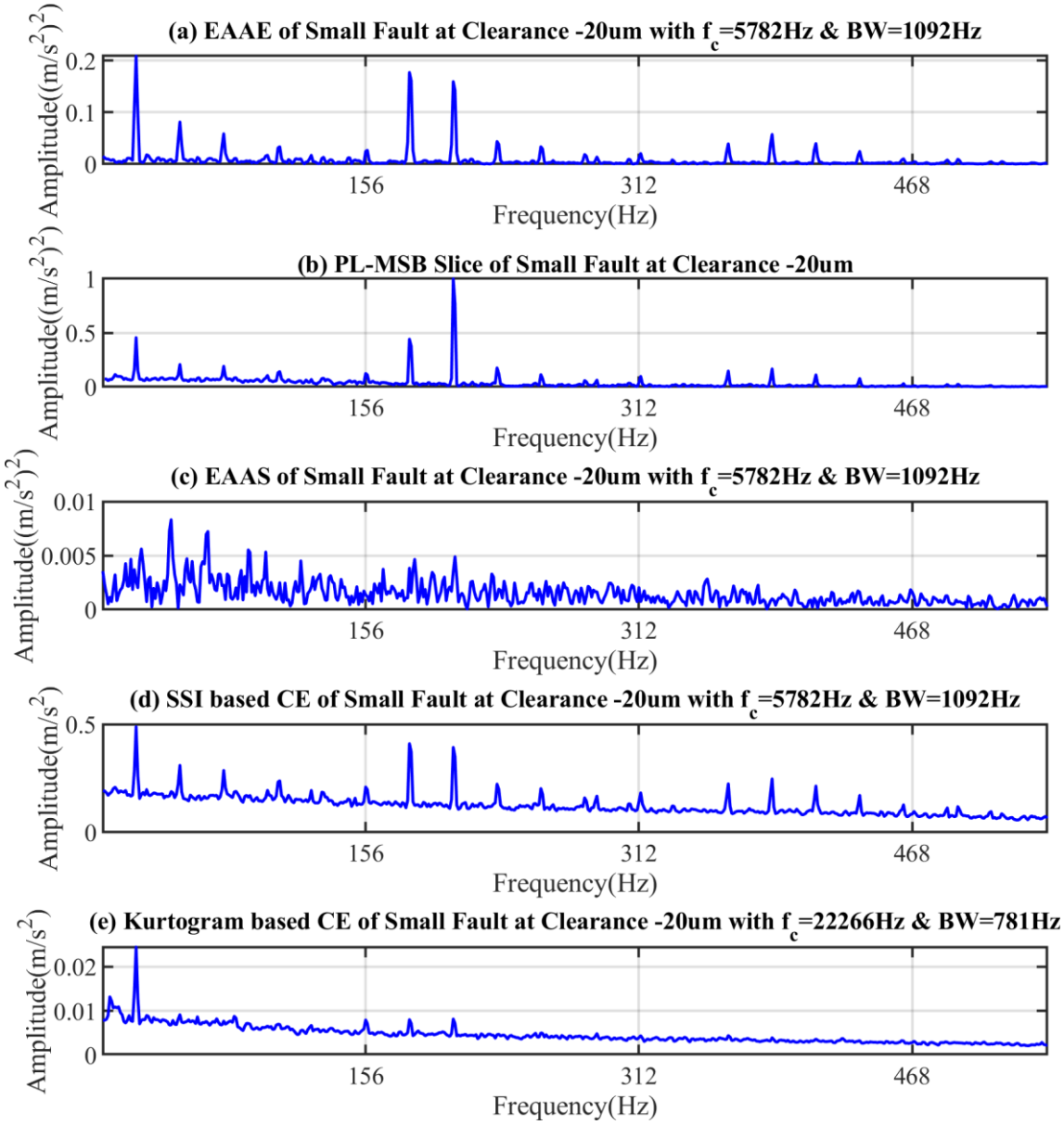


Figure 7-34 EAAE, PL-MSB, EAAS and CE for small outer race fault at clearance -20um

The internal clearance of the bearing increases to the finger tight condition. Based on the frequency bands selected by the Kurtogram and SSI, the demodulation analysis is presented in the Figure 7-35. The fault frequency of the outer race defect is 157Hz at the 100% motor speed. The spectrums of EAAE and PL-MSB methods in the Figure 7-35 (a) and (b) display a manifest fault feature. Figure 7-35 (d) displays that the SSI based CE can extract the fault frequency from the noisy signals although the noise floor is significantly high in the spectrum. The Kurtogram based

CE in Figure 7-35 (e) fails to find the fault information because the frequency band used is not rich of fault information.

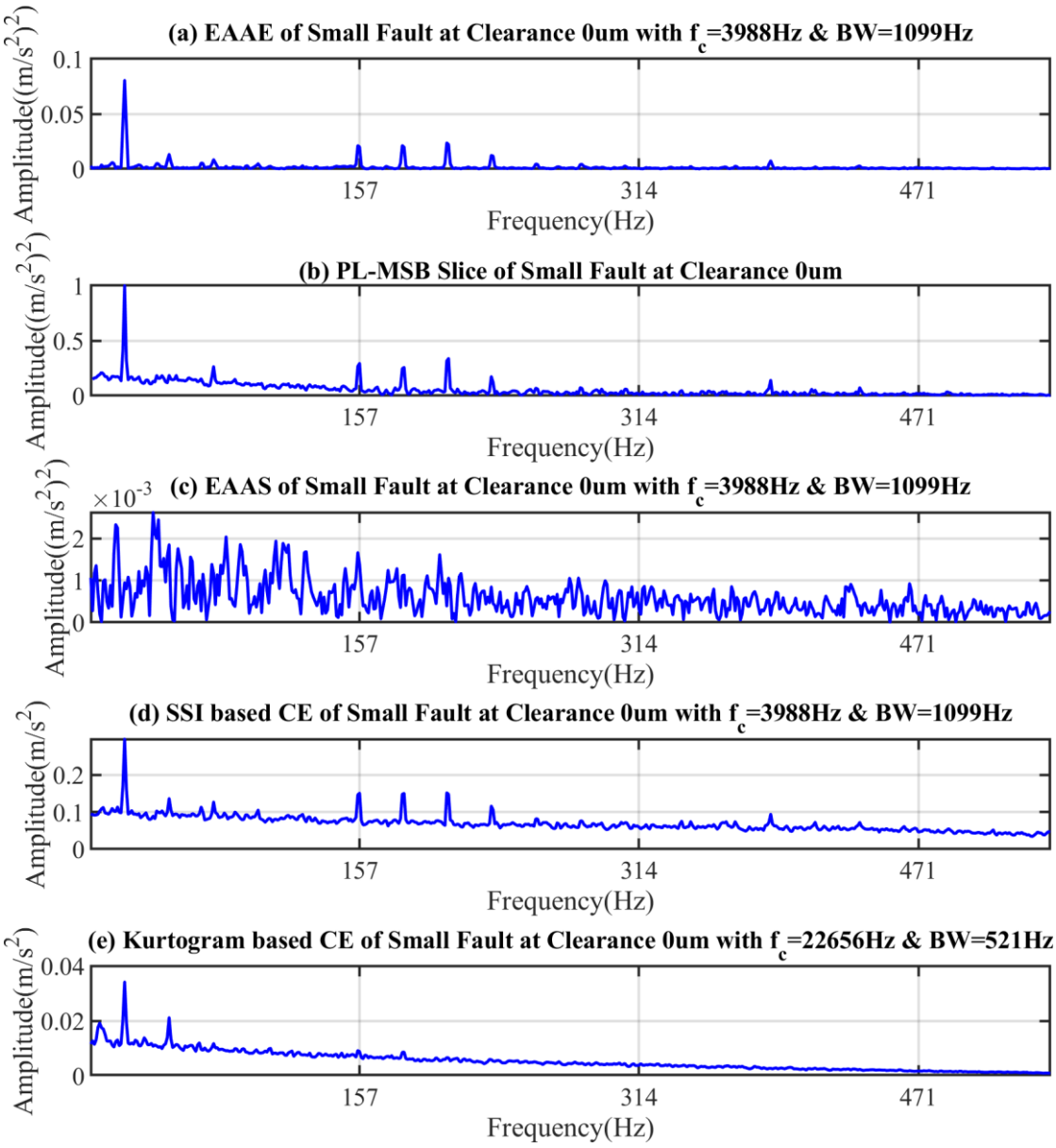


Figure 7-35 EAAE, PL-MSB, EAAS and CE for small outer race fault at clearance 0um

The next scenario of the bearing test is the internal clearance of +20um. The small outer race defect at the clearance +20um is a challenging case for the fault detection. The Figure 7-36 displays the demodulation analysis of the bearing faults. The EAAE and PL-MSB in Figure 7-36 (a) and (b) tell the fundamental frequency of the outer race defect has a very tiny amplitude. The most obvious component is the shaft rotating frequency. As shown in Figure 7-36 (d), the SSI based CE is hardly to highlight the fault signature from the background noise. The Kurtogram based CE shows the same strength of the fault signatures but the noise floor contains too random peaks, which confuses the final diagnostic result.

