uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Behroz Behnam

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Systems Science)

GRADE / DEGREE

Systems Science

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Bearing Fault Detection and Classification by Wavelet-Artificial Neural Network and Wavelet-Energy Singular Value Ratio

TITRE DE LA THÈSE / TITLE OF THESIS

M. Liang

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

D. Necsulescu

T. Yeap

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Bearing Fault Detection and Classification by Wavelet-Artificial Neural Network and Wavelet-Energy Singular Value Ratio

By
Behroz Behnam

Directed by
Ming Liang

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Systems Science at the University of Ottawa

University of Ottawa
Ottawa, Ontario, Canada
January 2009

Your file Votre référence
ISBN: 978-0-494-58198-8
Our file Notre référence
ISBN: 978-0-494-58198-8

Abstract

In this dissertation, Wavelet-ANN (Artificial Neural Network) and Wavelet-ESVR (Energy Singular Value Ratio) are proposed for detection and classification of faults in systems with periodical characteristics. Bearings in particular display such periodical characteristics. Bearings are one of the most critical components in rotary machinery and the majority of failures arise from defective bearings. Early warning in bearing deterioration is essential and it can prevent substantial machine downtime. In the first method an ANN is trained to diagnose bearing health by extracting information from discrete Wavelet coefficients. In the second approach, ESVRs are used to detect variations in the continuous Wavelets coefficients as symptoms of bearing health. Computer-simulated data and real bearing vibration data were then applied to perform initial testing and validation of these approaches. The test results show that the proposed methods are effectively detecting different bearing faults.

Table of Contents

List of Figures

List of Tables

Acronyms

ANN	Artificial Neural Networks
BFPI	Ball Pass Frequency Inner Race
BFPO	Ball Pass Frequency Outer Race
BSF	Ball Spin Frequency
CWRS	Case Western Reserve University
db	Decibel
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
FD	Fault Diameter
FT	Fourier Transforms
Hz	Hertz
KHz	Kilo Hertz
LCL	Lower control limit
MLP	Multilayer Perceptron
NN	Neural Network
NNs	Neural Networks
RPM	Revolution Per Minute
SK	Spectral Kurtosis
STFT	Short Time Fourier Transform
SNR	Signal to Noise Ratio
SVD	Singular Value Decomposition
ESVR	Energy Singular Value Ratio
ESVRs	Energy Singular Value Ratios
UCL	Upper Control Limit
Wavelet-ANN	Wavelet Artificial Neural Network
Wavelet-ESVR	Wavelet Energy Singular Value Ratio

Nomenclature

a	scale factor
b	translation factor in time
BFPI	Ball Pass Frequency Inner Race
BFPO	Ball Pass Frequency Outer Race
BSF	Ball Spin Frequency
C	damping constant
C_g	admissibly constant
$c(k)$	scaling lowpass filter coefficients
$d(k)$	Wavelet highpass filter coefficients
$D_{m,n}$	Wavelet coefficient at scale index
DN	D: Daubechies Wavelets, N: order of the filter
E_i	energy of monitored bearing at scale index i
E_m	Wavelet energy spectrum at scale m,
En_i	energy of a normal bearing at scale index i.
f_0	center frequency of the selected Wavelet function
$f_p^{(i)}$	peak frequency at scale index i
F_k	power of impact
H1	Hihpass filter
Ks	stiffness of the structure
K	kurtosis
K_i	kurtosis of the monitored bearing at scale index i
Kn_i	average kurtosis of normal bearing at scale index i
L1	lowpass filter
M_a	mass
M	total number of scales
$n(t)$	noise
N	number of data samples
N_b	number of balls
P_d	pitch diameter
R	scale resolution
Rd	roller diameter
RK_i	relative kurtosis at scale index i.
RE_i	relative energy at scale index i.
S	standard deviation
S	shaft rotational speed

S_i	rotational speed of inner race
t	time
T	fault period
U	left singular vectors
V	right singular vectors
V^H	Hermitian adjoint of matrix V.
$W\ (a,b)$	Wavelet transform at scale a
x(n)	finite length signal
$x(t)$	measured signal at time t
X_i	amplitude of ^{ith} data point
$\bar{X}$	mean value
ψ^*	complex conjugate of ψ
ψ	the basis function
ϕ	contact angle
$\phi(t)$	scaling equation
ξ	damping ratio
Σ	diagonal matrix of singular values
σ	singular value
ω_0	angular frequency of the Morlet Wavelet
ω_d	frequency of damped oscillation
ω_n	natural circular frequency

Chapter 1. Introduction

1.1 Overview

Machine condition monitoring, early fault detection, diagnosis and classification are extremely important topics in the engineering field. Proper machine monitoring and fault detection methods will result in improved safety, improved reliability and reduction of the costs of different engineering system operations (Chow 1998). A traditional approach to prevent machine failure is based on planning, regular inspection, parts replacement and preventive maintenance. Although these methods are effective, they are all incapable of giving early warning for the machine conditions under operation. For this reason automatic fault detection and diagnosis methods that monitor the machine while it is operating is becoming popular. Many techniques such as temperature monitoring, process measurement, lubricant analysis, vibration and acoustic signal analysis have been used in fault detection and diagnostics. Among these techniques, vibration analysis is the most established and tangible technology, as almost all machines vibrate (Rao 1996). In rotary machines, bearings are crucial components and the majority of failures arise from defective bearings. The unprocessed vibration data collected from a rotary machine gives some information about the condition of the bearing; however, to detect fault at early stages, further processing of the vibration signal is necessary. In this thesis, Wavelet-ANN (Artificial Neural Network) and Wavelet-ESVR (Energy Singular Value Ratio) are proposed as two methods for detection and classification of faults in bearings. In the first

method discrete Wavelet transform (DWT) is used for signal decomposition and neural network (NN) performs bearing fault diagnosis based on the decomposed data. In the second method a combination of singular value decomposition (SVD) and Wavelet energy are used to detect defect periodicity hidden inside the coefficients obtained by continuous Wavelet transform. Computer-simulated data and real bearing vibration data were then applied to perform initial testing and validation of these approaches.

1.2 Objectives

The aim of this research is to develop methods for early fault detection and classification of rolling bearings or any system that produces short time high frequency periodic transient signals. Two specific objectives of this research are:

a) To detect and classify bearing faults by using Wavelet transform as a preprocessor and artificial neural network as an intelligent diagnosing tool, and
b) To detect and classify bearing faults by measuring the defect periodicities inside the Wavelet coefficients obtained by continuous Wavelet transform.

1.3 Thesis outline

Hereafter, the thesis is organized as follows. Chapter 2 presents a brief literature review of the commonly used techniques for rotary machine fault detection. Chapter 3 describes the bearing rotational frequencies and bearing resonance frequencies which are important concepts for bearing fault detection. The experimental setup and data used in this study are also elaborated in this chapter. In Chapter 4, the Wavelet-ANN method as well as its training and testing detailed. Chapter 5 reports a Wavelet-ESVR method for periodicity detection without a training process. The method is examined by both simulated and

experimental data. Chapter 6 draw conclusions from this research and recommendation for future work have been presented.

Chapter 2. Literature Review

As a general rule, rotating machines do not breakdown without early warning. Therefore, by analyzing vibration signals one can predict possible failure and determine the severity and type of the defect (Kessissoglou et al. 2003). Several methods in different domains have been implemented for interpreting the vibration signals. The commonly employed signal processing methods for analyzing measured data are time domain, frequency-domain, and time-frequency domain techniques. Each of these methods has been widely used for fault detection, but they all have some limitations and cannot be applied to all conditions. This chapter will provide a brief review of some of these methods.

2.1. Time-domain analysis

In time–domain, the interpretation of the signal is done through several parameters. Some of the parameters are: peak and root-mean-square (RMS) level, crest factor, probability density function, second, third and fourth order statistical moments which can be extracted from the vibration signal (Norton et al. 2003). The fourth order statistical moment, which is known as kurtosis, is widely used in machine condition monitoring and more specifically for detecting faults in gears. The kurtosis is an indicator which is sensitive to the impulsiveness of a signal. The kurtosis of white noise has a value of 3. A higher kurtosis value indicates possible defects in the system. Antoni (2006) and Antoni and Randall(2006) proposed spectral kurtosis (SK) for detecting and characterizing transient signals buried in additive noise. Sawalhi et.al.(2005) employed kurtosis of the

signal to select a time frequency frame with the highest impulsiveness. In this study, kurtosis will be used to extract level of impulsiveness in the Wavelet coefficients (As it will be explained in the chapters 4.1 and 5.1, in discrete wavelet transform wavelet coefficients are computed by passing through a series of low-pass and high-pass filters with a cut off frequency of $\omega_0 / 2$ which is half of maximum frequency of the signal. In continuous wavelet transform, coefficients are computed by convolving of a wavelet function with the signal).

Time-domain analysis has some limitations to identify faults in a specific component because a machine consists of many components and their individual contributions to the overall vibration signal are very difficult to identify. This is why researchers are interested in the frequency-domain analysis. In frequency domain analysis, the frequencies of major peaks can be associated with the specific components based on the geometric specification and the operating speed of the components (figure 2.1). By monitoring this frequency one can detect the defective component more easily.

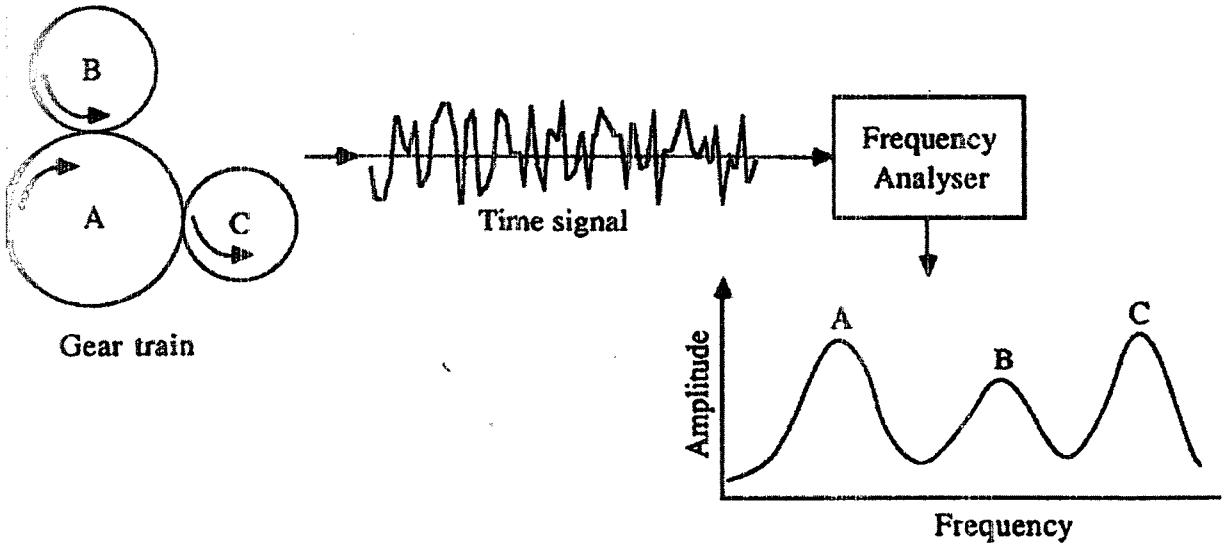


Figure 2.1. Time-domain and frequency domain representation of a vibration signal generated by a gear train, adapted from Norton et al. (2003), page 343

2.2 Frequency domain and time-frequency domain analysis

There are some techniques which can help analyze the vibration signal in the frequency-domain. Among them the most recognized is the Fourier transform. The Fourier transform is simply a method of expressing a periodic signal as a sum of sinusoids or exponentials of various frequencies. Several techniques such as power spectrum utilize Fourier transform as the basis of analysis. Auto-power spectrum of the signal is simply a Fourier transform of the signal multiplied by its complex conjugate. Cavacece et al. (2002) used auto-power and cross power spectrum to detect faults at early stages in aeronautical transmissions. One shortcoming of Fourier transform is that it is impossible to determine at what time the concerned frequencies occurred, in other words this transform is mostly appropriate for stationary signals. The vibration signals of defective bearings are mostly intermittent. These intermittent defects are in non-stationary form with time varying transient signals and frequency characteristics. The non-stationary property of the signal makes the Fourier transform an inappropriate signal processing tool as it does not provide time information. To overcome the lack of time information , some researchers including Li et al. (2000) have used both frequency and time-domain representation of a vibration signal to detect bearing faults.

To extend the application of Fourier transform to the non-stationary signals, Gabor (1946) employed the Fourier transform to process a small section of the signal each time segment (referred to as windowing). This windowing of the signal is called the Short-Time Fourier Transform (STFT). Nonetheless, a drawback of STFT is that once a particular size for the time window is selected, that window remains the same for all frequencies. The allocated window size may reveal one specific fault but not other faults

that do not have a good resolution for that window size. Wang et.al. (1993) used STFT to detect fault in gears. He showed that the recommended width of the window for the detection of gear tooth damage is related to the meshing period of one tooth.

2.2.1 Wavelet analysis

According to the uncertainty principle, it is not possible to know the exact time-frequency representation of a signal (resolution problem). Fourier transforms use non-localized bases in time domain which result in perfect localization in frequency domain and poor localization in time domain. STFT can overcome this problem by employing the Fourier transform to process a small section of the signal each time (referred to as windowing). Nonetheless, as mentioned earlier a drawback of STFT is that it uses a fixed window size which cannot fit all frequencies. The use of Wavelet transform will mitigate this problem by using scalable window. In the Wavelet transform, the window is shifted along the signal at every position the spectrum is calculated. This process is repeated with a shorter (or longer) window for every new cycle. The result will be a collection of time-frequency representations of the signal. Wavelet transform can be performed in discrete or continuous forms. Discrete Wavelet transform uses orthogonal bases and dyadic scaling and translation which result in coefficients without any redundancy. Wang et al., (1995) used DWT to detect faults in gears. Their results showed computational efficiency of DWT, however because of limitation of scales they were not able to capture all details of the signal. The second form of Wavelet transform, continuous Wavelet transform, can overcome the shortcoming of DWT by utilizing non-orthogonal bases.

