# MACHINE TOOL SPINDLE BEARING DIAGNOSTICS UNDER OPERATING CONDITIONS

A Dissertation
Presented to
The Academic Faculty

by

Aoyu Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mechanical Engineering

Georgia Institute of Technology
May 2017

# MACHINE TOOL SPINDLE BEARING DIAGNOSTICS UNDER OPERATING CONDITIONS

Approved by:


Dr. Thomas Kurfess, Advisor
School of Mechanical Engineering
*Georgia Institute of Technology*

Dr. Christopher Saldana
School of Mechanical Engineering
*Georgia Institute of Technology*

Dr. Steven Danyluk
School of Mechanical Engineering
*Georgia Institute of Technology*

Dr. Nagi Gebraeel
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Dr. Richard Cowan
School of Mechanical Engineering
*Georgia Institute of Technology*


Date Approved: March 14, 2017

# ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Thomas Kurfess for being encouraging and patient in the past four years. Without his support and guidance, I could not have improved as much in both research and life. I would also like to thank my parents for their forever support and endless love. I am also deeply grateful to all my committee members, Dr. Steven Danyluk, Dr. Richard Cowan, Dr. Christopher Saldana and Dr. Nagi Gebraeel, for helping me improve the quality of my work. Special thanks are also due to Mr. Lejun Cen, Mr. Vinh Nguyen, Mr. Svyatoslav Yorish, Mr. Brandon Royal, Mr. Nathaniel Mauldin and Mr. Steven Sheffield for their help and contribution to this project. I would also like to thank all my friends and labmates for making my life at Georgia Tech wonderful.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# NOMENCLATURE

| | |
|---|---|
| $\alpha$ | Balancing parameter of VMD |
| $\alpha_0$ | Contact angle of the angular contact ball bearing |
| $a_0$ | Intersection area between ball and inner race under initial contact force |
| $a_{min}$ | Minimum value of the acceleration entry signal |
| $a_e$ | Absolute value of the local minimum in the entry signal |
| $a_l$ | Peak-to-peak value of the low frequency noise in the entry signal |
| $a_n$ | Peak-to-peak value of the high frequency noise in the entry signal |
| b | Intersection area between ball and outer race |
| $b_0$ | Intersection area between ball and outer race under initial contact force |
| $c_1$ | Damping ratio of the shaft |
| $c_2$ | Damping ratio of the housing |
| d | Defect width |
| $d_m$ | Pitch diameter |
| D | Ball diameter |
| E | Young’s modulus of the bearing material |

| | |
|---|---|
| $f_{bpor}$ | Ball-pass outer race frequency |
| $f_{bpir}$ | Ball-pass inner race frequency |
| $f_r$ | Ball frequency |
| $f_s$ | Sampling frequency |
| $f_{tft}$ | Fundamental train frequency (cage frequency) |
| $F$ | Contact force between ball and both races |
| $F_0$ | Initial contact force between ball and both races |
| $F_{12}$ | Resultant force of the other 12 balls |
| $k_1$ | Stiffness of the shaft |
| $k_2$ | Stiffness of the outer race |
| $k_3$ | Load-deformation constant |
| $K$ | Number of modes in VMD |
| $m_1$ | Equivalent mass of the shaft and inner race |
| $m_2$ | Equivalent mass of the housing and outer race |
| $\eta$ | Angle between two adjacent balls |
| $n$ | Spindle speed in RPM |
| $N_i$ | Fundamental frequency (shaft frequency) |

$\theta$ Ball center angular position relative to defect entry edge

$\theta_0$ Initial ball center angular position relative to defect entry edge

$\theta_1$ Initial ball center angular position relative to defect entry edge

$\theta_2$ Ball center angular position relative to defect entry edge when contact force decreases to zero

$\theta_d$ Central angle of the defect

$\phi$ Normalized spindle speed with respect to 1000 rpm

r Ball diameter

R Distance from inner race center to ball center without deflection

$R_1$ Inner race radius

$R_2$ Ball radius

$R_3$ Distance between inner race center to defect entry edge

$R_i$ Inner race groove diameter

$R_o$ Outer race groove diameter

$SNR_h$ Signal-to-noise ratio of the entry signal with respect to the high frequency noise

$SNR_l$ Signal-to-noise ratio of the entry signal with respect to the low frequency noise

$t_e$ Entry to peak time

$t_p$ Peak to impact time

$t_t$ Total time from entry to impact

$\nu$ Poisson's ratio of the bearing material

$v_b$ Ball center velocity

$\omega_c$ Cage angular velocity (Ball angular velocity)

$\Delta x_0$ Total deflection of ball under initial contact force

$x_1$ Displacement of the inner race

$x_2$ Displacement of the outer race

$x_b$ Distance from inner race center to ball center with deflection

$x_{cross}$ Zero-crossing point index of the fitted line in the velocity function

$x_m$ Local minimum index of the acceleration function

$Z$ Number of balls

# SUMMARY

Bearing diagnostics provide valuable information related to a bearing's health and facilitate Condition Based Maintenance (CBM) for rotary machines. This is an effective method to decrease unnecessary cost and downtime resulting from unanticipated machine spindle failure. Defective signatures can be extracted from the corresponding vibration signals through both the time and frequency domain signal processing procedures. However, techniques to effectively evaluate bearing damage severity from these extracted features are still a significant challenge, because the relationship between the bearing damage severity and the extracted feature is not well understood. Moreover, previous methods are mostly tested under constant loading conditions, and are not suitable for bearing diagnostics during machining operations. In this thesis, a time-domain-based bearing defect size estimation method is proposed for the inner and outer race defects. This new approach is built on the bearing system nonlinear dynamic model and the Hertzian contact defect size estimation model. The new defect size estimation model is independent of the contact force between ball and raceway, and all the required information for defect estimation can be obtained from the vibration signal. The signal processing method is developed to automatically extract the time information from the vibration signal for defect size estimation. Statistical analysis is performed on the time information and the results support the proposed bearing models. A test system designed for CNC-based bearing diagnostics was fabricated to validate the new method. Experiments in the speed range 500-3000 rpm were performed under both no-cutting and cutting conditions with different feed rates. Experiment results are consistent under different operational conditions

(speeds/feeds), and they are agreeable to both the bearing system dynamic model and the defect estimation model. The estimation results are close to the true defect size with relative error of approximately 10%, validating the proposed method.

# CHAPTER 1. INTRODUCTION

## 1.1 Problem statement

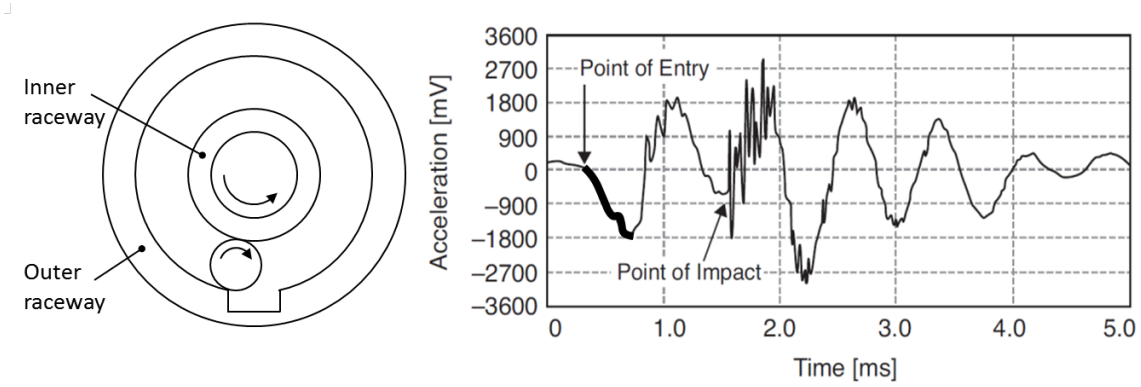
Rolling Element Bearings (REBs) are critical components in rotary machinery. However, bearing failure consists of a large percentage of machine shutdowns and can result in significant repair costs [1, 2]. Therefore, bearing diagnostics have become a topic of interest since the 1960's due to its potential application towards Condition Based Maintenance (CBM) for rotary machines, which is an effective tool to decrease loss when unanticipated bearing issues occur [3-5]. Though many types of bearing defects exist, they can be categorized as either distributed or localized. The defect's gradual increase in size gives rise to heat generated by friction between bearing's rolling elements and raceways. The defect eventually causes spindle seizures and damage the rotary machinery. A large number of techniques to detect bearing damage before permanent damage occurs have been developed by monitoring phenomena corresponding to bearing defects, including temperature monitoring, debris measurement, vibration and acoustic analysis [6-10]. Among these methods, vibration based monitoring is the most widely employed technique, due to its straightforward interpretable features, easily processed signals and simplicity in measurement [1]. For the purpose of CBM, the most critical task is to estimate the bearing damage severity according to the vibration signals.

Techniques to effectively evaluate bearing damage severity are still a significant challenge. Previous research using time or frequency domain analysis have demonstrated the extraction and recognition of fault signatures from strong background noise [11, 12], but the relationship between the bearing damage severity and the extracted feature is not

well understood [13]. Based on these signal processing approaches, other researchers have estimated damage severity by correlating the change of fault signatures using run to failure tests, also known as Trend Analysis [14, 15]. Unfortunately, Trend Analysis is difficult to implement in practical environments because a large amount of historical vibration data is required for comparison with currently recorded data [13]. Moreover, this statistical method is sensitive to variations in operating conditions and, therefore, does not provide reliable information on the progression of damage, especially for localized defects. In addition, Trend Analysis is time-consuming and a universal damage standard is impractical to implement among different bearing systems. Therefore, simply extracting the fault signature and trending its development over time is not sufficient for the purpose of developing an optimal CBM schedule [16].

A better solution than Trend Analysis is Absolute Analysis. Absolute Analysis involves direct calculation of the bearing defect size using time series information with a physics based model. Thus, Absolute Analysis is desirable because the vibration signal is directly related to damage severity and no baseline data are needed. Bearing manufacturers have stated that the standard of bearing damage severity should be represented by the spall size located on bearing raceways and rollers (for example, the Timken Company defines a failure as defect size reaches 6.45 $mm^2$ [17]). Therefore, if the spall size can be calculated, the bearing damage severity can be determined accordingly. Thus, a quantitative failure standard can be implemented in a CBM algorithm. To estimate the spall size, various methods have been demonstrated to work for specific spindle speeds and loads [17] while other methods depend on the RMS value of the vibration signal, which can be an unreliable metric [16]. Other efforts have focused on the localized spall-like defect, which is known

to be a typical bearing failure mode. Researchers have proposed that when a rolling element passes a spall on the raceways, a repeatable response pattern corresponding to entry and exit events arises in the vibration signal. Therefore, the time separation between these two events can be used to estimate the size of the defect [18-21], as shown in Figure 1. Using this method, the absolute value of the defect size can be directly estimated and the damage severity can be determined from the estimation. Therefore, baseline data are not required in Absolute Analysis. This method is a promising candidate to quantify the severity of a spall-like damage for a CBM algorithm. However, the nature of the vibration response signal can vary across different test platforms. Therefore, a deeper understanding of the processes corresponding to the generation and transmission of vibrations within the bearing system is necessary. Another difficulty of the absolute time domain method is accurately identifying the entry and exit points from a noisy vibration signal. Because the entry event is a continuous destressing process and, can be difficult to identify, especially due to the existence of noise; thus, the time domain estimation for a constant spall defect can be incorrectly identified due to variations in spindle speed [12, 22]. For the exit point, previous preferred methods are threshold sensitive, and therefore difficult to calculate a reliable threshold value for automatic diagnostics. To address these aforementioned issues, a dynamic shaft-bearing-housing model is required to interpret the vibration response due to a spall-like defect. In addition, a more reliable signal processing method is needed to extract the entry and exit points for the automatic bearing diagnostics. Therefore, one of the major tasks of this thesis is to perform absolute bearing diagnostics by solving the problems described previously.




(a) (b)

Figure 1. Typical bearing defective response due to a localized defect [13].

More importantly, an absolute method can facilitate real-time spindle diagnostics, which is critical during machining operations including milling, turning and drilling. Performing bearing diagnostics during machining can yield various benefits for an industrial manufacturer, as serious bearing damage or severe defect growth during machining operations can be monitored in real-time and catastrophic results can be avoided. Such recognition is particularly critical for machining processes with higher spindle speeds and cutting forces than typical rotary bearing operations. In addition, the machine tool spindle typically runs under operating conditions that maximize production rate, but bearing diagnostics typically occur when the machine is not cutting. Altering the machine's schedule to facilitate diagnostic when the spindle is spinning but not cutting reduces productivity to unacceptable levels. In addition, limited data exist during non-machining operations for effective diagnostics. Therefore, conducting diagnostic testing during actual machining is desirable for optimizing production rate while simultaneously collecting data to improve bearing diagnostic results. However, such diagnostics are not possible for Trend Analysis due to its sensitivity to operating conditions. Bearing loads change frequently in machining, and therefore comparison between the current data and

baseline data is impractical. Thus, diagnostics in prior literature are performed when the spindle is purely spinning. In addition, most of the previous work use vibration data collected from bearing test rigs, rather than actual machine tools. Due to the aforementioned limitations, bearing diagnostics during machining operations is lacking. Thus, another important task of this thesis is to investigate the performance of absolute diagnostic methods during machining operations. However, spindle diagnostics during machining can be difficult to perform. For experimental test rigs, artificial defects can be seeded on the bearings and a constant external radial load can be applied. However, these conditions are not applicable on machine tool spindle bearing systems, because frequent disassembly and reassembly of the spindle may result in the failure of the machine tool. Therefore, disassembling the spindle structure and seeding a defect on the bearings is impractical. In addition, the excessive vibration noise caused by the cutting forces during machining further increase the complexity of extracting precise defect size from the vibration signal.

In conclusion, this thesis focuses on developing a time-domain spall size tracking method and signal processing algorithm for bearing diagnostics during machining operations. Even though this new Absolute Analysis is able to directly estimate spall size, it is recommended as an augment to the Trend Analysis rather than a replacement. Trend Analysis has been demonstrated for distributed bearing defect and can defect bearing defect in incipient stage. When the bearing entry and exit signal are detected, the Absolute Analysis can be applied to further improve the diagnostic results.

## 1.2 Research objectives

According to the previously described discussions, the objectives of this research include: 1) Fabricate a new bearing test system with the capability of performing machining operations, 2) Develop a novel shaft-bearing-housing system dynamic model to describe the defective signature of the vibration signal due to a spall-like artificial defect seeded on the bearing races, 3) Estimate bearing spall size based on the defective signature when the spindle is running at various spindle speeds, 4) Develop a signal processing method and algorithm to estimate bearing spall size during machining operations.

## 1.3 Thesis outline

In this thesis, Chapter 2 introduces the comprehensive background and literature review of bearing diagnostics closely related to the research of this thesis. The topics covered include typical bearing failure modes, bearing test stands and sensors, bearing system dynamic models, signal processing methods, and application in production operations.

In Chapter 3, the bearing test system to estimate the size of a spall-like defect on bearing raceways is described in detail. Because artificial defects cannot be seeded inside the original machine spindle, an external set of bearings is mounted to the current machine spindle. Line-spall defects of various widths can be seeded separately on the inner and outer races of the add-on bearings using Electronic Discharge Machining (EDM). An accelerometer and a dynamometer with a DAQ system are used to collect the vibration and force signals in the experiment. Based on vibration signals at different spindle speeds and widths, repeatable defective features can be recognized and extracted. These features are used to estimate defect size (width).

To understand the mechanism of the defective feature, Chapter 4 describes a dynamic model of the bearing test system. Then, three defect size estimation models are developed to describe the relationship between the vibration signal and the defect size.

Chapter 5 describes an empirical entry signal model that is implemented in a new signal processing method to extract the time information from the vibration signal. The performance of the proposed signal processing method is analyzed.

Chapter 6 describes the evaluation of the signal processing method developed in Chapter 5 when the spindle is purely spinning. Statistical properties of the vibration feature are analyzed and the estimation method is validated with experiments.

To perform bearing diagnostics during cutting operations, Chapter 7 describes the application of the signal processing method from Chapter 5 to extract the defective feature corrupted by the vibration due to cutting forces. Cutting experiments under varying operating conditions are performed for validation.

Chapter 8 summarizes the conclusions and main contributions from this thesis and suggest recommendations for future work.

# CHAPTER 2. BACKGROUND AND LITERATURE REVIEW

## 2.1 Introduction

Background and previous work addressed in this section include: 1) typical bearing failure modes, 2) sensors and bearing test systems, 3) modeling of bearing vibration due to a localized spall defect, 4) signal processing methods and damage severity estimation, and 5) application of advanced bearing diagnostic approaches in machine tools.

## 2.2 Bearing failure modes

In some applications, REBs can operate under improper conditions, such as extremely heavy or complex loading, high speed or accelerations, significantly high or low temperatures, insufficient lubrication, and improper machinery assembly. Even under normal operating conditions, bearings suffer from material imperfection and fatigue failure. The previously mentioned faults result in earlier bearing failure than the designed bearing lifetime under normal operating conditions [23]. Bearing damage can be categorized into distributed and localized defects.

### *2.2.1 Distributed bearing defect*

Distributed defects are bearing damage in which “the magnitude of the ball-race contact force varies continuously and periodically as the bearing rotates” [24]. This type of defect exhibits a relatively large damage area and shallow depth.

Reasons for the distributed defects include thermal unbalance-induced overheating, excessive thrust loading or heavy radial preloading, corrosion pitting and oxidation of

rolling contact surfaces due to moisture in the lubricant, misalignment of bearing components and indentations caused by hard particle contaminants. Some researchers conclude that the distributed bearing defect produce sinusoidal vibrations. In addition, the frequency and amplitude of the sinusoidal signal have been investigated [25]. However, when distributed defects occur, excessive vibration and heat already exist in the spindle and the quality of the product has been compromised. Therefore, it is better to detect bearing defect in a more incipient failure stage as introduced in the ensuing section.

### *2.2.2 Localized bearing defect*

Localized defects are small isolated areas of damages on the bearing raceways or rolling elements [13], also known as the spall. Different from the distributed defects, localized defects have sharp discontinuities and larger depth in the rolling surface that produce impulsive type vibrations [26].

Brinnelling can cause this type of defect, which is defined as plastic deformation caused by either sudden impact loading during bearing operation or heavy loading while the bearing is not rotating [23]. A typical situation when brinnelling occurs is during a spindle crash. The result is a small dent on a ball bearing or a line spall for roller bearings left on the surface of the bearing raceway. In either case, when the rolling element passes by the dent, extremely high-pressure ridges form in the vicinity of the dented region. Thus, a spall will start to form at the ridge on the trailing edge of that dent. The spall will shorten the bearing lifetime and may finally lead to rapid bearing failure.

Another cause of a localized defect is the subsurface-initiated spall. This phenomenon is unavoidable even for an accurately manufactured and well-lubricated

properly mounted bearing, because subsurface material quality is difficult to measure after the bearings have been manufactured. Figure 2 shows the formation of a subsurface-initiated spall on the outer raceway. In this formation, a hard oxide point exists in the subsurface layer of the material as the imperfection. When the bearing operates under normal operational conditions with cyclic Hertzian stresses, subsurface material fatigue occurs. This fatigue promotes crack nucleation and growth, and eventually subsurface cracks propagate to the surface. The material loosened by cracking can then break-away resulting in a raceway discontinuity as a spall [27].

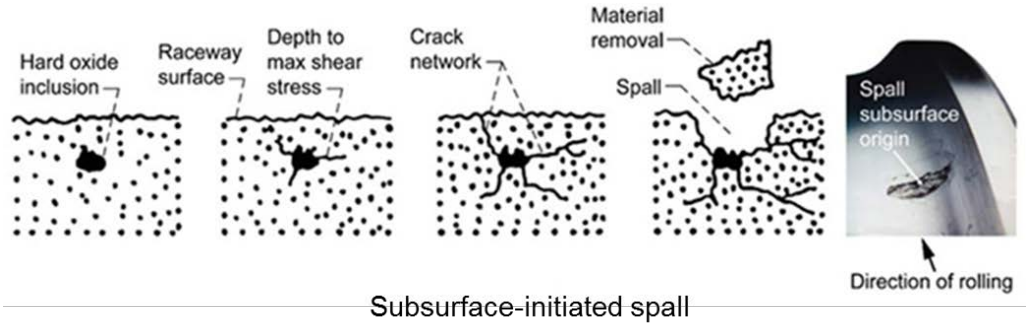



Figure 2. Formation of a spall on bearing raceway.

The spall is not considered as a catastrophic failure but its existence indicates the incipient failure. After some time, the spall may propagate in area and depth and cause excessive vibrations or heat that may lead to bearing seize. Thus, detecting bearing damage at this stage is critical and a majority of bearing diagnostics approaches have been focusing on this type of defect. Therefore, the scope of this thesis is focused on the localized defect.

## 2.3 Sensors and bearing test systems

Since bearing diagnostics during machining operations has not been the focus of previous research, typical bearing test systems usually consist of sensors attached outside the bearing test rig close to the defective test bearing with/without artificially initiated defects. The sensors and test rigs used for typical bearing test will be introduced in this section.

### *2.3.1 Sensors*

Vibrations and acoustic signals have been used by previous researchers to diagnose localized bearing defect. Therefore, accelerometers and acoustic emission (AE) sensors have been identified as two potential candidates for this research.

#### 2.3.1.1 Piezoelectric accelerometer

A piezoelectric accelerometer is a sensor that transfers mechanical vibrations (acceleration) into electric charge. These electric charges are amplified and the output is a voltage signal proportional to its acceleration. Piezo accelerometers have wide bandwidth, high sensitivity and large measurement range with minimal noise and non-linearity. Because of these characteristics, some researchers have used piezo accelerometer to estimate the size of an artificially generated line-spall on bearing raceways [28]. However, manufacturing cost of the piezoelectric accelerometer is relatively high and usually an external charge amplifier is needed for signal conditioning. Therefore, the overall cost for the piezo accelerometer can be a limiting factor in production implementation.

#### 2.3.1.2 MEMS accelerometer

Accelerometers can be fabricated as micro electro-mechanical systems (MEMS). A MEMS accelerometer transfers mechanical vibrations (acceleration) into the change of the capacitance between the moveable plates and the fixed plates. A simple circuit can then be used to process the signal and output a voltage proportional to its acceleration. Recently, MEMS accelerometers have become a viable sensing system compared to a piezo accelerometer due to its cheaper price, simplicity of circuitry, and increasing product quality. However, due to the non-linearity and noise of MEMS accelerometers, they still cannot measure the size of a localized bearing defect, because the vibration signal can be distorted and masked. As MEMS technology develops, the MEMS accelerometer can possibly be used to estimate the spall size.

#### 2.3.1.3 Acoustic emission sensor

Acoustic Emission (AE) sensors are typically piezoelectric sensors with elements made of special ceramic. They are sensitive to high frequency (100 kHz to 1 MHz) elastic stress waves that propagate through the sensor. The local dynamic material displacement caused by the elastic stress waves can be converted to an electrical signal. Therefore, each time when the rolling element passes a spall, the elastic stress of the bearing raceway changes according to the relative position and a corresponding electric signal can be obtained. Using this phenomenon, Mba [29] measured the burst duration of the acoustic signal to approximate defect size. However, the defect estimation result can be 20%-60% larger than the true defect size. As suggested by the author, this is attributed to the decay characteristics of the AE transient bursts. The start of the burst is earlier than the ball entry

and the end of the burst is later than the ball exit. In addition, the burst to noise ratio depends on the defect size. When the defect size becomes smaller, the burst to noise ratio decreases significantly. Therefore, this method has limited performance for defect size smaller than 2mm. Hemmati [30] used the duration between two peaks in the acoustic signal to estimate defect size (the first peak is related to entry event and the second peak is related to exit event). However, according to Mba [31], the existence of the peaks are inconsistent due to a protrusion generated during artificial defect seeding process. Thus, AE sensors cannot provide a reliable and practical method for defect size estimation.

According to the above discussions, the characteristics of these different sensors are compared and summarized in Table 1. As the result, piezo accelerometer is selected as the optimal sensor for this research.

Table 1. Comparison of different sensors.

| Sensor | Linearity | Sensitivity /Bandwidth | RMS Noise | Complexity of Circuitry | Signal Interpretation | Cost |
|---|---|---|---|---|---|---|
| Piezo | Good | Good | Low | Complex | Simple | High |
| MEMS | Bad | Bad | High | Simple | Simple | Low |
| AE | Good | Fair | Fair | Complex | Complex | Fair |

### *2.3.2 Bearing test rigs*

A typical bearing test rig is shown in Figure 3. Two or three sets of bearings are mounted on the shaft and one of the bearing sets is the tested bearing seeded with artificial defects. The bearings outer raceway is fixed by a housing that encloses the core of the test

system. Constant loading in radial and thrust directions can be applied to the loading structure. During the bearing test, the sensor is attached to the housing close to the tested bearing and the shaft rotates at a specific speed driven by a servo motor.

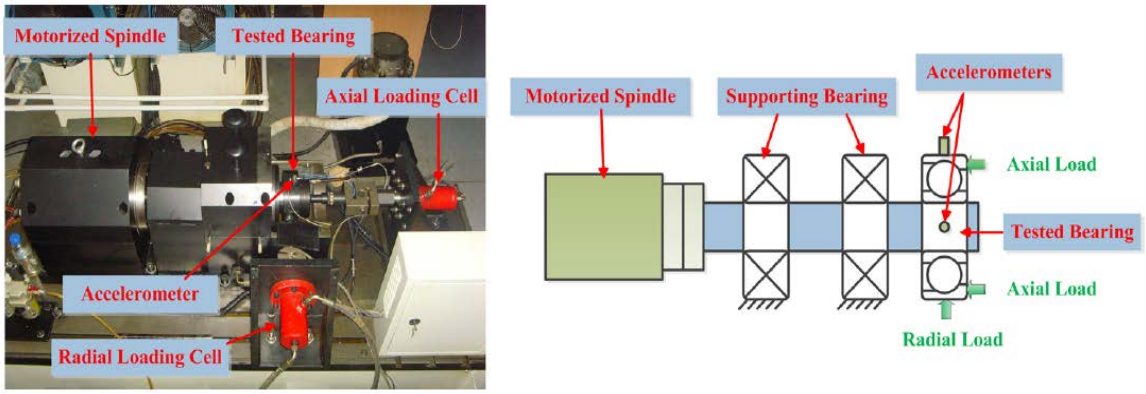



Figure 3. A typical bearing test rig [19]

The design of the test rig facilitates replacement of test bearings and applications under specific constant thrust or radial loading. The first feature is very useful, because usually in the bearing test, replacement of bearings is frequent. The constant loading is excellent to compare bearing test results under different bearing loads, but this loading condition is quite different from an actual CNC machine tool during machining operations. When the CNC spindle is machining, the loading conditions in the bearing system change frequently due to the cutting forces, which cannot be simulated by the test rig. Therefore, even though prior research has presented bearing test results based on the test rig, these results cannot be directly applied to machine tools.

The ideal solution to validate bearing fault detection is to perform tests directly within a machine tool's spindle. This enables the ability to perform bearing diagnostics during machining operations. In addition, results from bearings that operate under

conditions similar to real industrial environments are more applicable in manufacturing. However, a challenge of this setup is the seeding of the defect in the spindle bearing and replacing the bearings when necessary. Frequent disassembly and reassembly of the spindle may result in the failure of the machine tool to satisfy industrial precision requirements. Therefore, opening the spindle structure and seeding a defect on the bearing is impractical. As a compromised solution, the only way is to combine the test rig with a CNC machine tool. A new design with advantages from both the test rig and the CNC machine tool is proposed in Chapter 3.

## 2.4 Modeling of bearing vibration

Researchers have proposed that a consistent relationship exists between the defective response and the defect size [18, 26, 28]. However, before the defect size can be precisely estimated, dynamic modeling of the bearing system is necessary. This is because: 1) the process when a rolling element passes a localized spall on the raceway has not been fully understood, 2) difficulties exist in determining the point of entry, and 3) bearing defective responses can differ from platform to platform due to different system dynamics. Therefore, modeling of bearing vibration response due to raceway defects is an important element for both quality inspection and bearing condition monitoring [32]. Using a dynamic model, bearing system responses under different operating conditions (*e.g.*, spindle speeds/loads) and with varying defect sizes can be simulated. Research of bearing modeling are summarized in the next several paragraphs.

### *2.4.1 Bearing nonlinear stiffness and damping model*

Bearing modeling initially starts with describing smaller defects. Bearing nonlinear stiffness and damping have been modeled based on Hertzian contact theory and dynamic system modeling [25, 33, 34]. Given the loading condition, the deflection of the bearing components can be determined and the contact forces can be derived. A similar model for thrust loading condition is developed by F. Wardle [25] and the damping in this system has been investigated by T. Walford [34]. Then, a two degree of freedom system can be solved. Using this model, vibration due to bearing waviness has been simulated and satisfactory results were achieved when compared with experimental data. However, this model assumes that the ball and raceways are always loaded, which may not be the case for a localized spall defect [35]. Even though this model cannot be directly used for a spall-like defect, the basic theory used in this model can be applied to the discontinuous case.

### *2.4.2 The bearing impulse train model*

McFadden and J.D. Smith have developed a bearing impulse train model for single and multiple spall bearings [36, 37]. Randall [38] improved this model by introducing non-stationary random slip between rolling elements and raceways. In these models, the impact force between rolling element and bearing raceways due to the spall is simplified as a square or triangular pulse [20, 39-42]. The amplitudes of these pulses are assumed proportional to the bearing loads. Even though the pulse train model can precisely reveal frequency components including the corresponding localized defects, shaft unbalance and frequency modulation, the defect propagation cannot be predicted reliably, because the contact force between the ball and raceways is over simplified [20] and the entry event is

not modeled. Therefore, this method cannot be used to estimate defect size unless the contact force can be modeled precisely.

### *2.4.3 The bearing dynamic model for localized defect*

Other researchers have employed multi-body dynamic models. The first non-linear multi-body dynamic bearing model was derived by Harsha in 2005 [43]. Randall improved this model by considering the slip between bearing components and applied this model towards estimation of the bearing response due to a localized defect [44]. However, since the entry event was not considered in the signal processing method, speed dependent errors appeared in the defect size estimation result. Moazen and Ahmadi [32, 35] further improved this model by considering the finite size of the rolling elements, and the results demonstrated an increased accuracy when compared to previous models. However, the effects of shaft speed and bearing load conditions, which significantly affect bearing dynamic response during the entry event, were not fully investigated. In addition, even though the angular contact ball bearing is the most frequently used bearing type in machine spindles, most of the previous models are built for roller bearings or deep groove ball bearings. Therefore, the dynamic model for the angular contact ball bearing is necessary for enhancing spindle diagnostic capabilities.

Epps [26] has suggested that future work for localized bearing defect diagnostics lies in developing signal processing method to extract the useful information for defect size estimation. The following section comprehensively introduces the general signal processing methods for bearing diagnostics. Then, the signal processing methods for localized spall is addressed.

## 2.5 Signal processing methods and damage severity estimation

Great efforts have been focused on extracting the weak defective feature from the strong background noises to detect the existence of a local bearing defect. The extracted features are then used in Trend Analysis approaches to evaluate damage severity. The mainstream signal processing methods developed before 1990's are summarized as traditional methods in Section 2.5.1. Signal processing methods used to identify the entry and exit events for defect size estimation are introduced in Section 2.5.2.

### *2.5.1 Traditional signal processing methods*

Bearing defects exhibit specific features in both the time and frequency domain. These features are extracted from the vibration signals via a variety of signal processing methods. Based on the assumption that increasing damage level yields more dominant defect features, these features can be used to estimate bearing damage severity. The traditional signal processing methods are categorized into time domain, frequency domain, and time-frequency domain analysis.

#### 2.5.1.1 Time domain analysis

The earliest time domain technique known as "shock pulse counting" was proposed by Gustaffs and Tallian in 1962. Other researchers compared statistical moments, such as Root Mean Square (RMS), mean, kurtosis and crest factor with thresholds to determine damage severity [11, 16].

RMS

RMS represents the average power of a zero-centered signal, which is defined as:

$$RMS = \sqrt{\frac{\sum x_i^2}{N}} \tag{2.1}$$

Where N is the number of data points and $x_i$ represents the signal value at the corresponding index. As bearing defects propagate, the RMS value should increase. By comparing the RMS value of the current data and the baseline data, the damage severity can be correlated and analyzed.

Skewness and Kurtosis

Skewness and kurtosis are the third and fourth order statistical moments defined in Eq (2.2) and (2.3):

$$Skewness = \frac{\sum(x_i - \bar{x})^3}{N - 1} \tag{2.2}$$

$$Kurtosis = \frac{(N-1)\sum(x_i - \bar{x})^4}{\left(\sum(x_i - \bar{x})^2\right)^2} \tag{2.3}$$

Where N is the number of data points and $x_i$ represents the signal value at the corresponding index, $\bar{x}$ is the average of all data points. Skewness characterizes the degree of asymmetry of a distribution around its mean. Kurtosis is a statistical measure used to describe the distribution of observed data around the mean. Compared to healthy bearings, damaged bearings are more likely to produce asymmetric statistical distributions and that can be detected by the skewness. Kurtosis is generally normalized by variance and a high quality surface finish corresponds to a theoretical kurtosis of 3 [11]. Thus, kurtosis is a potential absolute method for bearing diagnostics because no baseline data is needed.

Peak Value and Crest Factor

Peak value in time domain is the maximum of the absolute value of the signal. Crest factor is defined as peak value over RMS.

$$Peak\ value = \max|x_i| \tag{2.4}$$

$$Crest\ factor = \frac{\max|x_i|}{RMS} \tag{2.5}$$

Prior work has shown that both parameters increase as defect size increases [45]. However, time domain analysis cannot utilize robust thresholds or are sensitive to noises and disturbances. Therefore, time domain analysis cannot give reliable results when used alone.

2.5.1.2 Frequency domain analysis

Since bearing defects yield well defined frequency based features, frequency domain analysis is a commonly used diagnostic method [17]. Depending on the location of the defect, the impulses excited by the localized defect has specific frequencies named as ball-pass frequencies as shown in Eq (2.6) through Eq (2.9):

$$f_{bpor} = Z\frac{N_i}{2}\left(1 - \frac{D}{d_m}cos\alpha_0\right) \tag{2.6}$$

$$f_{bpir} = Z\frac{N_i}{2}\left(1 + \frac{D}{d_m}cos\alpha_0\right) \tag{2.7}$$

$$f_r = \frac{N_i d_m}{2D}\left[1 - \left(\frac{D}{d_m}cos\alpha_0\right)^2\right] \tag{2.8}$$

$$f_{ftf} = \frac{N_i}{2}\left(1 - \frac{D}{d_m} cos\alpha_0\right) \tag{2.9}$$

Where Z is number of balls, $N_i$ is the shaft frequency, D is the ball diameter, $d_m$ is the pitch diameter, $\alpha_0$ is the contact angle. The ball-pass frequency from Eq (2.6) through Eq (2.9) represents outer race, inner race, ball/roller, and cage frequencies correspondingly. If any of these frequencies appear and increase rapidly in amplitude, the corresponding bearing damage may occur.

High Frequency Resonance Technique (HFRT)

The High Frequency Resonance Technique (HFRT), also known as Demodulated Resonance Analysis or Envelop Detection, is the most widely used frequency domain method to extract the defective signal masked by strong background noise [46, 47]. The essence of this method is to recover the low frequency component excited by the defect from the modulations produced in high frequency vibrations.

A Fast Fourier Transform (FFT) analysis is applied to the "enveloped" signal output by using HFRT. To denoise the signal, the spectrum obtained by HFRT is usually averaged to remove white noise. Cepstrum analysis, which is defined as the spectrum of the spectrum, is used to collect defect related frequencies into one signature to increase signal-to-noise ratio. Other techniques, such as adaptive filtering (Adaptive Line Enhancer), matched filtering, and Wiener filtering, were implemented to further improve HFRT. However, a limitation of HFRT is the ambiguity of the relationship between spectrum amplitude and damage severity. Therefore, HFRT is an appropriate method for detecting the existence of an incipient damage but not for evaluation of damage level.