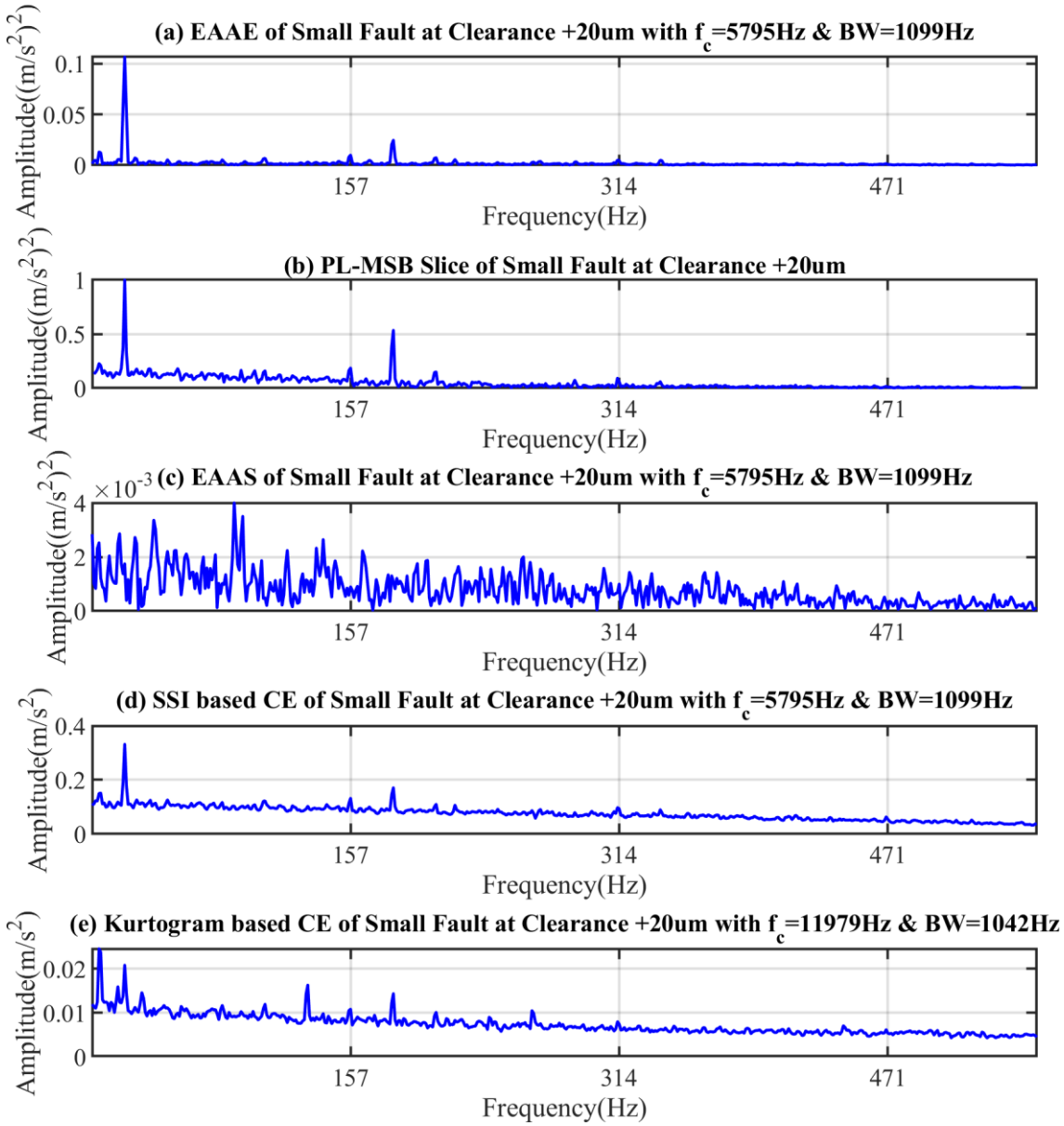


Figure 7-36 EAAE, PL-MSB, EAAS and CE for small outer race fault at clearance +20um

The internal clearance of the bearings increases to +40um. The slippages between rollers and raceways would increase to a high level and the vibration signals are quite difficult to be analysed by the stationary signal processing methods. The input of the bearing system is not as regular as the pure rotating condition. The fault diagnosis of the outer race fault is a tough task because the impacts induced by the localised faults show a flat spectrum in the frequency domain, which is a typically cyclostationary signal. Figure 7-37 describes the demodulation analysis based on the resonant modulation mechanisms. The EAAE in Figure 7-37 (a) shows a clean spectrum, in which the fault frequency of 157Hz is obvious and the rotating frequency sideband is also visible due to the imbalance of the shaft. The high amplitude of fundamental frequency of the rotating shaft denotes the dominate signals in the demodulated waveform. The PL-MSB in Figure 7-37 (b) displays the method can extract the fault information from the noisy signals however the noise is

not sufficiently suppressed due to the instantaneous phase not being accurately estimated. It is not beyond expectation that the EAAS is not effective for this nonstationary signal. With the assistance of the proper frequency band, the SSI based CE can obtain the information of the outer race fault but the fault signature does not exceed the noise carpet and it can be easily neglected because of the tiny amplitude. The spectrum in the Figure 7-37 (e) shows that the Kurtogram based CE cannot find the fault which occurred on the outer race.