Unlike Fourier transform, which always uses complex exponential basis, Wavelet transform uses a localized oscillatory function called mother Wavelet. The choice of Wavelet basis has a great importance and many researchers have tried to find the best one for their applications.

Wang et al., (1996) used continuous Wavelet transform (CWT) to analyze the vibration signals produced by a gear in a helicopter gearbox. He suggested that among different possible choices of bases, the Gaussian-enveloped oscillating Wavelet is well suited for detecting various sizes of gear faults. He also suggested that for detecting transients from a faulty signal, basic types of tapered-rectangular and double-sided (figure 2.2) exponential Wavelets possess high sensitivity and resolution (Wang et al., 2001). Lin and Qu (2000) used Morlet Wavelet which has a close similarity to the impulse function generated by a faulty mechanical system. Lin and Zuo (2003) used the right-hand side of Morlet Wavelet as the basis. Parts 'a' and 'b' of figure 2.2 illustrate the time and frequency representation of the single and double sided exponential Wavelets respectively.

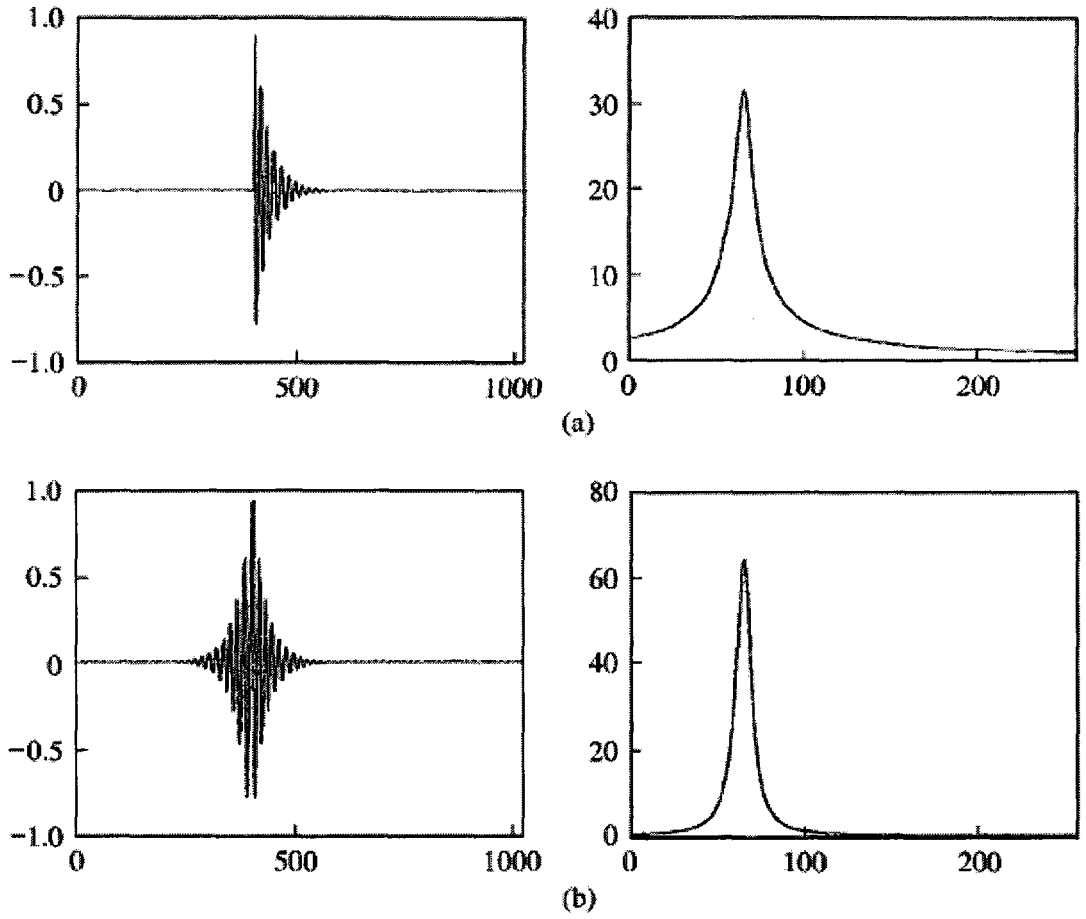


Figure 2.2 (a) Single-sided exponential oscillation (single-sided exponential Wavelet). (b) Double-sided exponential oscillation (double-sided exponential Wavelet), adapted from Wang (2001)

In brief, converting the vibration signal into time-frequency domain by Wavelet transform reveals some definite features of the original signal and hence one can detect faults by visual reading. Although in many cases faults can be identified by the visual reading of Wavelet plots, a much more reliable method is needed to diagnose the signals especially when the fault advancement is so small. Another promising technique to automatically analyze the vibration signal reading is Neural Network whose performance highly depends on extracting features.

2.3 Artificial neural networks (ANN) and feature extraction

The interpretation of Wavelet coefficients is done by several methods including visual inspection and intelligent systems. One important method of intelligent systems is ANN. ANNs with a remarkable ability to extract information from data which are too complex to interpret have been widely used in many fields such as automatic fault detection and classification. Lou et. al, (2004) employed discrete Wavelet transform as a preprocessor to generate feature vectors and they used a neural-fuzzy inference system (ANFIS) as a diagnostic classifier. Paya and Esat (1997) employed Back-Propagation Neural Network for classification. They used Wavelet transform to decompose the signal into a time-frequency domain and extract 10 Wavelet numbers and their 10 corresponding amplitudes as features or input vector to the ANN. Their results show that an overall average classification rate of 96% can be achieved. Devaney and Eren (2004) used a Radial Basic Neural Network to classify two common sources of bearing faults by suppressing the more dominant power system harmonics and using Wavelet-Packet-Decomposition for the remaining signal. Chen and Wang (2002) employed a Multi-Layer Perceptron Neural Network for automated interpretation of the Wavelet map. Lin and Qu (2000) used soft-thresholding to extract the impulse component as the feature from the decomposed vibration signal by Wavelet transform. Lou et al (2004) decomposed the signal by discrete Wavelet transform and then they used vector distance and correlation coefficient as a feature vector to train a neuron-fuzzy system.

In the first part of this thesis, a simple yet effective method will be introduced which is capable of detecting the condition of a bearing precisely. This method is based on the training of a feed-forward ANN which uses extracted information from Wavelet coefficients. It will be shown that the percentage of Wavelet coefficients' energy and kurtosis can effectively determine the condition of a bearing.

2.4 Periodicity detection

Visual inspection of Wavelet plots is of limited use as a fault detection method because in the early stage of faults even an experienced person may not be able to detect the faults. Optimizing the Wavelet parameters can increase the visibility of impact impulses in the Wavelet plots. Nevertheless, even at the optimum parameters the transients may still be hidden in the background noise. For instance, vibration signals generated by the Case Western Reserve University (CWRS) have been widely used by researchers as the test signals. Figure 2.3 shows four bearing conditions. Parts (a) and (c) of figure 2.3 clearly show the sign of the impact impulses due to the defective outer race and inner race in the time signal. The visibility of transients totally disappears when faults develop in the defective outer race, see part (b). On the other hand, in many cases the fault characteristic frequencies may also be observed in the normal bearing which leads to the misinterpretation of the signal. For example, figure 2.4 presents the time signal and Wavelet plot of a normal bearing. Though the roller fault characteristic frequency cannot be easily observed, it can be clearly seen in the Wavelet plot. Therefore, this bearing would be treated as a defect based on the Wavelet result alone.

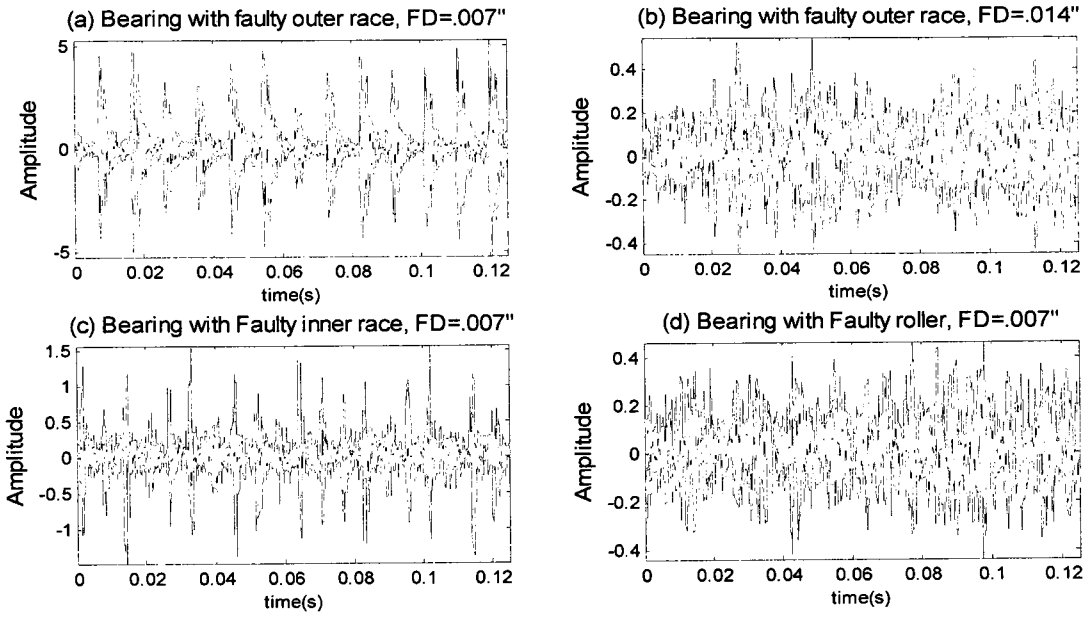


Figure-2.3 Vibration signals generated by CWRS

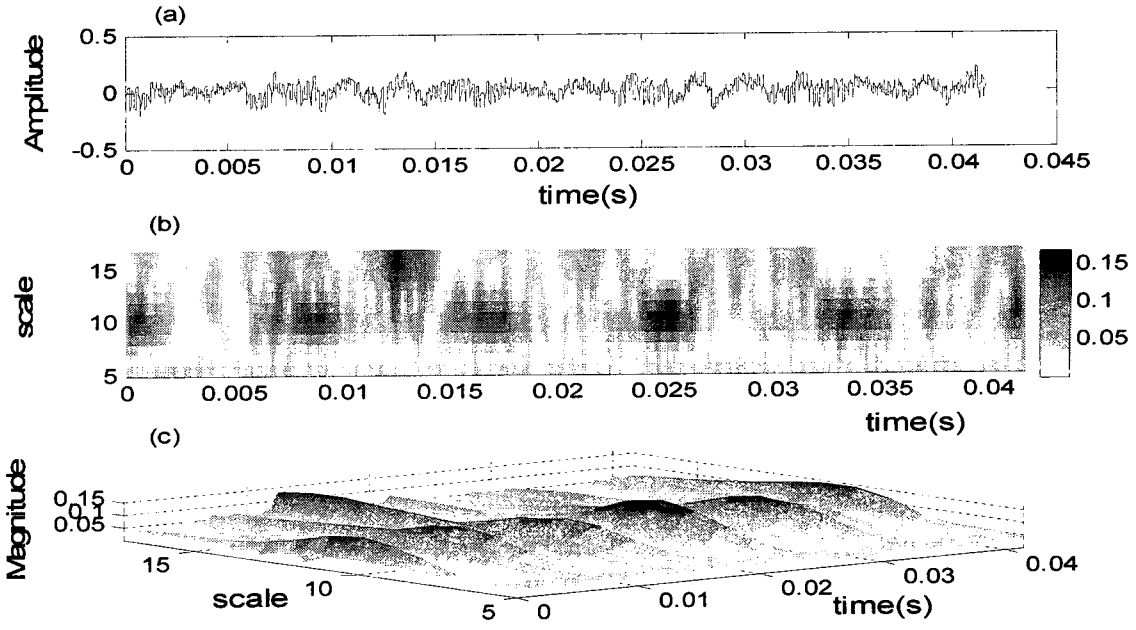


Figure 2.4 a) Signal of healthy bearing, b) 2-D Wavelet plot, and c) 3-D Wavelet plot (Absolute coefficients)

Most of the researchers who used data provided by CWRS, i.e., Al-Raheem et.al. (2007) and Hong et al. (2005) limited their testing to the outer and inner race faults where the characteristic bearing frequencies are evident in the time signal.

One way to detect very weak transients in the faulty signals or to detect the status of a normal bearing with a sign of strong fault characteristic frequency is to train an intelligent system such as ANN to detect and classify faults. Abbasion et. al., (2007) trained a SVM (Support vector machine) network and they used data from CWRS. Their system could successfully detect faulty roller bearing and faulty inner and outer race bearings. As stated in the previous section, in this thesis Wavelet-ANN will be presented and it will be shown that the system, if properly trained, can effectively detect and classify faults. However, the need of training data limits their applications.

To avoid the difficulty present in training, a new method will be introduced in this thesis to detect all the bearing conditions without going through the training process. The important point is that transient impulses produce variations in the Wavelet coefficients in one or more scales. The amount of variation depends on the fault specifications and the energy of impulses. By measuring the variations in the repeating pattern of the Wavelet coefficients one can detect and classify faults easily. This method utilizes singular value decomposition (SVD) and energy of the scales to measure the variations of the Wavelet coefficients. The measured variations are compared to the normal condition through a

pre-established X-bar chart. The experimental results verify that the Wavelet-ESVR method can detect three major bearing faults precisely.