#### 2.5.1.3 Time-frequency domain analysis

Since the time-frequency domain analysis includes information from both time and frequency domain, it is suitable for the analysis of non-stationary signals related to machinery defects.

Short Time FFT (STFT)

The STFT is a method that decomposes a non-stationary signal into many small segments, which can be assumed to be locally stationary, and applies FFT to each of the segment. The length of each segment is determined by a window function. For each segment of data, a different spectrum can be calculated and by combining all of these spectrums, the time-frequency distribution function, called a Spectrogram, is obtained. Thus, information in both time and frequency domain can be revealed.

However, increasing time resolution decreases frequency resolution and vice-versa. A large window length provides high frequency resolution but poor time resolution while a small window length provides good time resolution but poor frequency resolution. More importantly, since displaying of multi-resolution in time-frequency distribution diagram is an important requirement, STFT has the limitation of possessing the same resolution at all locations in the time-frequency plane due to the same window through the whole signal. This limitation can be conquered by the Wavelet Transform as described in the next section.

Wavelet Analysis

Over the last two decades, great progress has been made in the theory and applications of wavelets in the fields of fault diagnosis [48]. The Wavelet Transform (WT) is a mathematical tool that converts a time domain signal into time-scale domain using a wavelet function. There are two types of WT, continuous and discrete.

The Continuous Wavelet Transform (CWT) is defined in the following equation:

$$X(s,\tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} \mathrm{x(t)}\psi\left(\frac{\mathrm{t}-\tau}{s}\right) dt \qquad (2.10)$$

Where s and τ represent scale and translation, ψ(t) is the wavelet. The CWT is similar to the Fourier Transform except that the base functions sine and cosine are replaced by the wavelet function. The CWT of a signal will produce redundancy as s and τ are changing all the time. The redundancy requires a significant amount of computation time.

To save computation time and cost, the Discrete Wavelet Transform (DWT) is preferred since it provides sufficient information with a significant reduction in computation time [49]. DWT can be realized by applying filter banks to the time domain signal to obtain the detailed and approximate signal in each level. However, in each decomposition the time resolution is halved, which is not suitable for analysis with high time resolution requirement.

The Wavelet Packet Transform (WPT) further decomposes the detail information of DWT in the high frequency region. The enhanced signal decomposition capability facilitates identification defective transient components in the bearing vibration signal. In

addition, the WPT can identify frequency shifts associated with the phase transitions of defect growth. Even though a WPT method can provide sufficient information, the technique cannot provide physical reasoning for the machinery condition analysis because of the complexity of most mechanical systems.

In conclusion, even though agreeable results are observed using these traditional methods, the relationship between the damage severity and the defective feature is uncertain. Therefore, a new method that estimates the defect size is desired.

### *2.5.2 Signal processing methods to identify the entry and exit events*

Recently the entry and exit event in the bearing signals due to a spall-like defect has attracted the attention of researchers [43]. Prior work has suggested that the time difference between the entry and exit event can be used to estimate spall size. Therefore, this method corresponds to the time domain analysis.

Kumar [50] and Khanam [21] applied discrete wavelet transform on the vibration signal to eliminate random noise and separate the entry and exit events to estimate the spall size. However, the literature has not reached a consensus on how the distance between the entry and exit edges of the spall is related to the entry and exit events of the signal. Also, when the entry event is masked by strong noise and disturbances, this method may fail. In addition, the effect of spindle speeds/loads is lacking.

Randall proposed a semi-automated algorithm to separate the entry and exit event by applying pre-whitening and the wavelet analysis to the vibration data [18]. Then the power cepstrum is calculated to estimate the average separation of the two pulses. Randall

considered the most dominant peak in the entry signal as the moment when the ball starts to enter the defect zone. However, other work has demonstrated this assumption to be inaccurate for larger defects. Prior work has shown the full entry signal starts earlier than the most dominant peak in the entry signal. Since the time delay can be a significant percentage of the entry-to-exit time, Randall's estimation results were shown to be smaller than the true defect. In addition, this method suffers from a large speed dependent error.

Smith proposed an alternative method to identify the entry point [22]. He demonstrated a consistent roll-off after the entry point in the pre-impact vibration signal and presented two methods to estimate the entry point. In the first method, the zero-crossing point of the low-pass filtered acceleration gradient function is considered as the entry point. However, under the influence of noise, the gradient function is not guaranteed to cross zero at an entry point. Therefore, in addition to difficult automatic implementation, the method is sensitive to noise and threshold. In the second method, the peak location of the low-pass filtered acceleration signal is considered as the entry point. However, this assumption is only correct when the acceleration "hump" in the pre-impact signal coincides with the entry event.

Ahmadi determined the correct position for the entry and exit point in the vibration signal through both observations and dynamic modeling [28, 32]. Ahmadi proposed that Randall's result needed to be compensated by a correction factor, which is a function of bearing maximum load and stiffness. The correction factor increases defect estimation accuracy. However, the bearing load and stiffness is difficult to measure and changes over time. Therefore, a new method independent of the bearing load and stiffness measurement is required.

## 2.6 Application of advanced bearing diagnostic approaches in machine tools

The advanced bearing models and diagnostic approaches previously discussed have been investigated on a bearing test rig, rather than an actual machine tool. Therefore, applications of these advanced bearing diagnostics methods towards real machine tools are limited. The few cases that have been investigated are listed as follows.

Saravanan [51] monitored a lathe spindle bearing system under different spindle speeds, feed rates, and depths of cut. Workpiece surface finish, vibration, and acoustic signal in time and frequency domain were used to indicate damage severity. While this research shows potential for in-process bearing monitoring, the focused was on lubricant pollution rather than the behavior of a localized defect.

Werner [52] developed an early warning spindle bearing failure monitoring system. This system was applied on an Okuma CNC lathe and a vertical machining center. The signal was analyzed in both time and frequency domain to trend the propagation of bearing damage. However, an ultrasonic sensor was used as the primary measurement system, which has been demonstrated as a flawed choice for vibration detection. Also, the relationship between the damage severity and vibration signal was not sufficiently shown. In addition, a reasonable method to select the threshold for failure detection was not given.

Bangcheng Zhang [53] monitored the spindle bearing system of a CNC machining center by enveloping the first layer of wavelet detail data, and then extracting the ball-pass outer race frequency in frequency domain. This research focused on fault signature extraction, and the effect of cutting condition was not fully investigated.

Yuxue Chen [54] developed an online automatic diagnostic procedure to detect scratch defects on bearings. Peak amplitude from the enveloped spectrum and RMS in the time domain were used to quantify damage severity. However, this method was not robust due to susceptibility of the RMS value to noise and sensor location. Also, more defect sizes have to be tested to further validate the effectiveness of this research.

Kozochkin [55] periodically monitored vibration signal of a lathe under no external load over an 8 month period to trend the propagation of bearing damage in different stages. The peak factor value of the time domain signal was used to quantify damage severity and indicate the damage stage of the bearing. However, experimental validation under other operating conditions was not investigated. In these machine tool spindle diagnostic cases, a model is further needed to relate the damage severity to the vibration signals. The effects of spindle speeds/loads on the vibration were not clearly illustrated. Also, diagnostics under machining operations are lacking.

## 2.7 Summary

This chapter introduces the background and literature review of bearing diagnostics including different bearing failure modes, sensors and bearing test systems, modeling of bearing vibration due to a localized spall defect, signal processing methods and damage severity estimation, and application of advanced bearing diagnostic approaches in machine tools. Therefore, the scope of this thesis is localized bearing defect analysis using vibration based method. The test setup is a combination of the traditional test rig and a CNC mill. To obtain a better understanding of the defective signature, a shaft-bearing-housing dynamic model is developed. Based on the experiment results, signal processing methods

are implemented to extract the defective information for defect size estimation during no cutting and cutting conditions.

# CHAPTER 3. EXPERIMENTAL APPARATUS AND TEST EQUIPMENT

## 3.1 System Description

This chapter describes the experimental test system to validate the proposed defect size estimation method. The complete test system consists of a vertical CNC mill machine tool with an add-on bearing housing, sensors for vibration and force measurement, and a data acquisition system including charge amplifiers and Data Acquisition Unit (DAQ). A schematic and pictures of the overall system are shown in Figure 4, Figure 5, and Figure 6. This chapter describes the design, specifications, and capability of each subsystem in detail.

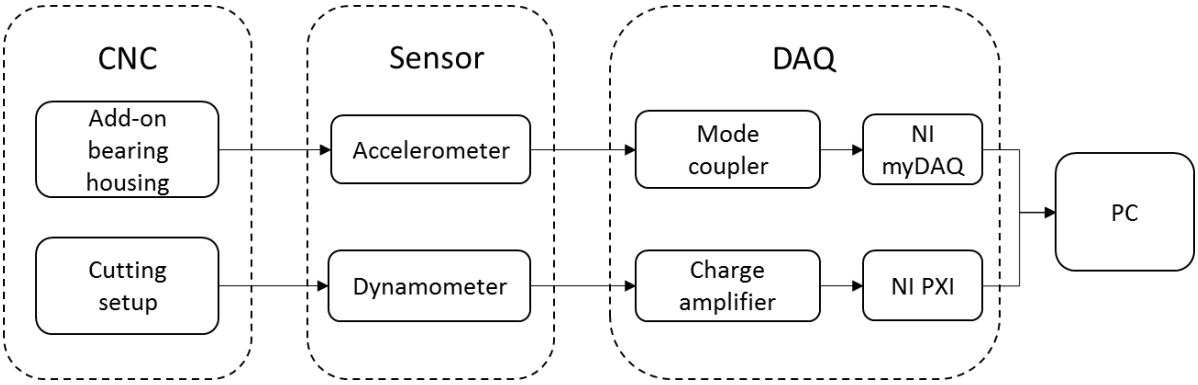



Figure 4. The experimental schematic.

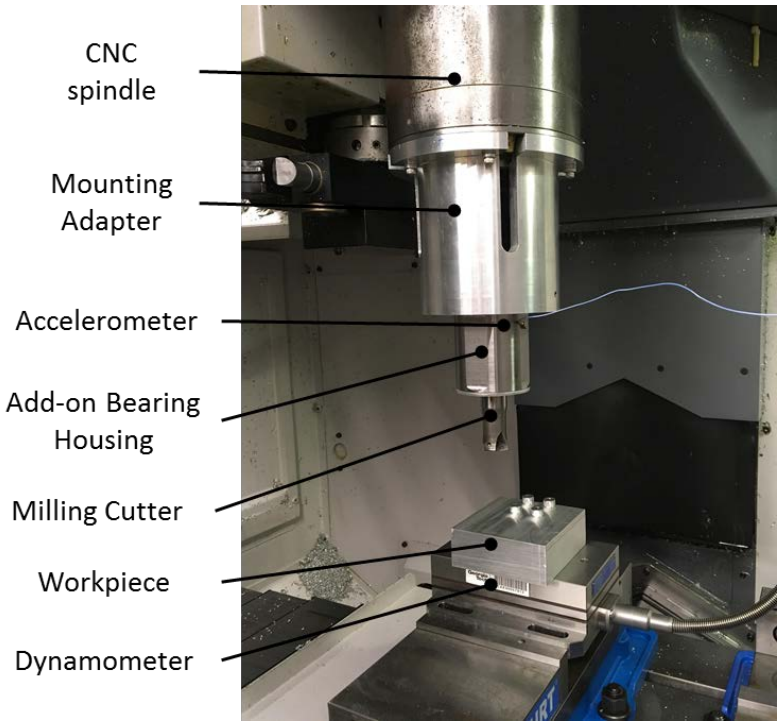



Figure 5. The CNC-based bearing test system and sensors.

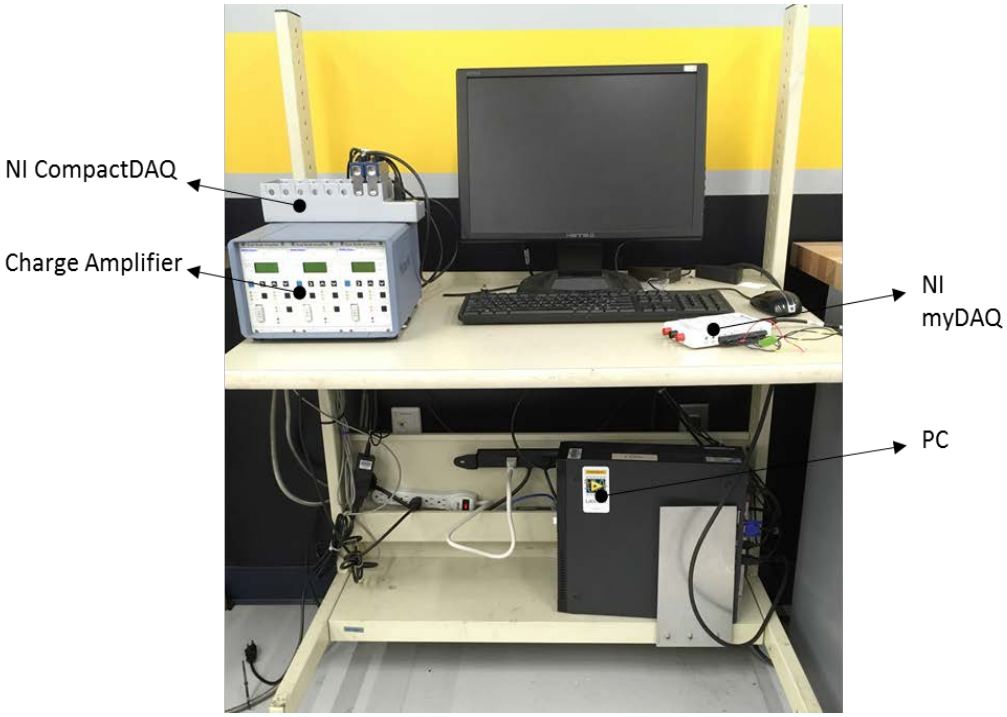



Figure 6. The data acquisition system.

## 3.2 The CNC-based bearing test system

### *3.2.1 The add-on bearing housing*

Because frequent disassembly and reassembly of the spindle may result in the failure of the machine tool, the original spindle of the machine cannot be disassembled and seeded with defects. Therefore, an external bearing housing with the tested angular contact ball bearing is mounted onto the original machine spindle by a mounting adapter as shown in Figure 5. Four flat planes perpendicular to each other are machined on the surface of the add-on housing to provide a robust support for the vibration sensor. The housing was fabricated from steel for possible magnetic mounting. A section view of the setup including the mounting adapter and the add-on housing is shown in Figure 7. The aluminum mounting adapter is fixed to the CNC by six bolts and a setscrew is used to fix the outer race with respect to the inner race. During machining operations, the mounting adapter will increase the stiffness of the add-on structure. The shaft of the add-on bearing system is also the milling cutter. Therefore, the add-on test system can be directly driven by the spindle of the 3-axis Okuma CNC mill. One set of two angular contact ball bearings are installed in the housing. The housing and the outer race are stationary, while the shaft and the inner race are rotating. The top bearing is seeded with defect and the outer race has a transition match for easy replacement. The outer races of the bearing set are supported by the housing shoulders and the inner races of the bearing set are supported by a spacer in between them. The top cap is machined with fine thread and thus the pre-load of the bearing can be adjusted by the screw on the top cap. A section view of the housing without the cutter is shown in Figure 8. This setup provides operating conditions closer to an actual machine tool than the traditional bearing test rigs developed in previous bearing work [12, 56]. In

addition, the new setup has the capabilities to perform bearing diagnostics during machining operations.

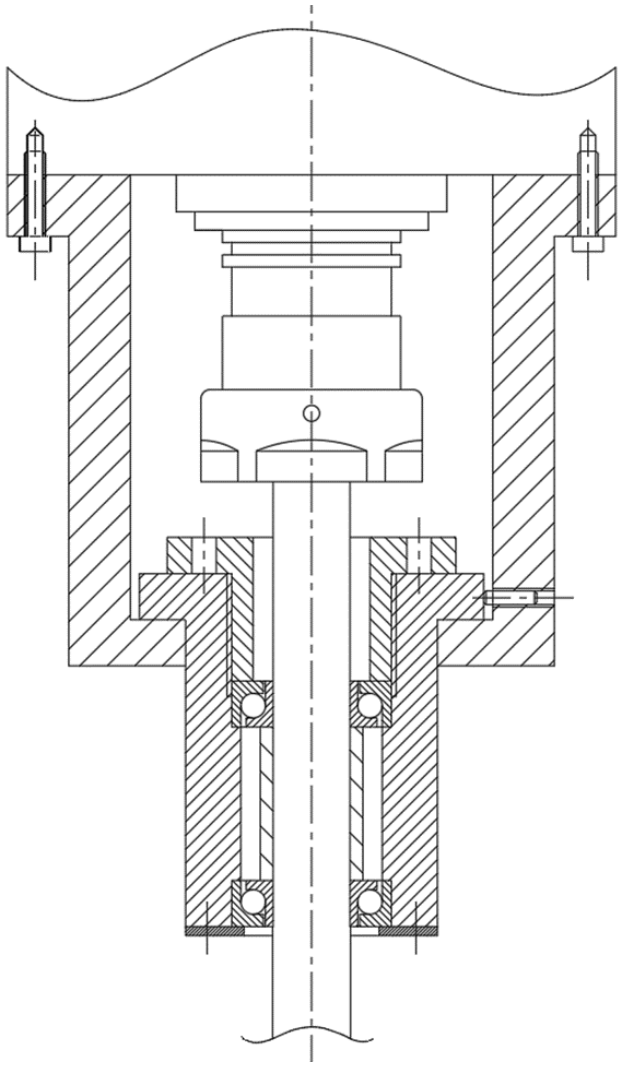


Figure 7. Section view of the bearing test stand.

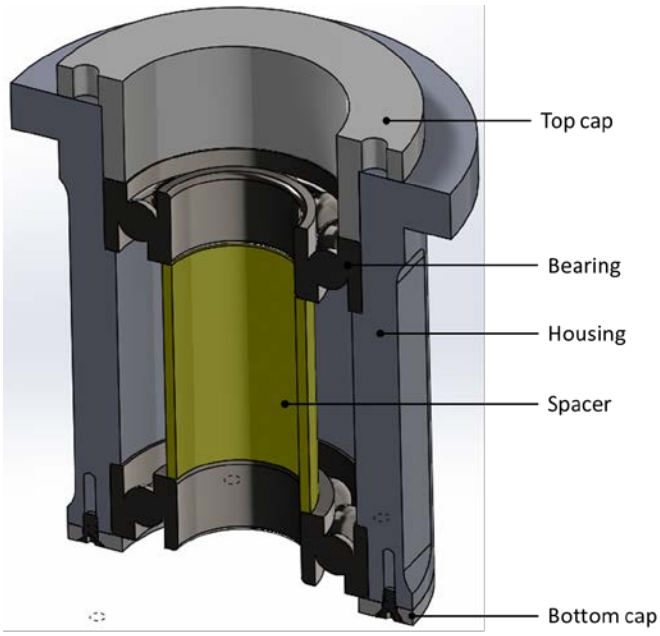



Figure 8. Section view of the add-on bearing housing system.

### *3.2.2 The tested bearings and defect generation*

#### 3.2.2.1 Tested bearings

The tested bearings used in this thesis are SKF 7205BEP single row angular contact ball bearings, as shown in Figure 9. The outer/inner races and the balls are made of bearing steel and the cage is plastic. The bearing geometry and property are listed in Table 2.

Figure 9. The tested bearing.

Table 2. Bearing geometry.

| Geometry and property | SKF 7205BEP angular contact ball bearing |
|---|---|
| Outer Diameter (mm) | 52 |
| Inner Diameter (mm) | 25 |
| Pitch Diameter (mm) | 38.5 |
| Ball Diameter (mm) | 7.9 |
| Width (mm) | 15 |
| Contact Angle | 40 ° |
| Max Speed (rpm) | 15000 |
| Basic Dynamic Load (kN) | 14.8 |
| Basic Static Load (kN) | 9.3 |
| Mass (kg) | 0.13 |

Given the geometry information Table 2, the cage frequency, outer race ball-pass frequency and inner race ball-pass frequency can be calculated using Eq (2.9), (2.6) and (2.7):

$$f_{ftf} = \frac{N_i}{2}\left(1 - \frac{D}{d_m}cos\alpha_0\right) = 0.42N_i \tag{3.1}$$

$$f_{bpor} = Z\frac{N_i}{2}\left(1 - \frac{D}{d_m}cos\alpha_0\right) = 5.46N_i \quad (3.2)$$

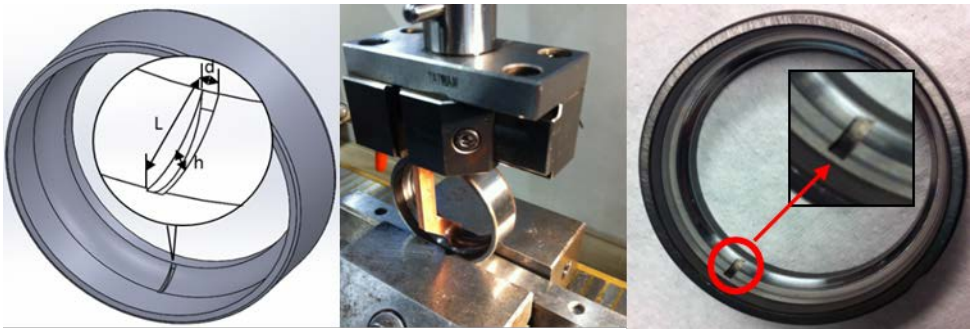
$$f_{bpir} = Z\frac{N_i}{2}\left(1 + \frac{D}{d_m}cos\alpha_0\right) = 7.54N_i \quad (3.3)$$

Eq (3.1) means that the cage frequency is 0.42 times the shaft frequency. In other words, the relative velocity between the cage and the outer race (stationary) is $0.42N_i$. Because the relative velocity between the shaft and the outer race is $N_i$, the relative velocity between the cage and the inner race is $(1\text{-}0.42)N_i=0.58N_i$. Since the ball moves together with the cage, $0.42N_i$ and $0.58N_i$ are also the relative velocities between the ball/outer race and ball/inner race. Because the velocity of the ball will directly affect the defect size estimation, the coefficients 0.42 and 0.58 are very important. From Eq (3.2) and (3.3) it can be seen that 0.42 is the ratio between outer race ball-pass frequency and shaft frequency divided by number of balls, so 0.42 is defined as the outer race ball-pass frequency ratio; 0.58 is the ratio between inner race ball-pass frequency and shaft frequency divided by number of balls, so 0.58 is defined as the inner race ball-pass frequency ratio. Both are validated in the experiments by comparing the theoretical calculation in this chapter and the experimental results.

#### 3.2.2.2 Defect generation

Three different defect widths, 0.794 mm, 1.135 mm and 1.530 mm, are separately seeded on the outer and inner races of the tested bearing. A line-spall defect is shown in Figure 10. The length of the defect along axis direction L is maximized, meaning that L is through the whole groove part of the outer race. The depth of the defect h is considered

large with respect to the width d, and therefore the ball cannot contact the bottom of the defect as it passes the defect zone. Thus, the values of L and h do not affect the defective signature. The primary focus of this research is the defect width, d. The defect is seeded on the bearing races, as shown in Figure 10b), using an Electrical Discharge Machining (EDM) machine with a copper electrode. Figure 10c) shows the 1.530 mm width defect on the outer race.




a) b) c)

Figure 10. a) A line-spall defect on the outer race b) Seed defect with EDM c) 1.530mm outer race defect.

### *3.2.3 The cutting setup*

A two-flute 25.4 mm milling cutter is used in the experiment as the shaft and as well as to perform bearing diagnostics during machining aluminum workpieces. The geometry and the parameters of the cutter are shown in Figure 11 and Table 3. The material property of the aluminum workpieces is also listed in Table 3.

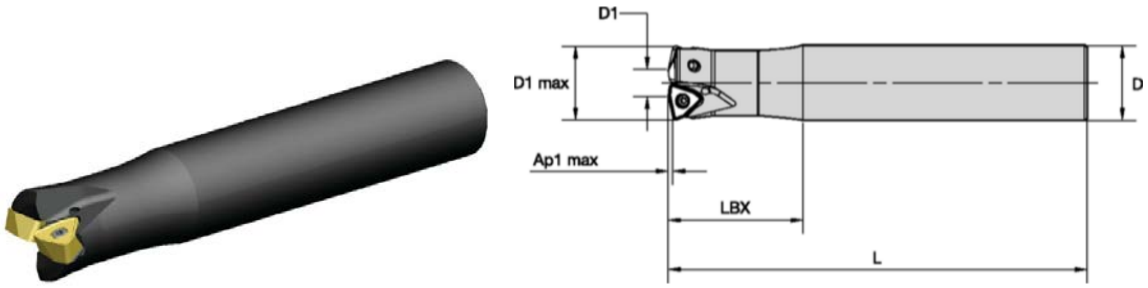



Figure 11. Cutter geometry.

Table 3. Cutter and workpieces parameters.

| | | |
|---|---|---|
| Cutter | L (mm) | 203.2 |
| | Number of Teeth | 2 |
| | D1 max (mm) | 25.4 |
| | Shank Diameter (mm) | 25.4 |
| | Ap1 max (mm) | 1.5 |
| | Max RPM | 36600 |
| | Insert | WOEJ090512 |
| | Mass (kg) | 0.72 |
| Workpieces | Material | Aluminum 6061 |
| | Young's modulus | 70GPa |
| | Poisson's ratio | 0.35 |
| | Density | 2.70g/cm$^3$ |

## 3.3 Sensors

### *3.3.1 Accelerometer*

The PCB 352A21 single-axis accelerometer is used to obtain the vibration signature of the bearing system. The sensitivity of the accelerometer is 10 mV/g. The frequency response at different mounting conditions of the accelerometer are shown in Figure 12. In this research, the accelerometer is waxed to the flat surface of the add-on bearing housing as shown in Figure 5. Therefore, the resonance frequency for the accelerometer is approximately 10 kHz.

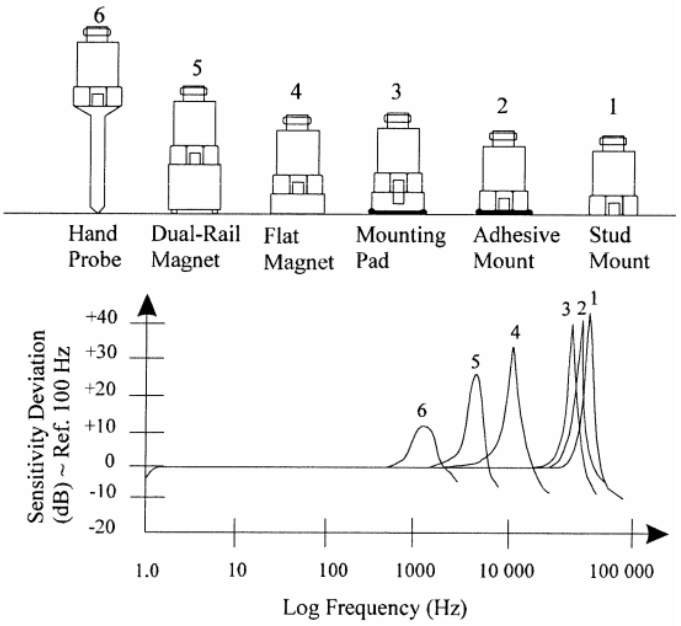



Figure 12. Mounting configurations and their effects on high frequency [57].

### *3.3.2 Dynamometer*

The tri-axis Kistler 9257B dynamometer is used to measure the three orthogonal components of the cutting force during the experiments. The dynamometer has a natural frequency of 3.5 kHz. Its high resolution (3.33 mV/N) enables high sensitivity to smallest dynamic changes in large forces. The dynamometer consists of four three-component force sensors fitted under high preload between a baseplate and a top plate. Each sensor contains three pairs of quartz plates, one sensitive to pressure in the z direction and the other two responding to shear in the x and y directions. The dynamometer is mounted on the machine tool table with clamps as shown in Figure 5.

## 3.4 The data acquisition system

### *3.4.1 Amplifier*

The accelerometer signal was amplified by a Type 5134 Kistler power supply/coupler. The force signal is amplified by a three channel Kistler dual mode amplifier type 5010. The accelerometer and dynamometer signal while the machine is in an idling state are shown in Figure 13 and Figure 14, respectively. As shown, the noise amplitude is minimal when compared to the defective response.

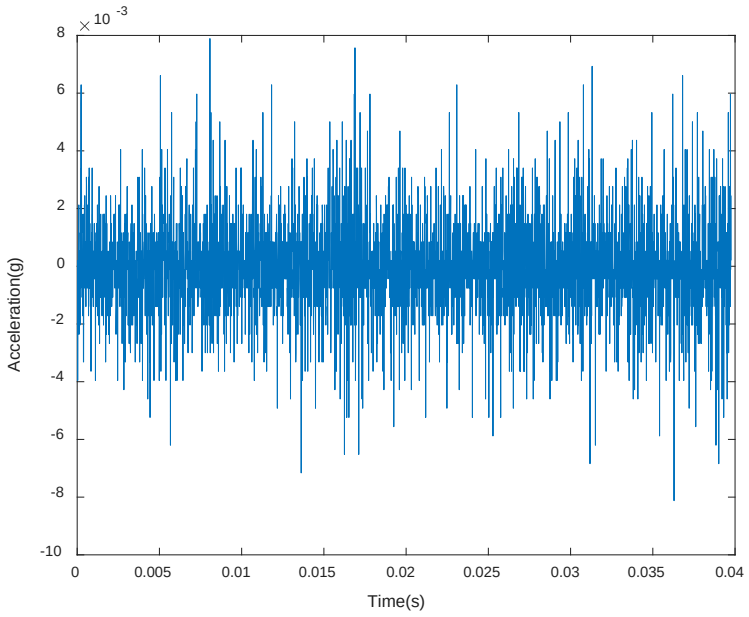



Figure 13. The accelerometer signal during machine idling.

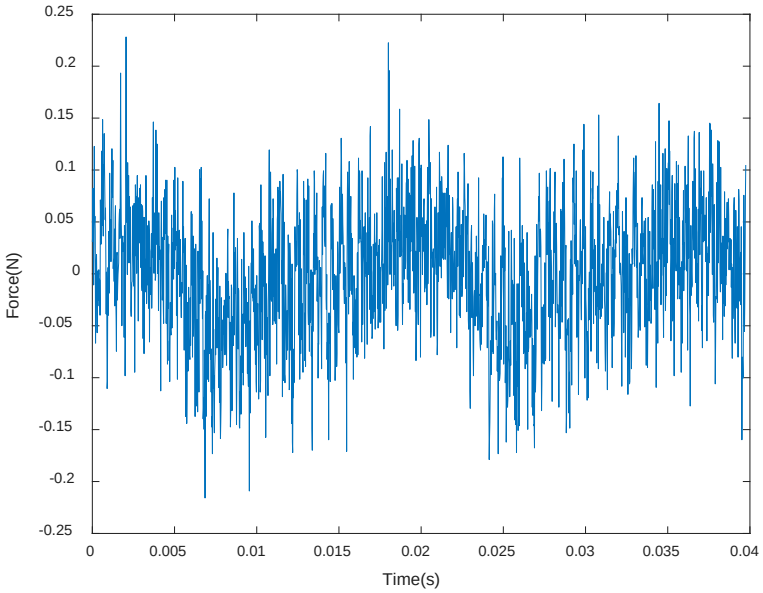



Figure 14. The dynamometer signal during machine idling.

### *3.4.2 Data acquisition*

The forcing signals are collected at 50 kHz sampling rate by the National Instrument data acquisition device NI compactDAQ with two 3-channel C Series dynamic signal acquisition modules. The vibration signal is collected at 100 kHz sampling rate by NI myDAQ. The LabVIEW code was programmed and run on the PC to collect the data.

## 3.5 Summary

This chapter introduces the experimental apparatus and test equipment for the validation of the defect size estimation method proposed in this thesis. The schematic and pictures of the overall system are shown in 3.1. The test system consists of a vertical CNC mill with the add-on bearing housing, sensors for vibration and force measurement, and a

data acquisition system. Each of the subsystem is described in detail from 3.2 through 3.4 including the design and specifications.

# CHAPTER 4. BEARING MODELS

The vibration signals generated from a spall-like defect exhibit a repeatable pattern and can be used to estimate the defect size. However, to calculate an appropriate estimation from the vibration signal, an understanding of the physical process relating the vibration pattern to the defect size is required. Therefore, the first section of this chapter introduces the bearing system dynamic model used to describe the bearing vibration signal due to a spall-like defect. To describe the relationship between the vibration signal and the defect size, the following section provides three models to estimate the defect size based on the vibration signals.

## 4.1 Bearing dynamic model

Recently, prior work has proposed that when a rolling element passes a spall on the raceways, the entry and exit points in the vibration signal can be identified and the time difference between these two points can be used to estimate the size of the defect [18-21]. The proposed technique is a promising method to quantify the severity of a spall-like damage for a CBM program. To obtain a sufficiently accurate estimation result, a dynamic model is proposed for a better understanding of the defective response with respect to the physical process. The proposed model also facilitates recognizing and extracting defective patterns from strong background noises and disturbances. Kinematic analysis is used as the foundation of the bearing dynamic model. Then, the Hertzian contact theory is applied to derive the nonlinear system equations, which are solved numerically. Finally, the simulation result is discussed and the defective pattern is summarized.

### *4.1.1 Kinematic analysis*

When the ball rolls into the defect zone, the destressing contact between the ball and the outer race causes the bearing system to vibrate [58]. A repeatable pattern in the vibrations signal is shown in Figure 15.

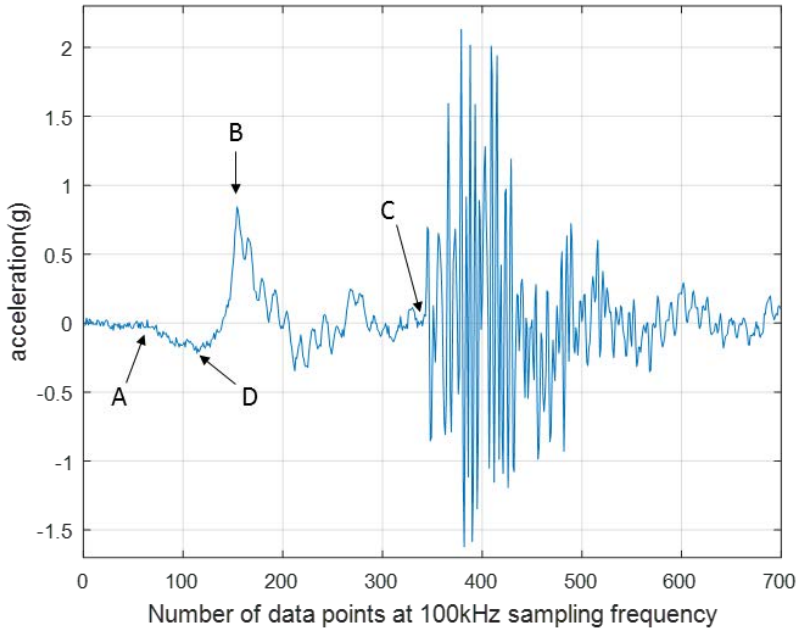



Figure 15. Typical defective response due to a localized defect.