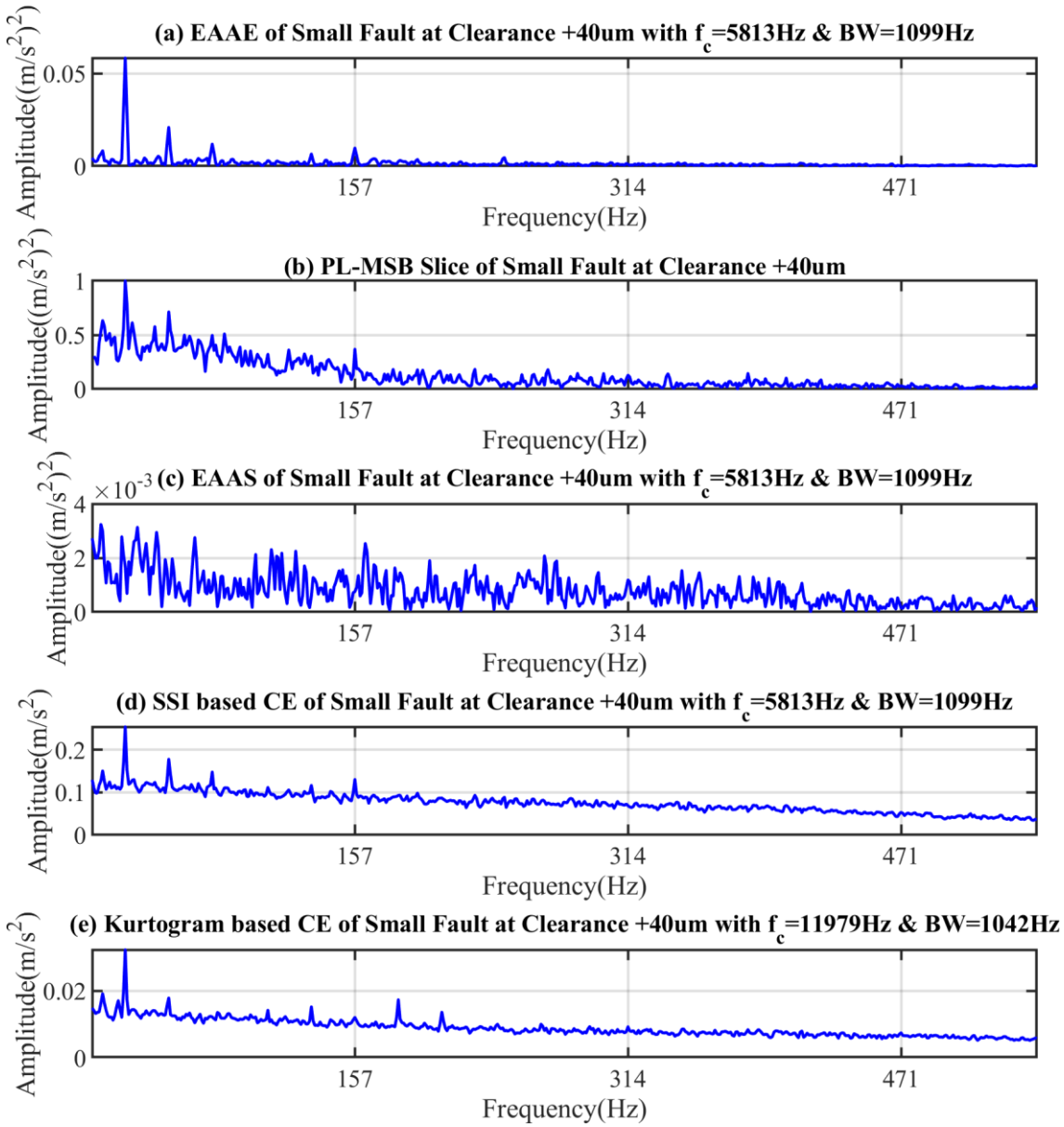


Figure 7-37 EAAE, PL-MSB, EAAS and CE for small outer race fault at clearance +40um

The most challenging signals in the tapered roller bearing tests are the internal clearance of +60um. The clearance between the inner ring and the outer ring leads to a significantly low contact stiffness which means the rollers suffer from a high level of random slippages. The cyclostationary bearing signals analysed by the five methods and the results are shown in Figure 7-38. The EAAE method clearly shows the defect frequency of the outer race. The rotating frequency is still the most

pronounced component in the demodulated signals. The PL-MSB delivers a best diagnostic spectrum with a significantly obvious fault frequency, which clearly shows the outer race fault. The conventional envelope is not reliable at this condition because the fault signature is so weak that the fault features are submerged into the background noise.

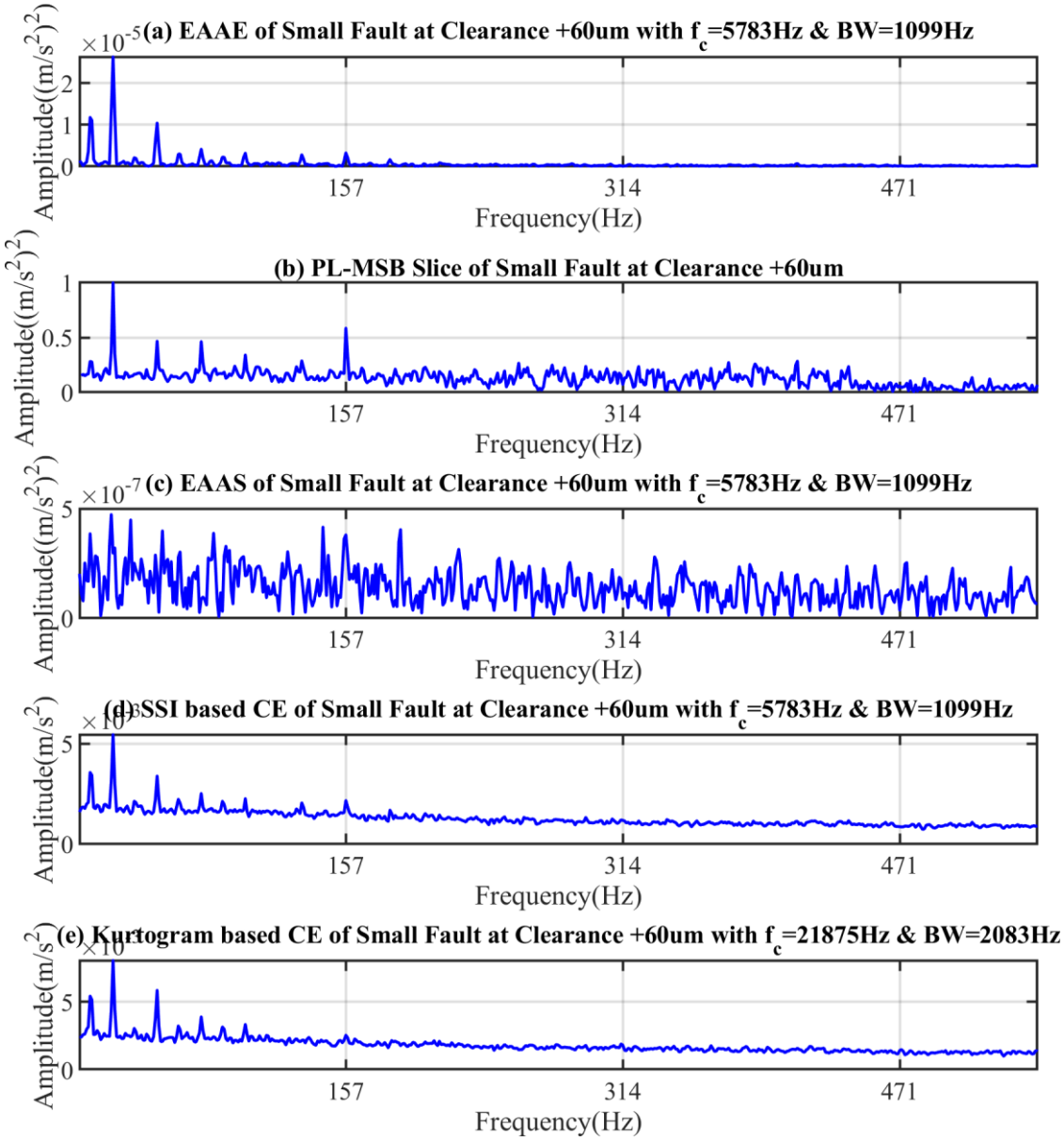


Figure 7-38 EAAE, PL-MSB, EAAS and CE for small outer race fault at clearance +60um

7.3.3 Quantification of Bearing Fault Signatures

To quantify the practical performance of five demodulation methods, the signature indicators $SNR_{signature}$ extracted in processing the experimental vibration signals are shown in Figure 7-39. The fault signatures extracted by the EAAE and PL-MSB are more pronounced than the other three methods. Although the fault size is large, the fault information extracted by the EAAS is attenuated substantially with the increase of the internal bearing clearance, which can show the

vibration signals of the tapered bearing is cyclostationary. In spite of the EAAS, the other four approaches can detect the occurrence of the outer race fault. The developed EAAE and PL-MSB are the most effective and the strength of the fault features is more than 18dB at all five clearances. The SSI based CE indicate better diagnostic results than the Kurtogram based CE because the frequency band selected by the SSI is more reliable.

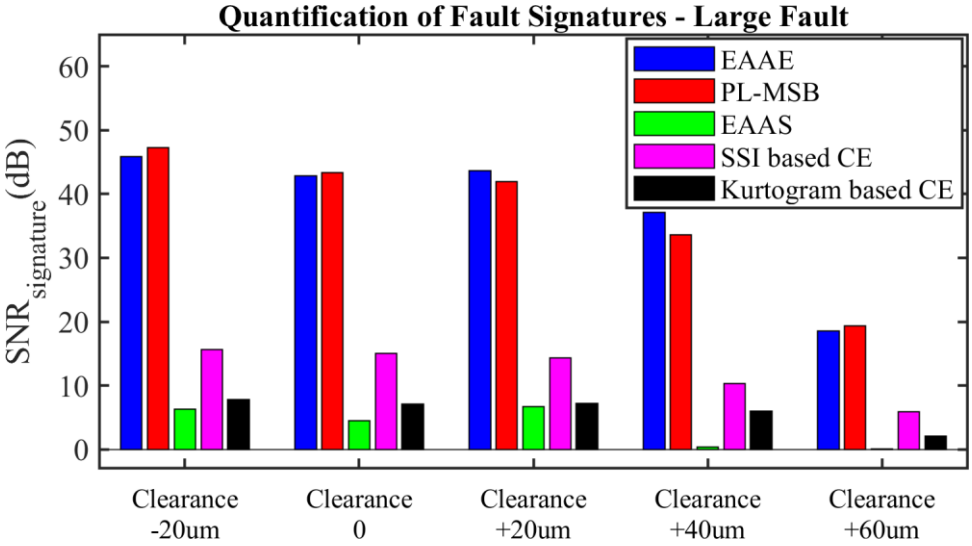


Figure 7-39 Signature indicators $SNR_{signature}$ of large outer race faults

For the very tiny fault on the outer ring of the tapered bearing, the nonstationary fault impacts are weak and the induced responses are of poor SNR. The fault signatures extracted by the five methods are shown in Figure 7-40. The EAAE and PL-MSB demonstrate the effectiveness on early fault detection and diagnosis.