2.5 Summary

The time-domain analysis has perfect time resolution but very poor frequency resolution which limits its application to identify faults in a specific component in a complex machine. The frequency domain analysis has perfect frequency resolution but very poor time resolution that makes it inappropriate for non-stationary signals. Short time Fourier transform has been used by researchers to overcome shortcoming of Fourier transform; however, it has window size problem (Wang et.al., 1993). The use of Wavelet transform will mitigate the shortcoming of STFT by using scalable window. Discrete Wavelet transform uses orthogonal bases and dyadic scaling and translation which result in fast processing and coefficients without any redundancy. However, because of limitation of scales the transient impulses may not be seen in the Wavelet coefficients by visual reading (Wang et al., 1995). To interpret the Wavelet coefficients a more robust system is required. In this thesis, discrete Wavelet transform is used to decompose the signal and ANNs is used to detect the condition of the machine automatically. It will be shown that the energy and kurtosis of the Wavelet coefficients can clearly identify the running state of a bearing. Employing ANN is not always applicable as generating faulty signals may not be always possible. The continuous Wavelet transform, can overcome the low resolution of DWT by utilizing non-orthogonal bases. However, at early stage of faults still visual inspection in the early stage of faults may not be able to detect the faults. Many researchers have tried to optimize the Wavelet parameters to improve the visibility of the impulses; however, even at the optimum Wavelet parameters, the transient

impulses might not be visible or might be visible in normal condition that makes it impossible to diagnose the condition of the machine. In this thesis a new method will be introduced to reveals the bearing defect frequency in Wavelet coefficients without going through the training process of ANNs. In this method an attempt has been made to extract hidden information from Wavelet coefficients instead of just optimizing the Wavelet coefficients. This method utilizes singular value decomposition (SVD) and energy of the scales to detect small variations in the Wavelet coefficients. The experimental results will verify that the weakest variations in Wavelets coefficients can be detected.

Chapter 3. Bearing Fault Analyses (BFA) and Experimental Data Used in This Thesis

3.1 Bearing characteristic frequencies (BCF)

Generally, a bearing consist of two rings, inner race and outer race with a set of rolling elements located in a cage which prevents any contact between rollers and also provides uniform spacing. There are many causes of bearing failures including material flaws, lubricant failure, misaligned load, excessive contact stress, electrical discharge, etc. In all cases the malfunction is the result of a defect in inner race, outer race or rolling elements. A vibration analyst can detect these failures by collecting enough information about the frequencies that a defective bearing might produce. Three important frequencies that defective bearings can generate are as follows:

1. Ball/Roller Pass Frequency of Outer race (BPFO): This frequency occurs when each ball/roller passes over the defective location in the outer race and can be calculated using equation 3.1 (Taylor and Kirkland 2004).
2. Ball/Roller Pass Frequency of Inner race (BPFI): This frequency occurs when each ball/roller pass over the defective location in the inner race and can be estimated by equation 3.2 (Taylor and Kirkland 2004).
3. Two times Ball/Roller Spin frequency (2*BSF): This frequency equals twice of the spinning frequency the balls/Rollers. This frequency occurs when a single defect on a

ball/roller contacts the inner race and outer race (that is why BSF is multiplied by 2). Equation 3.3 is used to calculate 2*BSF (Taylor and Kirkland 2004).

For computing BPFO, BPFI, and BSF of a bearing, geometry and rotational speeds of bearings are needed. Parts (a) and (b) of figure 3.1 depict cross-sectional and axial views of a bearing, respectively. All the necessary geometry and speed required to calculate bearing-related frequencies are shown in these figures.

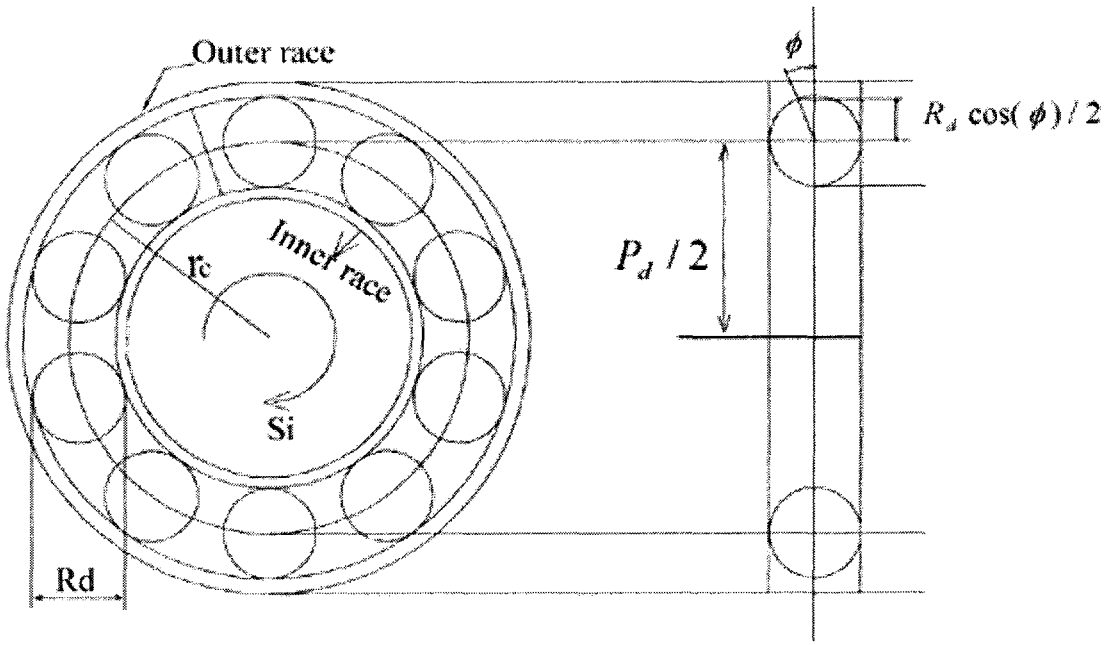


Figure 3.1: (a) Axial view of a generic bearing. (b) Cross-sectional view of a generic bearing (Taylor and Kirkland 2004).

S_i = Rotational speed of inner race
S = Shaft rotational speed, $S = S_i$
P_d = Pitch diameter
R_d =Ball diameter
N_b = Number of balls
ϕ = Contact angle

Table 3.1 shows formulas for calculating bearing fault frequencies.

Frequencies	*Formula*	*Comment*
Ball Pass Frequency of Outer race (BPFO)	$S\frac{N_b}{2}(1-\frac{R_d\cos\phi}{P_d})$ (3.1)	Defective outer race
Ball Pass Frequency of Inner race (BPFO)	$S\frac{N_b}{2}(1+\frac{R_d\cos\phi}{P_d})$ (3.2)	Defective inner race
Two times Ball spin frequency(2*BSF)	$S\frac{P_d}{R_d}(1-\frac{R_d^2\cos^2\phi}{P_d^2})$ (3.3)	Defective Ball

Table 3.1: Bearing fault characteristic frequency formulas (Taylor and Kirkland 2004)

3.2 Bearing Resonance Frequencies (BRF)

As detailed in the previous section the bearing fault characteristic frequencies can be estimated from bearing geometry and rotating speed. However, the resonance frequency of impulse or the oscillation frequency inside an impulse (or simply inside frequency) is very difficult to compute. The inside frequency is the result of excitation of resonance frequencies of the bearing elements and housing structure. Each time a ball rolls over a defect an impact impulse is generated which causes the bearing vibrates with a ringing frequency that depends on the bearing resonance frequencies. Though the resonance frequencies of uninstalled ball bearings could be calculated (Love 1994, Martin 1970, Tandon et.al.,1992), it is very difficult to compute the assembled bearing resonance frequency which has been highly affected by the housing structure, sensors and their associated resonant frequency and any coupled masses (Tandon et.al.,1992). These frequencies are high and independent of rotational speed; however, the shaft speed will increase the resonance amplitude (Martin 1970).

3.3 Experimental data

The experimental data used in this study are taken from Case Western Reserve University (CWRS).

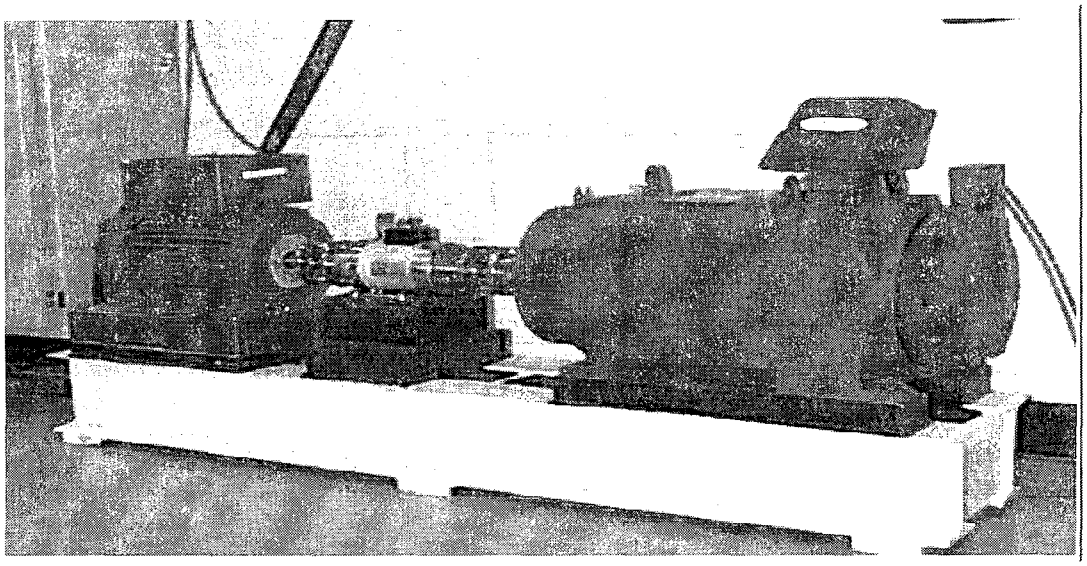


Figure 3.2: The test stand.(Adapted from CWRS website)

The following is the description of the experimental setup and data. Figure 3.2 shows the test stand which consists of a 2 hp motor (left), a torque transducer (center), and a dynamometer (right). In this experiment the SKF bearings support the motor shaft. Single point faults were introduced to the test bearings with a fault diameter of 7 mils (1 mil=0.001 inches) by using electro-discharge machines. Data are collected from drive-end bearing by accelerometers near the housing. The load is 1 horsepower at 1772 RPM and the sampling rate is 48000Hz. The bearing specifications and fault characteristic frequencies are shown in tables 3.2 and 3.3 respectively.

Bearing type	Inside Diameter	Outside Diameter	Thickness	Ball Diameter	Pitch Diameter
6205-2RS JEM SKF (deep groove ball bearing) (in inches)	0.09843	2.0472	0.5906	0.3126	1.537

Table 3.2: Bearing specification

Inner Ring	Outer Ring	Rolling Element
5.4152	3.5848	4.7135

Table 3.3: Bearing fault characteristic frequencies (Hz)

Chapter 4. Bearing Fault Diagnosis Algorithm Using Discrete Wavelet Transform and ANN (Wavelet-ANN)

In this chapter, a new technique for localized bearing fault diagnosis is developed by using the discrete Wavelet transform and ANN. The Wavelet-ANN method can be defined in two stages: preprocessing and diagnosis. In the preprocessing stage, experimental vibration signals are transformed to the time-frequency domain to build feature vectors. In the diagnostic stage ANN is used to detect and classify faults based on the information extracted from Wavelet coefficients. Figure 4.1 illustrates the block diagram of the Wavelet-ANN fault detection scheme. The details of the two stages will be elaborated in the next few sections.

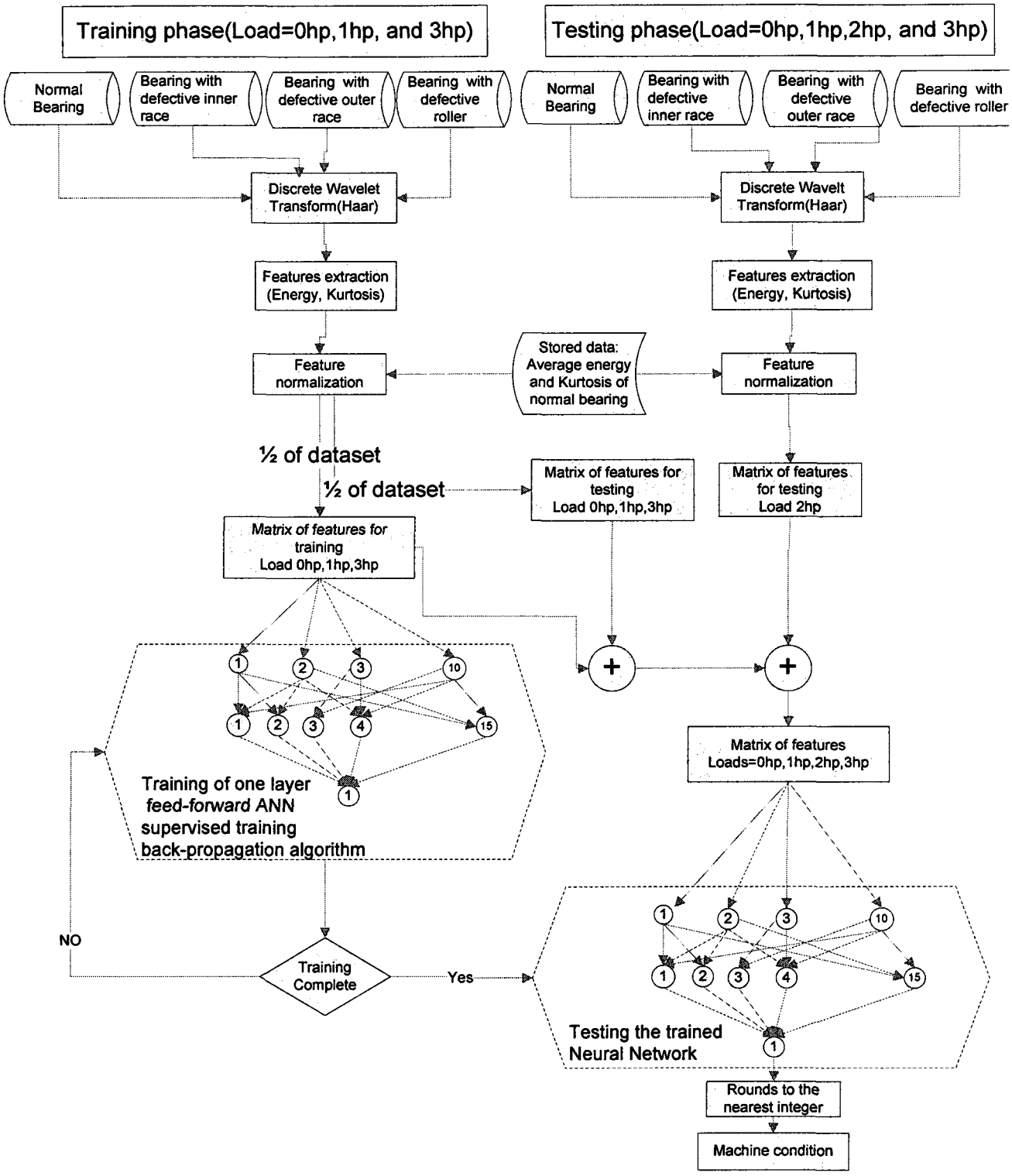


Figure 4.1 Schematic diagram of Wavelet-ANN fault detection and classification process

4.1 Wavelet Analysis

Wavelet analysis uses small wavelike functions known as Wavelets to transform a time domain signal (figure 4.2 (a)) into another representation which is called time-frequency domain (figure 4.2 (b))(Addison 2002). The new representation of the signal makes certain features of the time signal such as time-frequency contents visible.