At point A, the “destressing” process starts and then the vibration signal decreases to the local minimum point D. After passing through point D, the time series signal begins to increase and reaches the first peak at point B. Peak B is shown to be very dominant and repeatable in the response pattern. Then, the signal vibrates at a certain frequency until the beginning of a high frequency response at point C. Prior research estimates defect size using the time from B to C and the estimation result is much smaller than the true defect size [18]. This error occurs because the vibration pattern is not well related to the physical

model. Therefore, to improve the estimation, a dynamic model that describes the defective pattern is proposed. The kinematic analysis including the ball center path is introduced in this section as the foundation for the proposed dynamic model.

The bearing diagram is shown in Figure 16. The inner race rotates in the direction as labeled by the arrow. The outer race is stationary. Due to the preload in the bearing, the ball diameter is larger than the distance between the inner and outer races, which is represented by the area in shadow in Figure 16. $\theta_0$ is the angular position of the ball center when the ball just contacts the entry edge. The area of the intersection between inner race and the ball is $a_0$, and the area of the intersection between outer race and the ball is $b_0$, under the initial bearing load $F_0$. Note that both the deflection of ball and the defect size is exaggerated in Figure 16.

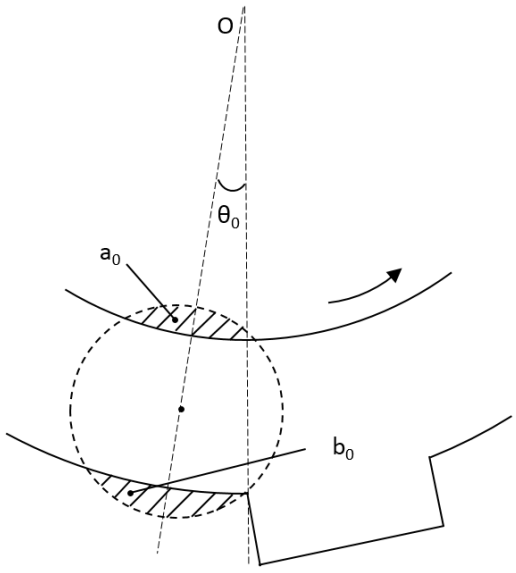
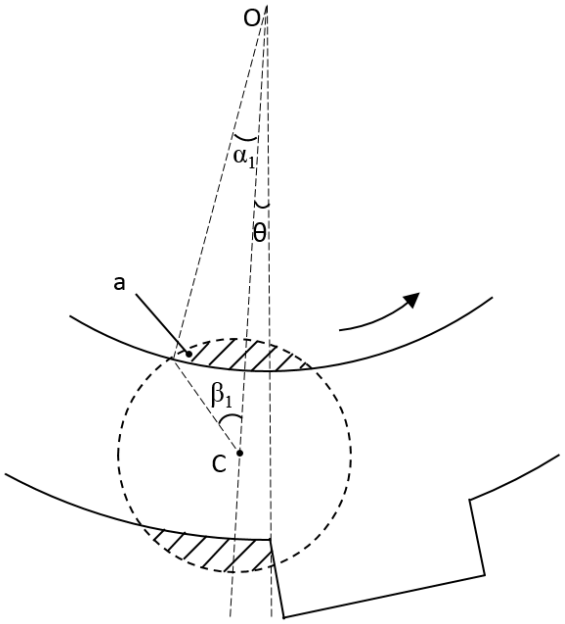
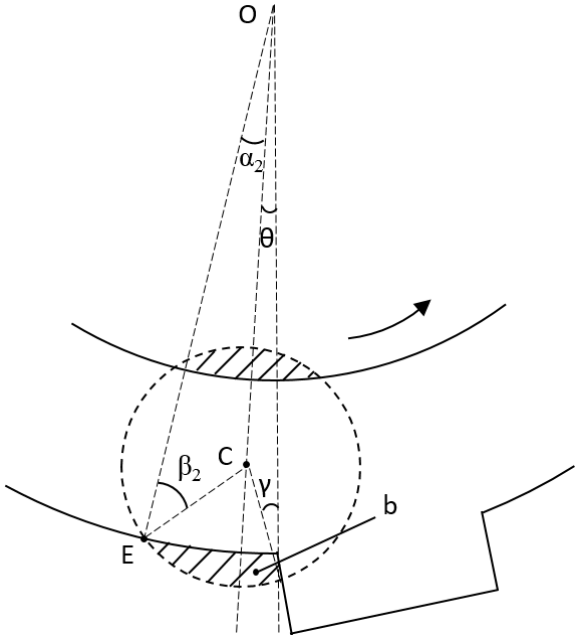



Figure 16. Deflection in ball-raceway contact.

When the ball enters the defect region, the relationship between a, b and the ball position is as shown in Figure 17.




(a)




(b)

Figure 17. Intersection area a and b with respect to their positions.

Before deriving the dynamic model, the path of the ball's center point needs to be determined. Note that the following assumptions are used: 1) the contact force between the

ball and the raceway is proportional to the intersection area, 2) the magnitude of the contact force between the ball/inner race and the ball/outer race is the same, and 3) the ball is assumed to rotate around the bearing center with the same angular velocity as the cage. Because the areas of a and b are equal, when the ball rolls into the defect zone, area b starts to decrease along with area a. Therefore, the distance between the inner race center O and the ball center C ($x_b$) at a specific time t can be calculated using the following equations:

$$\theta = \theta_0 - 2\pi \cdot 0.42 \cdot \frac{n}{60} \cdot t \tag{4.1}$$

$$a = \alpha_1 \cdot R_1^2 + \beta_1 \cdot R_2^2 - x_b \cdot R_1 \cdot sin\alpha_1 \tag{4.2}$$

$$\begin{aligned} b = {} & (\alpha_2 + \beta_2 + \theta + \gamma) \cdot R_2^2 - (\alpha_2 + \theta) \cdot R_3^2 + x_b \cdot R_3 \cdot sin\alpha_2 \\ & + 0.5x_b \left( x_b \cdot \cos\theta + \sqrt{R_2^2 - x_b^2 \sin^2\theta} \right) sin\theta \end{aligned} \tag{4.3}$$

$$a = b \tag{4.4}$$

Where $R_1$ is inner race radius, $R_2$ is ball radius, $R_3$ is the length of OE as shown in Figure 17 (b) and n is the spindle speed in RPM. Note that $\gamma$, $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$ can be derived from trigonometry. After $x_b$ is calculated, the intersection area a and b can be calculated and the contact force F can be calculated as:

$$F = F_0 \cdot \frac{a}{a_0} \tag{4.5}$$

This force will be used in the dynamic model described in the following section.

### *4.1.2 Bearing nonlinear dynamic model*

The dynamic bearing system diagram is shown in Figure 18. $m_1$ is the mass of the shaft with the inner race, and $m_2$ is the mass of the add-on bearing housing with the outer race. $k_1$, $k_2$, $c_1$, $c_2$ are the stiffness and damping coefficients respectively. F is the contact force between the ball that rolls into the defect and both races, as derived in 4.1.1. $k_3$ is defined as the load-deformation constant for single point contact [59, 60] and therefore represents the relationship between bearing load and deformation. $x_1$ is the displacement of the inner race and $x_2$ is the displacement of the outer race.

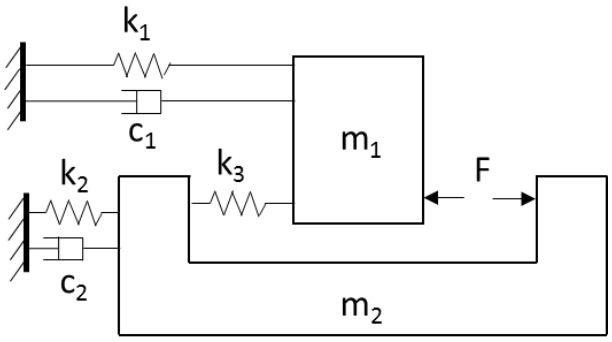



Figure 18. Diagram of the bearing dynamic model.

To formulate the equations of motion for the bearing system, the contact force between the races and balls must be derived. According to the Hertzian contact theory, the contact force between the raceway and the other 12 balls can be represented as:

$$F_{12} = -k_3 \sum_{i=1}^{12} \left( \Delta x_0 + \frac{(x_1 - x_2)\cos(i\eta)}{\cos 40^\circ} \right)^{1.5} \cos 40^\circ \cos(i\eta) \qquad (4.6)$$

Where $40^{\circ}$ is the contact angle of the angular contact ball bearing, η is the angle between two adjacent balls, (in this thesis, there are 13 balls in the bearing and therefore η =2π/13), and $k_3$ can be calculated as:

$$k_3 = \left(1.5 \cdot \frac{1-v^2}{E}\right)^{-1} \cdot \left[\left(\frac{1}{r} - \frac{1}{R_o}\right)^{\frac{1}{3}} + \left(\frac{1}{r} + \frac{1}{R_i}\right)^{\frac{1}{3}}\right]^{-1.5} \tag{4.7}$$

Where r is ball radius, $R_o$ is outer race groove radius, and $R_i$ is inner race groove radius, E is the Young's Modulus, and ν is the Poisson's ratio. The negative sign in Eq (4.6) indicates that the positive direction of $F_{12}$ is radially pointing from the rotating axis to the defect. $\Delta x_0$ is the total deflection between ball and races under the pre-load $F_0$, which can be calculated as:

$$\Delta x_0 = k_3^{-\frac{2}{3}} \cdot F_0^{\frac{2}{3}} \tag{4.8}$$

After the contact force between ball and races $F_{12}$ and F, respectively, are determined, the equations of motion for the bearing system can be represented as:

$$m_1\ddot{x}_1 + c_1\dot{x}_1 + k_1x_1 = F_{12} - F \cdot \cos 40^{\circ} \tag{4.9}$$

$$m_2\ddot{x}_2 + c_2\dot{x}_2 + k_2x_2 = -F_{12} + F \cdot \cos 40^{\circ} \tag{4.10}$$

Note that the equations are a system of second order nonlinear differential equations and can be solve numerically using a standard 4th order Runge-Kutta method. The results are presented in the following section.

### *4.1.3 Bearing defective response pattern due to a localized spall-like defect*

The dynamic model describes the bearing contact process starting from when the ball enters the defect at point A to when it collides with the exit edge at point C. Note that A and C are shown in Figure 15. The experimental and simulated vibration responses are shown in Figure 19. The spindle speeds are 500 rpm, 1000 rpm and 1500 rpm. In the experiment, the accelerometer was mounted above the defect on the outer race.

By comparing the two plots in Figure 19, the simulation is able to represent some key features in the experimental results, which can be summarized as:

1) Peak B is dominant in both the experimental and simulated vibration response.

2) The time from A to B is inversely proportional to spindle speed.

3) The local minimum at point D is proportional to the square of the spindle speed.

4) After point B, the signal vibrates at a particular frequency and decays.

Because the simulated result is agreeable with the experimental data, the dynamic model is validated and these key features can be used to estimate the defect size. In addition, the proposed dynamic model is applicable to inner race defects.

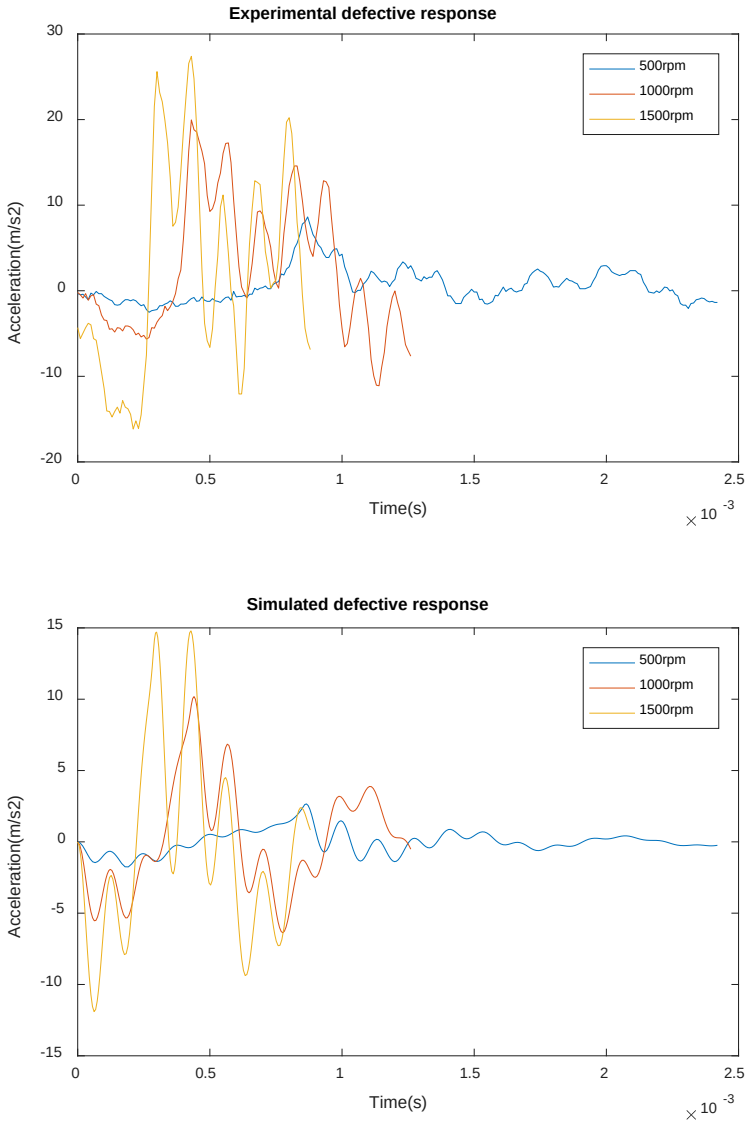



Figure 19. Experimental and simulated defective response.

The corresponding simulated contact forces are shown in Figure 20. By comparing Figure 19 with Figure 20, it can be inferred that peak B occurs at the moment when the ball and raceway lose contact with each other. In Figure 20, the initial contact force is 100 N. As predicted by the kinematic analysis, the time between A and B increases with the initial bearing load due to the larger deflection between ball and raceway.

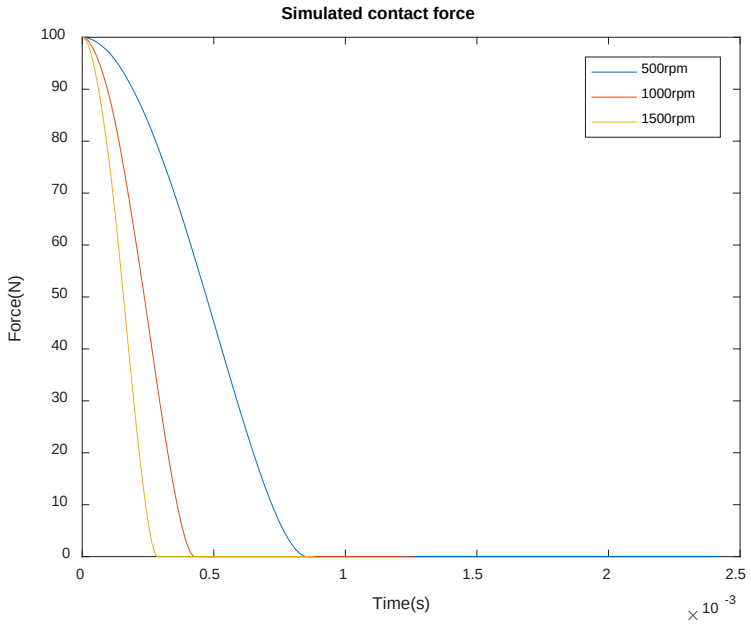



Figure 20. Simulated contact force.

Note that even though the simulation is agreeable with the experiment results in describing key features related to defect size estimation, some discrepancies exist as well. Even though the simulated and experimental responses both show decreasing trends from A to C, the corresponding shapes and vibration magnitudes are shown to differ. The shape from A to B depends on the contact force between the ball and raceway. Since the ball center path is calculated from the assumption that the intersection area between ball and raceways is the same, the contact force model can be prone to error when the assumption fails. The higher than expected level of vibration from A to B might due to the assumption that the damping in the ball is negligible. In addition, the phase in the vibration after point B is different. This might due to the slightly different damping ratio between the simulation and the actual system.

After the collision, the bearing “restressing” process is masked by the high frequency noise due to the impact. Note that the time information after C is difficult to extract for a reliable estimation, and therefore is not considered in the scope of the dynamic model developed in this thesis.

Finally, the defective response pattern in Figure 15 can be explained as follows. The time from entry point A to the first peak B of the entry event is defined as the entry-to-peak time $t_e$. The time from the first peak B to the impact point C is defined as the peak-to-impact time $t_p$. The time from the entry point to the impact point is defined as the total time $t_t$. The mechanism of the defective feature can be described as shown in Figure 21. Point A is when the bearing starts to enter the defect zone, which is called the entry event. During the entry event, the contact force between the ball and raceways decreases. At point B, the entry event ends and the ball loses contact with both races, and therefore the contact force decreases to zero. The loss of contact the outer race to vibrate at a certain frequency. Then, the ball continues to move forward with an initial speed and eventually collides with the exit edge of the defect at C. The amplitude of the vibration at C increases dramatically due to the impact.

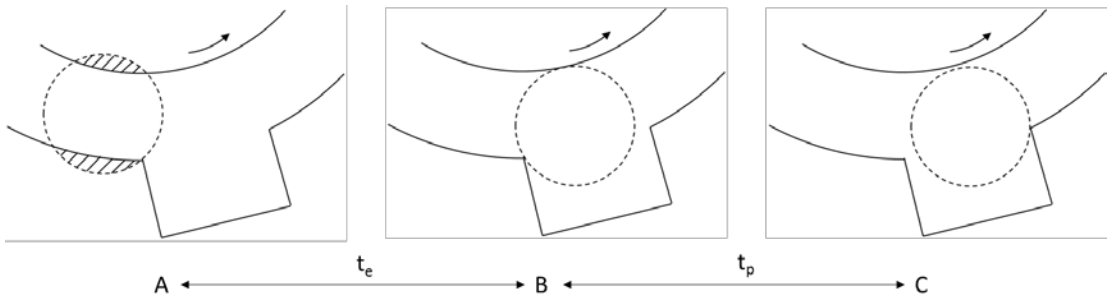



Figure 21. Key points when a ball passes a localized defect.

## 4.2 Defect size estimation model

The bearing vibration signal has been explained in the previous section. The next task is to estimate the defect size using the time information $t_e$ and $t_p$ obtained from the vibration signal. To solve this problem, three models are developed in the next several paragraphs from the simplest to the most comprehensive.

### *4.2.1 The isosceles trapezoidal model*

The isosceles trapezoidal model is the simplest defect-size-estimation model and the geometry relationship between the defect size d and the path traveled by the ball is as shown in Figure 22. O is the center of the inner/outer race, A is the ball center when the ball is at the entry edge C and B is the ball center when the ball is at the exit edge D.

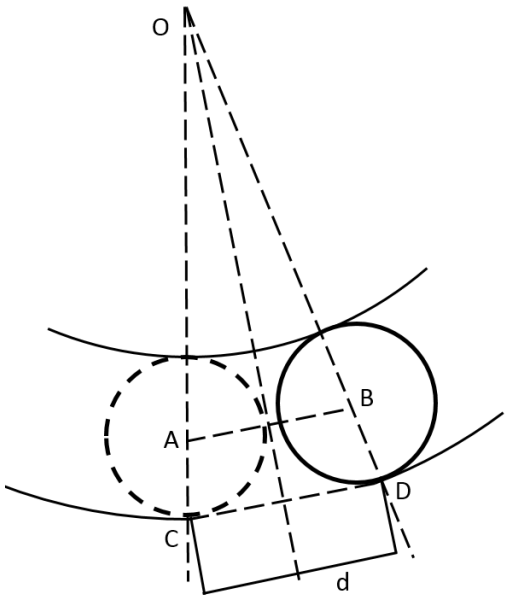



Figure 22. The isosceles trapezoidal model.

The ball is assumed to travel through AB with a constant ball center velocity before it enters the defect zone, which is calculated by Eq (4.11):

$$v_b = \frac{2 \times \pi \times 0.42 \times n}{60} \cdot R \tag{4.11}$$

Where n is the spindle speed in RPM. 0.42 is the outer race ball-pass frequency ratio as introduced in Chapter 3, and it is a constant determined by the bearing geometry. R is the distance from inner race center to ball center. The time from A to B can be measured from the signal, which is $t_e$+$t_p$. Then, the defect size d can be derived from the geometry in Figure 22:

$$d = v_b \times (t_e + t_p) \times \frac{R+r}{R} \tag{4.12}$$

Where r is the ball radius. For the inner race case, Eq (4.11) and (4.12) should be updated as:

$$v_b = \frac{2 \times \pi \times 0.58 \times n \times R}{60} \tag{4.13}$$

$$d = v_b \cdot (t_e + t_p) \cdot \frac{R - r}{R} \tag{4.14}$$

Where 0.58 is the inner race ball-pass frequency ratio. For the inner race defect, since the ball path is still constrained by the outer race when it passes the defect, this trapezoidal model is suitable. However, this model should provide an estimation of d smaller than the true defect size for the outer race defect, because the ball will collide with the exit edge before arriving position B as shown in Figure 22. An improved model is introduced in the next section for the outer race defect estimation.

### *4.2.2 The one-right-angle quadrilateral model*

The one-right-angle quadrilateral model considers the fact that the ball collides with the exit edge before the ball center arrives the exit edge. The geometry relationship between defect size d and the path traveled by the ball is shown in Figure 23. O is the center of the inner/outer race, A is the ball center at the entry edge C and B is the ball center when it collides with the exit edge D.

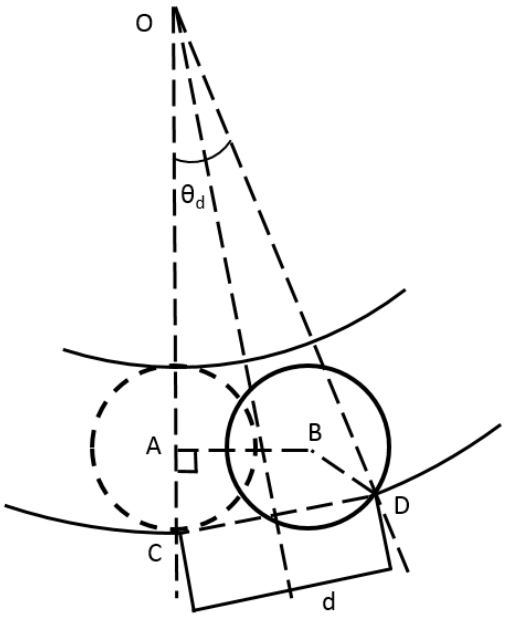



Figure 23. The one-right-angle quadrilateral model.

This model assumes that the ball loses contact with both races at point A and travels with the same velocity as before it enters the defect zone, as in Eq (4.11). The travelling time from A to B can be obtained from the vibration signal, which is $t_e$+$t_p$. Therefore, the length AB can be expressed as:

$$AB = \frac{2 \times \pi \times 0.42 \times n \times R \times (t_e + t_p)}{60} \tag{4.15}$$

Once AB is obtained, according to the trigonometry relationship as shown in Figure 23, the angle $\theta_d$ can be derived using the equation below:

$$[(R + r) \cdot \cos \theta_d - R]^2 + [(R + r) \cdot \sin \theta_d - AB]^2 = r^2 \tag{4.16}$$

Where R is the distance from inner race center O to ball center A, and r is the ball radius. For easy comparison with the isosceles trapezoidal model, Eq (4.16) can be simplified.

Since $\theta_d$ is a small angle, $\sin\theta_d \approx \theta_d$, $1-\cos\theta_d = 2\sin^2\theta_d/2 \approx \theta_d^2/4$. Then, in Eq (4.16), $\theta_d$ can be solved as:

$$\theta_d = \frac{R + r + \sqrt{(R+r)\cdot r}}{(R+r)\cdot R} \cdot AB \tag{4.17}$$

Then, the defect size d can be estimated as:

$$d = v_b \cdot (t_e + t_p) \cdot \frac{R + r + \sqrt{(R+r)\cdot r}}{R} \tag{4.18}$$

Note that between Eq (4.12) and Eq (4.18) that the coefficient has been updated. In the isosceles trapezoidal model, the coefficient is calculated to be 0.8407 and in one-right-angle quadrilateral model the coefficient is 1.589. Therefore, this model provides estimation results almost twice as large as the first model. Note that this model is not suitable for the inner race defect, because the outer race does not allow the ball travel in the horizontal direction as in the outer race defect. In other words, the ball is constrained by the outer race and its trajectory along a circular motion.

### *4.2.3 The Hertzian quadrilateral model*

This model is built on the one-right-angle quadrilateral model and considers ball-raceway deflection based on the Hertzian contact theory. Therefore, this is the most comprehensive model among the three.

4.2.3.1 The Hertzian contact theory of spheres

When two spheres contact with each other under a specific contact force F, a circular contact area will form between the contact surfaces as shown in Figure 24.

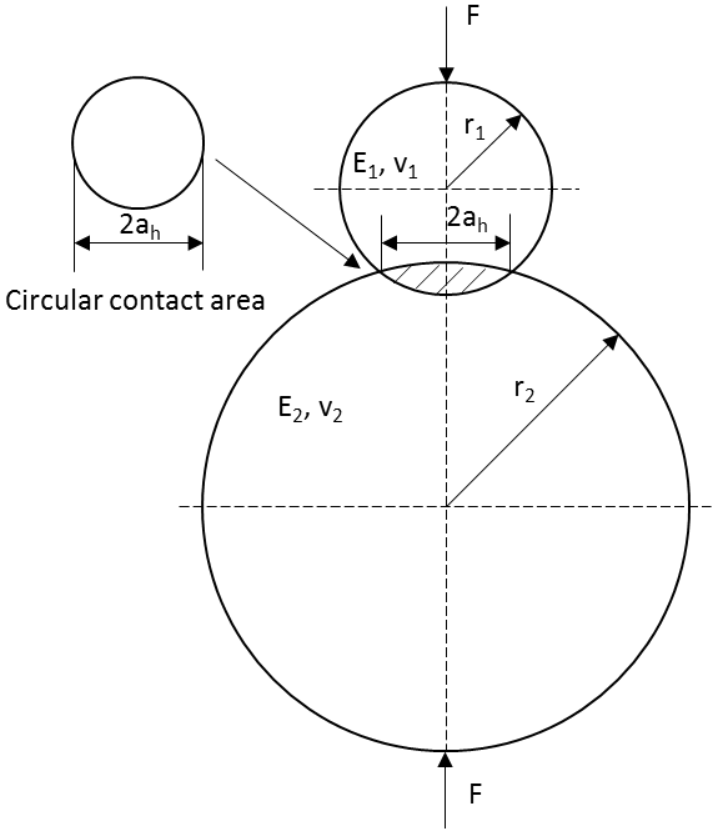



Figure 24. The Hertzian contact theory.

The radius of the contact area is given by Eq (4.19),

$$a_h = \sqrt[3]{\frac{3F\left[\frac{1-v_1^2}{E_1}+\frac{1-v_2^2}{E_2}\right]}{4\left(\frac{1}{r_1}+\frac{1}{r_2}\right)}} \tag{4.19}$$

Where F is the contact force between the two spheres, $E_1$ and $E_2$ are the Young's modules, $v_1$ and $v_2$ are the Poisson's ratios. The centers of the two spheres move closer to each other by δ and it can be expressed using the equation below:

$$\delta = a_h^2 \times \left(\frac{1}{r_1} + \frac{1}{r_2}\right) \tag{4.20}$$

For the ball-outer race contact, Eq (4.19) and (4.20) should be updated as:

$$a_h = \sqrt[3]{\frac{3F\left[\frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2}\right]}{4\left(\frac{1}{r_1} - \frac{1}{r_2}\right)}} \tag{4.21}$$

$$\delta = a_h \times \left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{4.22}$$

4.2.3.2 The Hertzian quadrilateral model

The geometry relationship between the defect size d and the path traveled by the ball is as shown in Figure 25. At the moment when the deflected area just arrives at the entry edge C, the ball center is at point A. On the outer race, the distance between the deflect area center F to the entry edge C is $a_1$. O is the center of the inner/outer race, and B is the ball center when the ball collides with the exit edge D. The ball loses contact with both races at E. r is the radius of the ball.

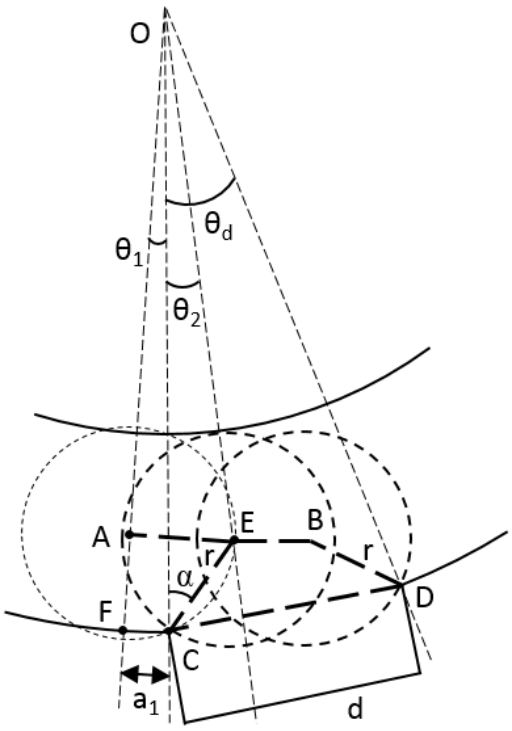



Figure 25. The Hertzian quadrilateral model.

At position A, given the initial contact force $F_0$, the radius of the ball-outer race and ball-inner race deflections $a_1$ and $a_2$, can be derived from Eq (4.19) and (4.21). Within the speed range from 500-3000 rpm, the centrifugal force is very small compared to the contact force, and therefore is not considered in this model. For simplicity,

$$a_1 = K_1 \cdot F_0^{\frac{1}{3}} \tag{4.23}$$

$$a_2 = K_2 \cdot F_0^{\frac{1}{3}} \tag{4.24}$$

Where $K_1$ and $K_2$ are constants given bearing geometry and material property. According to Eq (4.20) and (4.22), the ball center moves closer to the inner race center by $\Delta x_2$ and moves closer to the outer race by $\Delta x_1$:

$$\Delta x_1 = a_1^2 \cdot \left(\frac{1}{r} - \frac{1}{R+r}\right) \tag{4.25}$$

$$\Delta x_2 = a_2^2 \cdot \left(\frac{1}{r} + \frac{1}{R-r}\right) \tag{4.26}$$

If the distance from ball center to inner race center without deflection is R, then the length OC and OD in Figure 25 can be expressed as:

$$OC = OD = R + r - \Delta x_0 \tag{4.27}$$

Where $\Delta x_0$ is the total defection of the ball with both races and:

$$\Delta x_0 = \Delta x_1 + \Delta x_2 = K_3 \cdot F_0^{\frac{2}{3}} \tag{4.28}$$

$K_3$ is a constant given the bearing geometry and material property. When the ball loses contact with both races at E, OE is R and CE is r. Then the angle $\theta_2$ can be derived as:

$$cos\theta_2 = \frac{R^2 + (R + r - \Delta x_0)^2 - r^2}{2R(R + r - \Delta x_0)} \tag{4.29}$$

Then, $\theta_2$ can be further simplified as below:

$$\theta_2 = \frac{\sqrt{2(R+r)Rr\Delta x_0}}{R(R + r - \Delta x_0)} \tag{4.30}$$

From position E, the ball is assumed to travel in the horizontal direction with a constant speed $v_b$ from E to B as shown in Eq (4.11). According to the dynamic model, the

first peak in the entry signal occurs at E. Therefore, the time from E to B is $t_p$, which can be obtained from the vibration signal. Then, EB can be derived as:

$$EB = v_b \cdot t_p \tag{4.31}$$

When the ball collides with the exit edge D, the angle $\theta_d$ can be solved using the trigonometry relationship as shown in Figure 23:

$$[OD \cdot \cos\theta_d - R\cos\theta_2]^2 + [OD \cdot \sin\theta_d - R\sin\theta_2 - v_b \cdot t_p\,]^2 = r^2 \tag{4.32}$$

In Eq (4.32), $\Delta x_0$ and $\theta_2$ only depends on the initial contact force $F_0$, and $t_p$ can be obtained from the vibration signal. In other words, $\theta_d$ is a function of $F_0$ and $t_p$. However, since Eq (4.32) is more complicated than the previous models, $\theta_d$ cannot be represented explicitly like in Eq (4.17) but solved numerically. Then the defect size d can be estimated using Eq (4.33):

$$d = 2(R + r)\sin\frac{\theta_d}{2} \tag{4.33}$$

Therefore, the defect size d is only dependent on the initial contact force $F_0$ and the peak-to-impact time $t_p$. In other words, if the bearing load is known and $t_p$ can be obtained from the vibration signal, the defect size d can be estimated. However, the contact force $F_0$ is usually hard to determine. In this case, the contact force can be estimated using $t_e$ from the vibration signal. This method will be introduced in the next section.

4.2.3.3 Estimation of the initial contact force $F_0$

According to the dynamic model, the entry-to-peak time $t_e$ is closely related to the initial contact force $F_0$. When $F_0$ is larger, the time $t_e$ becomes longer. At the moment when the deflected area just arrives at the entry edge, the distance between the deflect area center F to the entry edge C is $a_1$. Then, the angle $\theta_1$ can be calculated as:

$$\theta_1 = \frac{K_1 \cdot F_0^{\frac{1}{3}}}{R + r} \tag{4.34}$$

The angle $\theta_2$ can referred from Eq (4.30). Both $\theta_1$ and $\theta_2$ are functions of $F_0$. Since the small deflection, $\theta_1$ and $\theta_2$ are very small. Therefore, it can be assumed that the velocity of the ball center is constant. Thus,

$$t_e = \frac{\theta_1 + \theta_2}{\omega_c} \tag{4.35}$$

Where $\omega_c$ is the cage angular velocity. From Eq (4.35), the relationship between $F_0$ and $t_e$ can be determined. An interesting thing to notice is that, the ratio between $\theta_1$ and $\theta_2$ is almost a constant value for such small defections. This is because the ratio $\theta_1/\theta_2$ can be derived as:

$$\frac{\theta_1}{\theta_2} = \frac{R \cdot \left(\frac{1}{r} - \frac{1}{R+r}\right)^{-\frac{1}{3}}}{\sqrt{2 \cdot (R+r) \cdot R \cdot r \cdot \left[\left(\frac{1}{r} - \frac{1}{R+r}\right)^{\frac{1}{3}} + \left(\frac{1}{r} - \frac{1}{R+r}\right)^{\frac{1}{3}}\right]}} \tag{4.36}$$

Therefore, the ratio $\theta_1/\theta_2$ only depends on the bearing geometry. Then, Eq (4.35) can be updated as:

$$t_e = K \cdot \frac{F_0^{\frac{1}{3}}}{n} \tag{4.37}$$

Where n is the spindle speed in RPM and K is a constant value that can be determined by the bearing geometry and material properties. Substituting Eq (4.37) into Eq (4.28):

$$\Delta x_0 = K_3 K^2 n^2 t_e^2 \tag{4.38}$$

Then, substituting Eq (4.38) into Eq (4.32), the defect size d can be obtained from the measurement result of $t_e$ and $t_p$. This relationship can be linearized as:

$$d = v_b \left( \left(2 - \sqrt{2}\right) \cdot \frac{R + r}{r} \cdot t_e + \frac{R + r + \sqrt{r(R + r) f(t_e/t_p)}}{R} \cdot t_p \right) \tag{4.39}$$

4.2.3.4 The defect size lower limit

The Hertzian quadrilateral model depends on the entry-to-peak time $t_e$ and the peak-to-impact time $t_p$. A prerequisite for this model is that the ball collides with the exit edge after it leaves the entry edge. The lower limit of the defect size that ensures the Hertzian quadrilateral model is suitable can be calculated from Eq (4.32). The ball is assumed to collide with the exit edge when it arrives at position E in Figure 26, which is when it just leaves contact with the entry edge.