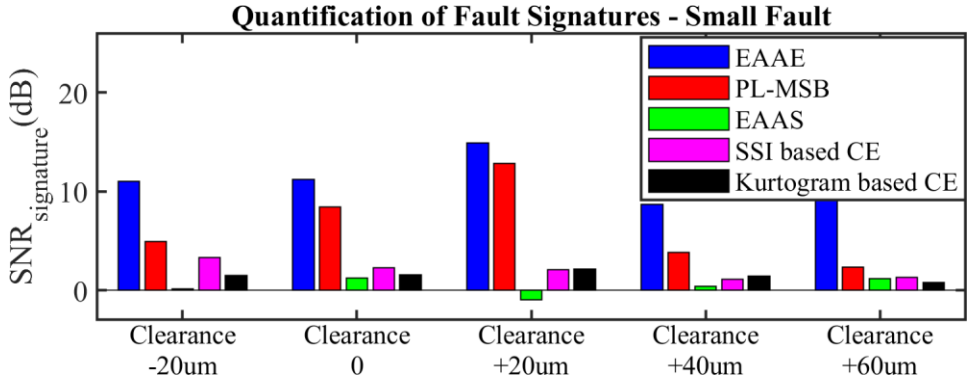


Figure 7-40 Signature indicators $SNR_{signature}$ of small outer race faults

7.4 Summary

The experimental studies on the tapered roller bearings verified the findings in the simulation investigation. The slippage phenomenon of rolling bearings introduced by Randal et al [110] were introduced in practical working conditions. The slippage within tapered bearings was controlled by the internal clearances. The vibration signals obtained at five clearance levels were collected to

examine the performance the methods on demodulating the resonant responses. Using the aperiodic impulses as the system input can lead to the cyclostationary output of the vibration signals. In this resonant modulation process, the demodulation methods for periodic signals are ineffective in this nonstationary condition. EAAE and PL-MSB are powerful methods to detect the incipient bearing faults even when the SNR of the signals is extremely low.

Chapter 8 Conclusions and Future Work

This chapter summarises the research activities and achievements for vibration-based machine condition monitoring in this thesis. The research achievements for the objectives are also summarised here. The contributions to knowledge are also included in this chapter. Finally, the future works are recommended for effective machine condition monitoring by using the resonant modulation and demodulation.

8.1 Conclusions

This research aims to exploit the resonant modulation in rotating machines for effective and efficient fault detection and diagnosis at early stages. By modelling and simulation based studies, several signal processing methods and a novel sensing technique were developed to extract fault signatures from extremely noisy vibration signals. The research objectives have been fulfilled and summarised as follows

(1) The typical demodulation methods, including envelope analysis, energy operator, modulation signal bispectrum (MSB), spectral correlation, were reviewed to present state-of-the-art usage of resonant modulation in vibration based condition monitoring of rotating machines.

(2) The dynamic responses of a linear system under different types of excitations were investigated to highlight the characteristics in resonant modulation, which leads to the basis for developing reliable demodulation methods.

(3) The stochastic subspace identification (SSI) was employed to determine the optimal central frequencies from vibration signals, showing more reliable results than the widely used Kurtogram method.

(4) The ensemble average of autocorrelation signals (EAAS) was developed to suppress the random noise and enhance the modulating signals in the resonant zones. The developed method can demodulate the fault information for the resonant behaviour under the very poor SNR up to -35dB.

(5) The ensemble average of autocorrelated envelopes (EAAE) and phase linearisation based modulation signal bispectrum (PL-MSB) were developed to characterise the aperiodic impacts induced resonant modulation signals and the second order cyclic properties can be identified by these two novel methods from very poor-quality signals.

The research activities carried out in the thesis are based on the resonant modulation in rotating machines and the findings can be summarised as follows:

(1) The review of the frequently used demodulation methods shows that resonant modulation and demodulation in vibration-based condition monitoring of rotating machines attract substantial attention but the insight of the resonant modulation has not been sufficiently investigated from the perspective of system identification.

(2) Based on the hypothesis of linear systems, the resonant modulation in outputs of the systems vary according to different input excitations. Three types of inputs (periodic, aperiodic impulsive, and quasi-stationary) can lead to different modulation characteristics. The effective demodulation analysis of these responses is the most beneficial part for utilising the resonant modulation in vibration based condition monitoring. The characteristics of the vibration signals are expected to be effectively analysed by different signal processing approaches for early fault detection and diagnosis at a very early stage.

(3) The stochastic subspace identification (SSI) approach is employed to select the optimal central frequencies from the viewpoint of system identification other than to select the most impulsive frequency band by using methods like Kurtogram. The simulation and experimental studies show that the SSI is more robust to the strong background noise.

(4) A novel method, named ensemble average of autocorrelation signals (EAAS), was developed to demodulate the periodic impacts induced resonant modulation signals and suppress the random noise for the purpose of effective and efficient early fault detection. Both simulation and experimental studies show that the EAAS is a more reliable and more accurate approach than the conventional envelope analysis for identifying the incipient ball bearing faults from severely contaminated vibration signals.

(5) The responses of the bearing under the approximate periodic impulsive excitations are cyclostationary. The random phase of aperiodic impacts is simulated by a Markov process and the cyclostationary responses under the approximately periodic excitations are characterised by two novel methods, EAAE and PL-MSB. Both simulation and experiment studies demonstrate that the proposed methods can detect the bearing defects from the extremely noisy cyclostationary responses by demodulating from the limited bandwidth vibration in resonant zones.

8.2 Contributions to Knowledge

The research activities in this thesis are concentrated on the resonant modulation and demodulation and the investigation has delivered new understanding on vibration based machine condition monitoring. The contributions to knowledge are listed as follows:

(1) The ensemble average of autocorrelation signals (EAAS) is a novel method to suppress the random noise and enhance the periodicity of the fault information. Combined with the central frequencies recommended by the stochastic subspace identification (SSI), the EAAS is able to deliver a promising and accurate fault diagnosis of early bearing faults.

(2) The ensemble average of autocorrelated envelopes (EAAE) developed for analysing impulsive cyclostationary signals is capable to identify the incipient defect from the severely contaminated nonstationary vibration signals.

(3) The phase linearisation based modulation signal bispectrum (PL-MSB) can eliminate the phase noise and tune the cyclostationary measurements into the deterministic components dominated signals. The PL-MSB is capable of characterising the second order cyclostationary vibration signals from the rolling element bearings and consequently gives effective and efficient detection and diagnosis of bearing failure at a very early stage.

8.3 Future Work

The resonant modulation in rotating machines has been investigated in this thesis and the modulation phenomena give potentials for effective and efficient detection and diagnosis at a very early stage. The continuation of the study could be considered from the following points.

(1) The SSI based optimal frequency bands can be improved to find the optimal natural frequency from the identified modal parameters. Moreover, the other operational modal analysis methods can be evaluated to find the appropriate methods for the different inputs induced resonant modulation behaviour.

(2) The background noise in raw vibration signals can be suppressed for the purpose of more accurately estimating the instantaneous phase of the carrier signal. On this basis, the PL-MSB can be applied more effectively for cyclostationary bearing signals.

(3) The conventional on-shell accelerometer is challenging in abstracting the desired information due to the contamination of the noise and the impacts from nearby machines or components. Novel sensing techniques can be explored to increase signal qualities. For instance, the On-Rotor Sensing (ORS) is a technique to directly obtain the rotor vibration signals by using an on-rotor accelerometer. The ORS can obtain high signal to noise ratio signal by nature because the ORS signals are exclusive of the complicated transmission path and the large quantities of background noise.

(4) Most of the machine failures can be attributed to lubrication failure. The early fault detection and diagnosis is recommended to monitor the lubrication conditions other than the existent defects. Monitoring lubrication conditions can lead to more effective and efficient detection and diagnosis of rotating machines.

Bibliography

[1] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*, 1 edition. Chichester, West Sussex, U.K. ; Hoboken, N.J: Wiley, 2011.