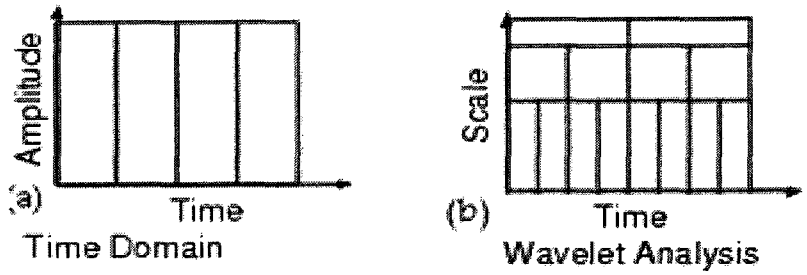


Figure 4.2: (a) Time based view, (b) Wavelet view of a signal (Adapted from Matlab help)

Wavelet transform can be performed in either a continuous or a discrete form. In continuous Wavelet transform a function $\psi(t)$ is used as basis and by taking inner product of a shifted and scaled of this function $\psi(\frac{t-b}{a})$ by a continuous signal, the frequency components of the signal at specific intervals can be computed (Addison 2002):

$$W\ (a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-b}{a})dt \qquad (4.1)$$

In the discrete Wavelet transform (DWT) instead of shifting and scaling Wavelet function, a band-limited signal decomposes the signal into approximation and details by passing through a series of low-pass and high-pass filters with a cut off frequency of

$\omega_0 / 2$ which is half of maximum frequency of the signal, see figure 4.3. The approximations consist of low frequency components and details consist of high frequency components of the signal. After filtering, the output has only half the frequency components of the original signal but the size of the signal is the same as the original, consequently and according to the Nyquist's rule half of the samples can be eliminated. This process will be accomplished by down sampling the output by a factor of two which simply means ignoring every other sample, see figure 4.3. The above process (decomposition) can be repeated for approximation of the signal until the signal is broken down into many lower resolution components and any further decomposition is impossible, see figure 4.3. This is called the Wavelet decomposition tree.

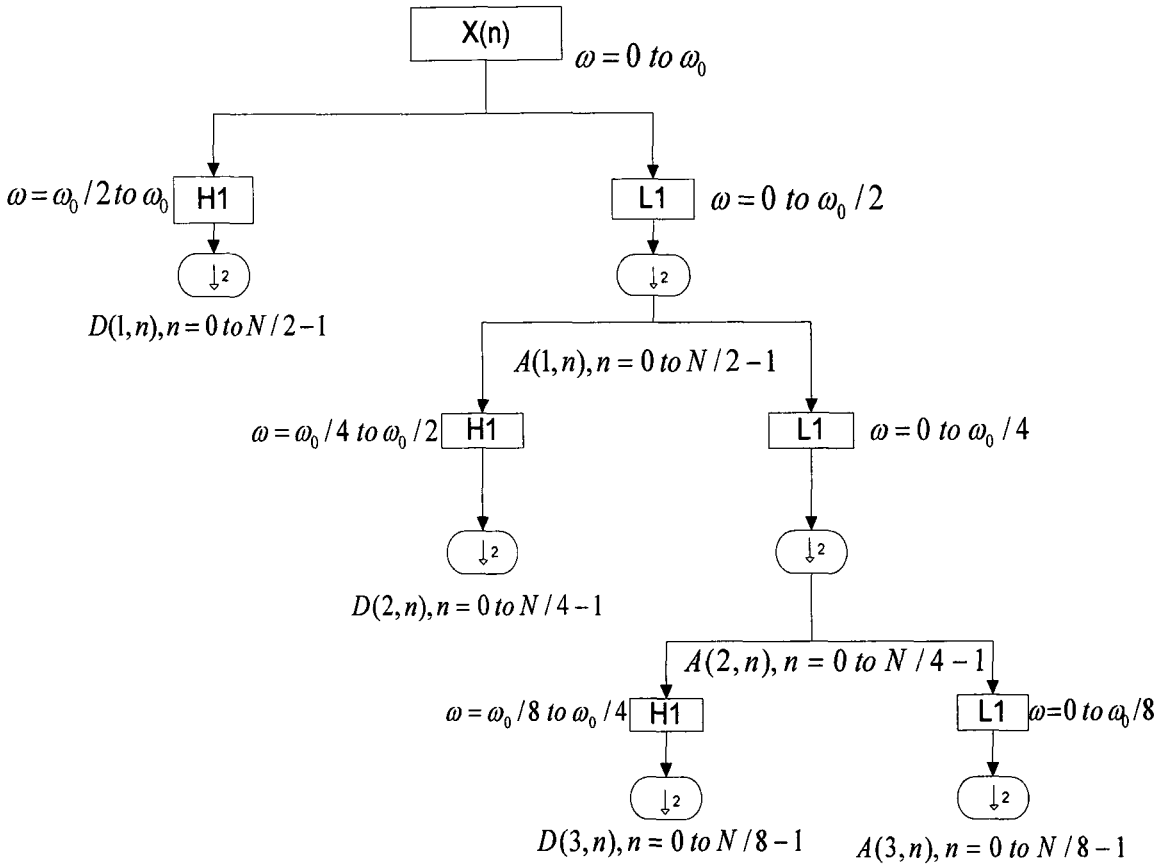


Figure 4.3. Discrete Wavelet multilevel decomposition

Corresponding to the low-pass filters, there might exist a continuous time scaling function $\phi(t)$, and corresponding to the high pass filters there might be a Wavelet function $w(t)$ (Strang et al. 1997). These functions can be derived from equations 4.2 and 4.3 below, which are respectively called scaling and Wavelet equations:

$$\phi(t) = \sum_{k=1}^{\infty} c(k)\phi(2t-k) \tag{4.2}$$

$$w(t) = \sum_{k=1}^{\infty} d(k)\phi(2t-k) \tag{4.3}$$

There are a wide variety of Wavelets to choose from. In this work three different discrete Wavelets, Haar, db4, and db8 are tested to evaluate the performance of the system. In the following sections the selected Wavelets will be discussed.

4.1.1 Haar Wavelets

The Haar Wavelet is the simplest, fastest, and only compactly supported orthonormal Wavelet that is linear phase. Its scaling function has only two nonzero coefficients c(0)=c(1)=1 and from equation 4.2 the scaling equation is defined by (Addison 2002)

$$\phi(t) = \phi(2t) + \phi(2t-1) \tag{4.4}$$

The solution of Haar scaling equation leads to the box function shown in figure 4.4 (a) and is defined as

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & elsewhere \end{cases} \tag{4.5}$$

Using Wavelet equation 4.3, the Haar Wavelet function is defined by

$$w(t) = \phi(2t) - \phi(2t-1) \tag{4.6}$$

The solution of Haar Wavelet function is shown in figure 4.4 (c) and is written as:

$$w(t) = \begin{cases} 1 & , \quad 0 \le t < 1/2 \\ -1 & , \quad 1/2 \le t < 1 \\ 0 & , \quad otherwise \end{cases} \tag{4.7}$$

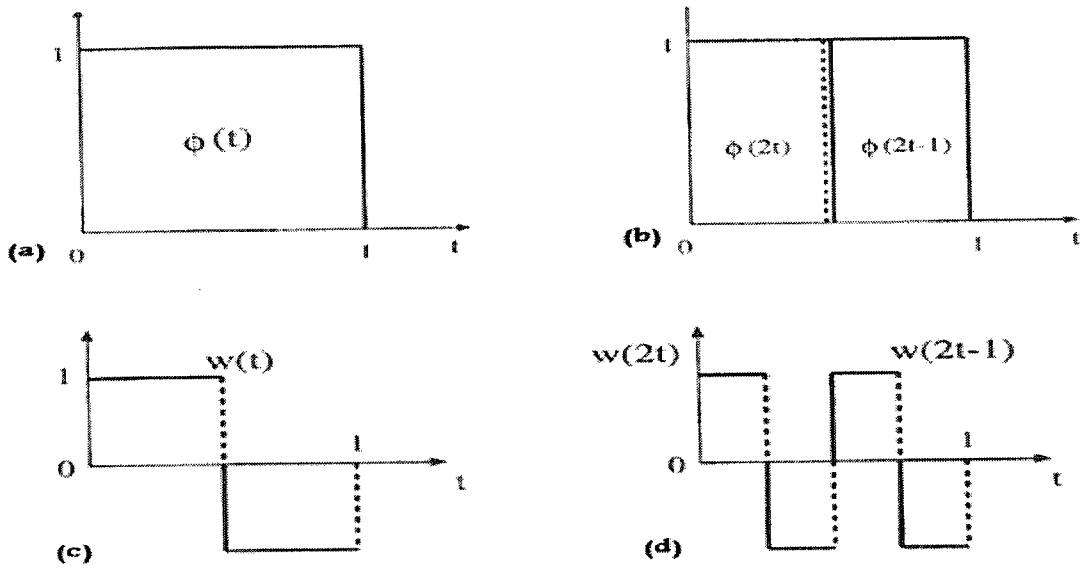


Figure 4.4: (a) Haar scaling function, (b) Haar scaling and translation, (c) Haar Wavelet, and (d) Scaled Wavelets w(2t) and w(2t-1), adapted from Strang et. al. 1997, page 23-24

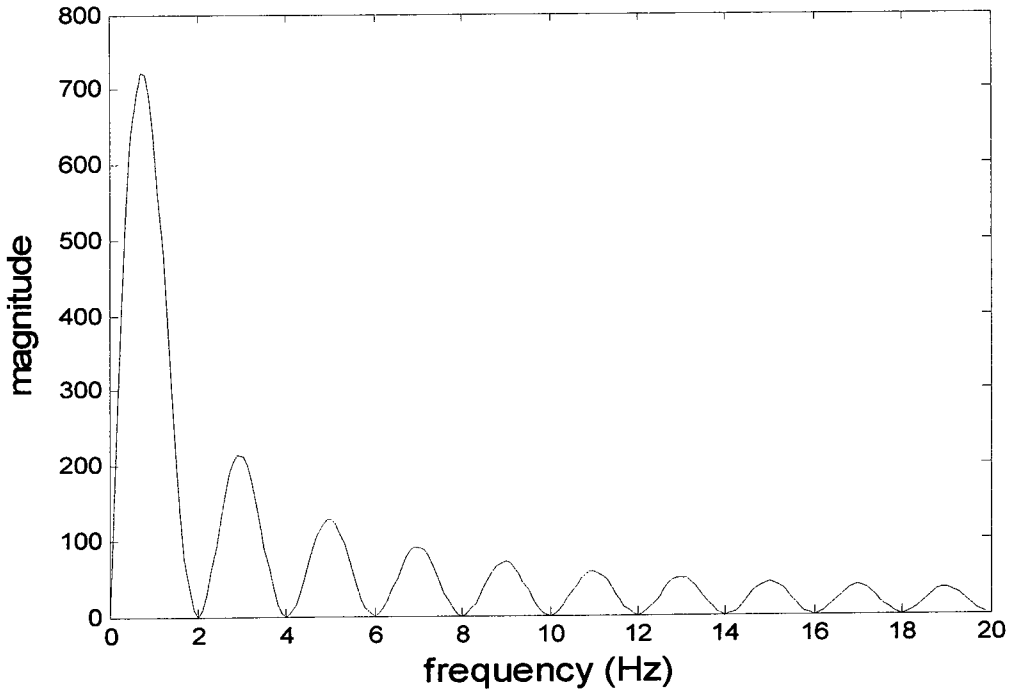


Figure 4.5: Fourier transform of Haar Wavelet

By scaling, the Haar Wavelet can be arbitrarily sharp and therefore localized in time; however, the frequency localization is not very good, see figure 4.5. Computational efficiency is one important advantage of Haar Wavelets. In some situations, the higher resolution provided by some other Wavelets is not worth the computational cost especially in online applications.

4.1.2 Daubechies Wavelets, D8, D16

Ingrid Daubechies (1992) invented compactly supported orthonormal Wavelets. Daubechies Wavelets are written in the form of DN in which D is the name of the Wavelet and N is the order of the filter or number of filter coefficients. The D2 correspond to Haar Wavelets. Daubechies Wavelets D8, D16 (Note: D8 and D16 are equivalent to db4 and db8 in MATLAB) and their scaling functions are illustrated in figure 4.6. The scaling function acts as a low-pass filter and Wavelet function acts as a high-pass filter. As the support length of Wavelet increases, the oscillatory nature of spectra decreases.