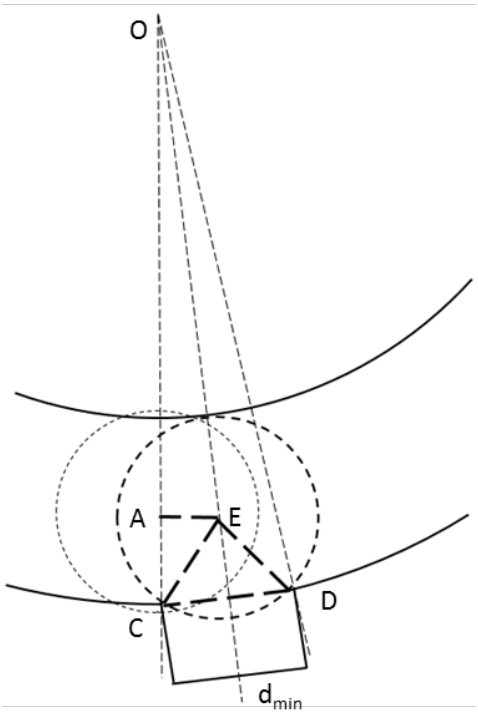



Figure 26. The lower limit of the defect size.

Then, Eq (4.32) should be updated with $t_p$=0 as:

$$[OD \cdot \cos\theta_d - R\cos\theta_2]^2 + [OD \cdot \sin\theta_d - R\sin\theta_2\ ]^2 = r^2 \qquad (4.40)$$

The angle $\theta_d$ can be approximated by the following equation:

$$\theta_d \approx 2\sqrt{2} \cdot \sqrt{\frac{r \cdot \Delta x_0}{R(R+r)}} \qquad (4.41)$$

Then the lower limit of the defect size that works for Hertzian quadrilateral model is as below:

$$d_{min} \approx 2\sqrt{2} \cdot \sqrt{\frac{r \cdot (R + r) \cdot \Delta x_0}{R}} \tag{4.42}$$

From Eq (4.42) it is obvious that $d_{min}$ only depends on the contact force between the ball and the raceways. When the contact force is larger, the deflection $\Delta x_0$ becomes larger accordingly, then the lower size limit for this model will increase. This size boundary is smaller than the smallest defect in the experiment, and therefore all the defects can be estimated based on the Hertzian quadrilateral model.

When the defect size is smaller than $d_{min}$, the geometry changes and the proposed three models will not work in that situation. As suggested by Randall, when the ball collides with the exit edge before it loses contact with the entry edge, the ball center is about at the center of the defect. Therefore, the following equation is a recommendation to estimate the defect smaller than the lower bound:

$$d = 2(R + r)\sin\left(t_t \cdot \omega_c - \frac{a_1}{R + r}\right) \tag{4.43}$$

Where $t_t$ is the total time from entry to impact. Validation of this equation is beyond the scope of this thesis, but it can be future work. The three defect-size-estimation models are introduced from the simplest to the most comprehensive. They will be used to estimate defect size based on the time information obtained from the vibration signals. The first model assumes a linear relationship between the defect size and the entry-to-impact time, which is $t_e$+$t_p$. The second model considers that the ball collides with the exit edge earlier before the ball center arrives the exit edge. The third model considers the contact deflection

between the ball and races. In Chapter 5, these three models will be used to estimate defect size and the experiment results will be compared.

## 4.3 Summary

This chapter proposes the nonlinear dynamic model for the bearing system to describe the vibration signal due to a localized defect on the raceway. The repeatable pattern is obtained and the simulation result is agreeable with experimental observations. Then, three defect size estimation models are introduced from the simplest trapezoidal model to the most complicated Hertzian quadrilateral model. In addition, the lower limit of the defect size that can be estimated using the Hertzian quadrilateral model is derived. These models are validated in Chapter 6 and Chapter 7 with the experimental results.

# CHAPTER 5. BEARING SIGNAL PROCESSING

This chapter describes the signal processing method to extract the time information from the vibration signal for defect size estimation. Then, the performance of this method is discussed with respect to the high and low frequency noises. For the high frequency noise, the Variational Model Decomposition (VMD) is used to remove the high frequency noise from the entry signal. For the low frequency noise, the lower bound of the spindle speed is determined. Finally, the complete signal-processing scheme for defect size estimation is summarized.

## 5.1 The signal processing methods to extract point A, B and C

In the defect estimation model proposed in Chapter 4, the time information between the entry and exit point is used to estimate the defect size. Therefore, the performance of this method relies on the precision in time domain, which can be affected by many factors in the actual bearing test system. One of them is ball skidding that occurs when the ball enters the defect zone, which cause variations in $t_e$. In addition, after the ball loses contact with both raceways, the no-external force assumption in the dynamic model may not always hold. In some instances, the cage and the raceways may still induce small impacts on the ball that leads to variations in the ball's velocity, which will affect the total time $t_t$ (from entry to impact). The skidding and small impacts occur randomly and cannot be foreseen. Therefore, the time information extracted from one impact response is not reliable. A more comprehensive approach is to estimate the defect size based on more impact responses collected during a longer period. The mean value of these multiple impact responses is a better estimation of the true mean (the real defect width). If large amount of

impacts need to be analyzed, it is necessary to develop an automatic algorithm that can extract the time information from the data to save time and effort. More importantly, the statistics of the time information can be analyzed to obtain deeper understanding of the bearing system and better estimation result.

### *5.1.1 Point B*

Point B is the peak at the end of the entry event and it represents the moment when the ball loses contact with the entry edge, as proposed in the dynamic model. By observation of long time sequence of the vibration signal, it can be seen that point B has two important features: 1) It is dominant in amplitude compared to the rest of the entry signal and the noise. 2) The time difference between point B and C ($t_p$) is a very reliable value with a small standard deviation. Also, each impulsive response is very repeatable. Therefore, point B can be identified by the cross-correlation algorithm. To use this method, a target signal starts at point B ends around the middle point between two adjacent impulses as shown in Figure 27 is needed. The cross-correlation between the target signal and the raw data is performed as:

$$\mathrm{h(k)} = \sum_{i=1}^{m} f(i)g(i+k), \qquad k = 1,2,\dots,n-m+1 \tag{5.1}$$

Where f is the target signal, g is the raw vibration signal, h is the cross-correlation between f and g at delay k, m is the length of the target signal, n is the length of the raw vibration signal. Since cross-correlation is a measure of the similarity of two series as a function of the displacement of one relative to the other, the maximum value of Eq (5.1) occurs when

the target signal perfectly lines up with any of the impacts in the raw data as shown in the bottom of Figure 27. Due to the good repeatability of the signal, the maximum value of Eq (5.1) for each impact in the raw data is within a close vicinity of point B. Therefore, by searching the maximum value within a small region around each maximum value of the cross-correlation signal, point B can be identified.

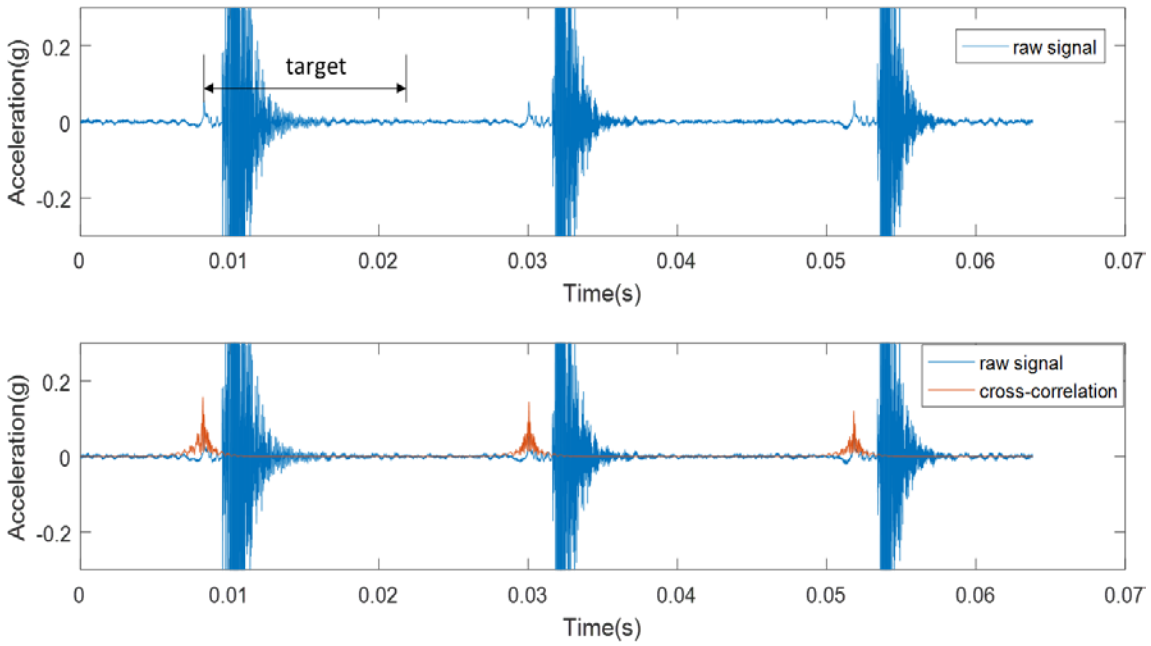



Figure 27. The cross-correlation between the target and the raw vibration signal.

### *5.1.2 Point C*

Point C is the impact point when the ball collides with the exit edge and excites a large high frequency impulsive response in the vibration signal. Therefore, the exit point is the initial point of the response and can be used to calculate the position of the exit edge. To identify the impact point C, some researchers use the maximum value in the response signal to mark the moment when the ball collides with the exit edge. This approach is straightforward to implement in an automatic system. However, results show that the

maximum value does not always occur right after the collision between the ball and the exit edge [58]. Sometimes the time delay between the maximum value and the collision can be very large. Therefore, this method can result in an estimation that is much larger than the real defect. In this thesis, point C is identified by using a threshold that separates the entry and exit events. Since the exit impact response consists of primarily high frequency components, point C can be visually identified from the vibration data as shown in Figure 28.

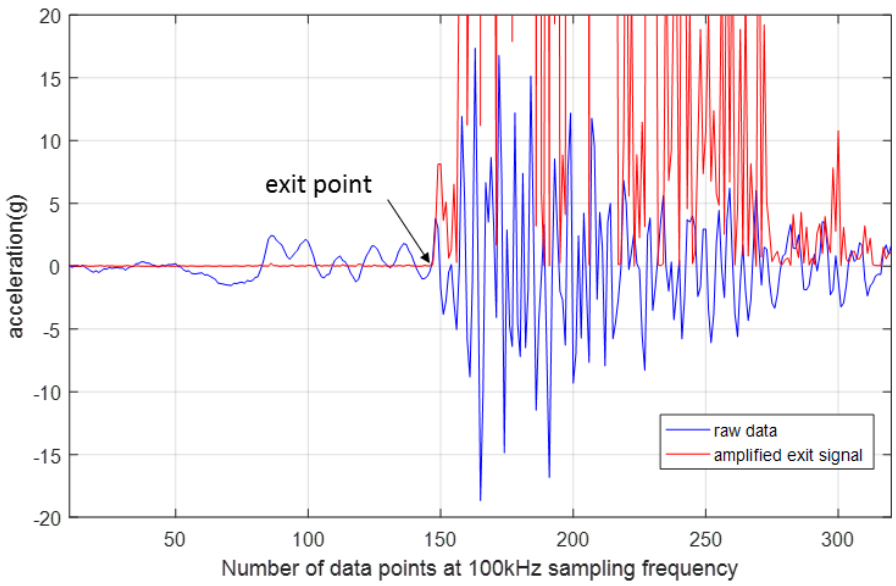



Figure 28. Method to find the exit point.

The vibration signal around point C exhibits the following properties: 1) The exit signal (after C) is made of much higher frequency components than the entry signal (before C). 2) The transition from low frequency to high frequency is very sharp. 3) The exit event has much more energy than the entry event. Therefore, a high-pass filter that amplifies the exit signal relative to the entry signal appears suitable. Then, a threshold can be set to identify the transition point C. However, in some cases the peak in the entry signal may

have a similar amplitude as the first part of the exit signal, thus making the estimation result sensitive to the threshold value. Therefore, the high-pass filter is required to be applied multiple times to achieve a satisfactory result. In this thesis, the vibration data is differentiated twice and squared as shown in Figure 28. The differentiation operation operates as a high-pass filter to identify the impact point C between the low frequency entry signal and high frequency exit signal as labeled in Figure 28. This method has been tested with a large number of impact responses and results show that a reliable region for a threshold value to consistently separate the entry and exit signal exists. Note that the response may differ among different bearing systems. Dynamic model-base simulations can be used to calculate the order of differentiation. The region for the threshold value for the nth impact response is defined from the maximum value of the processed entry signal ($L_n$) to the first peak value of the amplified exit signal ($U_n$). Therefore, final threshold for all observed impacts is from the maximum $L_n$ to the minimum $U_n$. The large difference between the entry and exit signal after the signal processing ensures the method is insensitive to the threshold value. If a reliable dynamic model cannot be obtained, this threshold value can also be tested directly from the data.

### *5.1.3 Point A*

Point A as shown in Figure 15 represents the moment when the ball contacts the entry edge and starts to enter the defect zone. According to Sawalhi and Randall [18], at this stage the contact force between the ball and races begin to decrease and this is call the "distressing process." During the destressing process, the vibration signal decreases slowly to a minimum level before increasing. Because the destressing process is continuous and lacks an obvious collision between the ball and the entry edge, the initial part of the bearing

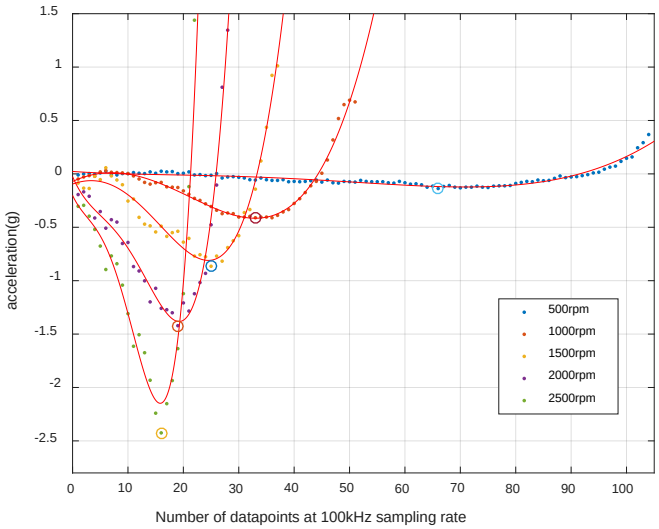
entry signal has a small slope and low signal-to-noise ratio. Therefore, precisely identifying the entry point is difficult. Moreover, sometimes the excessive background noises and disturbances distort the signal even further. Thus, a signal processing method that amplifies the entry signal and increases the signal-to-noise ratio is desired. In this thesis, the empirical model based method is used to identify the entry point A and they are introduced in the following paragraphs.

#### 5.1.3.1 Observation of the bearing entry signal

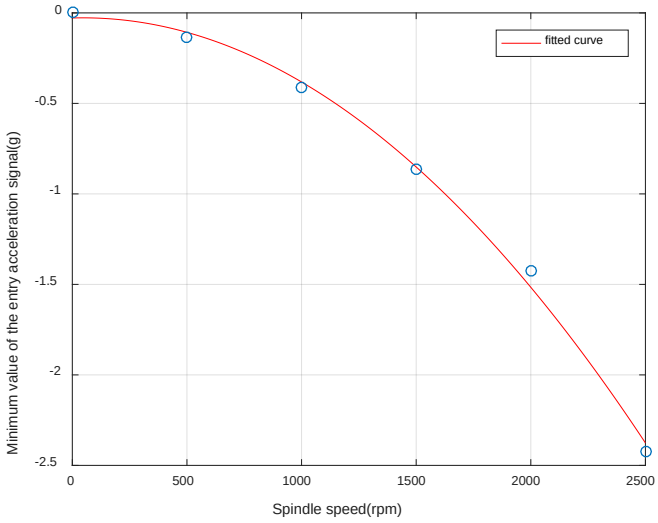
When the rolling element passes a line-spall defect with entry and exit edges, the vibration response has a repeatable pattern as shown in Figure 15. Because the entry signal has good repeatability, an empirical model can be used to calculate the entry point. This section introduces the empirical model and how to use this model to estimate the entry point A.

The entry signal exhibits a repeatable pattern at different spindle speeds and the observation result is shown in Figure 29(a). Through observations of multiple data sets at different spindle speeds, it is determined that the spindle speed only changes the x and y axis scale of the entry signal. The x axis is inversely proportional to the spindle speed while the y axis exhibits a quadratic relationship with the spindle speed as shown in Figure 29(b). Thus, the spindle speed only scales the signal in the x and y direction and does not affect the signal shape. This observation is agreeable with the simulation result based on the dynamic model developed in Chapter 4. Therefore, a “normalized” empirical function of the entry signal can be developed and used to estimate the entry point at different spindle

speeds. These observations are agreeable with the dynamic model as proposed in Chapter 4.




(a) Observation of bearing acceleration entry signal at different speeds



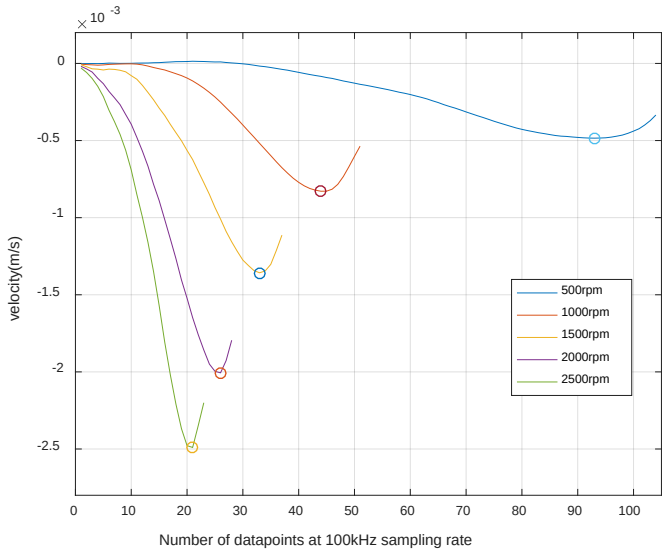

(b) The local minimum value of the acceleration entry signal at different speeds

Figure 29. The acceleration signal and its local minimum.

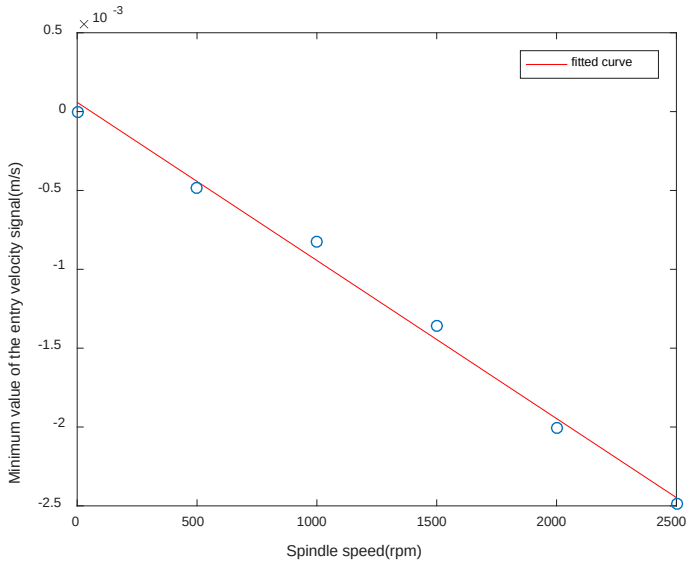
Since the entry signal is repeatable during each impact response, curve fitting can be used to obtain the empirical function of the entry signal. However, a consistent fitting

of the acceleration data is difficult because of its low signal-to-noise ratio. To resolve this issue, the raw vibration data is integrated to obtain the velocity. Integration operates as a low-pass filter by increasing the signal-to-noise ratio of the entry signal as shown in Figure 30(a). Integration is performed instead of a low-pass filter because digital-filter based techniques vary in performance depending on the spindle frequency. After integration, a fourth order polynomial with stable and repeatable coefficients can be fitted to the velocity entry signal. Eq (5.1) shows the curve fitting result at 1000rpm. The values of the constants $p_1$, $p_2$, $p_3$, $p_4$, and $p_5$ with 95% confidence interval are 1.3E-6, -1.1E-4, 2.7E-3, -3.3E-2 and 4.0E-2 respectively. The variable $x$ is the index of a particular data point. Note that even though a polynomial is used in this work for its computational simplicity, other functions can be implemented in the empirical model.

$$v = p_1 x^4 + p_2 x^3 + p_3 x^2 + p_4 x + p_5 \tag{5.2}$$




(a) The velocity vibration entry signal at different speeds




(b) The local minimum value of the velocity entry signal at different speeds

Figure 30. The velocity signal and its local minimum.

Figure 30(b) shows that the local minimum of the entry velocity exhibits a linear relationship with spindle speed. Therefore, if a velocity curve at a specific spindle speed is determined, the velocity curve at other spindle speeds can be obtained by scaling the x and y axes. Here, the normal spindle speed is selected as 1000 rpm. Then, the empirical model of the entry signal at other speeds can be expressed as:

$$v = \varphi(p_1\varphi^4x^4 + p_2\varphi^3x^3 + p_3\varphi^2x^2 + p_4\varphi x + p_5) \tag{5.3}$$

Where $p_1$, $p_2$, $p_3$, $p_4$, and $p_5$ are the same coefficients corresponding to 1000rpm. Therefore, $\varphi$ is n/1000, and $n$ is the current spindle speed in RPM. The $\varphi$ term outside the parenthesis in Eq (5.3) ensures that the velocity function is proportional to the spindle speed. To validate the velocity function, the acceleration function is formulated by

differentiating Eq (5.3). After differentiating, the expression of the acceleration is shown in Eq (5.4).

$$a = \varphi(4p_1\varphi^4x^3 + 3p_2\varphi^3x^2 + 2p_3\varphi^2x + p_4\varphi) \tag{5.4}$$

Therefore, the acceleration function reaches a local minimum at point $x_m$, which is expressed as:

$$x_m = \frac{-3p_2 + \sqrt{9p_2^2 - 24p_1p_3}}{12p_1\varphi} = \frac{\sigma_1}{\varphi} \tag{5.5}$$

Where $\sigma_1$ is a constant value determined by $p_1$, $p_2$, and $p_3$. The minimum value $a_{min}$ can be expressed as:

$$a_{min} = \varphi(4p_1\varphi^4x_m^3 + 3p_2\varphi^3x_m^2 + 2p_3\varphi^2x_m + p_4\varphi) = \sigma_2\varphi^2 \tag{5.6}$$

Where $\sigma_2$ is a constant and can be determined by $p_1$, $p_2$, $p_3$, and $p_4$. Note that Eq (5.6) shows that $a_{min}$ has a quadratic relationship with spindle speed, which agrees with the observation in Figure 29. Therefore, at least in the speed range of 500-3000 rpm, the empirical model represents the major properties of the entry signal through use of the velocity function. For the operational speed lower than this range, signal processing method that helps remove low frequency noise is suggested and for the operational speed higher than this range, a higher sampling frequency should be considered.

5.1.3.2 Signal processing method to identify point A

This section describes the empirical model-based signal processing method used to estimate the entry point. Then, the signal processing method based on differentiation used to identify the exit point is described. Finally, the methodology to calculate the defect size from the entry and exit points is presented. The first step in entry point estimation is solving for $a = 0$ in Eq (5.4), thus resulting in roots $x_1 < x_2 < x_3$. The index of the entry point can be expressed as $N - x_3 + x_1$ where $N$ is the index of the zero-crossing point of the entry signal as shown in Figure 15. Note that $N - x_3 + x_1$ is always positive. However, this calculation is prone to error due to the low signal-to-noise ratio in the acceleration model around the entry point. Therefore, a more accurate solution is to use a further point in the signal with a higher signal-to-noise ratio. Figure 31(a) shows that the velocity function has a large linear region before reaching the local minimum. The index of the midpoint in the linear region of the velocity function corresponds to the local minimum point index of the acceleration function $x_m$. At $x = x_m$, a line can be created with the slope $a_{min} = c_2\varphi^2$. This line is shown in Figure 30 and can be represented as:

$$y = c_2\varphi^2\left(x - \frac{c_1}{\varphi}\right) - k\varphi \tag{5.7}$$

Where $k$ is a constant and can be derived by substituting $x_m$ into Eq (5.3). The intersection of the line at $y = 0$ can be derived as:

$$x_{cross} = \frac{c_1}{\varphi} + \frac{k\varphi}{c_2\varphi^2} = \left(c_1 + \frac{k}{c_2}\right)\frac{1}{\varphi} \tag{5.8}$$

Figure 31 shows the relationship of these variables. According to Eq (5.8), the distance between the intersection index $x_{cross}$ and the origin is proportional to $\frac{1}{s}$. After substituting all constant values into Eq (5.7), $x_{cross}$ is calculated to be ~7% of the entry-to-exit time for defect size of 1.530 mm. When applied to real data, $x_{cross}$ can be obtained by fitting a straight line to multiple points of the velocity profile. Therefore, $x_{cross}$ can be used to estimate the entry point. However, a speed related compensation referring to Eq (5.8) needs to be considered, particularly for smaller defect widths where $x_{cross}$ is a significant portion of the entry-to-exit time.

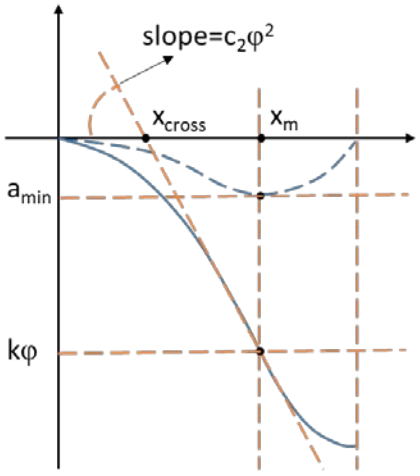



Figure 31. Method to estimate the entry point.

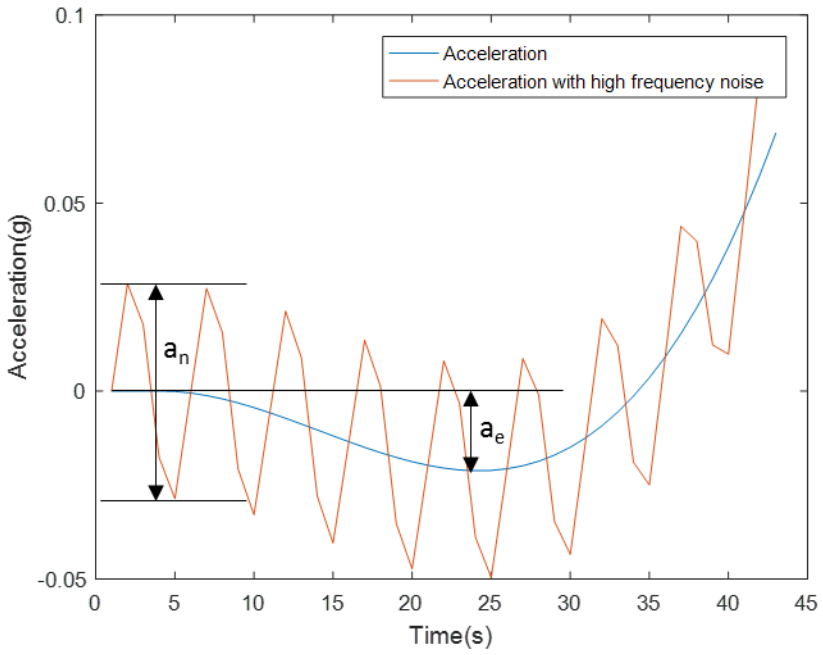
## 5.2 Discussion of the signal processing method with respect to noises

### *5.2.1 High frequency noise*

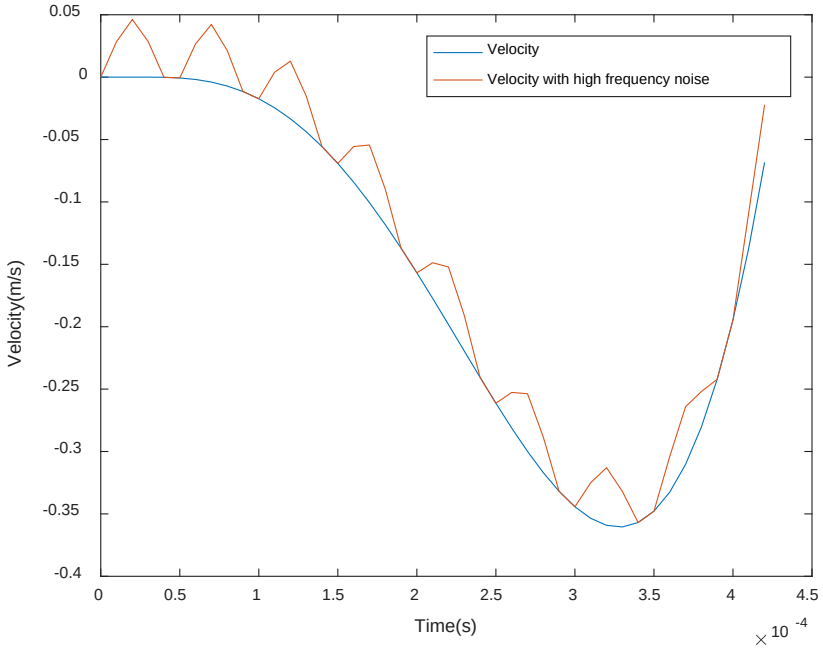
The empirical method described in the previous section is a powerful tool to extract point A, because it provides a large linear region in the velocity signal that is reliable through all the spindle speeds in this thesis. However, when the entry signal is exposed to

very strong high frequency background noises and disturbances, the signal processing method in 5.1.3 is not sufficient enough for denoising, especially at lower spindle speeds when the amplitude of the entry signal is very small. Figure 32(a) shows the simulation result of an entry signal with/without the high frequency noise. The noise is a 20 kHz sine wave with peak-to-peak value 0.06 g. When the acceleration signal is integrated to obtain the velocity signal as shown in Figure 32(b), the high frequency noise still exists in the integrated signal. If the data points in the linear region are used to extract the entry point, the result are significantly affected by the high frequency noise. To determine when the signal processing method in 5.1.3 can be directly used to extract point A, the SNR of the entry signal with respect to the high frequency noise is defined in Eq (5.9).

---




(a) The acceleration entry signal




(b) The velocity entry signal

Figure 32. Simulation of the entry signal and the high frequency noise.

$$SNR_h = \frac{a_e}{a_n} \tag{5.9}$$

Where $a_e$ is the absolute value of the local minimum in the entry signal, and $a_n$ is the peak-to-peak value of the high frequency noise, as labeled in Figure 33. It is recommended by this thesis that when the $SNR_h$ is smaller than 1, a different denoising algorithm is required before applying the empirical model method to identify point A. In this research, the Variational Mode Decomposition (VMD) is used in bearing diagnostics to denoise the entry signal.

#### 5.2.1.1 Variational Mode Decomposition (VMD)

VMD is a methodology for adaptive and quasi-orthogonal signal decomposition developed by Konstantin Dragomiretskiy and Dominique Zosso [61]. The goal of VMD is to decompose an input signal into a discrete number of sub-signals (modes). Each of the modes are considered compact around a corresponding center frequency $\omega_k$. $\omega_k$ is determined by the algorithm automatically along the decomposition. This is achieved by solving the following constrained variational problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) e^{-j\omega_k t} \right] \right\|_2^2 \right\} \tag{5.10}$$

subject to

$$\sum_{k=1}^{K} u_k(t) = f(t) \tag{5.11}$$

Where f(t) is the original signal and $u_k$ is the signal of each mode after decomposition around the corresponding center frequency $\omega_k$. The bandwidth of $u_k$ is assessed by means of $H^1$ Gaussian smoothness. Konstantin Dragomiretskiy and Dominique Zosso suggested solving this problem by the augmented Lagrangian method to render the problem unconstrained as presented below:

$$\mathcal{L}(\{u_k\},\{\omega_k\},\lambda)$$

$$= \alpha \sum_{k=1}^{K} \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) e^{-j\omega_k t} \right] \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \rangle \tag{5.12}$$

Now the solution to the original variational problem is found as the saddle point of the augmented Lagrangian as shown in Eq (5.12). Then, the modes $u_k$ and the center frequency $\omega_k$ are updated as:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{5.13}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{5.14}$$

Where α is the balancing parameter of the data-fidelity constraint. α will affect the bandwidth of each mode. When α is smaller, the bandwidth is larger; conversely, when α is larger, the bandwidth is smaller.

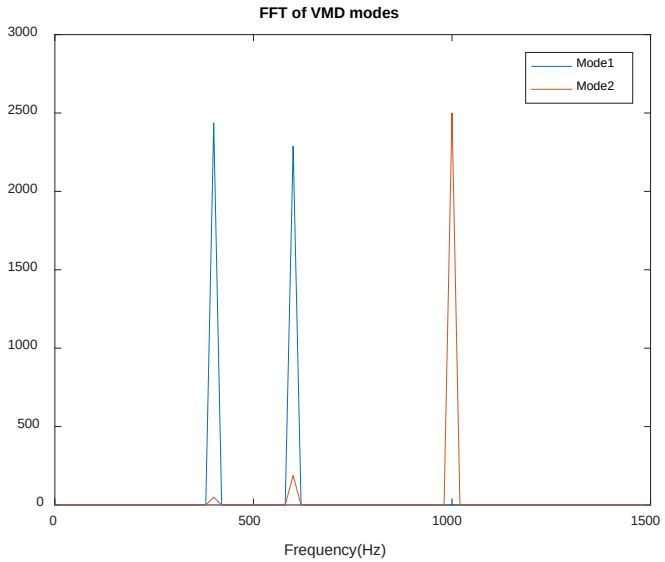
5.2.1.2 Discussion of VMD parameters

To apply VMD, the number of modes K and the balancing parameter α need to be determined. Both K and α affect the VMD output. To obtain the desired signal in the decomposition, these parameters can be selected by examining the spectrum of the signal.

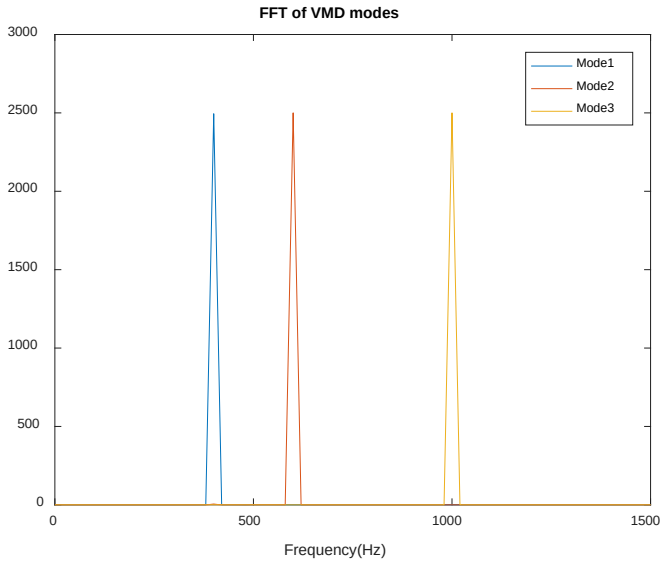
Figure 34 and Figure 35 show two examples of the VMD output with different setup of K and α.

In Figure 33, the input signal is simulated by the sum result of three sine waves at 400 Hz, 600 Hz and 1000 Hz. The wanted signal is the 400 Hz sine wave with the smallest amplitude. If K=2, only the 1000 Hz signal is decomposed correctly as shown in Figure 33(a). The 400 Hz signal is still mixed with the 600 Hz signal. This means that two modes cannot differentiate the 400 Hz signal from the 600 Hz signal. If K=3, the three signals can be decomposed correctly into three separate modes, as shown in Figure 33(b). Therefore, if the desired signal is closer to another narrow band signal in frequency, K should be large enough to differentiate the difference between the desired signal and the narrow band noise. However, K cannot be too large as larger K requires more computational efforts. The optimal method is to select the smallest K value that can differentiate the desired signal.