[2] A. Davies, Ed., *Handbook of condition monitoring*. Dordrecht: Springer Netherlands, 1998.

[3] A. K. Tangirala, *Principles of system identification: theory and practice*. Taylor & Francis, 2014.

[4] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. Valente de Oliveira, and R. E. Vásquez, 'A review on data-driven fault severity assessment in rolling bearings', *Mechanical Systems and Signal Processing*, vol. 99, no. Supplement C, pp. 169–196, Jan. 2018, doi: 10.1016/j.ymssp.2017.06.012.

[5] S. A. Abdusslam, 'Detection and Diagnosis of Rolling Element Bearing Faults Using Time Encoded Signal Processing and Recognition', doctoral, University of Huddersfield, 2012.

[6] Z. Chang, Q. Jia, X. Yuan, and Y. Chen, 'Main failure mode of oil-air lubricated rolling bearing installed in high speed machining', *Tribology International*, vol. 112, pp. 68–74, Aug. 2017, doi: 10.1016/j.triboint.2017.03.024.

[7] Y. Xu, D. Zhen, J. X. Gu, K. Rabeyee, F. Chu, F. Gu, and A. D. Ball, 'Autocorrelated Envelopes for early fault detection of rolling bearings', *Mechanical Systems and Signal Processing*, vol. 146, p. 106990, Jan. 2021, doi: 10.1016/j.ymssp.2020.106990.

[8] J. S. Rao, *History of rotating machinery dynamics*, vol. 20. Dordrecht: Springer Netherlands, 2011.

[9] D. J. Pedregal and M. Carmen Carnero, 'Vibration analysis diagnostics by continuous-time models: A case study', *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 244–253, Feb. 2009, doi: 10.1016/j.ress.2008.03.003.

[10] M. S. Darlow, R. H. Badgley, and G. W. Hogg, 'Application of High-Frequency Resonance Techniques for Bearing Diagnostics in Helicopter Gearboxes.', No. MTI-74TR25. MECHANICAL TECHNOLOGY INC LATHAM NY, Oct. 1974.

[11] A. Daraz, S. Alabied, A. Smith, F. Gu, and A. D. Ball, 'Detection and diagnosis of centrifugal pump bearing faults based on the envelope analysis of airborne sound signals', in *2018 24th International Conference on Automation and Computing (ICAC)*, 2018, doi: 10.23919/IConAC.2018.8749053.

[12] X. Tang, Y. Xu, F. Gu, A. D. Ball, and G. Wang, 'Fault detection of rolling element bearings using the frequency shift and envelope based compressive sensing', in *2017 23rd International Conference on Automation and Computing (ICAC)*, 2017, pp. 1–6, doi: 10.23919/IConAC.2017.8082063.

[13] J. F. Kaiser, 'On a simple algorithm to calculate the "energy" of a signal', in *International Conference on Acoustics, Speech, and Signal Processing*, 1990, pp. 381–384 vol.1, doi: 10.1109/ICASSP.1990.115702.

[14] I. S. Bozchalooi and M. Liang, 'Teager energy operator for multi-modulation extraction and its application for gearbox fault detection', *Smart Materials and Structures*, vol. 19, no. 7, p. 075008, Jun. 2010, doi: 10.1088/0964-1726/19/7/075008.

[15] M. Liang and I. Soltani Bozchalooi, 'An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection', *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1473–1494, Jul. 2010, doi: 10.1016/j.ymssp.2009.12.007.

[16] H. Liu, X. Wang, and C. Lu, 'Rolling bearing fault diagnosis based on LCD–TEO and multifractal detrended fluctuation analysis', *Mechanical Systems and Signal Processing*, vol. 60–61, pp. 273–288, Aug. 2015, doi: 10.1016/j.ymssp.2015.02.002.

[17] Y. Imaouchen, M. Kedadouche, R. Alkama, and M. Thomas, 'A Frequency-Weighted Energy Operator and complementary ensemble empirical mode decomposition for bearing fault detection', *Mechanical Systems and Signal Processing*, vol. 82, pp. 103–116, Jan. 2017, doi: 10.1016/j.ymssp.2016.05.009.
[18] R. B. Randall and W. A. Smith, 'Uses and mis-uses of energy operators for machine diagnostics', *Mechanical Systems and Signal Processing*, vol. 133, p. 106199, Nov. 2019, doi: 10.1016/j.ymssp.2019.06.017.
[19] J. Antoni, 'The Spectral Kurtosis: A Useful Tool for Characterising Non-Stationary Signals', *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 282–307, Feb. 2006, doi: 10.1016/j.ymssp.2004.09.001.
[20] J. Antoni and R. B. Randall, 'The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines', *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 308–331, Feb. 2006, doi: 10.1016/j.ymssp.2004.09.002.
[21] Y. Lei, J. Lin, Z. He, and Y. Zi, 'Application of an improved kurtogram method for fault diagnosis of rolling element bearings', *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1738–1749, Jul. 2011, doi: 10.1016/j.ymssp.2010.12.011.
[22] F. Gu, X. Tian, Z. Chen, T. Wang, I. Rehab, and A. Ball, 'Fault severity diagnosis of rolling element bearings based on kurtogram and envelope analysis', in *Proceedings of the International conference on advances in civil structural and mechanical engineering*, 2014.
[23] T. Barszcz and A. JabŁoński, 'A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram', *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 431–451, Jan. 2011, doi: 10.1016/j.ymssp.2010.05.018.
[24] J. Antoni, 'The infogram: Entropic evidence of the signature of repetitive transients', *Mechanical Systems and Signal Processing*, vol. 74, pp. 73–94, Jun. 2016, doi: 10.1016/j.ymssp.2015.04.034.
[25] Y. Wang and M. Liang, 'Identification of multiple transient faults based on the adaptive spectral kurtosis method', *Journal of Sound and Vibration*, vol. 331, no. 2, pp. 470–486, Jan. 2012, doi: 10.1016/j.jsv.2011.08.029.
[26] H. Li, X. Lian, C. Guo, and P. Zhao, 'Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination', *J Intell Manuf*, vol. 26, no. 1, pp. 189–198, Apr. 2013, doi: 10.1007/s10845-013-0772-8.
[27] S. Chatterton, P. Borghesani, P. Pennacchi, and A. Vania, 'Optimal Frequency Band Selection for the Square Envelope Spectrum in the Diagnostics of Rolling Element Bearings', p. V008T11A003, Aug. 2014, doi: 10.1115/DETC2014-35088.
[28] D. Wang, P. W. Tse, and K. L. Tsui, 'An enhanced Kurtogram method for fault diagnosis of rolling element bearings', *Mechanical Systems and Signal Processing*, vol. 35, no. 1, pp. 176–199, Feb. 2013, doi: 10.1016/j.ymssp.2012.10.003.
[29] Y. C. Kim and E. J. Powers, 'Digital Bispectral Analysis and Its Applications to Nonlinear Wave Interactions', *IEEE Transactions on Plasma Science*, vol. 7, no. 2, pp. 120–131, Jun. 1979, doi: 10.1109/TPS.1979.4317207.
[30] I. M. Howard, 'Higher-order spectral techniques for machine vibration condition monitoring', *Proceedings of the IMechE*, vol. 211, no. 4, pp. 211–219, Apr. 1997, doi: 10.1243/0954410971532622.
[31] L. Saidi, 'The deterministic bispectrum of coupled harmonic random signals and its application to rotor faults diagnosis considering noise immunity', *Applied Acoustics*, vol. 122, pp. 72–87, Jul. 2017, doi: 10.1016/j.apacoust.2017.02.007.
[32] F. Gu, Y. Shao, N. Hu, B. Fazenda, and A. Ball, 'Motor current signal analysis using a modified bispectrum for machine fault diagnosis', in *2009 ICCAS-SICE*, 2009, pp. 4890–4895.
[33] X. Tian, F. Gu, I. Rehab, G. Abdalla, and A. Ball, 'A robust fault detection method of rolling bearings using modulation signal bispectrum analysis', presented at the 28th International