For the purposes of comparison, in this study three Wavelet bases, Haar, D8, and D16 are used for the decomposition of the signal. The experimental results will show that the simplest form of Wavelet, Haar Wavelets with only 2 filter coefficients, can successfully extract information from the time signal.

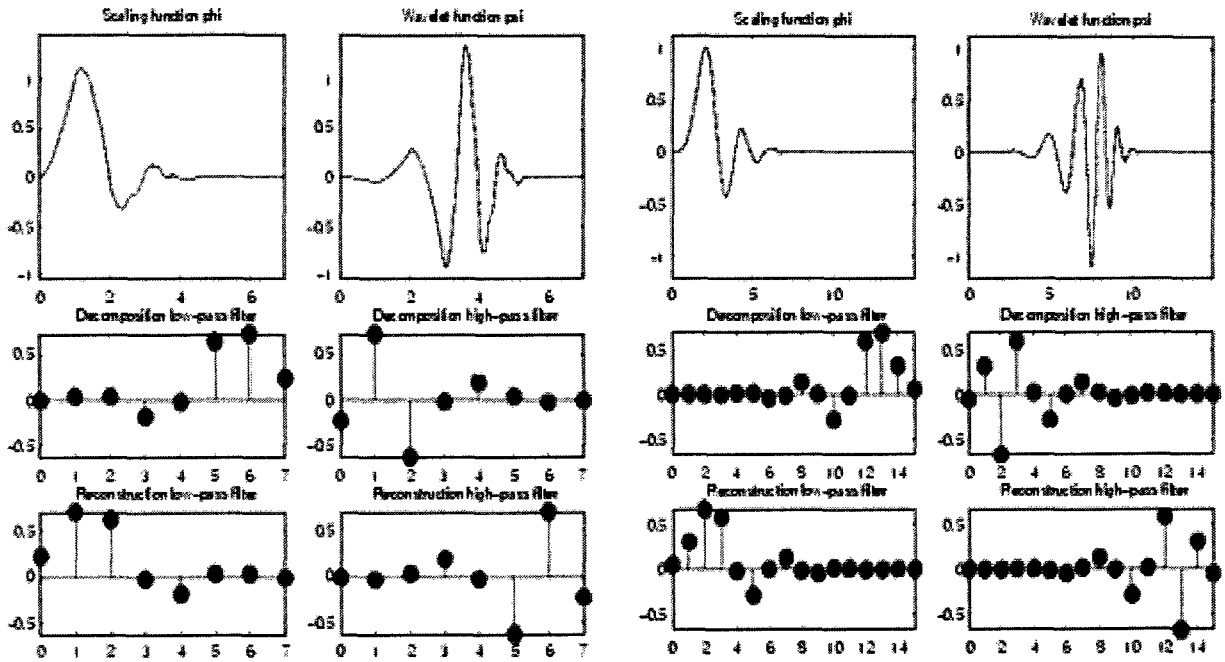


Figure 4.6: Daubechies Wavelet D8 on the left and D16 on the right (adapted from Matlab help)

4.2 Feature extraction

Transforming the time-domain signal to time-frequency domain by discrete Wavelet transform will result in a series of coefficients. To reduce the dimensionality of the Wavelet coefficients, some features which reflect the running state of the machinery should be extracted. In the next two sections it will be shown that the percentage of Wavelet coefficients' energy and kurtosis can effectively determine the condition of a machine.

4.2.1 Scale dependent Wavelet energy

A defective bearing generates periodic short duration pulses which increase the overall energy level of the vibration signal. The total energy of a discrete time signal with finite energy is defined by:

$$E = \sum_{n=0}^{N-1} x(n)^2 \tag{4.8}$$

where E is the energy, $x(n)$ is of finite length N.

In the time domain, individual contributions of energy from components in a machine to the overall machine vibration are very difficult, if not impossible, to identify. Transforming the signal to time-frequency domain by Wavelet-transform will let the transient impulse pass trough one of the Wavelet filters with the appropriate pass-band. So the impact impulses will increase the energy of the Wavelet coefficients at a specific scale.

In Wavelet transform the relative contribution of the signal energy contained at scale index m is given by (Addison 2002).

$$E_m = \sum_{n=0}^{2^{M-m}-1} (D_{m,n})^2 \tag{4.9}$$

where E_m is scale-dependent Wavelet energy spectrum, m is the scale index, $D_{m,n}$ is the Wavelet coefficient at scale index m and time location n, and M is the total number of scales which is defined by $N = 2^M$. Note that in the upper limit of the summation, 2^{M-m}, represents the total number of Wavelet coefficients in scale m. Obviously, different scale levels contain different number of Wavelet coefficients. Peaks in E_m show the dominant energetic scales in the signal.

4.2.2 Scale dependent Wavelet kurtosis

Kurtosis of Wavelet coefficients gives another metric which can identify whether coefficients at a specific scale are randomly varying over time or not. This metric can determine whether the data is coming from a random source or if there is a peak resulting

from a non-random source. Kurtosis is defined in statistics as the fourth order moment of the data set. The kurtosis of a data set is given as follows

$$K = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^4}{(N-1)S^4} \tag{4.10}$$

where X_i is the amplitude of *ith* data point, $\bar{X}$ is the mean value, S is the standard deviation, and N is the length of the signal sample.

Feature vectors constructed based on the energy and kurtosis of Wavelet coefficients are a powerful combination which can detect and classify faults very well. To illustrate, two experimental date sets are plotted in Figure 4.7. Part (a) in figure 4.7 shows a time signal consisting of noise, a high frequency transient impulse (1500Hz) and a low frequency transient impulse (50Hz). The durations of both transient impulses are 0.1sec. Part (c) shows the discrete Wavelet transform of the signal and figure 4.8 shows Approximation and details of the Wavelet coefficients.

Part (e) is the percentage of energy at different scales (calculated based on equation 4.9). Please note that, as there was no "normal" condition, the calculation was made based on equation 4.9. Therefore, energy instead of relative energy was plotted in the figure.). The plot shows that two dominant scales, i.e., scales 2 and 7, are related to the transient impulses at 1500 Hz and 50 Hz, respectively. Part (g) represents the kurtosis value of each scale. Similar to the energy graph there are two peaks in the kurtosis plot. The plot of the second data set is shown in figure 4.7(b) where the frequencies of the two transient signals are identical to those in the first data set, i.e., 1500 Hz and 50 Hz. However, the

duration of the high frequency transient signal is extended to 0.4 sec and the duration of the low frequency transient is reduced to 0.05 sec. Parts (d),(f) and (h) show respectively the Wavelet, energy, and kurtosis plots of the signal. In this case, the energy and kurtosis distributions are different from those of the first data set. The energy level now is very high at scale 2 but much lower at scale 7. On the other hand, the kurtosis shows an opposite trend, low at scale 2 but high at scale 7. Figures 4.8 and 4.9 show approximations and details of the Wavelet coefficients of the first and second signals, respectively.

Comparing figures 4.7 (e) and (f) indicates that the high energy level at scale 2 is caused by the extended high frequency signal duration whereas the low energy at scale 7 is due to the reduced duration of the low frequency signal. Therefore, both the signal transient nature and the long duration contribute to a high energy level. However, the energy level may not reflect the existence of short impulsive signals. In contrast, a comparison of figures 4.7 (g) and (h) suggests that the kurtosis is sensitive to impulsiveness or the short transient behavior of a signal but cannot clearly reveal a transient of prolonged duration. Nevertheless, both types of information could be important in fault detection. For rotational machinery, the periodic impulsiveness implies existence of a certain fault. The duration of a transient may help to distinguish the type, source or even the severity of the fault. Therefore, it is desirable to make full use of both energy and kurtosis for more informed fault detection decisions.

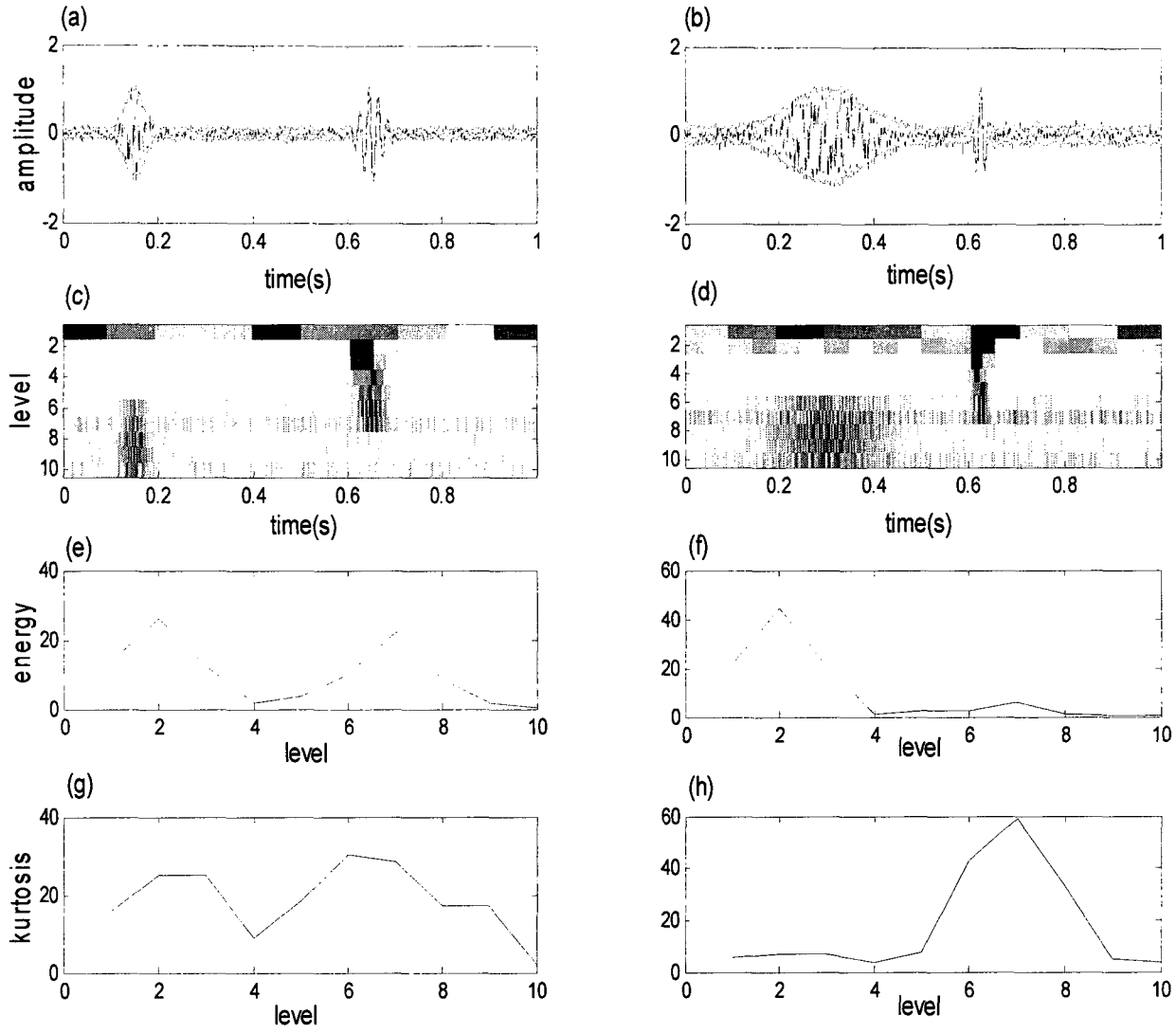


Figure 4.7: (a) The first signal (a transient impulse of 1500Hz appears between 0.1 to 0.2 seconds and another transient impulse of 50Hz appears between 0.6 to 0.7 seconds), (c) Discrete Wavelet transform of the first signal (absolute coefficients), (e) Wavelet energy distribution of the first signal, (g) kurtosis distribution at different scales of the first signal, (b) The second signal (a transient impulse of 1500Hz is seen between 0.1 to 0.5 seconds and a transient impulse of 50Hz developed from 0.6 to 0.65 seconds), (d) Discrete Wavelet transform coefficients of the second signal, (f) Wavelet energy distribution of the second signal, and (h) kurtosis distribution at different scales of the second signal.