---



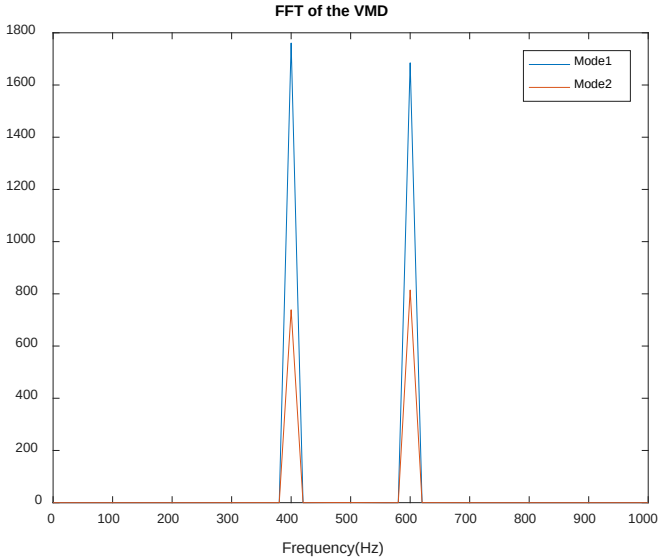

(a) K=2



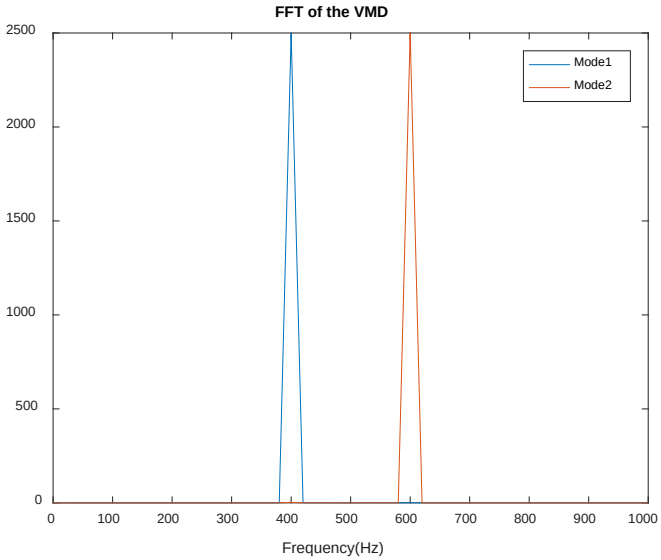

(b) K=3

Figure 33. The VMD output with different K.

Figure 34 shows the simulated input signal of two sine waves with frequencies close to each other. To separate these two sine waves into two modes of VMD, the balancing parameter α should be selected carefully. When α=1000, the decomposition result is poor because the bandwidth of each mode is too large and VMD cannot separate the two signals very well. When α=8000, VMD successfully decomposes the two signals into two different modes. Therefore, when a narrow band signal needs to be denoised from another narrow band noise with very close frequency, α should be large to narrow the bandwidth of the decomposition and therefore differentiate the signal from the noise. If the desired signal has a larger bandwidth, a small α is recommended to decrease the loss of the wanted signal.




(a) α=1000




(b) α=8000

Figure 34. The VMD outer put with different α.

5.2.1.3 VMD denoising

Figure 34 shows the result of applying VMD on the bearing entry signal. The blue line is the raw signal. The red line is the first mode (the lowest frequency mode) of the

VMD output. There are three modes altogether in this decomposition and the balancing parameter α is 2000. Figure 35 shows that the VMD effectively removes the high frequency noise from the vibration signal. This can be interpreted more clearly in the frequency domain as shown in Figure 36. The raw signal contains three dominant frequency within 0-10 kHz. The entry signal primarily consists of low frequency components, and thus, it dominates in the first mode. Therefore, by extracting the first mode from the VMD, the entry signal is separated from the high frequency noises and disturbances.

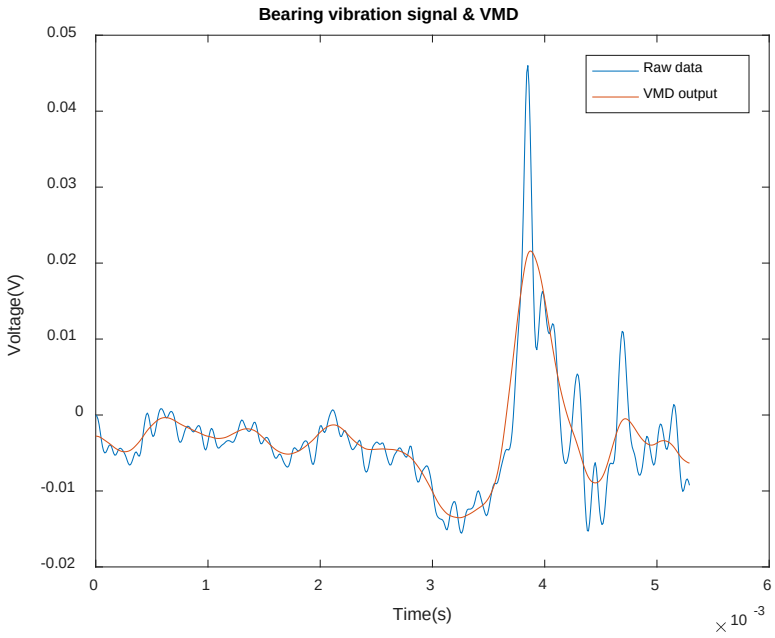



Figure 35. Bearing vibration signal with VMD denoising.

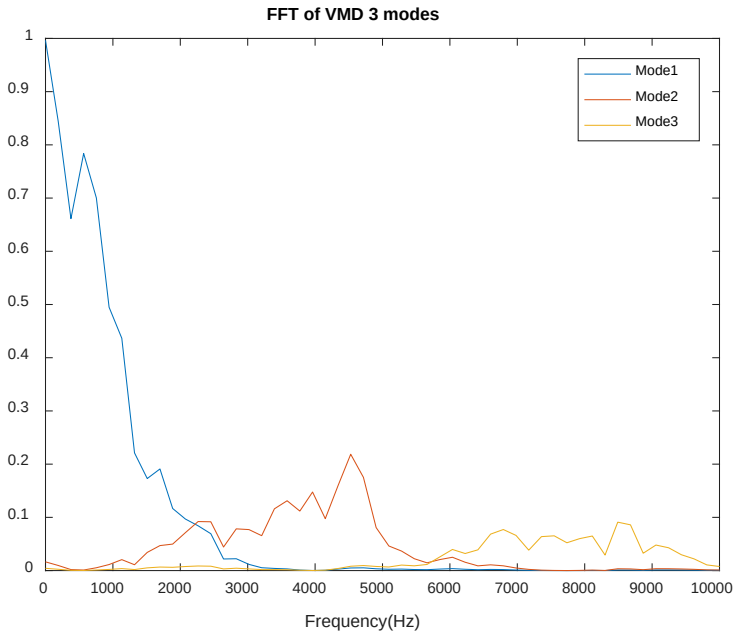



Figure 36. FFT of the 3 modes VMD.

### *5.2.2 Low frequency noise*

The VMD can efficiently remove the high frequency noise in the entry signal, but the low frequency vibration from the last impact still exists and is difficult to eliminate, as shown in Figure 37. This is because both the entry signal and the noise are mainly composed of low frequency components. In Figure 37, the low frequency noise is simulated by a sine wave at 4 kHz. The “low frequency” is relative to the entry signal, which means that the period of the noise is close or similar to the entry signal.

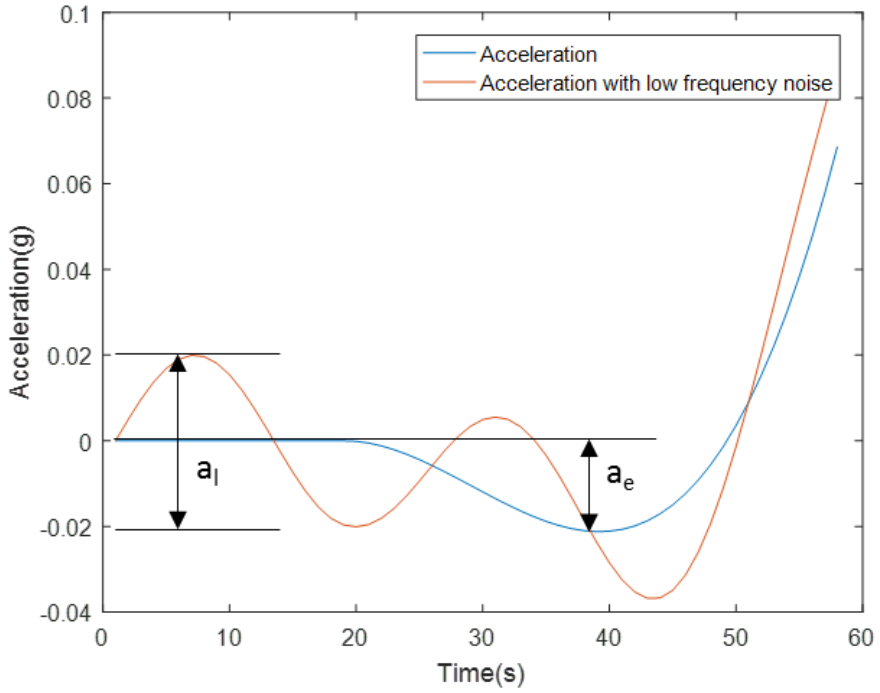



Figure 37. The entry signal with the low frequency noise.

At lower speeds, the amplitude of the entry signal can be close or even smaller than the low frequency noise. In this situation, the signal processing method proposed in 5.1.3 may not be able to provide accurate estimation of the entry point. The estimation error of the entry point depends on the time shift between the signal and the noise. In Figure 37, the entry event starts around the minimum of the sine wave. Then, the estimation error can be seen in Figure 38. It is obvious that the slopes of the linear region in the two plots of Figure 38 are quite different. This results in the measurement error in the entry point by six data points, which is about 30% of the total entry signal. At other time shift between the entry signal and the low frequency noise, the error is shown in Figure 39. As the entry signal swept through the signal wave, the error of the entry point is periodic and the bias is

above zero. Thus, when the low frequency noise is strong, the estimation result is usually smaller than the true defect.

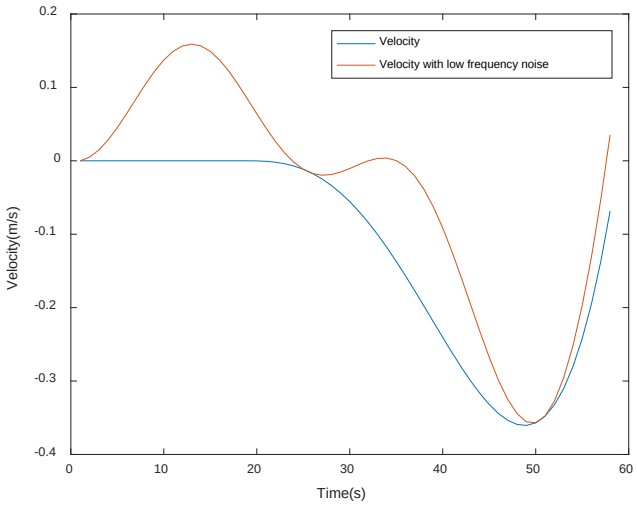



Figure 38. The velocity signal with/without the low frequency noise.

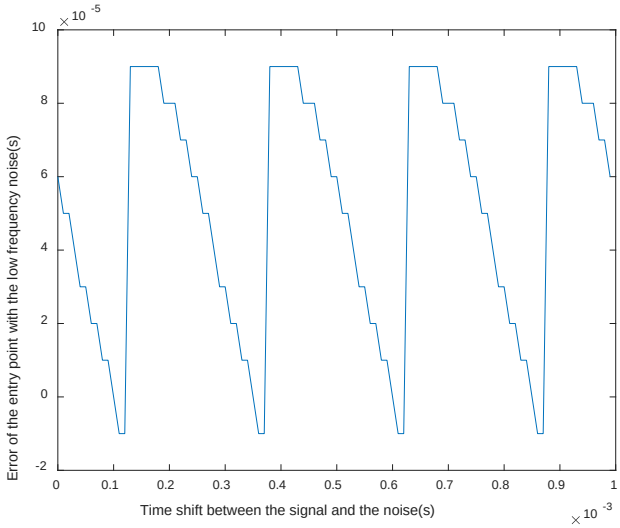



Figure 39. The error of the entry point with respect to time shift.

The error at lower spindle speed is difficult to remove because the entry point cannot even be identified visually. A better solution is to limit the lower speed bound,

which ensures that the estimation result is acceptable. To define such a bound, signal-to-noise ratio between the entry signal and the low frequency noise is defined as:

$$SNR_l = \frac{a_e}{a_l} \tag{5.15}$$

Where $a_e$ is the absolute value of the local minimum in the entry signal, and $a_l$ is the peak-to-peak value of the low frequency noise, as labeled in Figure 38. When $SNR_l \geq 1$, the average error is about 8% of the entry signal and the entry point can be detected using the signal processing method proposed in this thesis. When $SNR_l \leq 0.5$, the estimation result is more than 30% smaller than the true defect and the result is not acceptable. For a more precise estimation, a $SNR_l$ greater than 1.5 is recommended and the error is only 3% of the entry signal.

To determine the lower bound of the speed range, experiments were performed at different speeds. For all defect sizes, when the spindle speed is equal to or greater than 500 rpm, the $SNR_l$ is greater than 1. Therefore, the lower speed bound is set to be 500 rpm for all the tests in this thesis. For spindle speed lower than 500rpm, a signal processing method is proposed as follows. The low frequency noise can be considered as a decaying sinusoidal signal with certain period and decay factor. The period and decay factor can be solved by curve fitting to the vibration signal before the local minimum of the entry event. Then, the low frequency noise can be removed by subtracting the curve from the vibration signal.

## 5.3 The signal-processing scheme

After point A, B and C are successfully identified from the vibration signal, the entry-to-peak time $t_e$ and the peak-to-impact time $t_p$ can be determined. The process of the signal processing method to obtain $t_e$ and $t_p$ can be summarized as below:

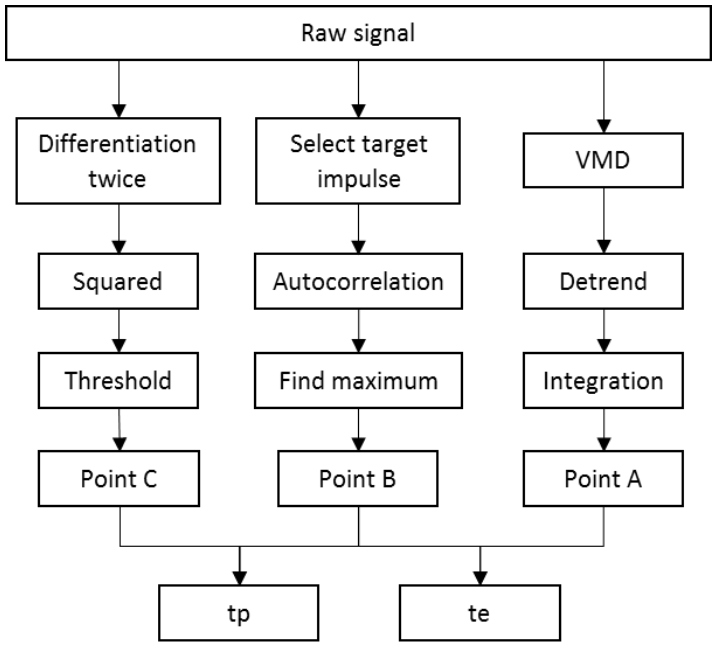



Figure 40. Flowchart of the signal processing procedure.

In Chapter 6 and Chapter 7, these signal processing methods is applied to the vibration signal and the defect estimation result is shown.

## 5.4 Summary

This chapter proposes the signal processing method to extract the key points A, B and C from the vibration signal. Point A is extracted using the empirical model of the bearing entry signal. Point B is extracted using cross-correlation. Point A is extracted using the differential algorithm. Then, the performance of the signal processing method is

evaluated with respect to high and low frequency noises. For the high frequency noise, the VMD is recommended to remove the noise from the entry signal. For the low frequency noise, the estimation results depend on the signal-to-noise ratio between the entry signal and the low frequency noise. When the signal-to-noise ratio is too small, the estimation result can be smaller than the true defect. Finally, the complete signal-processing scheme for defect size estimation is provided.

# CHAPTER 6. DEFECT SIZE ESTIMATION WITHOUT MACHINING

This chapter describes the validation of the bearing test system and the no-cutting experiment design. Then, the experiment results with observations under the no-cutting conditions are shown and discussed. Finally, the defect size estimation results for both inner and outer defects under different speeds are provided to validate the precision of the estimation method.

## 6.1 Validation of the test system and design of experiments

### *6.1.1 Validation of the test system*

Given the spindle speed, the velocity of each bearing component can be calculated according to Eq (2.6) through (2.9). However, when the spindle is running, sliding and friction between the balls and raceways may cause variations in the velocity of each bearing component comparing to theoretical calculation. Since the method used in this thesis relies on time domain information, the velocity of each bearing rotary component (inner race and balls) should be validated first. The validation is performed by comparing the actual ball-pass frequencies with the calculated bearing frequencies. Rotation test from 500-3000 rpm were performed and the results are shown in Figure 41 and Figure 42. There is a decreasing trend in the ball-pass frequency ratio as the speed increases. This means that at higher speeds, the bearing sliding becomes more prominent and the relative speeds between the cage and the outer/inner races decrease. Since the variation is only about 1%, it does not affect the final estimation results. Both Figure 41 and Figure 42 show that the experimental

ball-pass frequency ratios are similar to the calculated results 0.42 and 0.58. Therefore, the outer race ball-pass frequency ratio is determined to be 0.42 and the inner race ball-pass frequency ratio 0.58, as the theoretical calculation result in Chapter 3. This also means that the bearing in the test system is suitable and each component runs at velocities close to theoretical calculations.

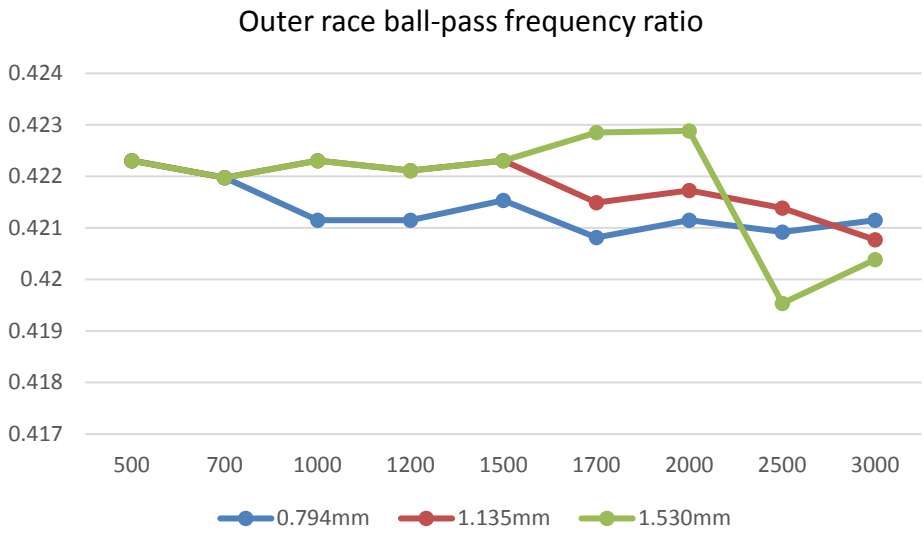



Figure 41. The ball-pass frequency ratio for outer race defect.

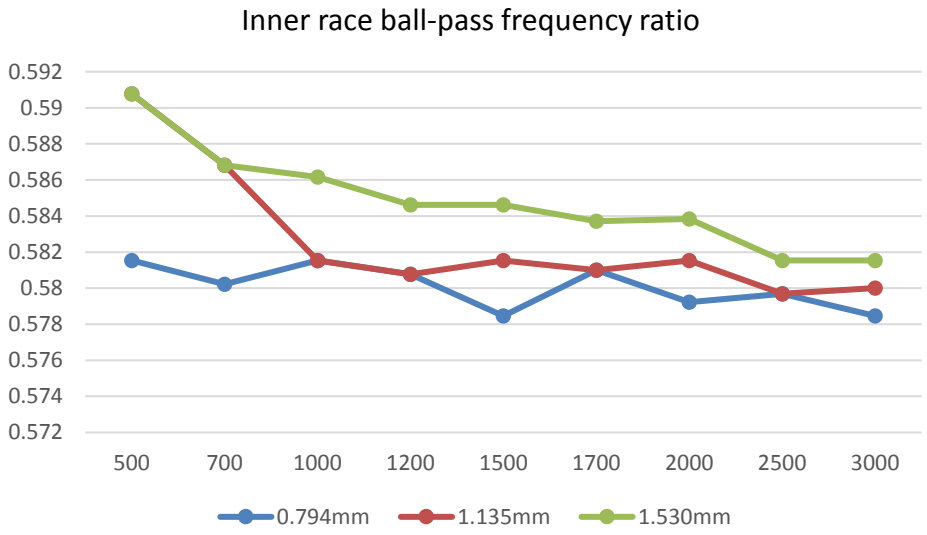



Figure 42. The ball-pass frequency ratio for inner race defect.

### *6.1.2 Design of the no-cutting experiments*

To validate the time domain bearing diagnostics method as proposed in this thesis, 3 different defect widths 0.794 mm, 1.135 mm and 1.530 mm are seeded on both inner and outer races. Each type of defect is tested under 9 different spindle speeds 500 rpm to 3000 rpm. This means that 9x3x2=54 different conditions are tested. For each condition, the vibration data are collected and analyzed using the signal processing method as proposed in Chapter 5. All the experiments are performed under the same bearing preload controlled by a torque wrench.

## 6.2 Results of $t_e$ and $t_p$ under no-cutting condition

As discussed in Chapter 4, for the outer race defect, the Hertzian quadrilateral model is used. Therefore, the measurement result of $t_e$ and $t_p$ for the outer race are shown separately. In each of the experiment, about 2 seconds of data at each speed is selected to

obtain $t_e$ and $t_p$ using the signal processing method introduced in 5.1. The results are averaged to get $t_e$ and $t_p$. While for the inner race defect, the first estimation model is used. Thus, only the total time $t_t$ from entry point A to impact point C is required to calculate the estimation result. The measurement result of $t_t$ for the inner race defect will be shown in a later section.

### *6.2.1 The measurement results of $t_e$ and $t_p$*

The measurement results of $t_e$ and $t_p$ at different speeds are shown in Figure 43 and 6.4. Both $t_e$ and $t_p$ show a linear trend with respect to 1/rpm. Figure 43 shows that the variances in $t_e$ among the three defect widths are small. This result is due to the fact that the bearing preload is constant throughout the experiments. This result agrees well with the assumption of the Hertzian quadrilateral model in Chapter 4, that $t_e$ is dependent on spindle speed and independent of the defect size. Figure 44 shows that $t_p$ is significantly affected by defect size. When the defect size is large, $t_p$ becomes longer; when the defect size is small, $t_p$ becomes shorter. Therefore, $t_p$ is determined by the speed and defect size. The relationship between $t_e$, $t_p$ and bearing load cannot be clearly seen from these two plots, but it is discussed later. So far, all of the observations from the experiment results for the outer race agree with the assumptions of the third defect estimation model.

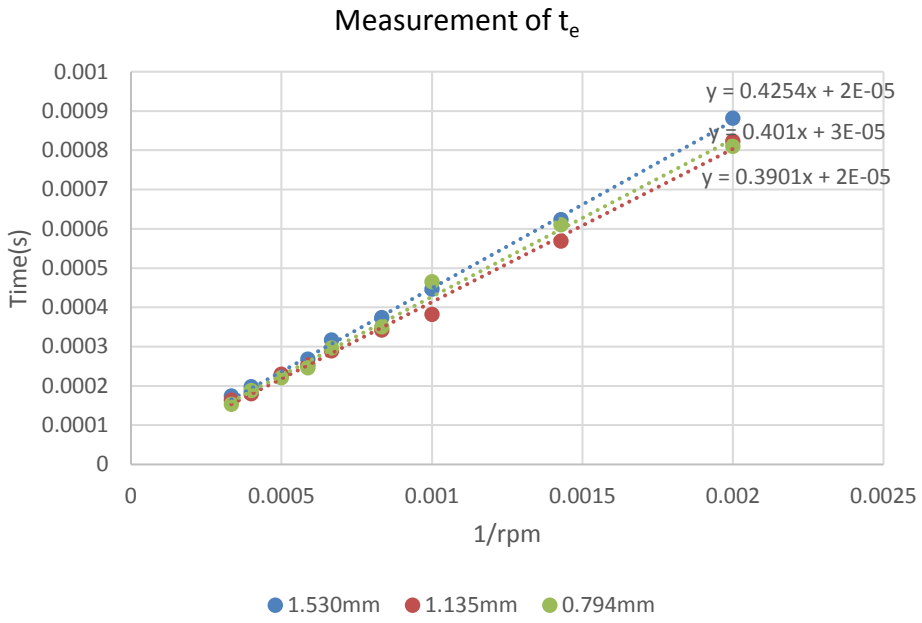



Figure 43. $t_e$ of outer race.

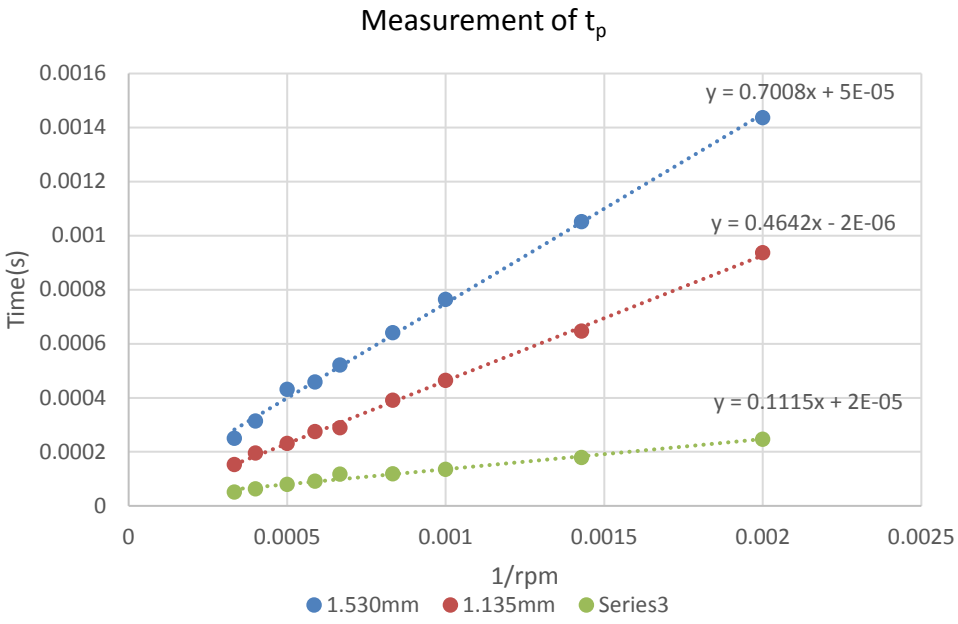



Figure 44. $t_p$ of outer race.

One of the advantages of the automatic signal processing method proposed in Chapter 5 is that, statistics of $t_e$ and $t_p$ can be analyzed based on the large data sets. Without the signal processing method, manual analysis is impractical. Therefore, this is the first time that the statistics of $t_e$ and $t_p$ are analyzed and discussed. More observations of $t_e$ and $t_p$ will be shown in the next several paragraphs and the results provide insight into the relationship between $t_e$, $t_p$ and bearing load that supports the Hertzian quadrilateral model.

#### 6.2.1.1 Time domain plot of $t_e$ and $t_p$

The time domain plots of both $t_e$ and $t_p$ are significantly affected by bearing load. Figure 45 shows an example of $t_e$ at 1700 rpm for the 1.530 mm outer defect. In this plot, $t_e$ vibrates between the upper and lower bound and looks like a sine wave with a specific period. Figure 46 shows the phenomenon more obviously in frequency domain with the most dominant frequency at the shaft frequency. This phenomenon can be consistently observed from the experiment results of the other two defect sizes as well. For $t_p$, similar result can be observed. Therefore, it can be concluded that both $t_e$ and $t_p$ are significantly affected by bearing loads. This is due to the change of the relative distance and bearing load between inner and outer races when the shaft rotates, as shown in Figure 47. Thus, when the ball enters the defect, if the relative distance between the inner and outer races is smaller, the contact force is also smaller. In addition, this means that the signal processing algorithm developed in Chapter 5 is precise enough to detect the effects of shaft unbalance on $t_e$ and $t_p$. This observation also supports the Hertzian quadrilateral model that the bearing contact force should be considered, because it significantly affects $t_e$ and $t_p$.

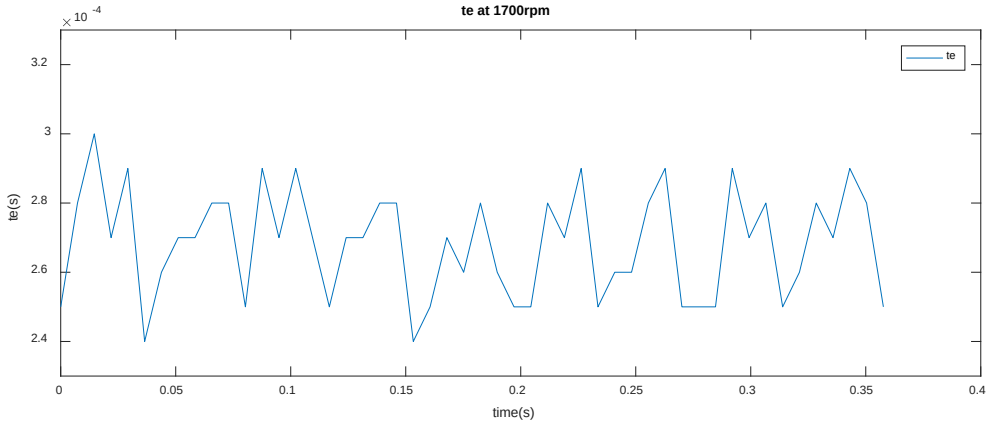



Figure 45. Time domain plot of $t_e$ at 1700rpm.

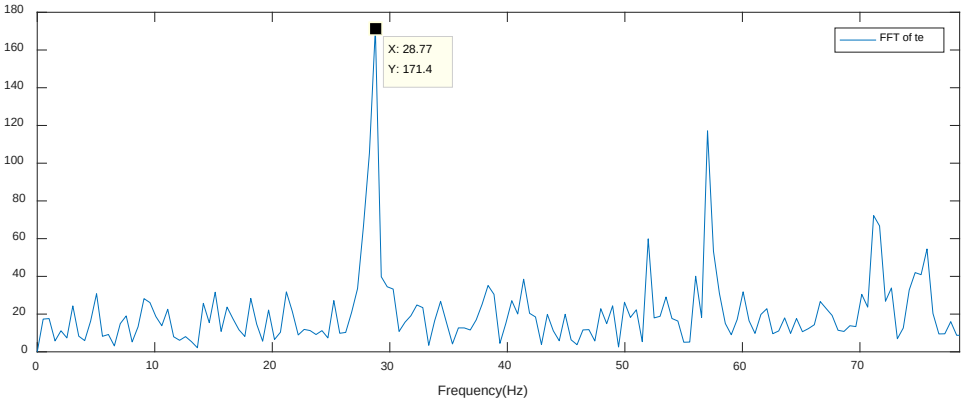



Figure 46. Frequency domain plot of $t_e$ at 1700rpm.

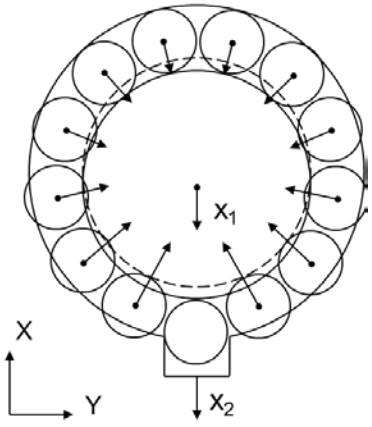



Figure 47. Load distribution in bearing.

#### 6.2.1.2 Statistics of $t_e$ and $t_p$

The statistics of $t_e$ and $t_p$ affect the precision of defect size estimation, and therefore will be quantified. The histograms of $t_e$ and $t_p$ at two different speeds are shown in Figure 48. At lower speeds, there are two separate peaks in the histogram. One of them centers on the upper bound of the time domain plot as shown in Figure 45, and the other centers around the lower bound. The result is very similar to two Gaussian distributions with an overlap between each other. This means that the local maximum and minimum in one period of the "sine wave" as shown in Figure 45 is close to a Gaussian distribution. Moreover, the distance between these two peaks is similar to the amplitude of the "sine wave" and proportional to the standard deviation of $t_e$ and $t_p$ as shown in Figure 49. At lower speeds, the standard deviation of $t_e$ and $t_p$ are larger, and therefore the two peaks in the histogram become farther away from each other. At lower speeds, the standard deviations of $t_e$ and $t_p$ become smaller, and the two peaks in the histogram starts to merge and the result becomes very similar to a Gaussian distribution.