Congress of Condition Monitoring and Diagnostic Engineering Management (COMADEM 2015), Buenos Aires, Argentina, 2015.
[34] X. Tian, J. Xi Gu, I. Rehab, G. M. Abdalla, F. Gu, and A. D. Ball, 'A robust detector for rolling element bearing condition monitoring based on the modulation signal bispectrum and its performance evaluation against the Kurtogram', *Mechanical Systems and Signal Processing*, vol. 100, pp. 167–187, Feb. 2018, doi: 10.1016/j.ymssp.2017.07.037.
[35] I. Rehab, X. Tian, F. Gu, and A. Ball, 'The fault detection and severity diagnosis of rolling element bearings using modulation signal bispectrum', presented at the Eleventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 2014.
[36] B. Huang, G. Feng, X. Tang, J. X. Gu, G. Xu, R. Cattley, F. Gu, and A. D. Ball, 'A performance evaluation of two bispectrum analysis methods applied to electrical current signals for monitoring induction motor-driven systems', *Energies*, vol. 12, no. 8, p. 1438, Jan. 2019, doi: 10.3390/en12081438.
[37] I. Antoniadis and G. Glossiotis, 'Cyclostationary analysis of rolling-element bearing vibration signals', *Journal of Sound and Vibration*, vol. 248, no. 5, pp. 829–845, Dec. 2001, doi: 10.1006/jsvi.2001.3815.
[38] R. B. Randall, J. Antoni, and S. Chobsaard, 'The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals', *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 945–962, Sep. 2001, doi: 10.1006/mssp.2001.1415.
[39] J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui, 'Cyclostationary modelling of rotating machine vibration signals', *Mechanical Systems and Signal Processing*, vol. 18, no. 6, pp. 1285–1314, Nov. 2004, doi: 10.1016/S0888-3270(03)00088-8.
[40] J. Antoni, 'Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions', *Journal of Sound and Vibration*, vol. 304, no. 3, pp. 497–529, Jul. 2007, doi: 10.1016/j.jsv.2007.02.029.
[41] J. Antoni, 'Cyclic spectral analysis in practice', *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 597–630, Feb. 2007, doi: 10.1016/j.ymssp.2006.08.007.
[42] A. Raad, J. Antoni, and M. Sidahmed, 'Indicators of cyclostationarity: Theory and application to gear fault monitoring', *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 574–587, Apr. 2008, doi: 10.1016/j.ymssp.2007.09.011.
[43] J. Antoni, 'Cyclostationarity by examples', *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 987–1036, May 2009, doi: 10.1016/j.ymssp.2008.10.010.
[44] D. Wang, X. Zhao, L.-L. Kou, Y. Qin, Y. Zhao, and K.-L. Tsui, 'A simple and fast guideline for generating enhanced/squared envelope spectra from spectral coherence for bearing fault diagnosis', *Mechanical Systems and Signal Processing*, vol. 122, pp. 754–768, May 2019, doi: 10.1016/j.ymssp.2018.12.055.
[45] J. Antoni, G. Xin, and N. Hamzaoui, 'Fast computation of the spectral correlation', *Mechanical Systems and Signal Processing*, vol. 92, pp. 248–277, Aug. 2017, doi: 10.1016/j.ymssp.2017.01.011.
[46] L. A. Zadeh, 'From circuit theory to system theory', *Proceedings of the IRE*, vol. 50, no. 5, pp. 856–865, 1962.
[47] P. Eykhoff, *System Identification: Parameter and State Estimation*. London ; New York etc.: Wiley-Blackwell, 1974.
[48] L. Ljung, 'Convergence analysis of parametric identification methods', *IEEE Transactions on Automatic Control*, vol. 23, no. 5, pp. 770–783, Oct. 1978, doi: 10.1109/TAC.1978.1101840.
[49] P. K.-J. Åström and D. T. Bohlin, 'Numerical Identification of Linear Dynamic Systems from Normal Operating Records', in *Theory of Self-Adaptive Control Systems*, P. H. Hammond, Ed. Springer US, 1966, pp. 96–111.

[50] K. J. Aström and P. Eykhoff, ‘System identification-A Survey’, *Automatica*, vol. 7, no. 2, pp. 123–162, Mar. 1971, doi: 10.1016/0005-1098(71)90059-8.
[51] L. Ljung, *System Identification Theory for the User*. Prentice Hall, 1987.
[52] J. Schoukens and R. Pintelon, *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*, 1 edition. Pergamon, 1991.
[53] B. L. K. Ho, ‘Effective construction of linear state-variable models from input/output functions.’, *REGELUNGSTECHNIK*, Jan. 1966.
[54] H. Akaike, ‘Stochastic theory of minimal realization’, *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 667–674, 1974, doi: 10.1109/TAC.1974.1100707.
[55] H. Akaike, ‘Markovian Representation of Stochastic Processes by Canonical Variables’, *SIAM Journal on Control*, vol. 13, no. 1, pp. 162–173, 1975, doi: 10.1137/0313010.
[56] M. Viberg, ‘Subspace-based methods for the identification of linear time-invariant systems’, *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995, doi: 10.1016/0005-1098(95)00107-5.
[57] V. Overchee and B. Moor, *Subspace identification for linear systems*. Kluwer Academic Publishers, Dordrecht, Holland, 1996.
[58] D. Bauer, ‘Asymptotic Properties of Subspace Estimators’, *Automatica*, vol. 41, no. 3, pp. 359–376, Mar. 2005, doi: 10.1016/j.automatica.2004.11.012.
[59] E. Reynders, ‘System Identification Methods for (Operational) Modal Analysis: Review and Comparison’, *Arch Computat Methods Eng*, vol. 19, no. 1, pp. 51–124, Feb. 2012, doi: 10.1007/s11831-012-9069-x.
[60] M. Gevers, ‘A Personal View of the Development of System Identification: A 30-Year Journey Through an Exciting Field’, *IEEE Control Systems*, vol. 26, no. 6, pp. 93–105, 2006.
[61] P. Ibáñez, ‘Identification of dynamic parameters of linear and non-linear structural models from experimental data’, *Nuclear Engineering and Design*, vol. 25, no. 1, pp. 30–41, 1973, doi: 10.1016/0029-5493(73)90059-9.
[62] S. F. Masri and T. K. Caughey, ‘A Nonparametric Identification Technique for Nonlinear Dynamic Problems’, *J. Appl. Mech*, vol. 46, no. 2, pp. 433–447, 1979, doi: 10.1115/1.3424568.
[63] G. Kerschen, K. Worden, A. F. Vakakis, and J.-C. Golinval, ‘Past, present and future of nonlinear system identification in structural dynamics’, *Mechanical Systems and Signal Processing*, vol. 20, no. 3, pp. 505–592, Apr. 2006, doi: 10.1016/j.ymssp.2005.04.008.
[64] T. K. Caughey, ‘Equivalent Linearization Techniques’, *The Journal of the Acoustical Society of America*, vol. 35, no. 11, pp. 1706–1711, Nov. 1963, doi: 10.1121/1.1918794.
[65] T. K. Caughey, ‘Response of Van der Pol’s oscillator to random excitation’, *Journal of Applied Mechanics*, vol. 26, no. 3, pp. 345–348, 1959.
[66] T. K. Caughey, ‘Random Excitation of a System With Bilinear Hysteresis’, *J. Appl. Mech*, vol. 27, no. 4, pp. 649–652, Dec. 1960, doi: 10.1115/1.3644077.
[67] W. D. Iwan, ‘A generalization of the concept of equivalent linearization’, *International Journal of Non-Linear Mechanics*, vol. 8, no. 3, pp. 279–287, 1973, doi: 10.1016/0020-7462(73)90049-8.
[68] W. D. Iwan and A. B. Mason, ‘Equivalent linearization for systems subjected to non-stationary random excitation’, *International Journal of Non-Linear Mechanics*, vol. 15, no. 2, pp. 71–82, 1980, doi: 10.1016/0020-7462(80)90001-3.
[69] H. J. Rice, ‘Identification of weakly non-linear systems using equivalent linearization’, *Journal of Sound and Vibration*, vol. 185, no. 3, pp. 473–481, 1995, doi: 10.1006/jsvi.1995.0393.
[70] R. Bouc, ‘The Power Spectral Density Of Response For A Strongly Non-linear Random Oscillator’, *Journal of Sound and Vibration*, vol. 175, no. 3, pp. 317–331, 1994, doi: 10.1006/jsvi.1994.1331.