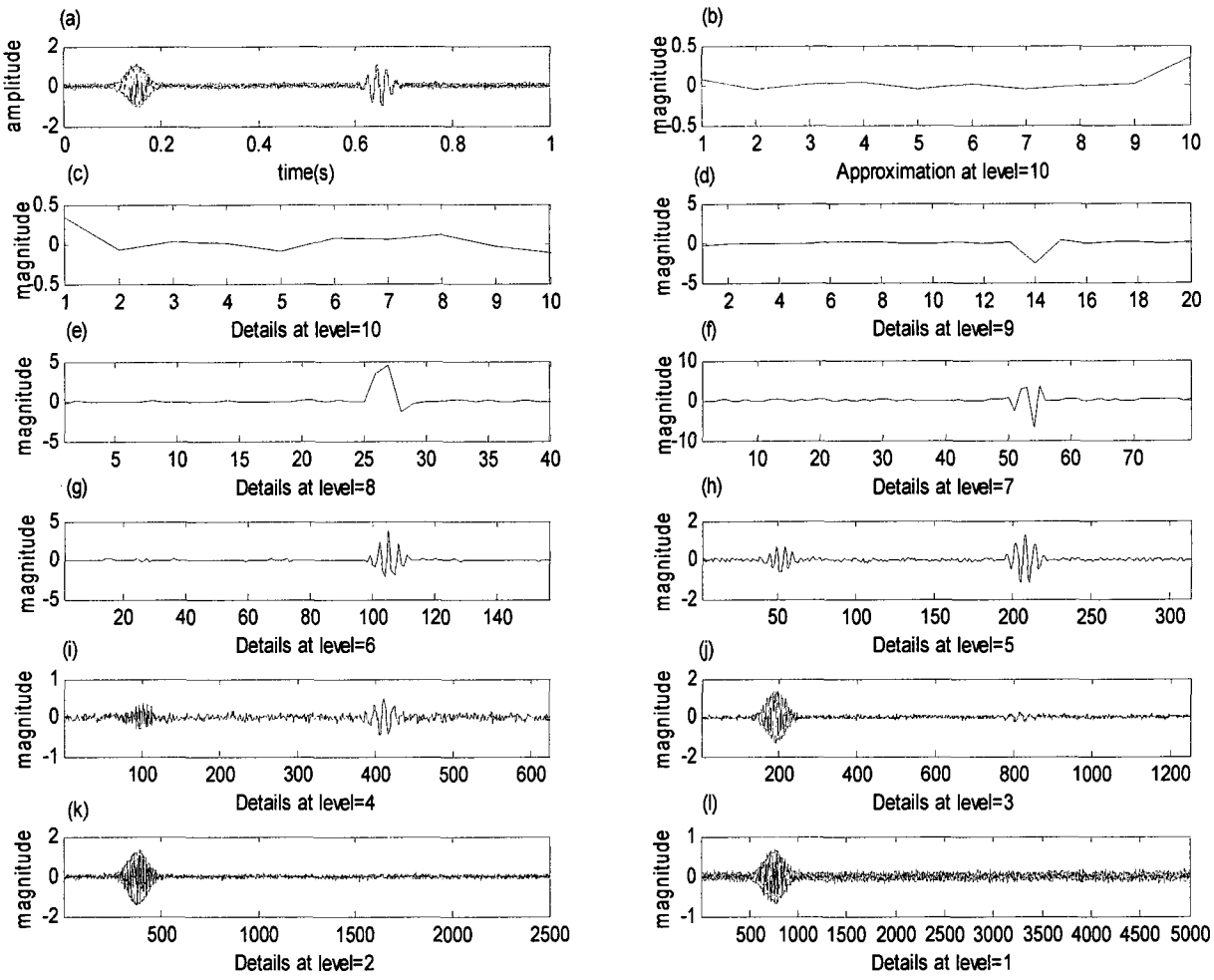


Figure 4.8: (a) The first experimental signal with two transient impulses of 1500 Hz and 50 Hz appearing from 0.1 to 0.2s and 0.6 to 0.7s, respectively), (b to l) DWT Approximation and details of the signal

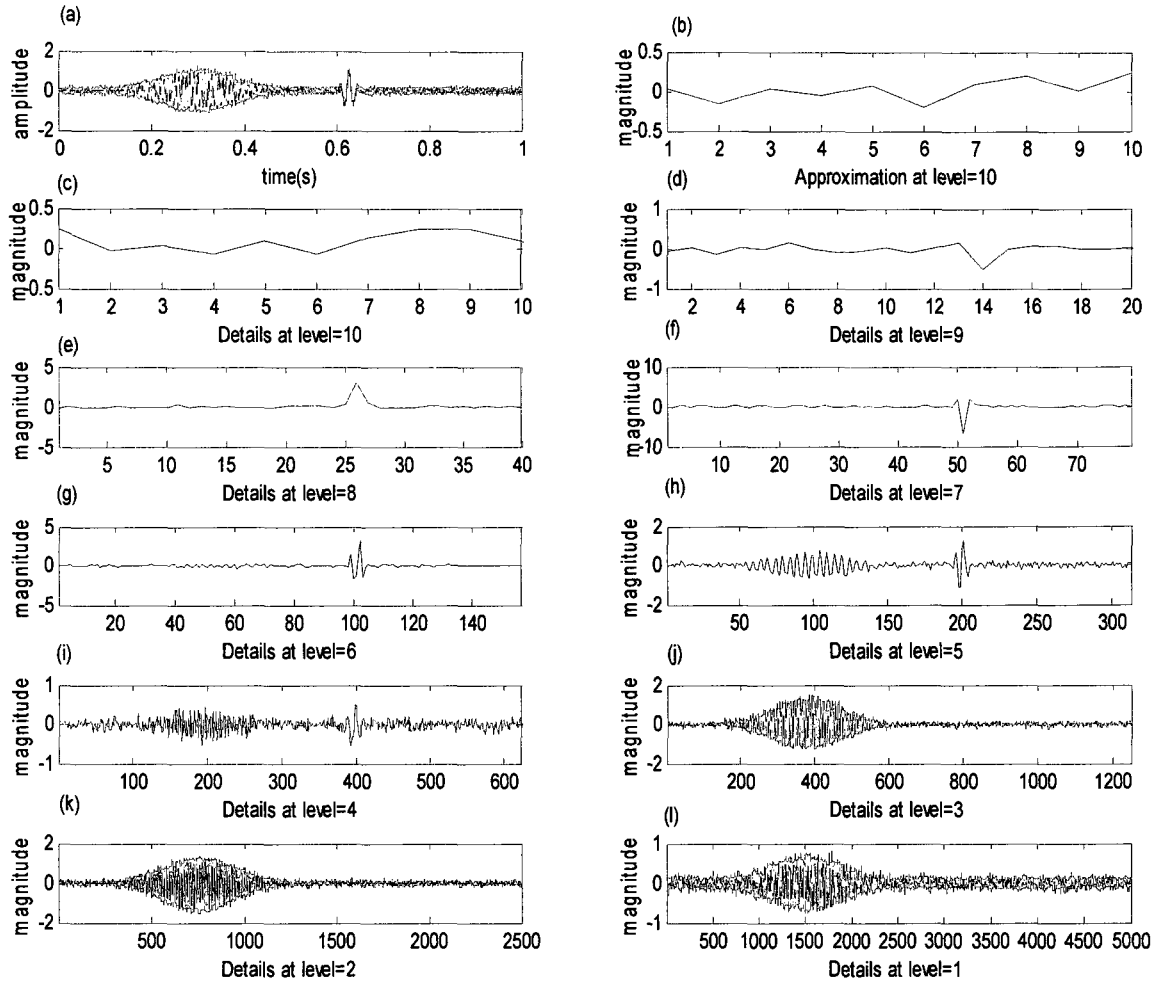


Figure 4.9: (a) The second experimental signal with two transient impulses of 1500 Hz and 50 Hz appearing from 0.1 to 0.5s and 0.6 to 0.65s, respectively), (b to l) DWT Approximation and details of the signal

4.2.3 ANN inputs

The above discussion has shown that the kurtosis and energy values can reflect the transient nature of a signal. Therefore their values will be used as the feature vectors as the inputs to the neural network. However, the direct use of the kurtosis and energy as the inputs to the ANN may not be appropriate because magnitude ranges of the two are quite different. Thus, the direct use of the kurtosis and energy values may unfairly amplify the

effect of one and discount the influence of the other. In this study, the relative kurtosis and energy values are used. They are illustrated in the following.

The relative kurtosis or energy is calculated by dividing the kurtosis or energy by the average kurtosis or energy of the normal condition, i.e.:

$$RK_i = \frac{K_i}{Kn_i} \tag{4.11}$$

where K_i, Kn_i and RK_i are respectively the kurtosis of the monitored bearing, average kurtosis of a normal bearing and relative kurtosis of scale index i.

$$RE_i = \frac{E_i}{En_i} \tag{4.12}$$

where E_i, En_i and RE_i are respectively the energy of the monitored bearing, energy of a normal bearing and relative energy of scale index i.

4.3 Artificial Neural Networks (ANN)

An ANN is a processing unit that acts like a biological nervous system. The structure of ANN is composed of a large number of interconnected neurons working together to solve a specific problem. One important property of ANN is that no details of the system are required. In other words, ANN considers the system as a black box and tries to mimic the behavior of the system through its computational elements connected by links with adjustable weights. An ANN consists of a set of nodes, usually organized into layers, and connected through weight elements called synapses. At each node, the weighted inputs are summed, thresholded and then inputted to the activation functions to produce output for that specific node. These operations are shown in figure 4.10.

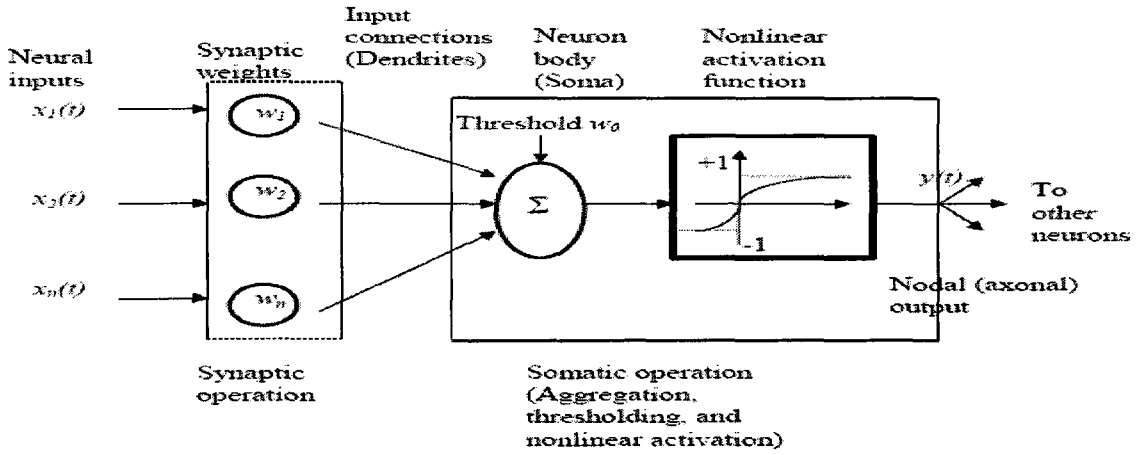


Figure 4.10: The operation at a node of a neural network.
Adapted from KARRAY et al. 2004, page 42

Feed-forward networks are considered one of the most important types of structures in the ANNs. In feed-forward networks, neurons are organized in layers and these layers are connected in one direction and there is no loop in the network. Through a training process, the parameters of an ANN are adjusted for a specific application. There are two kinds of learning algorithm: supervised and unsupervised. In supervised learning, by using target output a neural network are trained in a way that a particular input leads to a specific target output. One of the most important supervised learning algorithms is the Back-propagation Learning Algorithm which is mainly used to train multilayer neural networks. In back-propagation learning the network weights are updated in the direction of the gradient of the performance function. For mathematical details refer to KARRAY et al. 2004.

4.3.1 Neural network architecture for fault detection

For fault detection purpose, the extracted features, i.e., the relative energy and kurtosis values, are inputted to a one-layer feed-forward ANN with hyperbolic tangent sigmoid

neural transfer function (figure 4.11 (b)) for hidden layer and linear transfer function (figure 4.11 (a)) for output layer. The output of the ANN is a real number. This real number then will be rounded to the nearest integer number which indicates the condition of the machine and the type of the fault. The ANN is trained using the back-propagation algorithm.

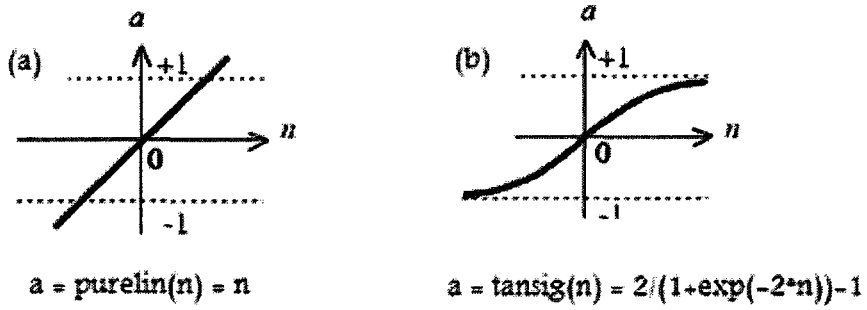


Figure 4.11: (a) linear transfer function, (b) hyperbolic tangent sigmoid (Adapted from Matlab Help)

4.4 Experimental results

Four types of bearing conditions: normal, defective inner race, defective outer race, and defective roller, with the following loads and speeds are used to train the neural network:

Load = 0 hp, speed 1797 RPM

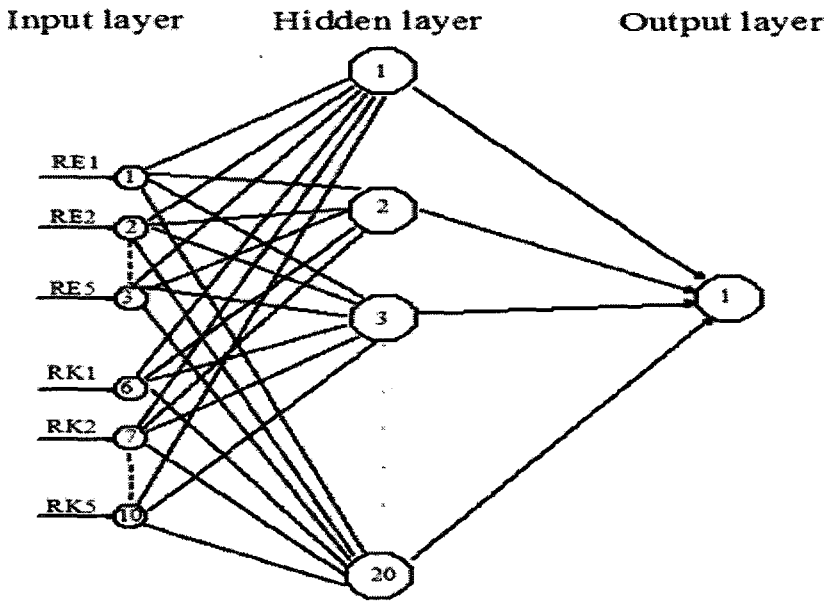
Load = 1 hp, speed 1772 RPM

Load = 3 hp, speed 1730 RPM

The experiment is performed on the vibration signals with duration of 0.5 seconds. To measure the effect of the number of samples on the performance of the system, the Wavelet-ANN approach is performed on the vibration signals segmented into sample sizes of 200, 400, 600, 800,1200, and 2400. To evaluate the effect of Wavelet bases on

the detection rate of the system, three Wavelet bases, Haar, D8, and D16 are used in the decomposition (preprocessing) stage.