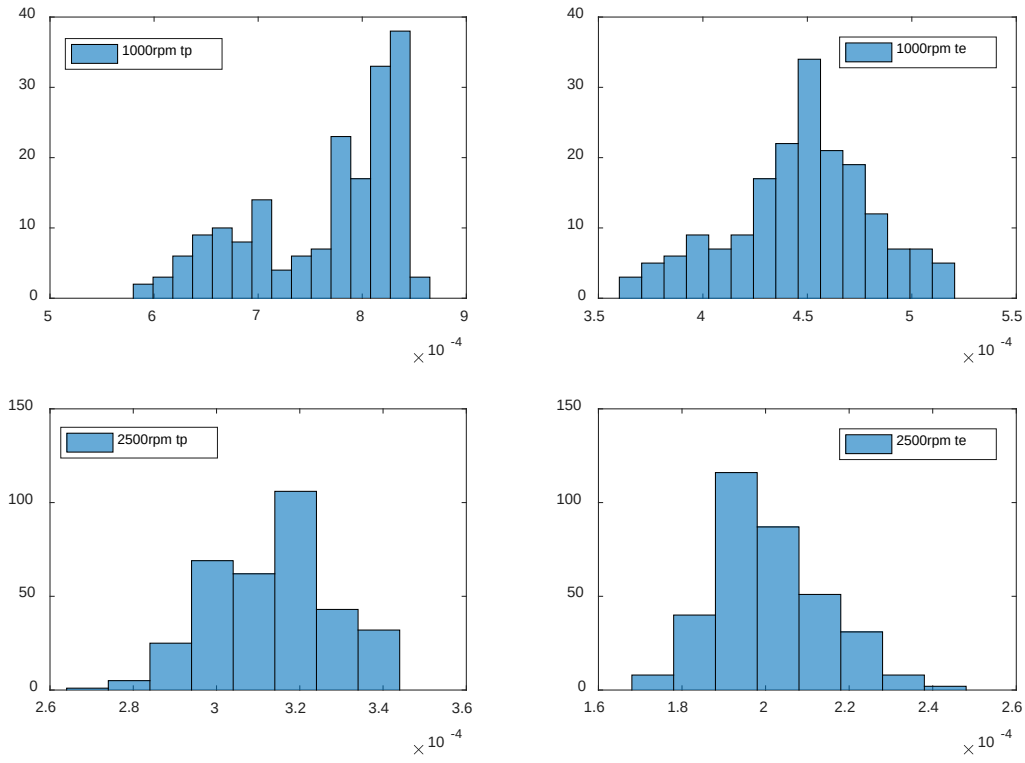
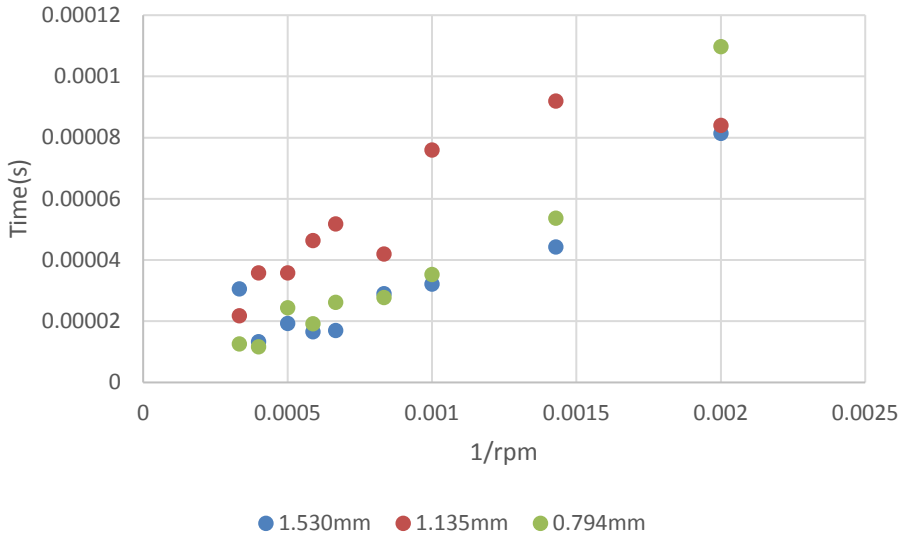
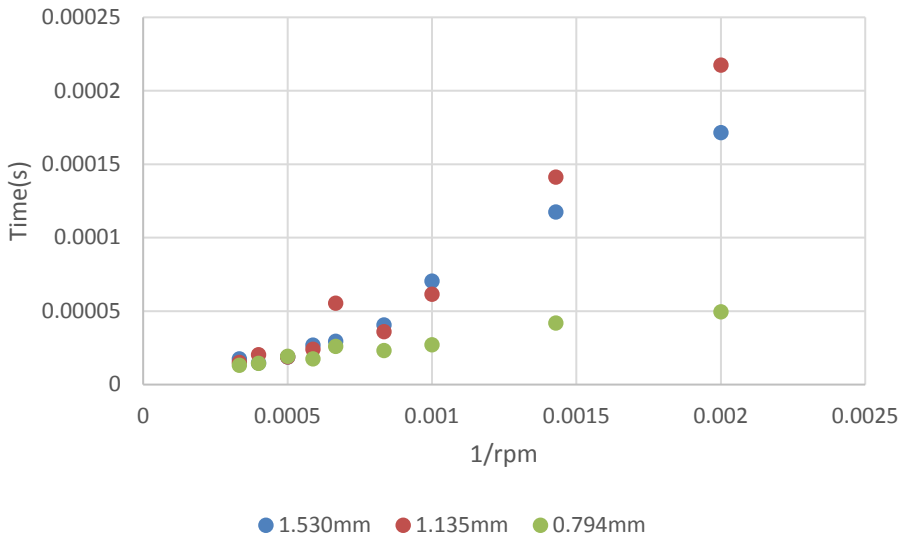



Figure 48. Histograms of $t_e$ and $t_p$.




(a) Standard deviation of $t_e$



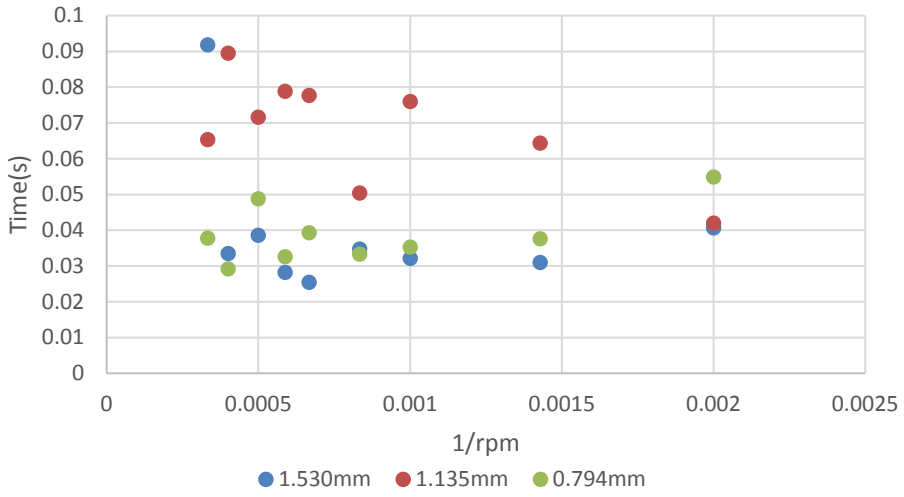

(b) Standard deviation of $t_p$

Figure 49. Standard deviation of $t_e$ and $t_p$ with respect to speeds.

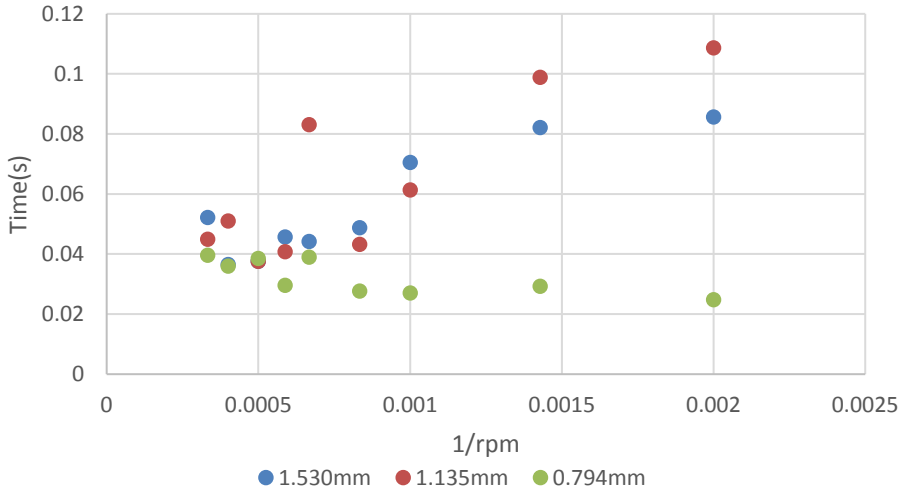
Figure 49 shows that as spindle speed increases, the standard deviation of $t_e$ and $t_p$ become smaller. However, this does not mean that the measurement becomes more precise. On the contrary, the uncertainty level might increase as speed increases because the same standard deviation at lower speeds takes larger portion in the measurement results at higher speeds, when both $t_e$ and $t_p$ become smaller. Therefore, it is important to define the normalized standard deviation with respect to speed as below to represent the precision of $t_e$ and $t_p$:

$$STD_n = STD/n \tag{6.1}$$

Where STD is the standard deviation of $t_e$ and $t_p$, and n is the spindle speed in RPM. Then, the normalized standard deviation of $t_e$ and $t_p$ are shown in Figure 50.




(a) Normalized standard deviation of $t_e$




(b) Normalized standard deviation of $t_p$

Figure 50. Normalized standard deviation of $t_e$ and $t_p$ with respect to speeds.

It can be seen that, the speed in the middle part of the test range, which is from 1000 rpm to 2000 rpm, has the smallest normalized standard deviation. At lower speeds, the signal to noise ratio of the entry signal is small, therefore, the normalized standard deviation is larger. Thus, the lower speed range is limited by the signal to noise ratio of the entry signal. At higher speeds, the normalized standard deviation increases due to the

smaller $t_e$ and $t_p$. To decrease the uncertainty level, a higher sampling frequency of the data acquisition system is required to increase the number of data points in $t_e$ and $t_p$. Therefore, the upper speed range is determined by the sampling frequency of the data acquisition system and the defect size (defect size affects $t_p$). If it is required that $t_e$ and $t_p$ should be represented by at least m data points at a specific spindle speed r, then the sampling frequency should be at least:

$$f_s \geq \frac{m}{\min(t_e, t_p)} \tag{6.2}$$

As derived in Eq (4.39), $t_e$ is a function of contact force and spindle speed, and $t_p$ is determined by defect size and spindle speed. Therefore, the lower bound of the sampling frequency depends on contact force, defect size and spindle speed. For this thesis, sampling frequency is selected as 100 kHz and this guarantees that for the smallest defect and at the highest speed, $t_p$ is still represented by at least five data points. Resolution of the estimation at different speeds for both outer and inner race defects can be calculated as:

$$Res_o = \frac{2\pi \cdot 0.42 \cdot n(R + r)}{60 \cdot f_s} \tag{6.3}$$

$$Res_i = \frac{2\pi \cdot 0.58 \cdot n(R - r)}{60 \cdot f_s} \tag{6.4}$$

At higher spindle speed and sampling frequency, the resolution of the test system is higher. Substituting sampling frequency $f_s$, spindle speed n, R and r into Eq (6.3) and (6.4), resolution of the test system is listed in Table 4. For applications with higher

precision requirement, the sampling frequency should be large enough to ensure that $t_e$ and $t_p$ are represented by enough number of data points.

Table 4. Resolution of the test system at 100 kHz sampling frequency.

| Spindle speed (RPM) | Resolution for outer race defect (mm) | Resolution for inner race defect (mm) |
|---|---|---|
| 500 | 0.0063 | 0.0063 |
| 700 | 0.0089 | 0.0088 |
| 1000 | 0.0127 | 0.0126 |
| 1200 | 0.0152 | 0.0152 |
| 1500 | 0.0190 | 0.0189 |
| 1700 | 0.0215 | 0.0215 |
| 2000 | 0.0253 | 0.0253 |
| 2500 | 0.0317 | 0.0316 |
| 3000 | 0.0380 | 0.0379 |

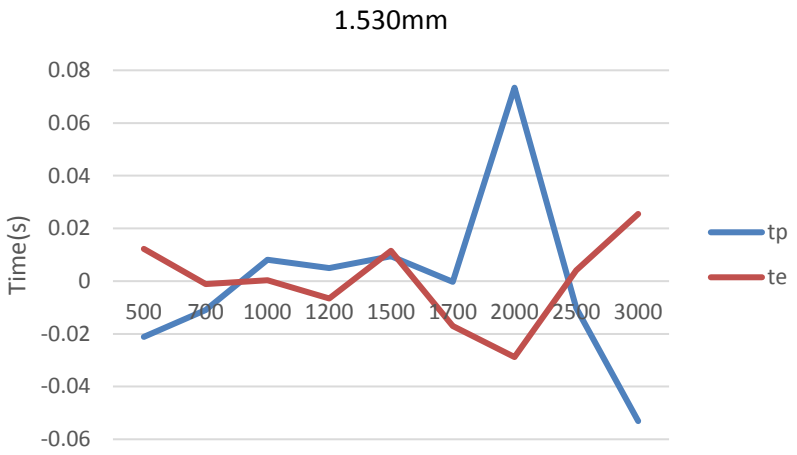
6.2.1.3 Correlation between $t_e$ and $t_p$

When $t_e$ increases, $t_p$ should decrease when the defect size is fixed, because the distance from the point the ball loses contact with the entry edge to the impact point becomes shorter. Therefore, $t_e$ and $t_p$ should have a negative correlation. This can be validated through the measurement result of $t_e$ and $t_p$. Firstly, the linear trend in the measurement results of $t_e$ and $t_p$ is removed, as shown in Figure 51. The negative trend between $t_e$ and $t_p$ can been seen obviously from Figure 51. To quantify the negative trend, the correlation coefficient is calculated between the detrended $t_e$ and $t_p$ as in Eq (6.5).

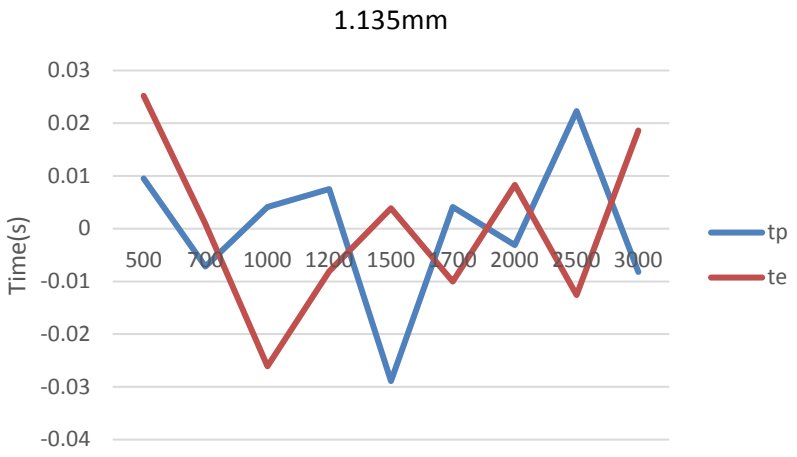
$$r_{ep} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{6.5}$$

To interpret the correlation coefficient, -1~-0.7 means a strong negative linear relationship, -0.7~-0.5 means a moderate negative linear relationship, and -0.5~-0.3 means a weak

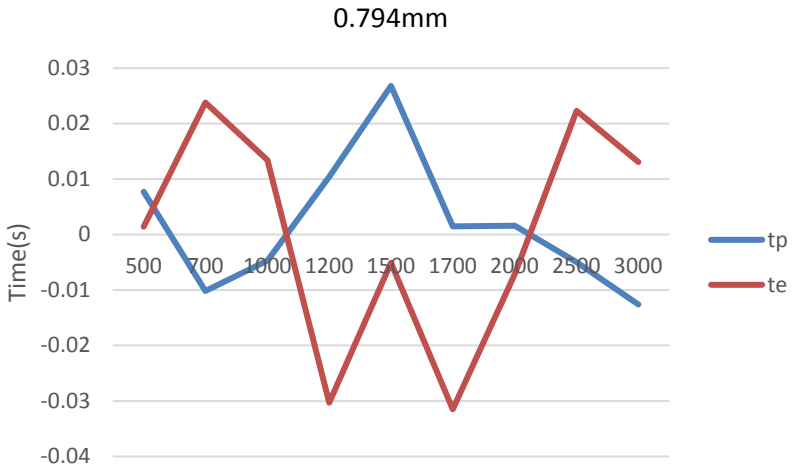
negative linear relationship. From all the tested bearings, the 1.530 mm defect size has the most obvious defective response with the best signal-to-noise ratio among the three defects, and it has the strongest negative correlation coefficient -0.84. For the 0.794 mm and 1.135 mm defects, the correlation coefficient is also negative, but not as strong as the 1.530 mm defect. This decrease in correlation may be due to the low signal-to-noise ratio at lower speed and uncertainty at higher speed. The negative correlation between $t_e$ and $t_p$ is another evidence to demonstrate the validity of the Hertzian quadrilateral model.




(a) 1.530mm




(b) 1.135mm




(c) 0.794mm

Figure 51. Detrended $t_e$ and $t_p$.

Table 5. Correlation coefficient of $t_e$ and $t_p$.

| Defect size (mm) | Correlation coefficient of $t_e$ and $t_p$ |
|---|---|
| 1.530 | -0.84 |
| 1.135 | -0.46 |
| 0.794 | -0.55 |

6.2.1.4 Discussion of $t_e$ and $t_p$ with respect to temperature

The above experiments and analysis are performed under room temperature. When the bearing runs for a longer period with a defect, excessive heat can be generated. If the rate of heat dissipation is less than the rate of heat generation, the temperature within the bearing will rise. Therefore, the discussion of the proposed method with respect to different temperature conditions is presented.

Under temperatures higher than the room temperature, the bearing races and ball will expand due to the heat. The higher temperature will increase bearing loads and

decrease the defect width due to thermal expansion. According to the derivation of $t_e$ and $t_p$ as shown in Eq (4.37) and (4.39), $t_e$ will increase and $t_p$ will decrease. Therefore, the defect size estimation result will decrease correspondingly as defect width decreases under heat expansion. Therefore, according to the bearing model, the measurement result of the defect width undergoing thermal expansion should be smaller than the defect size at room temperature. The defect width at room temperature can be calculated from the defect width undergoing thermal expansion using the temperature, bearing material property, bearing geometry and defect size.

#### 6.2.1.5 Discussion of $t_e$ and $t_p$ with respect to material property

One of the major causes of spall defect is brinnelling. This phenomenon is characterized by sudden impact loading during bearing operation, such as spindle crash, or heavy loading while the bearing is not rotating. The result is a small dent on a ball bearing or a line spall for roller bearings left on the surface of the bearing raceway. For either defect, when the rolling element passes by the dent, extremely high-pressure ridges form in the vicinity of the dented region. Therefore, the material of the defect's entry edge becomes the hardest location in the bearing. If the difference in hardness is large, the influence of this phenomenon on defect size estimation should be considered.

Assuming that the entry edge is harder, then the deflection between the ball and the entry edge when the ball just rolls into the defect region decreases. Therefore, $t_e$ will decrease and while $t_p$ remains constant. However, because the actual width of the defect does not change, the measurement result of the defect width will be smaller that the actual size. The error depends on the material property of the entry edge.

## 6.3 The total time $t_t$ from entry to impact

The measurement results of $t_e$, $t_p$ and some observations have been discussed in 6.2. The total time $t_t$ from the entry point A to the impact point C can be obtained by adding $t_e$ and $t_p$. The measurement result of $t_t$ for both outer and inner race defect is presented in this section.

### *6.3.1 Total time $t_t$ for the outer race defect*

The total time $t_t$ for the three outer race defects are shown in Figure 52. From the plot, a clear linear trend can be seen between $t_t$ and 1/rpm. The small intercepts and the large R factors of all the least-square fit lines mean that the linearity of the measurement is very good. By comparing Figure 52 with Figure 43 and Figure 44, it can be seen that the linearity of $t_t$ is better than the linearity of $t_e$ and $t_p$. The comparison result can be seen more clearly from Table 6 by comparing the R value of the linear estimation. This result is due to the strong negative correlation coefficient as shown in Table 5. Both $t_e$ and $t_p$ are sensitive to the variation in the bearing loads, but their sum, the total time $t_t$, is not. Thus, when the contact force increases, $t_e$ increases and $t_p$ decreases to make sure that the overall time $t_t$ is not affected by the variance of the contact force. Therefore, the measurement of $t_t$ is very consistent and not sensitive to bearing loads, and it is a suitable metric to estimate the defect size.

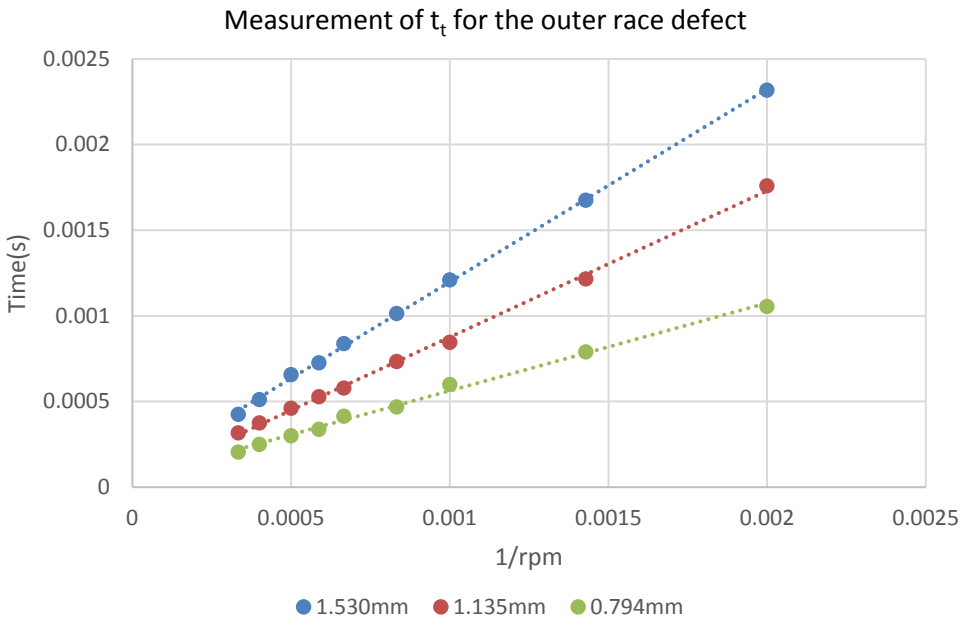



Figure 52. Total time $t_t$ for the outer race defect.

Table 6. The R value of the linear estimation for $t_e$, $t_p$ and $t_t$.

| Defect size (mm) | $t_e$ | $t_p$ | $t_t$ |
|---|---|---|---|
| 1.530 | 0.9989 | 0.9977 | 0.9995 |
| 1.135 | 0.9938 | 0.9983 | 0.9983 |
| 0.794 | 0.9950 | 0.9831 | 0.9954 |

### *6.3.2 Total time $t_t$ for the inner race defect*

The measurement result of the total time $t_t$ for the inner race defect is shown in Figure 53. Similar to the result of the outer race defect in Figure 52, a strong linear relationship exists between $t_t$ and 1/rpm, as suggested by the small intercepts and large R values.

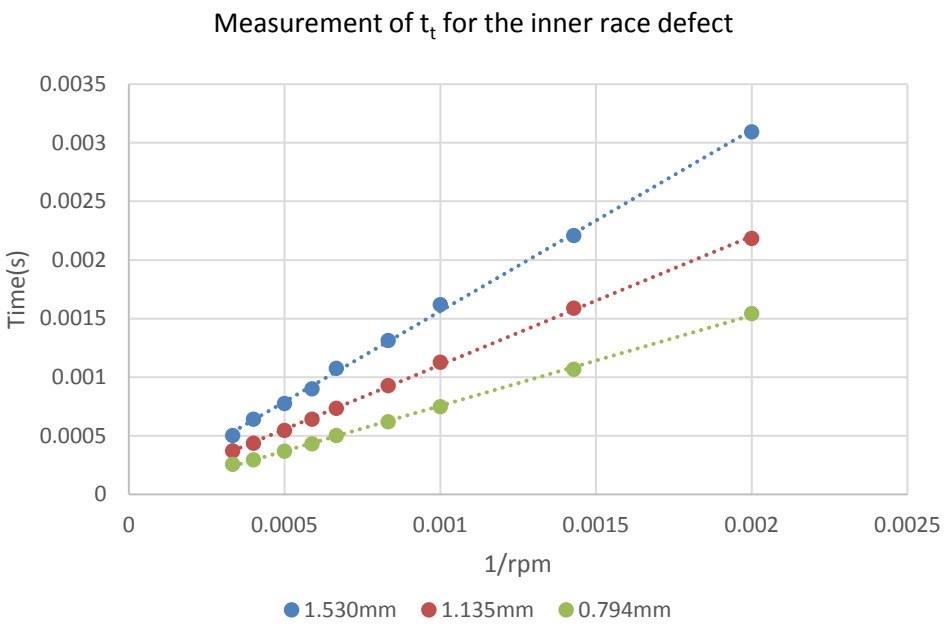



Figure 53. Total time $t_t$ for the inner race defect.

## 6.4 Defect size estimation results and discussions

After the total time $t_t$ is extracted, the defect size can be estimated based on the estimation model proposed in Chapter 4. The defect size estimation results for both outer and inner races at different speeds are shown in this section. For all the experiments, the estimation results are close to the true defect sizes. Therefore, the time domain estimation method proposed in this thesis is validated.

### *6.4.1 Outer race defect*

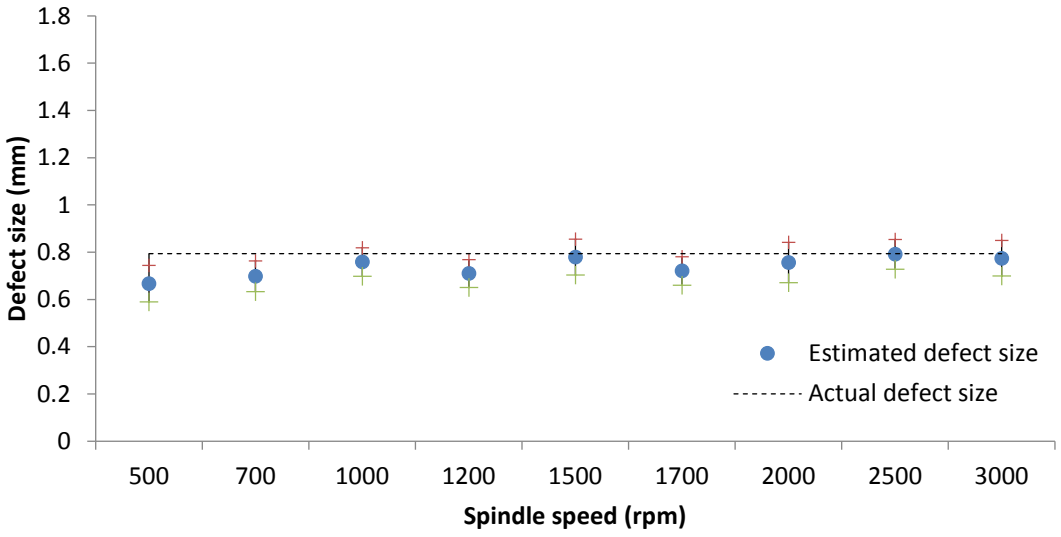
For the outer race defect, the defect size is estimated using the third estimation model. The results are shown in Figure 54(a) through (c). From the three plots, the following general trend can be obtained:

1. The estimation result within the speed range 1000 rpm to 2000 rpm are close to the true defect size.

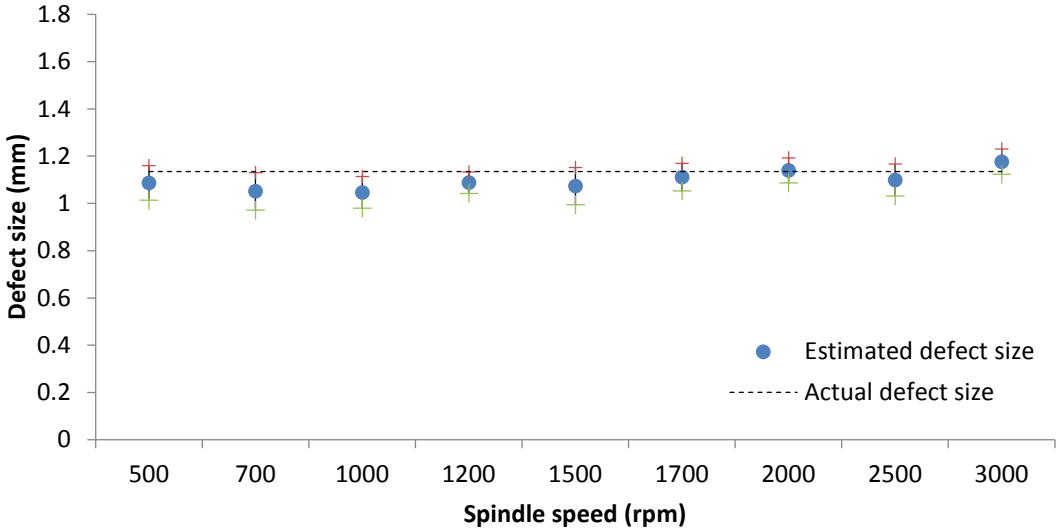
2. At lower speeds, the estimation result is a lower than the true defect size.

3. At higher speeds, the estimation result is a higher than the true defect size.

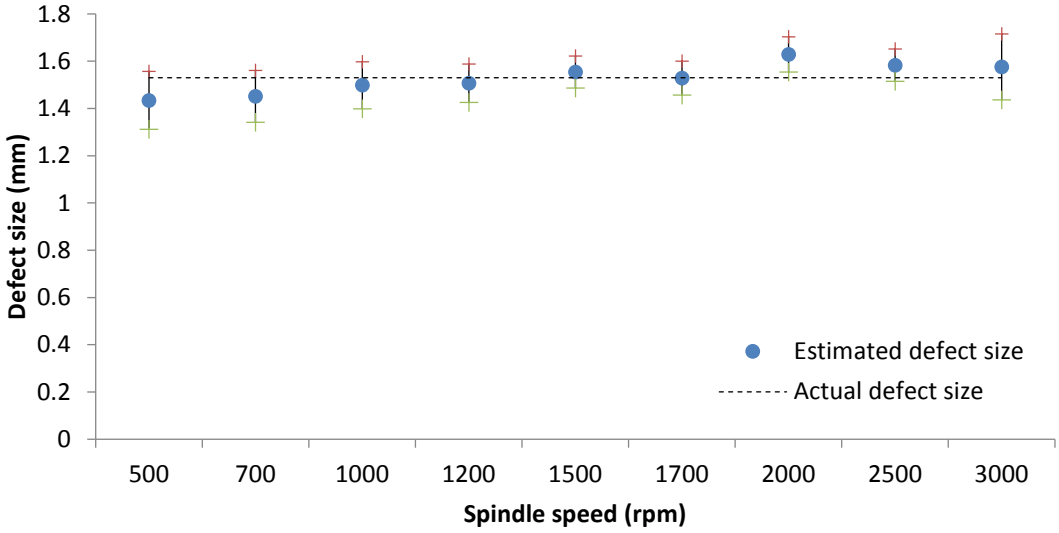
These observations agree with the statistics analysis in 6.2.1. At lower speeds, due to the smaller signal-to-noise ratio between the entry signal and the low frequency noise, the estimation result will be smaller than the true defect size. At higher speeds, because the time from entry to exit point is very short, the uncertainty in the measurement result will increase. In addition, the bearing sliding issue becomes more prominent at higher speeds as suggested in Figure 41 and Figure 42, and therefore, the estimation result will be a little bit larger than the true defect size.




(a) 0.794mm outer




(b) 1.135mm outer




(c) 1.530mm outer

Figure 54. Defect size estimation results for the outer race.

The estimation results and the relative error in percentage using different estimation models are listed in Table 7 and Table 8. It can be seen that the estimation result provided by isosceles trapezoidal model is about 25% smaller than the true defect size, while the

result of Hertzian quadrilateral model is about 30% larger than the true defect size. The estimation result given by Hertzian quadrilateral model has the smallest error, which is usually smaller than 8%. The average error of the estimation result of the three defects are shown in Table 10.

Table 7. Defect size estimation results of three different estimation models (mm).

| Spindle speed (RPM) | 0.794 mm | | | 1.135 mm | | | 1.530 mm | | |
|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 |
| 500 | 0.5114 | 0.9145 | 0.6672 | 0.852 | 1.5226 | 1.0875 | 1.1227 | 2.005 | 1.4344 |
| 700 | 0.5352 | 0.957 | 0.6987 | 0.8238 | 1.4723 | 1.0515 | 1.1359 | 2.0285 | 1.4518 |
| 1000 | 0.5816 | 1.0398 | 0.7593 | 0.8201 | 1.4658 | 1.0468 | 1.1721 | 2.0929 | 1.4982 |
| 1200 | 0.5459 | 0.9761 | 0.7104 | 0.8524 | 1.5233 | 1.088 | 1.1788 | 2.1048 | 1.5067 |
| 1500 | 0.6015 | 1.0754 | 0.7802 | 0.8404 | 1.5019 | 1.0733 | 1.217 | 2.1728 | 1.5547 |
| 1700 | 0.5551 | 0.9925 | 0.721 | 0.8705 | 1.5555 | 1.1112 | 1.1964 | 2.1361 | 1.5291 |
| 2000 | 0.5823 | 1.0412 | 0.7568 | 0.8928 | 1.5953 | 1.1401 | 1.2727 | 2.2718 | 1.6288 |
| 2500 | 0.6081 | 1.0872 | 0.7916 | 0.9087 | 1.6238 | 1.0989 | 1.2399 | 2.2135 | 1.5834 |
| 3000 | 0.5955 | 1.0647 | 0.7751 | 0.9209 | 1.6454 | 1.1765 | 1.2355 | 2.2056 | 1.5765 |

Table 8. Relative error of three different estimation models.

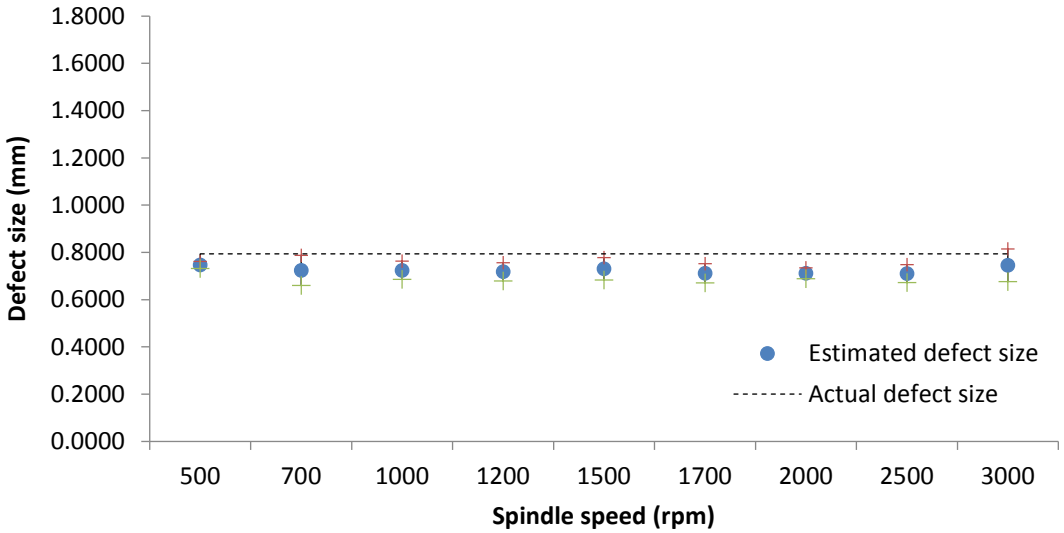
| Spindle speed (RPM) | 0.794 mm | | | 1.135 mm | | | 1.530 mm | | |
|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | M1 | M2 | M3 | M1 | M2 | M3 |
| 500 | 0.36 | 0.15 | 0.16 | 0.25 | 0.34 | 0.04 | 0.27 | 0.31 | 0.06 |
| 700 | 0.33 | 0.21 | 0.12 | 0.27 | 0.30 | 0.07 | 0.26 | 0.33 | 0.05 |
| 1000 | 0.27 | 0.31 | 0.04 | 0.28 | 0.29 | 0.08 | 0.23 | 0.37 | 0.02 |
| 1200 | 0.31 | 0.23 | 0.11 | 0.25 | 0.34 | 0.04 | 0.23 | 0.38 | 0.02 |
| 1500 | 0.24 | 0.35 | 0.02 | 0.26 | 0.32 | 0.05 | 0.20 | 0.42 | 0.02 |
| 1700 | 0.30 | 0.25 | 0.09 | 0.23 | 0.37 | 0.02 | 0.22 | 0.40 | 0.01 |
| 2000 | 0.27 | 0.31 | 0.05 | 0.21 | 0.41 | 0.01 | 0.17 | 0.48 | 0.06 |
| 2500 | 0.23 | 0.37 | 0.01 | 0.20 | 0.43 | 0.03 | 0.19 | 0.45 | 0.03 |
| 3000 | 0.25 | 0.34 | 0.02 | 0.19 | 0.45 | 0.04 | 0.19 | 0.44 | 0.03 |

Table 9. Average absolute and relative error of three different estimation models.