[71] C. Soize, ‘Stochastic linearization method with random parameters and power spectral density calculation’, presented at the Sixth International Conference on Structural Safety and Reliability, ICOSSAR’93, 1993, vol. 1, p. Pages: 217-222.
[72] C. soize, ‘Stochastic linearization method with random parameters for SDOF nonlinear dynamical systems: prediction and identification procedures’, *Probabilistic Engineering Mechanics*, vol. 10, no. 3, pp. 143–152, 1995, doi: 10.1016/0266-8920(95)00011-M.
[73] S. Bellizzi, R. Bouc, M. Defilippi, and P. Guihot, ‘Response Spectral Densities and Identification of a Randomly Excited Non-Linear Squeeze Film Oscillator’, *Mechanical Systems and Signal Processing*, vol. 12, no. 5, pp. 693–711, 1998, doi: 10.1006/mssp.1998.0170.
[74] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th Revised edition edition. Hoboken, N.J: Wiley-Blackwell, 2008.
[75] I. J. LEONTARITIS and S. A. BILLINGS, ‘Input-output parametric models for non-linear systems Part I: deterministic non-linear systems’, *International Journal of Control*, vol. 41, no. 2, pp. 303–328, 1985, doi: 10.1080/0020718508961129.
[76] I. J. LEONTARITIS and S. A. BILLINGS, ‘Input-output parametric models for non-linear systems Part II: stochastic non-linear systems’, *International Journal of Control*, vol. 41, no. 2, pp. 329–344, 1985, doi: 10.1080/0020718508961130.
[77] S. L. Lacy and D. S. Bernstein, ‘Subspace identification for non-linear systems with measured-input non-linearities’, *International Journal of Control*, vol. 78, no. 12, pp. 906–926, 2005, doi: 10.1080/00207170500214095.
[78] J. S. Bendat, *Nonlinear System Techniques and Applications*, 2 edition. New York: Wiley-Interscience, 1998.
[79] K. Yasuda, S. Kawamura, and K. Watanabe, ‘Identification of nonlinear multi-degree-of-freedom systems. 1st Report. Presentation of an identification technique.’, *ResearchGate*, vol. 53, no. 495, pp. 2191–2198, Nov. 1987, doi: 10.1299/kikaic.53.2191.
[80] K. Yasuda, S. Kawamura, and K. Watanabe, ‘Identification of nonlinear multi-degree-of-freedom systems. Identification under noisy measurements.’, *ResearchGate*, vol. 54, no. 500, pp. 816–824, Apr. 1988, doi: 10.1299/kikaic.54.816.
[81] H. J. Rice and J. A. Fitzpatrick, ‘A generalised technique for spectral analysis of non-linear systems’, *Mechanical Systems and Signal Processing*, vol. 2, no. 2, pp. 195–207, 1988, doi: 10.1016/0888-3270(88)90043-X.
[82] H. J. Rice and J. A. Fitzpatrick, ‘A procedure for the identification of linear and non-linear multi-degree-of-freedom systems’, *Journal of Sound and Vibration*, vol. 149, no. 3, pp. 397–411, 1991, doi: 10.1016/0022-460X(91)90444-O.
[83] D. E. ADAMS and R. J. ALLEMANG, ‘A Frequency Domain Method for Estimating the Parameters of a Non-Linear Structural Dynamic Model Through Feedback’, *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 637–656, 2000, doi: 10.1006/mssp.2000.1292.
[84] M. Haroon, D. E. Adams, Y. W. Luk, and A. A. Ferri, ‘A time and frequency domain approach for identifying nonlinear mechanical system models in the absence of an input measurement’, *Journal of Sound and Vibration*, vol. 283, no. 3–5, pp. 1137–1155, May 2005, doi: 10.1016/j.jsv.2004.06.008.
[85] W. J. Staszewski, ‘Analysis of non-linear systems using wavelets’, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 214, no. 11, pp. 1339–1353, Nov. 2000, doi: 10.1243/0954406001523317.
[86] R. M. Rosenberg, ‘On Nonlinear Vibrations of Systems with Many Degrees of Freedom’, in *Advances in Applied Mechanics*, vol. 9, H. L. D. G.G. Chernyi P. Germain, L. Howarth, W. Olszak, W. Prager, R. F. Probstein and H. Ziegler, Ed. Elsevier, 1966, pp. 155–242.
[87] R. M. Rosenberg, ‘The Normal Modes of Nonlinear n-Degree-of-Freedom Systems’, *J. Appl. Mech*, vol. 29, no. 1, pp. 7–14, 1962, doi: 10.1115/1.3636501.

[88] S. BELLIZZI, P. GUILLEMAIN, and R. KRONLAND-MARTINET, ‘Identification of Coupled Non-Linear Modes from Free Vibration Using Time-Frequency Representations’, *Journal of Sound and Vibration*, vol. 243, no. 2, pp. 191–213, 2001, doi: 10.1006/jsvi.2000.3407.
[89] W. Su, H. Yang, and Z. Liu, ‘System identification based on orthogonal polynomial regression analysis method’, in *Proceedings of the 29th Chinese Control Conference*, 2010, pp. 1441–1446.
[90] O. Giustolisi and D. A. Savic, ‘A symbolic data-driven technique based on evolutionary polynomial regression’, *Journal of Hydroinformatics*, vol. 8, no. 3, pp. 207–222, Jul. 2006, doi: 10.2166/hydro.2006.020.
[91] S. R. Chu, R. Shoureshi, and M. Tenorio, ‘Neural networks for system identification’, *IEEE Control Systems Magazine*, vol. 10, no. 3, pp. 31–35, 1990.
[92] J. Qiao, W. Li, X.-J. Zeng, and H. Han, ‘Identification of fuzzy neural networks by forward recursive input-output clustering and accurate similarity analysis’, *Applied Soft Computing*, vol. 49, pp. 524–543, 2016, doi: 10.1016/j.asoc.2016.08.009.
[93] T. A. Tutunji, ‘Parametric system identification using neural networks’, *Applied Soft Computing*, vol. 47, pp. 251–261, 2016, doi: 10.1016/j.asoc.2016.05.012.
[94] A. Yazdizadeh and K. Khorasani, ‘Identification of a class of nonlinear systems using dynamic neural network structures’, in , *International Conference on Neural Networks,1997*, 1997, vol. 1, pp. 194–198 vol.1, doi: 10.1109/ICNN.1997.611663.
[95] L. Dou, R. Ji, and J. Gao, ‘Identification of nonlinear aeroelastic system using fuzzy wavelet neural network’, *Neurocomputing*, 2016, doi: 10.1016/j.neucom.2016.07.021.
[96] X. Jiang, S. Mahadevan, and Y. Yuan, ‘Fuzzy stochastic neural network model for structural system identification’, *Mechanical Systems and Signal Processing*, vol. 82, pp. 394–411, 2017, doi: 10.1016/j.ymssp.2016.05.030.
[97] M. Scionti and J. P. Lanslots, ‘Stabilisation diagrams: Pole identification using fuzzy clustering techniques’, *Advances in Engineering Software*, vol. 36, no. 11–12, pp. 768–779, Nov. 2005, doi: 10.1016/j.advengsoft.2005.03.029.
[98] Y. Zhang, H. Zuo, and F. Bai, ‘Classification of fault location and performance degradation of a roller bearing’, *Measurement*, vol. 46, no. 3, pp. 1178–1189, Apr. 2013, doi: 10.1016/j.measurement.2012.11.025.
[99] X. Zhang, Y. Liang, J. Zhou, and Y. zang, ‘A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM’, *Measurement*, vol. 69, no. Supplement C, pp. 164–179, Jun. 2015, doi: 10.1016/j.measurement.2015.03.017.
[100] L. Saidi, J. Ben Ali, and F. Fnaiech, ‘Application of higher order spectral features and support vector machines for bearing faults classification’, *ISA Transactions*, vol. 54, no. Supplement C, pp. 193–206, Jan. 2015, doi: 10.1016/j.isatra.2014.08.007.
[101] X. Wei, L. Jia, and H. Liu, ‘A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements’, *Vehicle System Dynamics*, vol. 51, no. 5, pp. 700–720, May 2013, doi: 10.1080/00423114.2013.767464.
[102] M. Jesussek and K. Ellermann, ‘Fault detection and isolation for a full-scale railway vehicle suspension with multiple Kalman filters’, *Vehicle System Dynamics*, vol. 52, no. 12, pp. 1695–1715, Dec. 2014, doi: 10.1080/00423114.2014.959026.
[103] X. Y. Liu, S. Alfi, and S. Bruni, ‘An efficient recursive least square-based condition monitoring approach for a rail vehicle suspension system’, *Vehicle System Dynamics*, vol. 54, no. 6, pp. 814–830, Jun. 2016, doi: 10.1080/00423114.2016.1164869.
[104] F. Liu, H. Zhang, X. He, Y. Zhao, F. Gu, and A. D. Ball, ‘Correlation signal subset-based stochastic subspace identification for an online identification of railway vehicle suspension systems’, *Vehicle System Dynamics*, vol. 0, no. 0, pp. 1–21, Mar. 2019, doi: 10.1080/00423114.2019.1589534.