A five-scale discrete Wavelet transform is applied to each segment. The energy and Kurtosis of Wavelet coefficients at each scale for each bearing condition are computed and used as the feature vectors. The feature vectors are then normalized. The normalized feature vectors are used as the inputs of a one-layer feed-forward neural network with one hidden layer, 10 input nodes and one output node. Ten input nodes are associated with 10 extracted features (RE_1 - RE_5 and RK_1 - RK_5). The number of the neurons in the hidden layer is tested for 15 and 20 neurons, respectively. The structure of the ANN with 10 inputs, one hidden layer of 20 neurons, and 1 output neuron is shown in figure 4.12.



Figures 4.12: Structure of the ANN with 10 inputs, one hidden layer with 20 neurons, and 1 output neuron.

In the testing phase a new set of signals collected at load 2 hp is added to the test signals in addition to the load 0, 1 and 3 hp which has been used in the training phase. The speed at this load is 1750 RPM.

The vibration signals were transformed to the Wavelet domain using a MATLAB program. The training of the neural network with the data sets is performed in the MATLAB Neural Networks Toolbox. Tables 4.1, 4.2, 4.3, 4.4, 4.5, and 4.6 show the performance of the Wavelet-ANN for three Wavelet bases (HAAR, D8, and D16), two hidden layer sizes (15 and 20 neurons respectively) and four sample sizes (200,400,600, 800,1200, and 2400).

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
20	0	1797	240000	200	1200	5	0.996	1200	90	0.925	1200	144	0.88	1200	266	0.778	
20	1	1772	240000	200	1200	22	0.982	1200	80	0.933	1200	125	0.90	1200	359	0.701	
20	2	1750	240000	200	1200	30	0.975	1200	147	0.878	1200	161	0.87	1200	242	0.798	
20	3	1730	240000	200	1200	42	0.965	1200	149	0.876	1200	212	0.82	1200	121	0.899	
Total						99			466			642			988		0.886
20	0	1797	240000	400	600	0	1.000	600	31	0.948	600	2	1.00	600	19	0.968	
20	1	1772	240000	400	600	3	0.995	600	9	0.985	600	3	1.00	600	17	0.972	
20	2	1750	240000	400	600	2	0.997	600	28	0.953	600	23	0.96	600	18	0.970	
20	3	1730	240000	400	600	0	1.000	600	22	0.963	600	9	0.99	600	17	0.972	
Total						5			90			37			71		0.979
20	0	1797	240000	600	400	1	0.998	400	9	0.978	400	6	0.99	400	11	0.973	
20	1	1772	240000	600	400	2	0.995	400	12	0.970	400	3	0.99	400	5	0.988	
20	2	1750	240000	600	400	4	0.990	400	6	0.985	400	9	0.98	400	11	0.973	
20	3	1730	240000	600	400	1	0.998	400	4	0.990	400	3	0.99	400	7	0.983	
Total						8			31			21			34		0.985
20	0	1797	240000	800	300	0	1.000	300	5	0.983	300	0	1.00	300	3	0.990	
20	1	1772	240000	800	300	0	1.000	300	4	0.987	300	0	1.00	300	0	1.000	
20	2	1750	240000	800	300	0	1.000	300	4	0.987	300	6	0.98	300	0	1.000	
20	3	1730	240000	800	300	0	1.000	300	0	1.000	300	0	1.00	300	0	1.000	
Total						0			13			6			3		0.995
20	0	1797	240000	1200	200	0	1.000	200	0	1.000	200	0	1.00	200	1	0.995	
20	1	1772	240000	1200	200	0	1.000	200	0	1.000	200	2	0.99	200	0	1.000	
20	2	1750	240000	1200	200	0	1.000	200	0	1.000	200	4	0.98	200	0	1.000	
20	3	1730	240000	1200	200	0	1.000	200	0	1.000	200	1	1.00	200	0	1.000	
						0			0			7			1		0.998
20	0	1797	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
20	1	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
20	2	1750	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
20	3	1730	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
						0			0			0			0		1.000

Table 4.1: Wavelet-ANN performance, HAAR Wavelet, Number of Hidden neuron = 20

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
15	0	1797	240000	200	1200	7	0.994	1200	101	0.916	1200	155	0.87	1200	265	0.779	
15	1	1772	240000	200	1200	33	0.973	1200	99	0.918	1200	118	0.90	1200	428	0.643	
15	2	1750	240000	200	1200	37	0.969	1200	143	0.881	1200	206	0.83	1200	249	0.793	
15	3	1730	240000	200	1200	44	0.963	1200	138	0.885	1200	260	0.78	1200	149	0.876	
Total						121			481			739			1091		0.873
15	0	1797	240000	400	600	0	1.000	600	16	0.973	600	1	1.00	600	25	0.958	
15	1	1772	240000	400	600	1	0.998	600	11	0.982	600	4	0.99	600	12	0.980	
15	2	1750	240000	400	600	3	0.995	600	26	0.957	600	22	0.96	600	13	0.978	
15	3	1730	240000	400	600	0	1.000	600	18	0.970	600	9	0.99	600	16	0.973	
Total						4			71			36			66		0.982
15	0	1797	240000	600	400	2	0.995	400	11	0.973	400	2	1.00	400	10	0.975	
15	1	1772	240000	600	400	0	1.000	400	6	0.985	400	4	0.99	400	5	0.988	
15	2	1750	240000	600	400	1	0.998	400	7	0.983	400	9	0.98	400	6	0.985	
15	3	1730	240000	600	400	2	0.995	400	3	0.993	400	3	0.99	400	7	0.983	
Total						5			27			18			28		0.988
15	0	1797	240000	800	300	0	1.000	300	3	0.990	300	0	1.00	300	3	0.990	
15	1	1772	240000	800	300	0	1.000	300	5	0.983	300	4	0.99	300	0	1.000	
15	2	1750	240000	800	300	0	1.000	300	5	0.983	300	7	0.98	300	0	1.000	
15	3	1730	240000	800	300	0	1.000	300	3	0.990	300	7	0.98	300	0	1.000	
Total						0			16			18			3		0.992
15	0	1797	240000	1200	200	0	1.000	200	1	0.995	200	0	1.00	200	2	0.990	
15	1	1772	240000	1200	200	1	0.995	200	0	1.000	200	2	0.99	200	0	1.000	
15	2	1750	240000	1200	200	0	1.000	200	0	1.000	200	1	1.00	200	0	1.000	
15	3	1730	240000	1200	200	0	1.000	200	0	1.000	200	0	1.00	200	1	0.995	
Total						1			1			3			3		0.998
15	0	1797	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	1	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	2	1750	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	3	1730	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
Total						0			0			0			0		1.000

Table 4.2: Wavelet-ANN performance, HAAR Wavelet, Number of Hidden neuron = 15

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
20	0	1797	240000	200	1200	1	1.00	1200	24	0.98	1200	170	0.86	1200	324	0.73	
20	1	1772	240000	200	1200	3	1.00	1200	86	0.93	1200	95	0.92	1200	177	0.85	
20	2	1750	240000	200	1200	4	1.00	1200	107	0.91	1200	98	0.92	1200	172	0.86	
20	3	1730	240000	200	1200	1	1.00	1200	108	0.91	1200	123	0.90	1200	73	0.94	
Total						9			325			486			746		0.918
20	0	1797	240000	400	600	0	1.00	600	4	0.99	600	10	0.98	600	22	0.96	
20	1	1772	240000	400	600	1	1.00	600	7	0.99	600	3	1.00	600	8	0.99	
20	2	1750	240000	400	600	0	1.00	600	13	0.98	600	8	0.99	600	1	1.00	
20	3	1730	240000	400	600	0	1.00	600	14	0.98	600	5	0.99	600	2	1.00	
Total						1			38			26			33		0.990
20	0	1797	240000	600	400	0	1.00	400	3	0.99	400	0	1.00	400	3	0.99	
20	1	1772	240000	600	400	2	1.00	400	3	0.99	400	0	1.00	400	0	1.00	
20	2	1750	240000	600	400	1	1.00	400	11	0.97	400	0	1.00	400	0	1.00	
20	3	1730	240000	600	400	0	1.00	400	4	0.99	400	0	1.00	400	2	1.00	
Total						3			21			0			5		0.995
20	0	1797	240000	800	300	0	1.00	300	0	1.00	300	0	1.00	300	1	1.00	
20	1	1772	240000	800	300	0	1.00	300	0	1.00	300	0	1.00	300	2	0.99	
20	2	1750	240000	800	300	0	1.00	300	1	1.00	300	0	1.00	300	0	1.00	
20	3	1730	240000	800	300	0	1.00	300	0	1.00	300	0	1.00	300	1	1.00	
Total						0			1			0			4		0.999
20	0	1797	240000	1200	200	0	1.00	200	1	1.00	200	0	1.00	200	0	1.00	
20	1	1772	240000	1200	200	0	1.00	200	1	1.00	200	0	1.00	200	0	1.00	
20	2	1750	240000	1200	200	0	1.00	200	2	0.99	200	0	1.00	200	0	1.00	
20	3	1730	240000	1200	200	0	1.00	200	2	0.99	200	0	1.00	200	0	1.00	
Total						0			6			0			0		0.998
20	0	1797	240000	2400	100	0	1.00	100	0	1.00	100	0	1.00	100	0	1.00	
20	1	1772	240000	2400	100	0	1.00	100	0	1.00	100	0	1.00	100	0	1.00	
20	2	1750	240000	2400	100	0	1.00	100	0	1.00	100	0	1.00	100	0	1.00	
20	3	1730	240000	2400	100	0	1.00	100	0	1.00	100	0	1.00	100	1	0.99	
Total						0			0			0			1		0.999

Table 4.3: Wavelet-ANN performance, D8 Wavelet, Number of Hidden neuron = 20

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
15	0	1797	240000	200	1200	1	0.999	1200	36	0.970	1200	188	0.84	1200	304	0.747	
15	1	1772	240000	200	1200	1	0.999	1200	105	0.913	1200	111	0.91	1200	303	0.748	
15	2	1750	240000	200	1200	5	0.996	1200	141	0.883	1200	134	0.89	1200	181	0.849	
15	3	1730	240000	200	1200	1	0.999	1200	139	0.884	1200	145	0.88	1200	75	0.938	
15						8			421			578			863		0.903
15	0	1797	240000	400	600	0	1.000	600	5	0.992	600	2	1.00	600	14	0.977	
15	1	1772	240000	400	600	0	1.000	600	14	0.977	600	0	1.00	600	8	0.987	
15	2	1750	240000	400	600	0	1.000	600	18	0.970	600	2	1.00	600	3	0.995	
15	3	1730	240000	400	600	0	1.000	600	17	0.972	600	0	1.00	600	5	0.992	
15						0			54			4			30		0.991
15	0	1797	240000	600	400	0	1.000	400	1	0.998	400	0	1.00	400	3	0.993	
15	1	1772	240000	600	400	1	0.998	400	1	0.998	400	1	1.00	400	1	0.998	
15	2	1750	240000	600	400	2	0.995	400	13	0.968	400	0	1.00	400	0	1.000	
15	3	1730	240000	600	400	0	1.000	400	3	0.993	400	0	1.00	400	1	0.998	
15						3			18			1			5		0.996
15	0	1797	240000	800	300	0	1.000	300	0	1.000	300	0	1.00	300	1	0.997	
15	1	1772	240000	800	300	0	1.000	300	0	1.000	300	0	1.00	300	2	0.993	
15	2	1750	240000	800	300	0	1.000	300	2	0.993	300	1	1.00	300	0	1.000	
15	3	1730	240000	800	300	0	1.000	300	0	1.000	300	2	0.99	300	1	0.997	
						0			2			3			4		0.998
15	0	1797	240000	1200	200	0	1.000	200	1	0.995	200	0	1.00	200	0	1.000	
15	1	1772	240000	1200	200	0	1.000	200	1	0.995	200	0	1.00	200	0	1.000	
15	2	1750	240000	1200	200	0	1.000	200	2	0.990	200	0	1.00	200	0	1.000	
15	3	1730	240000	1200	200	0	1.000	200	1	0.995	200	0	1.00	200	0	1.000	
						0			5			0			0		0.998
15	0	1797	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	1	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	2	1750	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	3	1730	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	1	0.990	
						0			0			0			1		0.999

Table 4.4: Wavelet-ANN performance, D8 Wavelet, Number of Hidden neuron = 15

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
20	0	1797	240000	200	1200	6	0.995	1200	33	0.973	1200	278	0.77	1200	306	0.745	
20	1	1772	240000	200	1200	4	0.997	1200	125	0.896	1200	161	0.87	1200	307	0.744	
20	2	1750	240000	200	1200	6	0.995	1200	207	0.828	1200	140	0.88	1200	371	0.691	
20	3	1730	240000	200	1200	2	0.998	1200	169	0.859	1200	154	0.87	1200	185	0.846	
Total						18			534			733			1169		0.872
20	0	1797	240000	400	600	0	1.000	600	14	0.977	600	43	0.93	600	41	0.932	
20	1	1772	240000	400	600	0	1.000	600	34	0.943	600	28	0.95	600	35	0.942	
20	2	1750	240000	400	600	0	1.000	600	71	0.882	600	21	0.97	600	37	0.938	
20	3	1730	240000	400	600	0	1.000	600	32	0.947	600	27	0.96	600	18	0.970	
Total						0			151			119			131		0.958
20	0	1797	240000	600	400	0	1.000	400	0	1.000	400	2	1.00	400	10	0.975	
20	1	1772	240000	600	400	0	1.000	400	14	0.965	400	0	1.00	400	3	0.993	
20	2	1750	240000	600	400	0	1.000	400	24	0.940	400	3	0.99	400	4	0.990	
20	3	1730	240000	600	400	0	1.000	400	5	0.988	400	4	0.99	400	5	0.988	
Total						0			43			9			22		0.988
20	0	1797	240000	800	300	0	1.000	300	0	1.000	300	5	0.98	300	1	0.997	
20	1	1772	240000	800	300	0	1.000	300	2	0.993	300	2	0.99	300	0	1.000	
20	2	1750	240000	800	300	0	1.000	300	12	0.960	300	1	1.00	300	6	0.980	
20	3	1730	240000	800	300	0	1.000	300	6	0.980	300	1	1.00	300	3	0.990	
						0			20			9			10		0.992
20	0	1797	240000	1200	200	0	1.000	200	0	1.000	200	0	1.00	200	3	0.985	
20	1	1772	240000	1200	200	0	1.000	200	2	0.990	200	2	0.99	200	0	1.000	
20	2	1750	240000	1200	200	0	1.000	200	9	0.955	200	0	1.00	200	0	1.000	
20	3	1730	240000	1200	200	0	1.000	200	3	0.985	200	3	0.99	200	3	0.985	
						0			14			5			6		0.992
20	0	1797	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	1	0.990	
20	1	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
20	2	1750	240000	2400	100	0	1.000	100	2	0.980	100	0	1.00	100	0	1.000	
20	3	1730	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	1	0.990	
						0			2			0			2		0.998