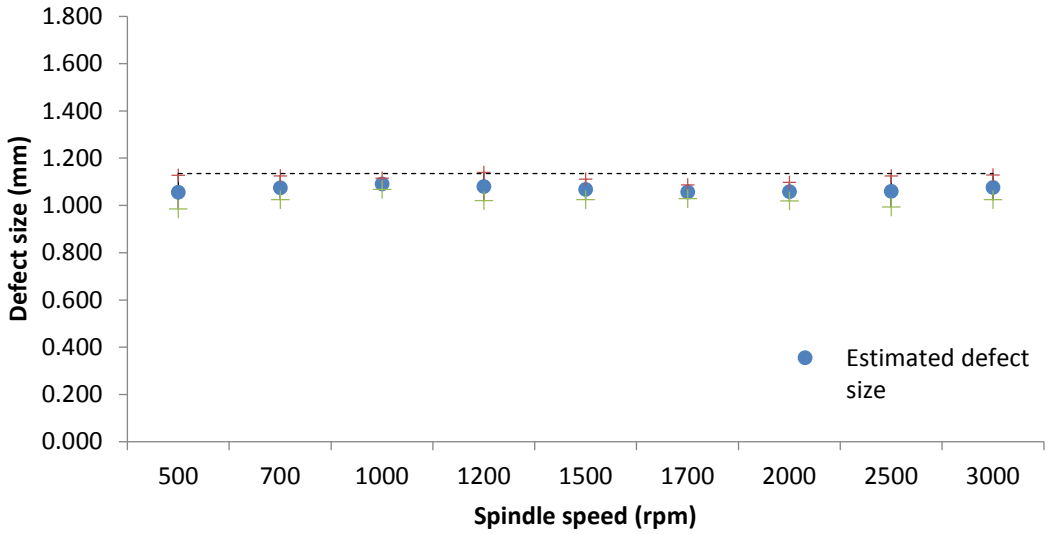
| Defect size (mm) | M1 | | M2 | | M3 | |
|---|---|---|---|---|---|---|
| | Absolute (mm) | Relative | Absolute (mm) | Relative | Absolute (mm) | Relative |
| 0.794mm | 0.225 | 27% | 0.222 | 28% | 0.054 | 7% |
| 1.135mm | 0.270 | 24% | 0.410 | 36% | 0.048 | 4% |
| 1.530mm | 0.333 | 22% | 0.607 | 40% | 0.050 | 3% |

### *6.4.2 Inner race defect*

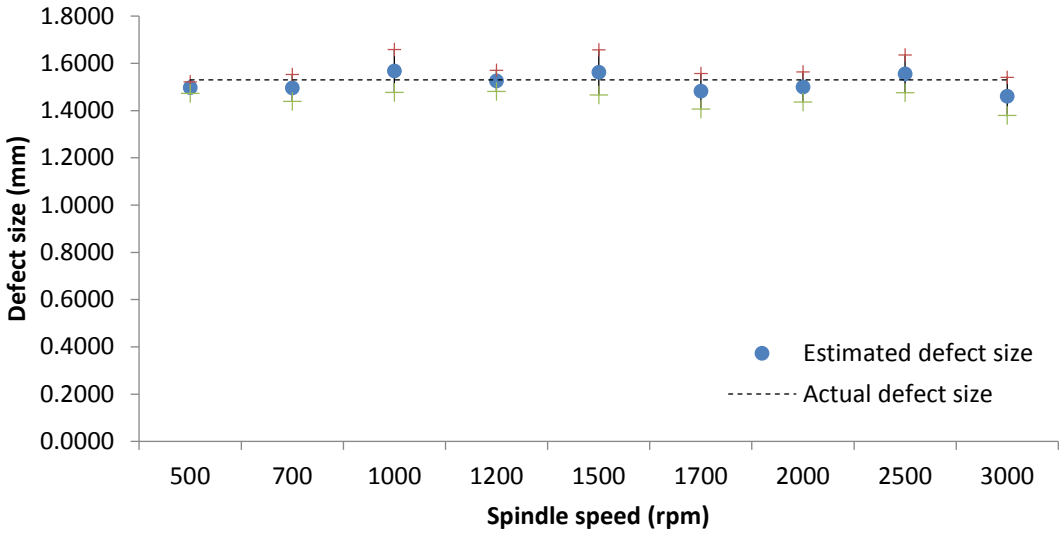
For the inner race defect, the defect size is estimated using the first estimation model. The results are shown in Figure 55(a) through (c). Since on average, the signal-to-noise ratio between the entry signal and the low frequency noise is larger than the outer race defects, the estimation results for the inner race defect at lower speeds is better than the outer race defect. The estimation results, relative error and average error are shown in Table 10 and Table 11 respectively.




(a) 0.794mm inner




(b) 1.135mm inner




(c) 1.530mm inner

Figure 55. Defect size estimation results for the inner race.

Table 10. The defect size estimation results and relative error for the inner race.

| Spindle speed (RPM) | 0.794 mm | | 1.135 mm | | 1.530 mm | |
|---|---|---|---|---|---|---|
| | Result (mm) | Relative error | Result (mm) | Relative error | Result (mm) | Relative error |
| 500 | 0.7468 | 0.06 | 1.056 | 0.07 | 1.4970 | 0.02 |
| 700 | 0.7239 | 0.09 | 1.075 | 0.05 | 1.4959 | 0.02 |
| 1000 | 0.7246 | 0.09 | 1.091 | 0.04 | 1.5672 | 0.02 |
| 1200 | 0.7186 | 0.09 | 1.080 | 0.05 | 1.5254 | 0.01 |
| 1500 | 0.7313 | 0.08 | 1.068 | 0.06 | 1.5614 | 0.02 |
| 1700 | 0.7121 | 0.10 | 1.057 | 0.07 | 1.4818 | 0.03 |
| 2000 | 0.7120 | 0.10 | 1.058 | 0.07 | 1.4995 | 0.02 |
| 2500 | 0.7110 | 0.10 | 1.059 | 0.07 | 1.5551 | 0.02 |
| 3000 | 0.7459 | 0.06 | 1.077 | 0.05 | 1.4598 | 0.05 |

Table 11. Average absolute and relative error for inner race defect.

| Defect size (mm) | Absolute (mm) | Relative |
|---|---|---|
| 0.794 | 0.09 | 10.9% |
| 1.135 | 0.06 | 5.1% |
| 1.530 | 0.02 | 1.5% |

## 6.5 Summary

This chapter validates the defect size estimation method proposed in this thesis with the no-cutting experiments. At the beginning of this chapter, the ball-pass frequency ratios of the bearing test system are validated at all tested speeds. Then, the experimental design is described. After that, the experiment results including measurement of $t_e$, $t_p$ and $t_t$ with other observations under the no-cutting conditions are shown and discussed. The statistics of the measurement results are performed due to the automatic signal-processing algorithm. All the results and observations agree with the assumptions in the Hertzian-quadrilateral defect estimation model as proposed in Chapter 4. The defect size estimation results for both inner and outer defects are close to the true defect size with the maximum error of 7% for the outer race and 10.9% for the inner race.

# CHAPTER 7. DEFECT SIZE ESTIMATION DURING MACHINING

During machining operations, spectrum of the vibration signal can be significantly affected by the cutting forces and excessive noises. Because variations in cutting conditions can produce different cutting forces, to use the traditional spectrum comparison based diagnostic method, it is necessary to collect the baseline data for every cutting condition. However, this method is impractical. Therefore, the traditional diagnostic method is not suitable under operating conditions. In Chapter 6, experimental results have shown that the new absolute method as proposed in this thesis can provide good estimation result under no-cutting conditions. In this chapter, the same method is used to estimate defect size when the machine is cutting. First, the effects of cutting forces on bearing vibration signals will be shown and discussed. Then, the signal processing method for cutting conditions will be introduced. Finally, the cutting experiment and estimation results based on the new time domain method are presented.

## 7.1 Effects of cutting forces on bearing vibration signals

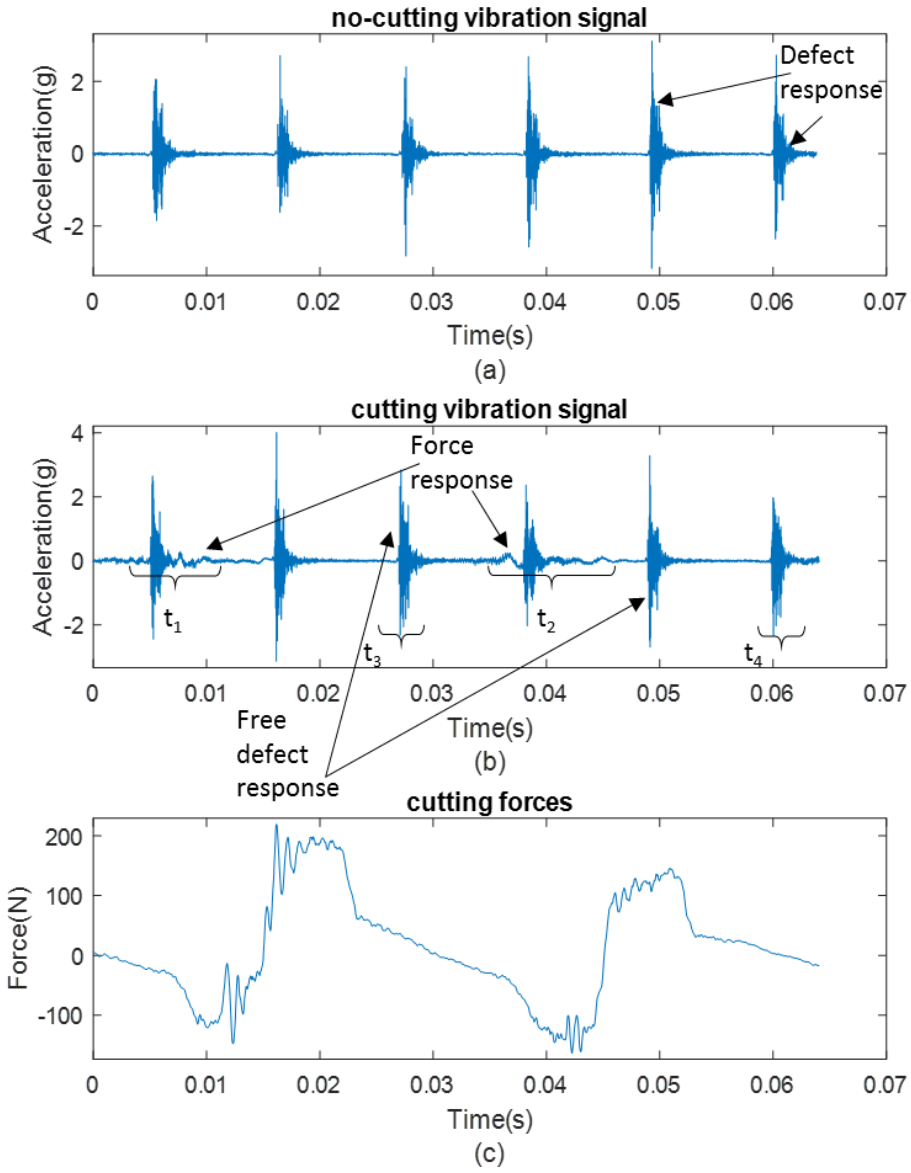



Figure 56. Vibration signals no-cutting/cutting and cutting force.

When the spindle is purely rotating, the vibration signal is dominated by the impulsive responses excited by the bearing defect, defined as the "defect response," as shown in Figure 56(a). When the spindle performs cutting operations, the responses excited by the cutting forces, defined as the "force response," appears as ripples in the vibration

signal as shown in Figure 56(b). These ripples tend to last longer than one period of the defect response with more significant amplitudes than the bearing entry signal. In addition, they may overlap and "destroy" the entry signal of the defect response. Therefore, the force response is a prevalent noise for defect size estimation. The cutting force is given in Figure 56(c) as a reference. The time range in Figure 56 is about one shaft revolution. The cutting conditions for this cut are shown in Table 12.

Table 12. Cutting condition for Figure 7.1 (c).

| | |
|---|---|
| Number of tooth | 2 |
| Spindle speed | 1000 rpm |
| Depth of cut | 0.02 in |
| Feed rate | 3 in/min |
| Radio immersion rate | 100% |

By comparing Figure 56(b) and Figure 56(c), it can be seen that the vibration responses during $t_1$ and $t_2$ are excited by the sharp decrease in the cutting forces correspondingly. When the cutting force changes smoothly, such as during $t_3$ and $t_4$, the defect response is not affected much by the force response. The cutter has two indexed teeth and the outer race ball-pass frequency is about 5 times the shaft frequency. Therefore, at about every two peaks, the defect response is free from contaminated by the force response, which is defined as the "free defect response." Therefore, the free defect response can be used to estimation defect size during machining operations. However, this is only suitable when the ball-pass frequency is larger than the number of teeth of the cutter. A more general signal processing method will be introduced in 7.2 for defect estimation during machining operations.

## 7.2 The signal processing method for cutting operations

Figure 57 provides a zoomed window in the entry signal contaminated by the force response. It can be seen that even though the force response is dominant and the entry event is significantly affected by the noise, the first peak in the entry signal (point B in Chapter 4) can still be seen as labeled in Figure 57.

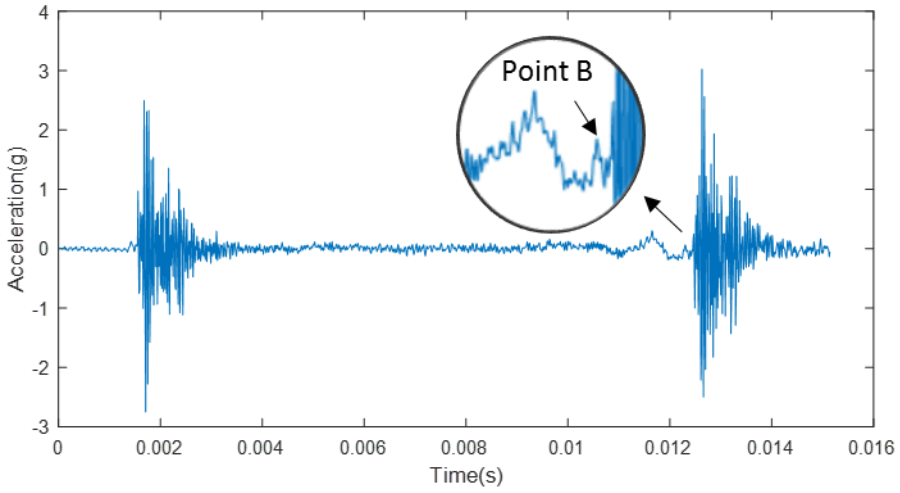



Figure 57. Detailed vibration signals during machining.

Therefore, the information of $t_p$ can be extracted from the cutting data. The signal processing method to extract $t_p$ is the same as the no-cutting condition in Chapter 5. To identify point B, the cross-correlation between a target impulsive response and the vibration signal is used. The maximum value in the cross-correlation signal corresponds to the index of point B. The impact point C is identified based on differentiation and threshold. Then, $t_p$ can be extracted from the cutting data. To estimate the defect size, $t_e$ is still needed. $t_e$ can be obtained by two methods as follows:

Method 1

If the bearing initial contact force $F_0$ is known, $t_e$ can be calculated using Eq (4.39). Then, the defect size can be estimated as:

$$d = f(F_0, t_p) \tag{7.1}$$

Where f can be determined from the estimation model in Chapter 4.

Method 2

If the bearing load is unknown and the prerequisite for the frequency defect response is not satisfied, $t_e$ can be extracted from the no-cutting data. The no-cutting data can be collected during the small time slot after the tool finishes its pass, as shown in Figure 58.

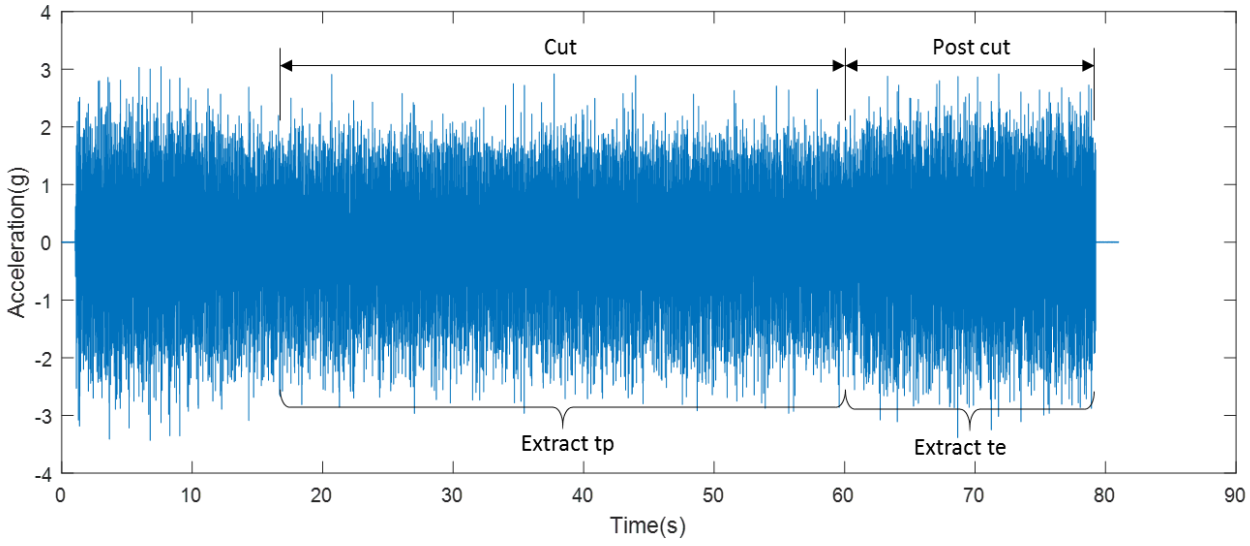



Figure 58. The whole cutting data set including the after-cut data.

Then, the defect size can be estimated as:

$$d = f(t_e, t_p) \tag{7.2}$$

Where $t_e$ is the entry to peak time extracted from the post cutting data, and $t_p$ is obtained from the cutting data. f can be determined from the estimation model in Chapter 4. The second method is more convenient because it is not constrained by the ball-pass frequency or the number of teeth. In addition, all the information can be obtained from the vibration signal. However, a prerequisite for this method is that $t_p$ is constant during and after the machining process and this prerequisite need to be validated. The comparison of $t_p$ during and after machining process is shown in the next section. The design of the experiment to validate this method is introduced in the next section.

## 7.3 Design of the cutting experiments

To validate the time domain bearing diagnostics method under cutting conditions, the same experiment setup described in the previous chapter was used in these cutting experiments. To test the robustness of the estimation method under cutting conditions, all of the experiments in Chapter 6 were repeated again under two different feed rates. The feed rate with other cutting parameters are listed in Table 13.

Table 13. Cutting experiment conditions.

| Feed rate (in/minutes) | Depth of cut (inches) | Radio immersion rate |
|---|---|---|
| 2.36 | 0.02 | 100% |
| 3 | 0.02 | 100% |

Therefore, 9 different speeds, 3 defect sizes for both inner and outer races, 2 different cutting conditions were tested. Therefore, the total number of cuts $n_{total}$ was calculated to be:

$$n_{total} = 9 \times 3 \times 2 \times 2 = 108 \quad (7.3)$$

For each of the 108 cutting conditions, the vibration data during and after cut are collected and analyzed using the signal processing method as proposed in 7.1.2. The cutting forces in three axes orthogonal to each other are also collected. All the experiments were performed under the same bearing preload controlled by a torque wrench.

## 7.4 Measurement result of $t_t$

To estimate the defect size under cutting conditions without knowing the bearing load, the $t_e$ in the vibration signal after cutting can be used. However, a prerequisite that $t_p$ during and after cutting does not change much needs to hold, as mentioned in 7.1.2. Therefore, 7.3 first validates the prerequisite by comparing the $t_p$ during and after cutting. Then, the correlation coefficient of $t_e$ and $t_p$ is analyzed to show that the assumption of the third estimation model holds as well under cutting conditions. Finally, the measurement result of $t_t$ is shown for both the outer and inner race defects.

### *7.4.1 Comparison of $t_p$ during cut and post-cut*

The $t_p$ for both outer and inner defects during and after machining operations are shown in Figure 59 and Figure 60. Each of the data points in this plot are the averaged $t_p$ based on about 2 seconds of the vibration signal during/after machining. It can be seen that the measurement result of $t_p$ during machining is close to the result after machining for both feed rates. The variation of $t_p$ in percentage can be referred from Table 14. The maximum variation of $t_p$ occurs at 2500rpm for the 0.794mm inner race defect, which is 16%. The average variation of $t_p$ for the outer race is 3.03%, and 3.99% for the inner race.

Thus, the variation of $t_p$ caused by the cutting forces is very small. Therefore, $t_e$ after machining and $t_p$ during machining can be added together to provide estimation of the defect size.

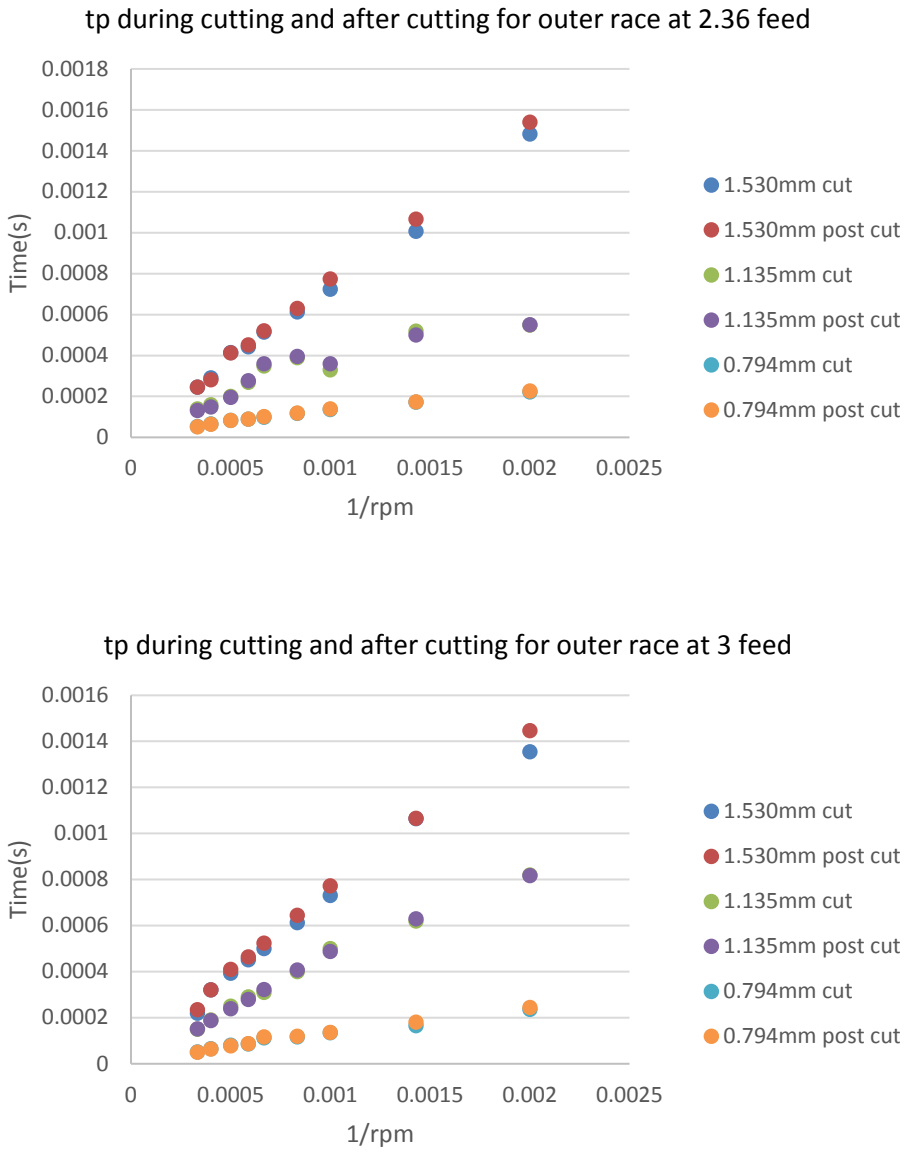



Figure 59. $t_p$ during and after cutting for the outer race.

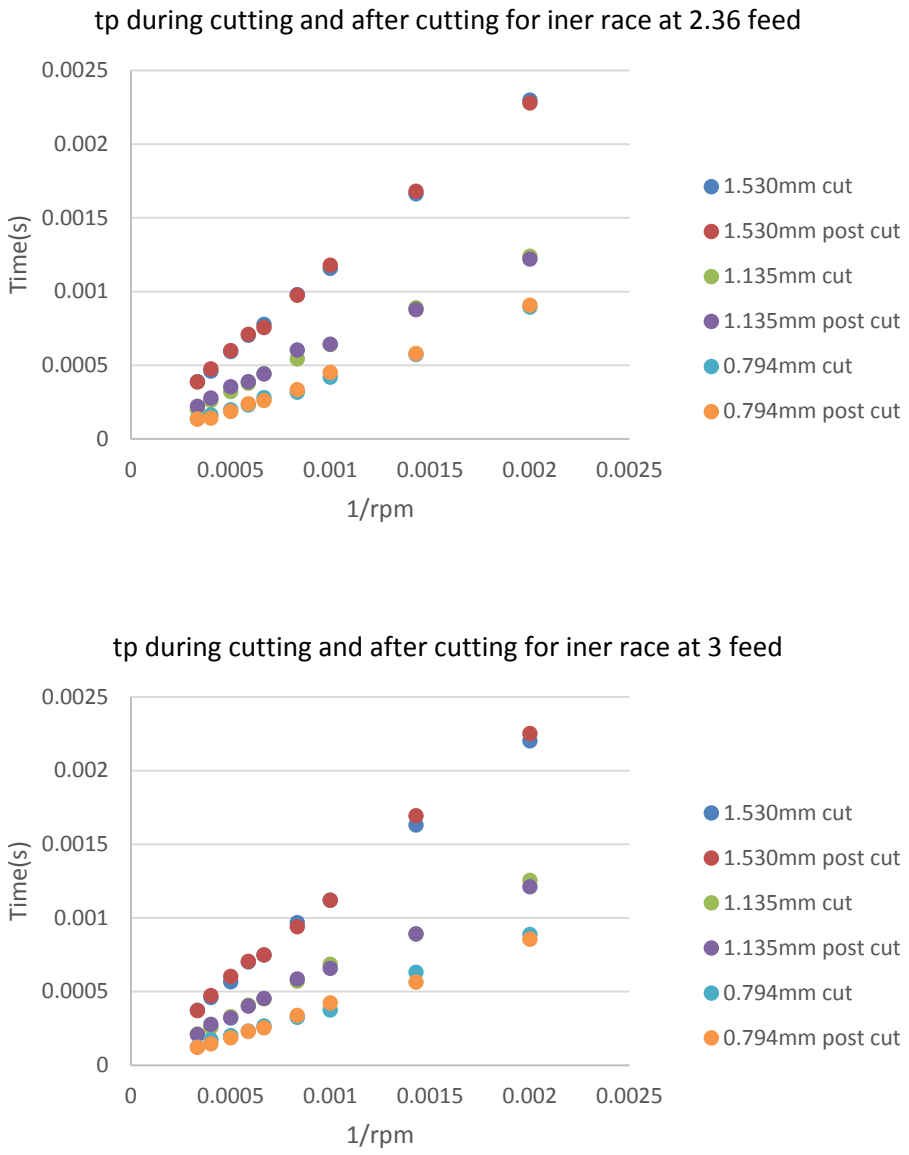


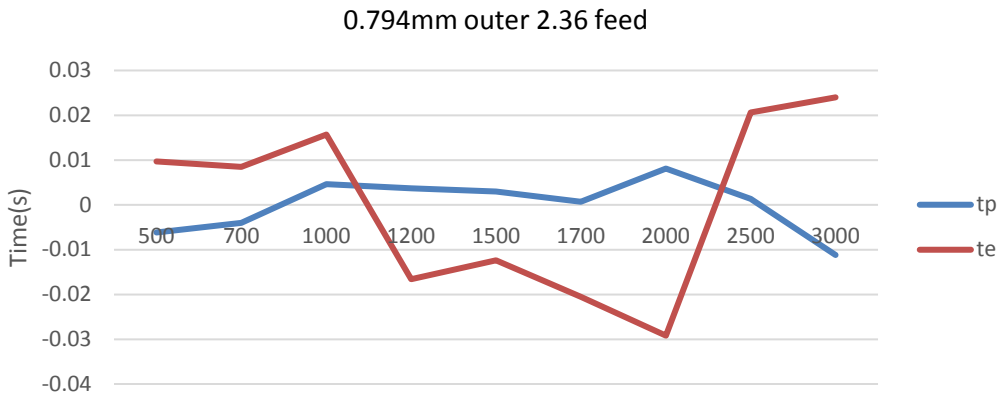

Figure 60. $t_p$ during and after cutting for the inner race.

Table 14. Variation of $t_p$ during and after cutting (%).

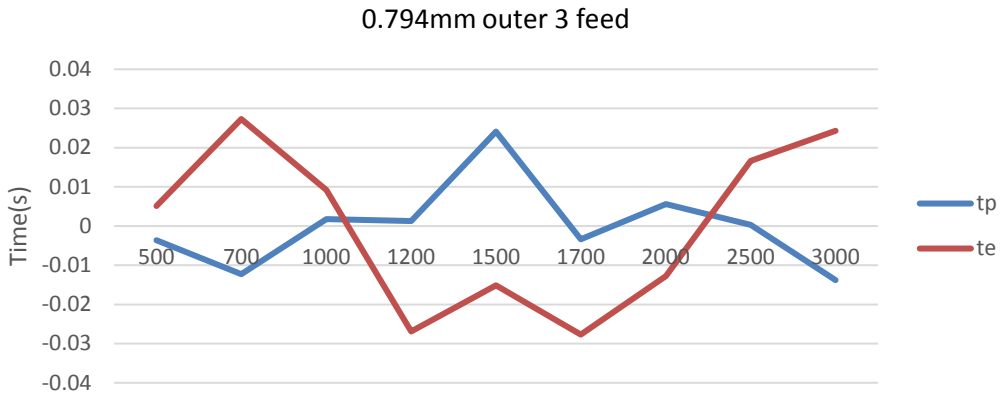
| Spindle Speed (RPM) | Outer race | | | | | | Inner race | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.794 mm | | 1.135 mm | | 1.530 mm | | 0.794 mm | | 1.135 mm | | 1.530 mm | |
| | 2.36 feed | 3 feed | 2.36 feed | 3 feed | 2.36 feed | 3 feed | 2.36 feed | 3 feed | 2.36 feed | 3 feed | 2.36 feed | 3 feed |
| 500 | 2.2 | 3.1 | 0.2 | 0.4 | 4.0 | 6.8 | 1.4 | 3.8 | 1.6 | 3.5 | 0.9 | 2.3 |
| 700 | 0.8 | 9.5 | 3.6 | 1.4 | 5.8 | 0.1 | 1.3 | 10.6 | 1.5 | 0.3 | 1.0 | 3.8 |
| 1000 | 2.8 | 0.4 | 9.1 | 2.3 | 6.9 | 5.7 | 7.7 | 12.9 | 0.5 | 4.0 | 1.8 | 0.1 |
| 1200 | 0.6 | 2.7 | 1.7 | 1.7 | 2.7 | 5.2 | 5.5 | 4.4 | 11.0 | 2.6 | 0.6 | 2.8 |
| 1500 | 3.2 | 4.1 | 3.1 | 3.8 | 1.2 | 4.6 | 7.1 | 4.3 | 0.8 | 0.6 | 2.9 | 0.1 |
| 1700 | 0.3 | 1.9 | 3.0 | 3.9 | 2.1 | 2.6 | 3.7 | 1.2 | 3.3 | 1.8 | 1.0 | 0.6 |
| 2000 | 0.4 | 3.2 | 2.1 | 4.4 | 0.2 | 4.1 | 5.3 | 7.4 | 10.1 | 3.0 | 1.3 | 6.3 |
| 2500 | 3.7 | 1.7 | 6.4 | 1.4 | 3.7 | 0.1 | 12.3 | **16.5** | 6.7 | 7.8 | 3.5 | 2.7 |
| 3000 | 1.7 | 3.3 | 6.1 | 0.6 | 0.3 | 6.9 | 1.8 | 2.7 | 11.3 | 2.4 | 0.6 | 0.7 |

### *7.4.2 Correlation coefficient of $t_e$ and $t_p$*

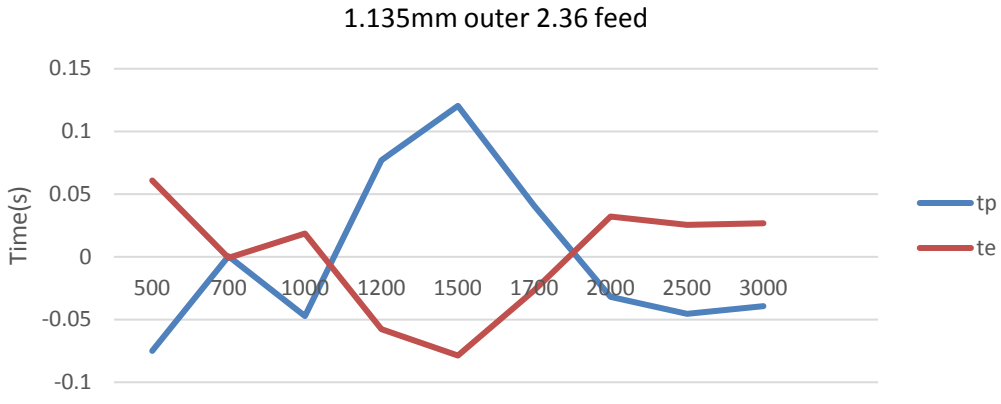
To validate the assumption of the estimation Hertzian quadrilateral model in Chapter 4 holds under cutting conditions, the correlation coefficient of $t_e$ and $t_p$ is analyzed, as shown in Figure 61. It can be seen that, exactly like the no-cutting condition, there is still a strong negative correlation between $t_e$ and $t_p$ under cutting condition, as listed in Table 15. The strong negative correlation means that, the total time $t_t$ is not sensitive to the bearing load even during machining operations, and it can be used to provide good estimation of the defect size. Therefore, the assumption in the Hertzian quadrilateral model works for both cases, either no-cutting or cutting.


0.794mm outer 2.36 feed
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
Time(s)
500
700
1000
1200
1500
1700
2000
2500
3000
tp
te

(a)

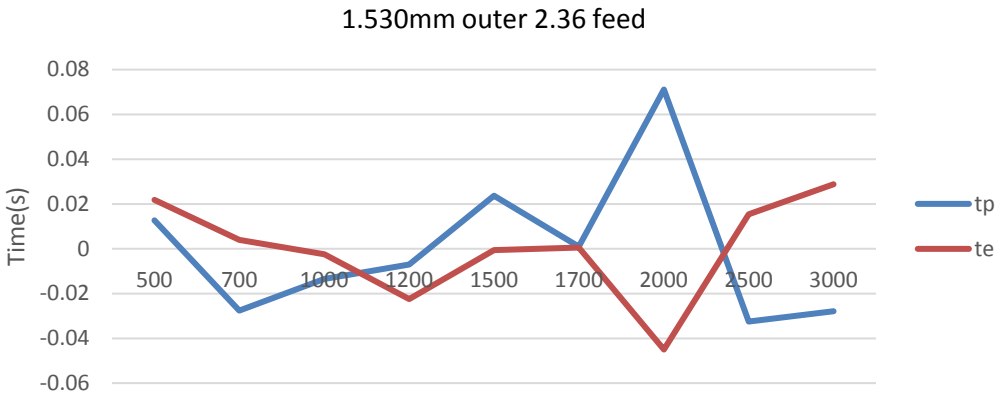

0.794mm outer 3 feed
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
Time(s)
500
700
1000
1200
1500
1700
2000
2500
3000
tp
te

(b)

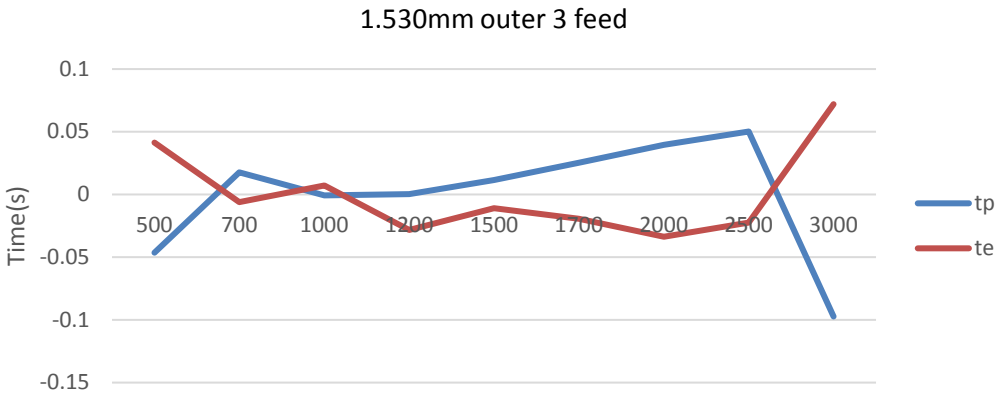

1.135mm outer 2.36 feed
0.15
0.1
0.05
0
-0.05
-0.1
Time(s)
500
700
1000
1200
1500
1700
2000
2500
3000
tp
te

(c)


(d)




(e)




(f)

Figure 61. The correlation coefficient of $t_e$ and $t_p$.