[105] W. Wang, ‘Early detection of gear tooth cracking using the resonance demodulation technique’, *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 887–903, Sep. 2001, doi: 10.1006/mssp.2001.1416.
[106] T. Wang, F. Chu, Q. Han, and Y. Kong, ‘Compound faults detection in gearbox via meshing resonance and spectral kurtosis methods’, *Journal of Sound and Vibration*, vol. 392, no. Supplement C, pp. 367–381, Mar. 2017, doi: 10.1016/j.jsv.2016.12.041.
[107] T. Wang, F. Chu, and Q. Han, ‘Fault diagnosis for wind turbine planetary ring gear via a meshing resonance based filtering algorithm’, *ISA Transactions*, vol. 67, pp. 173–182, Mar. 2017, doi: 10.1016/j.isatra.2016.11.008.
[108] S. Tong, Y. Huang, Y. Jiang, Y. Weng, Z. Tong, N. Tang, and F. Cong, ‘The identification of gearbox vibration using the meshing impacts based demodulation technique’, *Journal of Sound and Vibration*, vol. 461, p. 114879, Nov. 2019, doi: 10.1016/j.jsv.2019.114879.
[109] ‘Resonance’, *Wikipedia*. 08-Dec-2019.
[110] R. B. Randall, J. Antoni, and S. Chobsaard, ‘A comparison of cyclostationary and envelope analysis in the diagnostics of rolling element bearings’, in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, 2000, vol. 6, pp. 3882–3885 vol.6, doi: 10.1109/ICASSP.2000.860251.
[111] Y.-F. Wang and P. J. Kootsookos, ‘Modeling of Low Shaft Speed Bearing Faults for Condition Monitoring’, *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 415–426, May 1998, doi: 10.1006/mssp.1997.0149.
[112] I. A. M. Rehab, ‘The optimization of vibration data analysis for the detection and diagnosis of incipient faults in roller bearings’, doctoral, University of Huddersfield, 2016.
[113] P. Gupta and M. K. Pradhan, ‘Fault detection analysis in rolling element bearing: A review’, *Materials Today: Proceedings*, vol. 4, no. 2, pp. 2085–2094, Jan. 2017, doi: 10.1016/j.matpr.2017.02.054.
[114] R. Brincker and C. Ventura, *Introduction to Operational Modal Analysis*, 1st ed. John Wiley and Sons, 2015.
[115] P. Van Overschee and B. L. De Moor, *Subspace Identification for Linear Systems*. Springer Science & Business Media, 2012.
[116] B. Peeters, J. Lau, J. Lanslot, and H. van der Auweraer, ‘Automatic Modal Analysis-Myth or Reality?’, *Sound and Vibration*, vol. 42, no. 3, p. 17, 2008.
[117] G. Y. LUO, D. OSYPIW, and M. IRLE, ‘Real-Time Condition Monitoring by Significant and Natural Frequencies Analysis of Vibration Signal with Wavelet Filter and Autocorrelation Enhancement’, *Journal of Sound and Vibration*, vol. 236, no. 3, pp. 413–430, Sep. 2000, doi: 10.1006/jsvi.2000.2996.
[118] K. F. Al-Raheem, A. Roy, K. P. Ramachandran, D. K. Harrison, and S. Grainger, ‘Rolling element bearing faults diagnosis based on autocorrelation of optimized: wavelet de-noising technique’, *Int J Adv Manuf Technol*, vol. 40, no. 3–4, pp. 393–402, Jan. 2009, doi: 10.1007/s00170-007-1330-3.
[119] J. Rafiee and P. W. Tse, ‘Use of autocorrelation of wavelet coefficients for fault diagnosis’, *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1554–1572, Jul. 2009, doi: 10.1016/j.ymssp.2009.02.008.
[120] J. M. Ha, B. D. Youn, H. Oh, B. Han, Y. Jung, and J. Park, ‘Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines’, *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 161–175, Mar. 2016, doi: 10.1016/j.ymssp.2015.09.040.
[121] J. Singh, A. K. Darpe, and S. P. Singh, ‘Rolling element bearing fault diagnosis based on over-complete rational dilation wavelet transform and auto-correlation of analytic energy operator’, *Mechanical Systems and Signal Processing*, vol. 100, no. Supplement C, pp. 662–693, Feb. 2018, doi: 10.1016/j.ymssp.2017.06.040.

[122] F. Gu, T. Wang, A. Alwodai, X. Tian, Y. Shao, and A. D. Ball, 'A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals', *Mechanical Systems and Signal Processing*, vol. 50–51, pp. 400–413, Jan. 2015, doi: 10.1016/j.ymssp.2014.05.017.
[123] R. Zhang, F. Gu, H. Mansaf, T. Wang, and A. D. Ball, 'Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis', *Mechanical Systems and Signal Processing*, vol. 94, pp. 202–213, Sep. 2017, doi: 10.1016/j.ymssp.2017.02.037.
[124] F. Gu, Y. Shao, N. Hu, A. Naid, and A. D. Ball, 'Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment', *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 360–372, Jan. 2011, doi: 10.1016/j.ymssp.2010.07.004.
[125] E. Bechhoefer and M. Kingsley, 'A Review of Time Synchronous Average Algorithms', in *Annual conference of the prognostics and health management society*, 2009, vol. 1, pp. 1–10.
[126] F. Combet and L. Gelman, 'An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor', *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2590–2606, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.006.
[127] Y. Hu, X. Tu, F. Li, H. Li, and G. Meng, 'An adaptive and tacholess order analysis method based on enhanced empirical wavelet transform for fault detection of bearings with varying speeds', *Journal of Sound and Vibration*, vol. 409, pp. 241–255, Nov. 2017, doi: 10.1016/j.jsv.2017.08.003.
[128] Y. Wang, G. Xu, A. Luo, L. Liang, and K. Jiang, 'An online tacholess order tracking technique based on generalized demodulation for rolling bearing fault detection', *Journal of Sound and Vibration*, vol. 367, pp. 233–249, Apr. 2016, doi: 10.1016/j.jsv.2015.12.041.
[129] H. Cao, D. He, S. Xi, and X. Chen, 'Vibration signal correction of unbalanced rotor due to angular speed fluctuation', *Mechanical Systems and Signal Processing*, vol. 107, pp. 202–220, Jul. 2018, doi: 10.1016/j.ymssp.2018.01.030.
[130] S. Lu, Q. He, and J. Zhao, 'Bearing fault diagnosis of a permanent magnet synchronous motor via a fast and online order analysis method in an embedded system', *Mechanical Systems and Signal Processing*, vol. 113, pp. 36–49, Dec. 2018, doi: 10.1016/j.ymssp.2017.02.046.
[131] G. He, K. Ding, W. Li, and X. Jiao, 'A novel order tracking method for wind turbine planetary gearbox vibration analysis based on discrete spectrum correction technique', *Renewable Energy*, vol. 87, pp. 364–375, Mar. 2016, doi: 10.1016/j.renene.2015.10.036.
[132] A. C. Singer and M. Feder, 'Universal linear prediction by model order weighting', *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2685–2699, Oct. 1999, doi: 10.1109/78.790651.
[133] W. Wang and A. K. Wong, 'Autoregressive model-based gear fault diagnosis', *J. Vib. Acoust*, vol. 124, no. 2, pp. 172–179, Mar. 2002, doi: 10.1115/1.1456905.
[134] B. Li and X. Zhang, 'A new strategy of instantaneous angular speed extraction and its application to multistage gearbox fault diagnosis', *Journal of Sound and Vibration*, vol. 396, pp. 340–355, May 2017, doi: 10.1016/j.jsv.2017.02.043.
[135] S. Olhede and A. T. Walden, 'A generalized demodulation approach to time-frequency projections for multicomponent signals', *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 461, no. 2059, pp. 2159–2179, Jul. 2005, doi: 10.1098/rspa.2005.1455.
[136] K. Rodopoulos, C. Yiakopoulos, and I. Antoniadis, 'A parametric approach for the estimation of the instantaneous speed of rotating machinery', *Mechanical Systems and Signal Processing*, vol. 44, no. 1–2, pp. 31–46, Feb. 2014, doi: 10.1016/j.ymssp.2013.02.011.