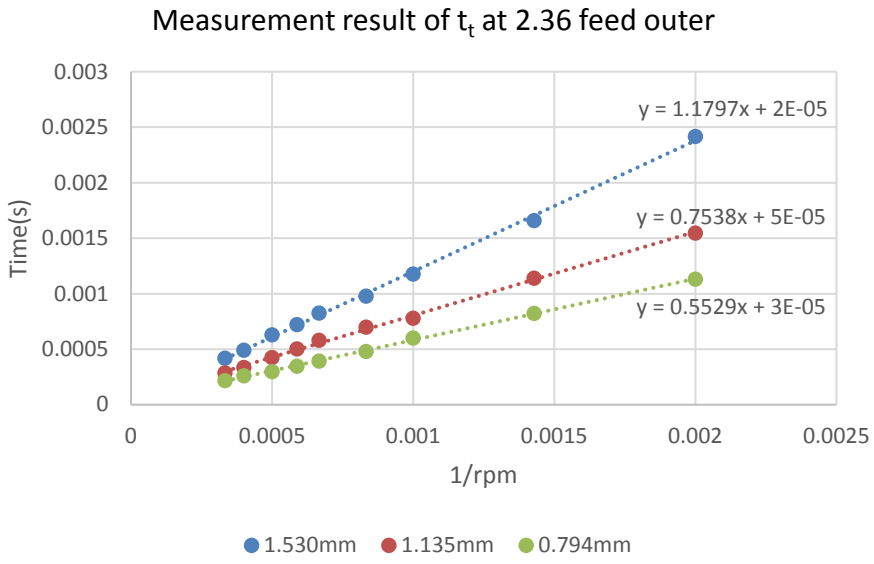
Table 15. Correlation coefficient of $t_e$ and $t_p$ for outer/inner race defects.

| Defect size (mm) | Outer race | | Inner race | |
|---|---|---|---|---|
| | 2.36 feed | 3 feed | 2.36 feed | 3 feed |
| 0.794 | -0.644 | -0.553 | -0.689 | -0.377 |
| 1.135 | -0.986 | -0.822 | -0.938 | -0.388 |
| 1.530 | -0.701 | -0.939 | -0.857 | -0.612 |

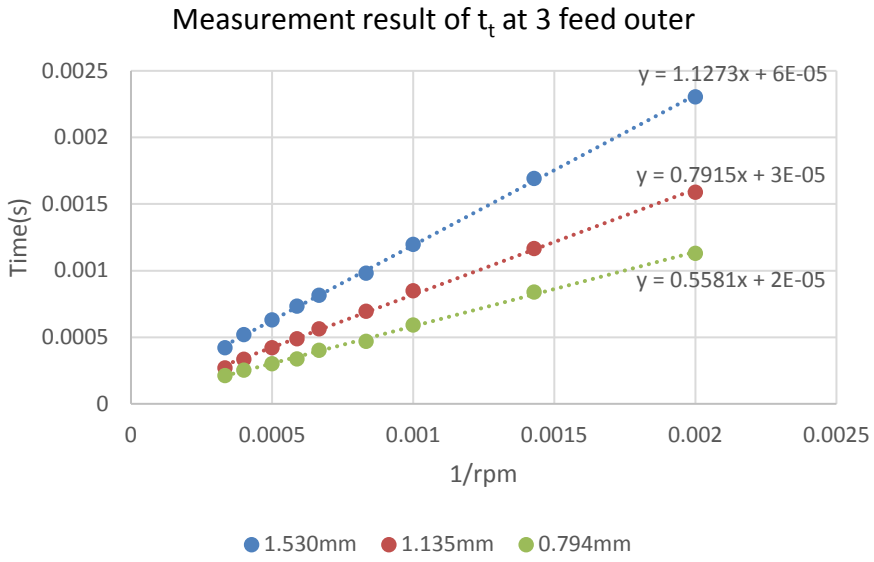
### *7.4.3 Measurement of $t_t$*

Since the time $t_p$ during and after the machining operation is fairly consistent and the total time $t_t$ is not sensitive to bearing loads, the defect size can be estimated based on $t_p$ during machining and $t_e$ after machining. The result for both outer and inner race defect is presented in the ensuing text.

The total time $t_t$ for both the outer and inner race defects under two different feed rates are shown in Figure 62 and Figure 63. Similar to the no-cutting experiment results, a clear linear trend can be seen between $t_t$ and 1/rpm. The small intercepts and the large R factors of all the least-square fit lines indicate good linearity of the measurement. By comparing Figure 62 and Figure 63 with Figure 59 and Figure 60, it is obvious that the linearity of $t_t$ is better than the linearity of $t_p$. This is due to the strong negative correlation coefficient as shown in Table 15. Therefore, the overall time $t_t$ is not affected by the variance of the contact force and it is a suitable metric to estimate the defect size even under cutting conditions. When feed rate increases by 27% from 2.36 in/min to 3 in/min, the estimation result is not affected much. This means that the estimation method is consistent with respect to cutting conditions.

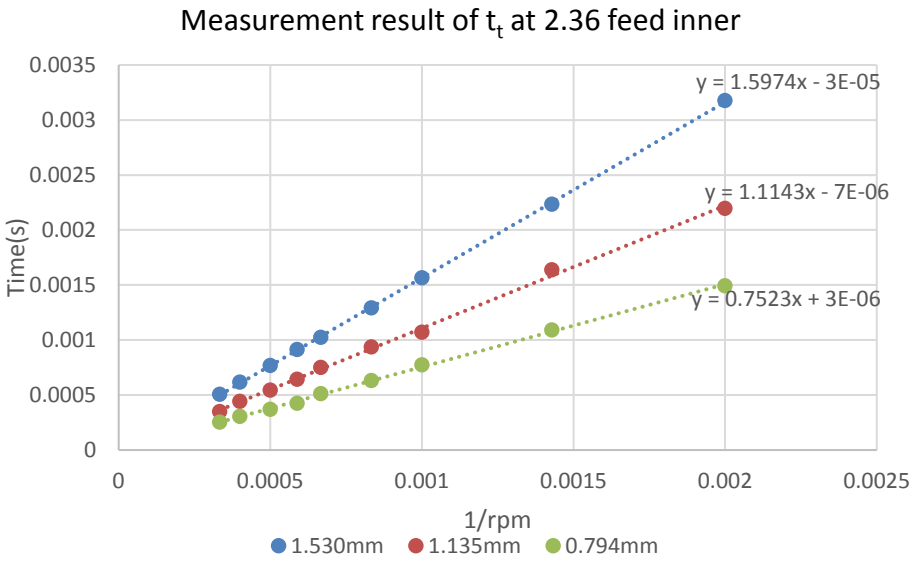



(a)

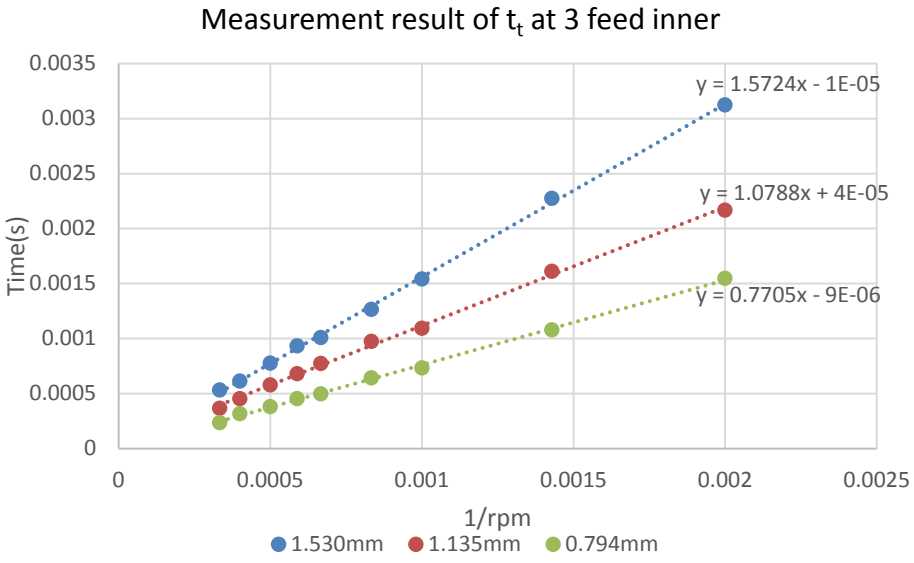



(b)

Figure 62. The measurement result of $t_t$ for outer race defect.




(a)




(b)

Figure 63. The measurement result of $t_t$ for inner race defect.

## 7.5 Defect size estimation result

### *7.5.1 Outer race defect*

The estimation results under different cutting conditions for the outer race defect are shown in Figure 64. It can be seen that the estimation results in the speed range 1000-2000rpm are very close to the true defect size. At lower speeds, the estimation result is smaller than the true defect. At higher speeds, uncertainty in the measurement result increases. This result is in agreement with the no-cutting experiment. When feed rate increases by 27% from 2.36 in/min to 3 in/min, the estimation result is not affected much. This means that the estimation method is consistent with respect to cutting conditions. The error of the estimated defect size is shown in Table 16. The best estimation result is for the 1.530 mm defect. This is because the signal-to-noise ratio of this test set is obviously higher than the other data sets. The largest error occurs in the outer race 1.135 mm defect at the lowest spindle speed. As discussed in Chapter 6, part of the reason is due to the small signal-to-noise ratio between the entry signal and the low frequency noise. Another reason that this result may occur is that the sensor is not located directly above the defect and this further decreases the signal-to-noise ratio.

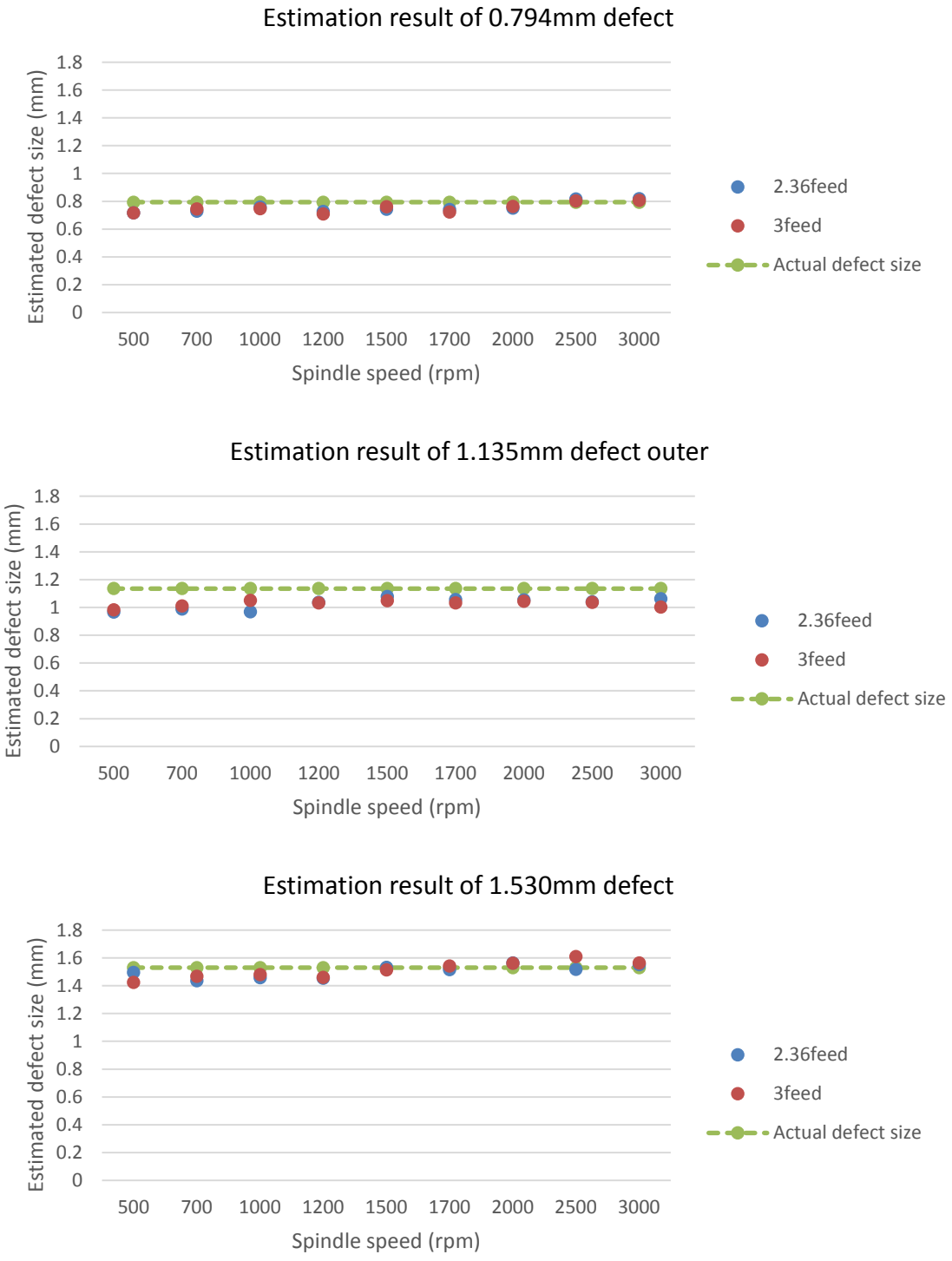



Figure 64. The defect estimation result for outer race defect.

Table 16. Average estimation error for the outer race defect.

| | 2.36 in/min feed | | 3 in/min feed | |
|---|---|---|---|---|
| Defect size (mm) | Absolute error (mm) | Relative error | Absolute error (mm) | Relative error |
| 0.794 | 0.0485 | 6.1% | 0.0460 | 5.8% |
| 1.135 | 0.1077 | 9.5% | 0.1093 | 9.6% |
| 1.530 | 0.0398 | 2.6% | 0.0514 | 3.4% |

### *7.5.2 Inner race defect*

Figure 65 shows the inner race defect estimation results under different cutting conditions. Similar trend exist in the inner race estimation result as the outer race defect. The estimation results in the speed range 1000-2000 rpm are very good. At lower speeds, the estimation result is smaller than the true defect, as shown in the 1.135 mm defect. At higher speeds, uncertainty in the measurement result increases. In general, the estimation results are better than the outer race defect, as in the no-cutting experiments. Part of the reason is the higher signal-to-noise ratio in the vibration data. Another reason might be due to the different mechanism of the inner/outer race defect. When a ball enters the inner race defect, its motion is regulated by the outer race. This decreases the uncertainty level of the ball's motion. While in the outer race defect, the ball loses contact with both races. Then, the motion of the ball suffers from more uncertainty caused by random factors, such as sliding and friction. Similar as the outer race defect, when feed rate increases by 27% from 2.36 in/min to 3 in/min, the estimation result is not affected much. This demonstrates that the estimation method is consistent with respect to cutting conditions. The error of the estimated defect size is shown in Table 17.

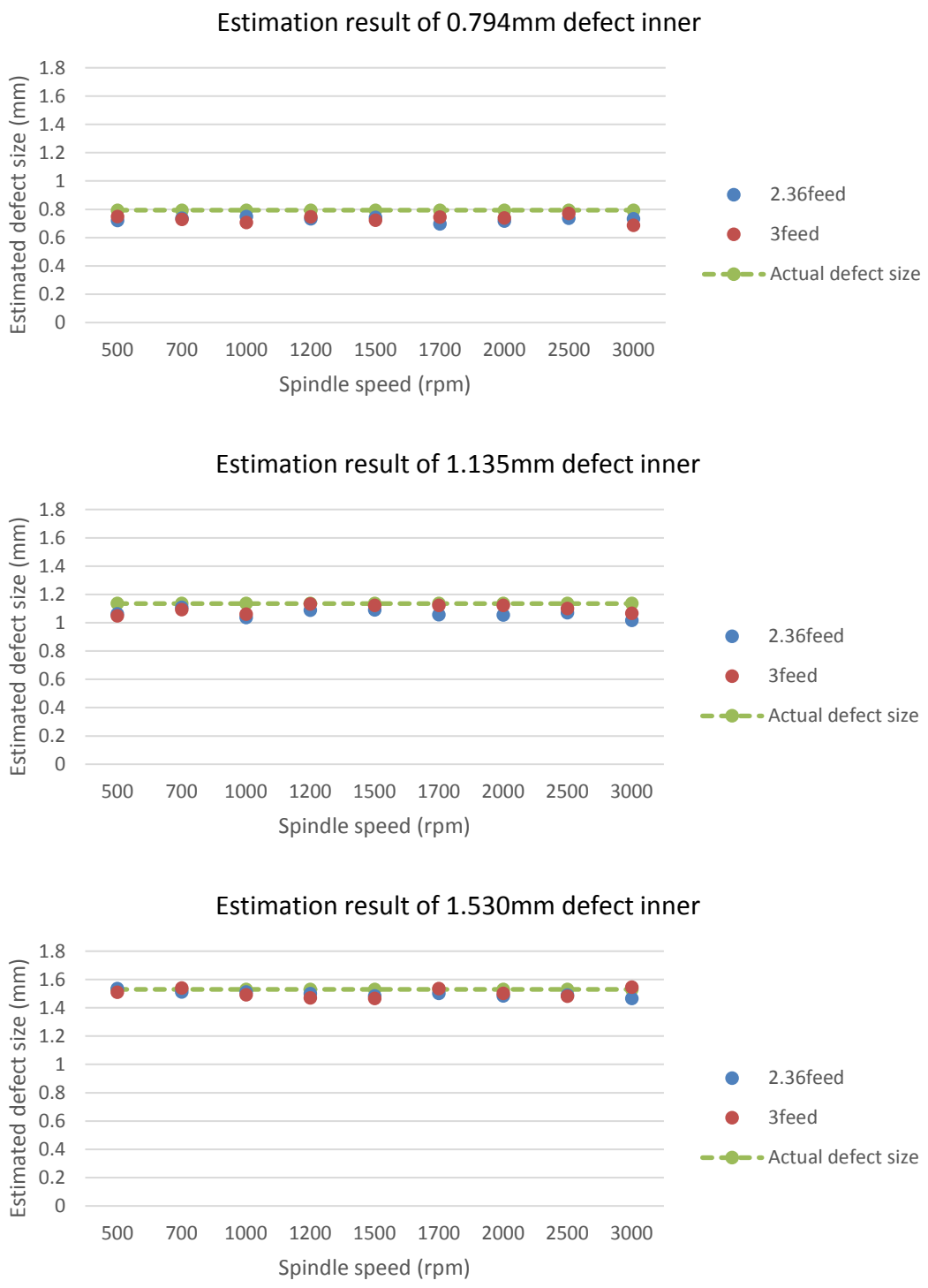



Figure 65. The defect estimation result for inner race defect.

Table 17. Average estimation error for the inner race defect.

| | 2.36 in/min feed | | 3 in/min feed | |
|---|---|---|---|---|
| Defect size (mm) | Absolute error (mm) | Relative error | Absolute error (mm) | Relative error |
| 0.794 | 0.0637 | 8.0% | 0.0610 | 7.7% |
| 1.135 | 0.0707 | 6.2% | 0.0685 | 6.0% |
| 1.530 | 0.0327 | 2.1% | 0.0320 | 2.1% |

## 7.6 Summary

This chapter describes the application of the defect estimation method in the cutting experiments. First, the influence of the cutting forces on the vibration signal is studied. Then, two signal processing methods are provided for defect estimation under cutting conditions. The second method is used in this thesis, which extracts $t_p$ from the cutting data and obtains $t_e$ from the post-cut data. The experiments in Chapter 6 are performed again in the cutting experiments under two different feed rate. Results show that the estimation method works well under cutting conditions with the maximum error 10% for the outer race defect and 8% for the inner race defect.

# CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS

An effective time-domain-based bearing defect size estimation method is proposed in this thesis. This chapter summarizes the contributions and conclusions of this research, and provides directions for future studies.

## 8.1 Contributions

A set of innovative methods and algorithms for bearing defect size estimation with/without cutting has been presented in this thesis. The contributions of this research are listed as below:

- A test system designed for CNC-based bearing diagnostics was proposed and fabricated. The setup facilitates bearing diagnostics during machining and modular replacement of test bearings without compromising the CNC spindle.
- A defect size estimation model quantifying the defective pattern in the vibration signal excited by the localized defect on raceways was proposed and validated. Rather than using contact force, the model uses the entry-to-peak and peak-to-impact times to quantify the defect size. Because the entry-to-peak and peak-to-impact times are measured directly from the current vibration signal, no baseline or historical data are required to determine the damage severity.
- A signal processing method was developed to automatically extract the time information from the vibration time sequence for defect size estimation. The entry point, first peak, impact point were isolated from the time series signal using empirical model, cross correlation, and twice differentiation, respectively. The

Variational Model Decomposition (VMD) is used to remove high frequency noise in the entry signal. These points were then used to calculate the entry-to-peak and peak-to-impact times. In addition, the signal processing methods were able to facilitate statistical analysis on the time series signal.

- The proposed bearing size estimation methods were experimentally validated during no-cutting and machining states.

## 8.2 Conclusions

The conclusions for each part of this research are summarized as below:

- The test system was demonstrated to accurately exhibit behaviors of a controlled bearing defect. The inner and outer ball-pass frequency ratios exhibited an approximately 1% decrease from 500 rpm to 3000 rpm. The decrease was determined to be minor and therefore the system was determined to be suitable as a bearing diagnostics experimental setup.
- The defect size estimation method was experimentally validated under no-cutting conditions. Within the tested spindle speeds, the defect size estimation results for both inner and outer defects were close to the true defect size with the maximum errors of 7% for the outer race and 10.9% for the inner race.
- Spectral analysis showed a significant effect of the load distribution on the entry-to-peak and peak-to-impact times, which agrees with the Hertzian quadrilateral model. The times were shown to be negatively correlated, thus showing that the model predicts a constant defect size under varying contact loads.

- The defect size estimation method during machining was validated via slot milling experiments. The peak-to-impact time was determined to exhibit sufficient resolution even under the influence of cutting. The peak-to-impact time was shown to be consistent before and after cutting. However, the entry-to-peak time was shown to be more sensitive to the influence of cutting. Therefore, entry-to-peak time under no-cutting conditions and peak-to-impact time under cutting conditions were applied to the method during machining.
- To improve the upper bound of spindle speeds for the bearing defect estimation method, a faster sampling is recommended. To improve the lower bound of spindle speeds, subtracting the curve fitted low frequency noise from the time domain signal is recommended. Note that to detect bearing defect sizes smaller than the spall width corresponding to simultaneous contact with the entry and exit edges, the Hertzian quadrilateral model cannot be utilized.

## 8.3 Future work and recommendations

This thesis proposes a new defect-size-estimation method based on the dynamic modeling for a line-spall defect on the raceway of a ball bearing. The dynamic model developed in this research can identify key features in the vibration signal, but the entry signal is not precisely represented. A more comprehensive model that possibly considers bearing lubricant conditions and a more advanced contact force model is recommended to describe the low frequency noise in the entry signal.

This research focuses on the ball bearing with a line-spall defect, but the actual bearing spall can be in different shapes (circle, rectangular) or dimensions (width, depth,

length). Therefore, research with respect to other defect shapes and geometries closer to a natural spall can be built on the model developed in this thesis. In addition, other types of defect (ball defect) and bearings (roller bearings) may have some special features in the defect response due to the kinematic and geometric difference. Therefore, modifying the dynamic model and applying it to ball defect and roller bearings can extend its applicability.

The defect size estimation method developed in this thesis can potentially be applied in shop floor automation applications for spall size tracking of a single defect on the inner or outer race, given the bearing geometries. For the purpose of automatic spall size measurement, the low frequency entry and high frequency exit signals should be identified and isolated before the other algorithms as proposed in this thesis are used. Depending on the different frequency components, signal processing procedures, such as low-pass/high-pass filters, Wavelet Transform and enveloping, can be applied toward automated spall size detection.

# REFERENCES


1. Shiroishi, J., et al., *Bearing condition diagnostics via vibration and acoustic emission measurements.* Mechanical systems and signal processing, 1997. **11**(5): p. 693-705.

2. Patidar, S. and P.K. Soni, *An overview on vibration analysis techniques for the diagnosis of rolling element bearing faults.* International Journal of Engineering Trends and Technology, 2013. **4**(5): p. 1803-1809.

3. Camci, F. and R.B. Chinnam, *Health-state estimation and prognostics in machining processes.* Automation Science and Engineering, IEEE Transactions on, 2010. **7**(3): p. 581-597.

4. Kurfess, T.R., S. Billington, and S.Y. Liang, *Advanced diagnostic and prognostic techniques for rolling element bearings*, in *Condition Monitoring and Control for Intelligent Manufacturing*. 2006, Springer. p. 137-165.

5. Jardine, A.K., D. Lin, and D. Banjevic, *A review on machinery diagnostics and prognostics implementing condition-based maintenance.* Mechanical systems and signal processing, 2006. **20**(7): p. 1483-1510.

6. Henao-Sepulveda, J.A., M. Toledo-Quinones, and Y. Jia. *Contactless Monitoring of Ball Bearing Temperature*. in *2005 IEEE Instrumentationand Measurement Technology Conference Proceedings*. 2005.

7. Dempsey, P.J., G. Kreider, and T. Fichter. *Investigation of tapered roller bearing damage detection using oil debris analysis*. in *Aerospace Conference, 2006 IEEE*. 2006. IEEE.

8. Tandon, N. and A. Choudhury, *A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings.* Tribology international, 1999. **32**(8): p. 469-480.

9. Choudhury, A. and N. Tandon, *Application of acoustic emission technique for the detection of defects in rolling element bearings.* Tribology international, 2000. **33**(1): p. 39-45.

10. Patel, V., N. Tandon, and R. Pandey. *Review of Defect Detection Techniques of Ball Bearings in the Presence of External Noise*. in *Ninth International Conference on Vibration Problems, IIT Kharagpur, India, January*. 2009.

11. Martin, H. and F. Honarvar, *Application of statistical moments to bearing failure detection.* Applied acoustics, 1995. **44**(1): p. 67-77.

12. Randall, R.B. and J. Antoni, *Rolling element bearing diagnostics—a tutorial.* Mechanical Systems and Signal Processing, 2011. **25**(2): p. 485-520.

13. Epps, I. and H. McCallion. *An investigation into the characteristics of vibration excited by discrete faults in rolling element bearings*. in *Annual Conference of the Vibration Association of New Zealand, Christchurch*. 1994.

14. Williams, T., et al., *Rolling element bearing diagnostics in run-to-failure lifetime testing.* Mechanical Systems and Signal Processing, 2001. **15**(5): p. 979-993.

15. Kotzalas, M.N. and T.A. Harris, *Fatigue failure progression in ball bearings.* Journal of tribology, 2001. **123**(2): p. 238-242.

16. Zhang, C., *Defect detection and life prediction of rolling element bearings.* 2001.

17. Shiroishi, J.W., *Bearing condition diagnostics via multiple sensors using the high frequency resonance technique with adaptive line enhancer.* 1996.

18. Sawalhi, N. and R. Randall, *Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size.* Mechanical Systems and Signal Processing, 2011. **25**(3): p. 846-870.

19. Niu, L., et al., *Dynamic modeling and vibration response simulation for high speed rolling ball bearings with localized surface defects in raceways.* Journal of Manufacturing Science and Engineering, 2014. **136**(4): p. 041015.

20. Behzad, M., A.R. Bastami, and D. Mba, *A new model for estimating vibrations generated in the defective rolling element bearings.* Journal of Vibration and Acoustics, 2011. **133**(4): p. 041011.

21. Khanam, S., N. Tandon, and J. Dutt, *Fault size estimation in the outer race of ball bearing using discrete wavelet transform of the vibration signal.* Procedia Technology, 2014. **14**: p. 12-19.

22. Smith, W.A., et al. *Vibration-Based Spall Size Tracking in Rolling Element Bearings*. in *Proceedings of the 9th IFToMM International Conference on Rotor Dynamics*. 2015. Springer.

23. Harris, T.A. and M.N. Kotzalas, *Advanced concepts of bearing technology: rolling bearing analysis*. 2006: CRC Press.

24. Meyer, L., F. Ahlgren, and B. Weichbrodt, *An analytic model for ball bearing vibrations to predict vibration response to distributed defects.* Journal of Mechanical Design, 1980. **102**(2): p. 205-210.

25. Wardle, F., *Vibration forces produced by waviness of the rolling surfaces of thrust loaded ball bearings Part 2: experimental validation.* Proceedings of the Institution

of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 1988. **202**(5): p. 313-319.

26. Epps, I., *An investigation into vibrations excited by discrete faults in rolling element bearings.* 1991.

27. Hawman, M. and W. Galinaitis. *Acoustic emission monitoring of rolling element bearings*. in *Ultrasonics Symposium, 1988. Proceedings., IEEE 1988*. 1988. IEEE.

28. Ahmadi, A.M., C.Q. Howard, and D. Petersen, *The path of rolling elements in defective bearings: Observations, analysis and methods to estimate spall size.* Journal of Sound and Vibration, 2016. **366**: p. 277-292.

29. Al-Ghamd, A.M. and D. Mba, *A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size.* Mechanical systems and signal processing, 2006. **20**(7): p. 1537-1571.

30. Hemmati, F., W. Orfali, and M.S. Gadala, *Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation.* Applied Acoustics, 2016. **104**: p. 101-118.

31. Al-Dossary, S., R.R. Hamzah, and D. Mba, *Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing.* Applied acoustics, 2009. **70**(1): p. 58-81.

32. Ahmadi, A.M., D. Petersen, and C. Howard, *A nonlinear dynamic vibration model of defective bearings–The importance of modelling the finite size of rolling elements.* Mechanical Systems and Signal Processing, 2015. **52**: p. 309-326.

33. While, M., *Rolling element bearing vibration transfer characteristics: effect of stiffness.* Journal of applied mechanics, 1979. **46**(3): p. 677-684.

34. Walford, T. and B. Stone, *The sources of damping in rolling element bearings under oscillating conditions.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 1983. **197**(4): p. 225-232.

35. Petersen, D., et al., *Analysis of bearing stiffness variations, contact forces and vibrations in radially loaded double row rolling element bearings with raceway defects.* Mechanical Systems and Signal Processing, 2015. **50**: p. 139-160.

36. McFadden, P. and J. Smith, *Model for the vibration produced by a single point defect in a rolling element bearing.* Journal of sound and vibration, 1984. **96**(1): p. 69-82.

37. McFadden, P. and J. Smith, *The vibration produced by multiple point defects in a rolling element bearing.* Journal of sound and vibration, 1985. **98**(2): p. 263-273.

38. Antoni, J. and R. Randall, *A stochastic model for simulation and diagnostics of rolling element bearings with localized faults.* Journal of vibration and acoustics, 2003. **125**(3): p. 282-289.

39. Tandon, N. and A. Choudhury, *An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect.* Journal of sound and vibration, 1997. **205**(3): p. 275-292.

40. Brie, D., *Modelling of the spalled rolling element bearing vibration signal: an overview and some new results.* Mechanical Systems and Signal Processing, 2000. **14**(3): p. 353-369.

41. Choudhury, A. and N. Tandon, *Vibration response of rolling element bearings in a rotor bearing system to a local defect under radial load.* Journal of Tribology, 2006. **128**(2): p. 252-261.

42. Cao, H., L. Niu, and Z. He, *Method for vibration response simulation and sensor placement optimization of a machine tool spindle system with a bearing defect.* Sensors, 2012. **12**(7): p. 8732-8754.

43. Harsha, S., *Nonlinear dynamic analysis of an unbalanced rotor supported by roller bearing.* Chaos, Solitons & Fractals, 2005. **26**(1): p. 47-66.

44. Sawalhi, N. and R. Randall, *Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults.* Mechanical Systems and Signal Processing, 2008. **22**(8): p. 1924-1951.

45. Tandon, N., *A comparison of some vibration parameters for the condition monitoring of rolling element bearings.* Measurement, 1994. **12**(3): p. 285-289.

46. Li, Y., et al., *Diagnostics and prognostics of a single surface defect on roller bearings.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2000. **214**(9): p. 1173-1185.

47. McFadden, P. and J. Smith, *Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review.* Tribology international, 1984. **17**(1): p. 3-10.

48. Yan, R., R.X. Gao, and X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications.* Signal Processing, 2014. **96**: p. 1-15.

49. Polikar, R., *The wavelet tutorial.* 1996.

50. Kumar, R. and M. Singh, *Outer race defect width measurement in taper roller bearing using discrete wavelet transform of vibration signal.* Measurement, 2013. **46**(1): p. 537-545.

51. Saravanan, S., G. Yadava, and P. Rao, *Condition monitoring studies on spindle bearing of a lathe.* The International Journal of Advanced Manufacturing Technology, 2006. **28**(9-10): p. 993-1005.

52. Werner, A., *An Early Warning Monitoring System for CNC Spindle Bearing Failure*. 2011, Clemson University.

53. Zhang, B.C., et al. *Research on Fault Diagnosis for Bearing of CNC Machine Tools*. in *Advanced Materials Research*. 2012. Trans Tech Publ.

54. Chen, Y., Z. He, and S. Yang, *Research on on-line automatic diagnostic technology for scratch defect of rolling element bearings.* International Journal of Precision Engineering and Manufacturing, 2012. **13**(3): p. 357-362.

55. Kozochkin, M., et al., *Vibrational diagnostics of roller bearings in metal-cutting machines.* Russian Engineering Research, 2013. **33**(8): p. 486-489.

56. Shen, J.-Y., C.-W. Tseng, and I. Shen, *Vibration of rotating disk/spindle systems with flexible housing/stator assemblies.* Journal of Sound and Vibration, 2004. **271**(3): p. 725-756.

57. Piezotronics, P. *Model 352A21*. 2013; Available from: https://www.pcb.com/contentstore/docs/PCB_Corporate/Vibration/Products/Manuals/352A21.pdf.

58. Moazenahmadi, A., et al. *Defect size estimation and analysis of the path of rolling elements in defective bearings with respect to the operational speed*. in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. 2014. Institute of Noise Control Engineering.

59. Wan, C. and W. Changsen, *Analysis of rolling element bearings*. 1991: John Wiley & Sons Incorporated.

60. Harris, T.A., *Rolling bearing analysis*. 2001: John Wiley and sons.

61. Dragomiretskiy, K. and D. Zosso, *Variational mode decomposition.* IEEE transactions on signal processing, 2014. **62**(3): p. 531-544.