Table 4.5:Wavelet-ANN performance, D16 Wavelet, Number of Hidden neuron = 20

#of Hidden Neron	Load	speed	Signal length	Sample Size	Normal			Inner race			Outer race			Roller			Total
	hp				# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	# of data sets	Not Detected	Accuracy rate	Accuracy rate
15	0	1772	240000	200	1200	6	0.995	1200	54	0.955	1200	290	0.76	1200	313	0.739	
15	1	1772	240000	200	1200	4	0.997	1200	141	0.883	1200	166	0.86	1200	327	0.728	
15	2	1750	240000	200	1200	6	0.995	1200	208	0.827	1200	156	0.87	1200	391	0.674	
15	3	1730	240000	200	1200	2	0.998	1200	155	0.871	1200	156	0.87	1200	190	0.842	
Total						18			558			768			1221		0.866
15	0	1772	240000	400	600	0	1.000	600	15	0.975	600	46	0.92	600	42	0.930	
15	1	1772	240000	400	600	0	1.000	600	28	0.953	600	29	0.95	600	19	0.968	
15	2	1750	240000	400	600	1	0.998	600	74	0.877	600	13	0.98	600	30	0.950	
15	3	1730	240000	400	600	0	1.000	600	30	0.950	600	17	0.97	600	11	0.982	
Total						1			147			105			102		0.963
15	0	1772	240000	600	400	0	1.000	400	2	0.995	400	3	0.99	400	12	0.970	
15	1	1772	240000	600	400	0	1.000	400	12	0.970	400	0	1.00	400	5	0.988	
15	2	1750	240000	600	400	0	1.000	400	30	0.925	400	0	1.00	400	7	0.983	
15	3	1730	240000	600	400	0	1.000	400	4	0.990	400	3	0.99	400	2	0.995	
Total						0			48			6			26		0.988
15	0	1772	240000	800	300	1	0.997	300	5	0.983	300	3	0.99	300	5	0.983	
15	1	1772	240000	800	300	0	1.000	300	3	0.990	300	3	0.99	300	1	0.997	
15	2	1750	240000	800	300	1	0.997	300	20	0.933	300	2	0.99	300	15	0.950	
15	3	1730	240000	800	300	0	1.000	300	7	0.977	300	4	0.99	300	2	0.993	
Total						2			35			12			23		0.985
15	0	1772	240000	1200	200	0	1.000	200	0	1.000	200	0	1.00	200	0	1.000	
15	1	1772	240000	1200	200	0	1.000	200	3	0.985	200	0	1.00	200	0	1.000	
15	2	1750	240000	1200	200	0	1.000	200	5	0.975	200	0	1.00	200	1	0.995	
15	3	1730	240000	1200	200	0	1.000	200	1	0.995	200	0	1.00	200	3	0.985	
Total						0			9			0			4		0.996
15	0	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	1	0.990	
15	1	1772	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	2	1750	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
15	3	1730	240000	2400	100	0	1.000	100	0	1.000	100	0	1.00	100	0	1.000	
Total						0			0			0			1		0.999

Table 4.6:Wavelet-ANN performance, D16 Wavelet, Number of Hidden neuron = 15

4.5 Discussion

The experiment results show that the combined DWT and ANN scheme could be a potential feature extraction tool for detection and classification of bearing faults. Tables 4.1 and 4.2 show Wavelet-Neural network performance computed by using Haar Wavelets. As the tables show, the highest accuracy rate is achieved at the sample size of 2400 where the total accuracy rate is 100%. The accuracy rate of 800 samples for the number of hidden layers of 20 and 15 are 0.995% and 0.992% respectively which is highly accurate for practical situation.

Tables 4.3 and 4.4 show that using D8 (filter with 8 coefficients) will improve the accuracy rate to 99.9% for the sample size of 800 and both 20 and 15 hidden neurons respectively. However, when employing 16 filter coefficients, the D16 Daubechies Wavelets decrease the accuracy rate of the system to 99.2% and 98.5% for 20 and 15 neurons, see tables 4.5 and 4.6.

Considering the computational cost of using 8 filter coefficients (D8) instead of using Haar Wavelets (2 filter coefficients), it can be concluded that the Haar Wavelet is the best choice of Wavelets.

The experiment results verify that the extracted features from Wavelet coefficients, energy and kurtosis, can effectively reflect the running state of the bearing. Figures 4.13,

4.14, 4.15, and 4.16 illustrate one segment of the signal for 4 machine conditions (each consist of 800 samples), their discrete Wavelet transforms, the energy, and kurtosis plots. As expected the kurtosis values of the normal bearing at all scales are close to 3 which indicates the lack of impulsiveness in the time signal. For the three faulty conditions, the kurtosis values in a critical scale increase to 6.5, 8 and 6 for the outer, inner and roller faults, respectively, due to the signal impulsiveness.

The experiment results show the influence of sample size on the performance of the system. By increasing the sample size, the correct classification rate of ANN can be improved within the scope of this study.

In all cases, the worst performance is obtained at the sample size of 200 where the total accuracy rate for Haar Wavelet is about 87%. At this sample size, the accuracy rate of the defective inner race is higher than the outer race. Despite what the experiment suggests to the contrary, the outer race generally generates stronger impulses and it should be the easiest type detected. This can be explained by the fact that the impulses generated by the defective inner race have lower periodicity (about 6.25ms) than outer race impulses (about 9.45ms). It means that the sample size of 200 is not sufficient to include outer race transients in all segments. In the sample size of 800 there is at least one impulse in any of the signal segments. Tables 4.1 and 4.2 verify that at higher sample size, the ANN can classify the outer race fault more accurately than the inner race fault. This result is consistent with the outcome of Wavelet-SVD approach which will be introduced later in this thesis.

Another observation from the tables is that increasing the number of hidden layer neurons will not always improve the performance of ANN. In the case of D16 Wavelet, the accuracy rate has been reduced from 96.3% to 95.8% when the number of hidden layer neurons increases from 15 to 20 for sample size 400. Considering all factors, using Haar Wavelets, a sample size of 800 and a feed-forward ANN with 15 hidden neurons can diagnose bearing health condition very accurately and that includes the signals that are not used in the training phase (2hp).

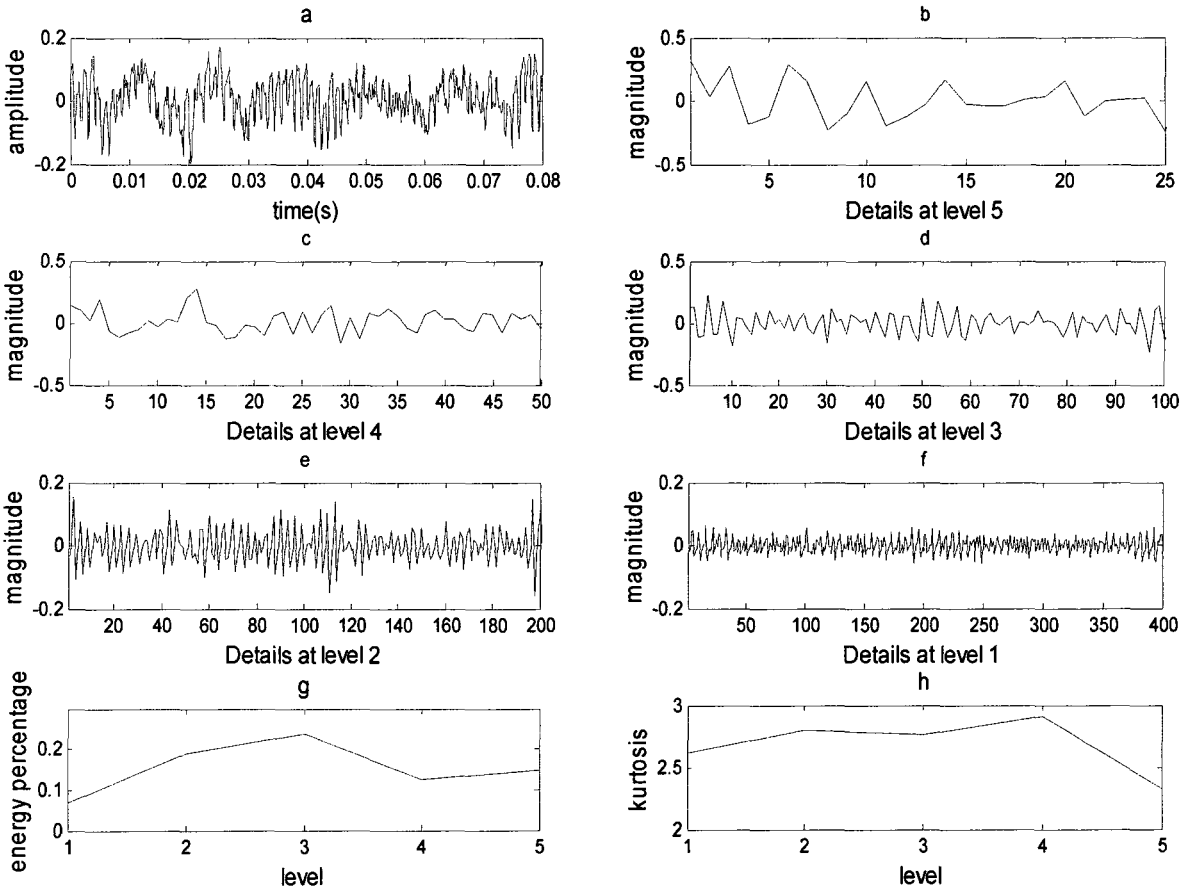


Figure 4.13: (a) Normal condition signal, (b)(c)(d)(e)(f) its discrete Wavelet transform details coefficients, (g) energy of Wavelet coefficients, (h) Kurtosis of Wavelet coefficients.

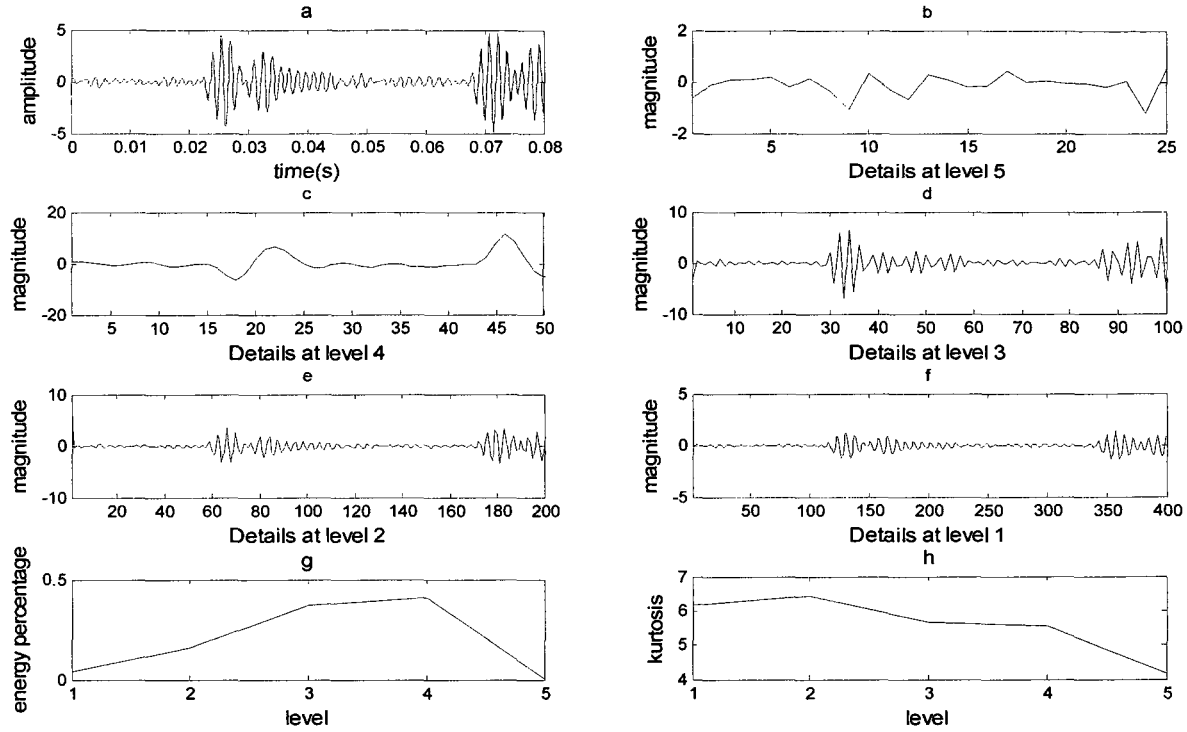


Figure 4.14: (a) Outer race faulty signal, (b)(c)(d)(e)(f) its discrete Wavelet transform details coefficients, (g) energy of Wavelet coefficients, (h) Kurtosis of Wavelet coefficients.

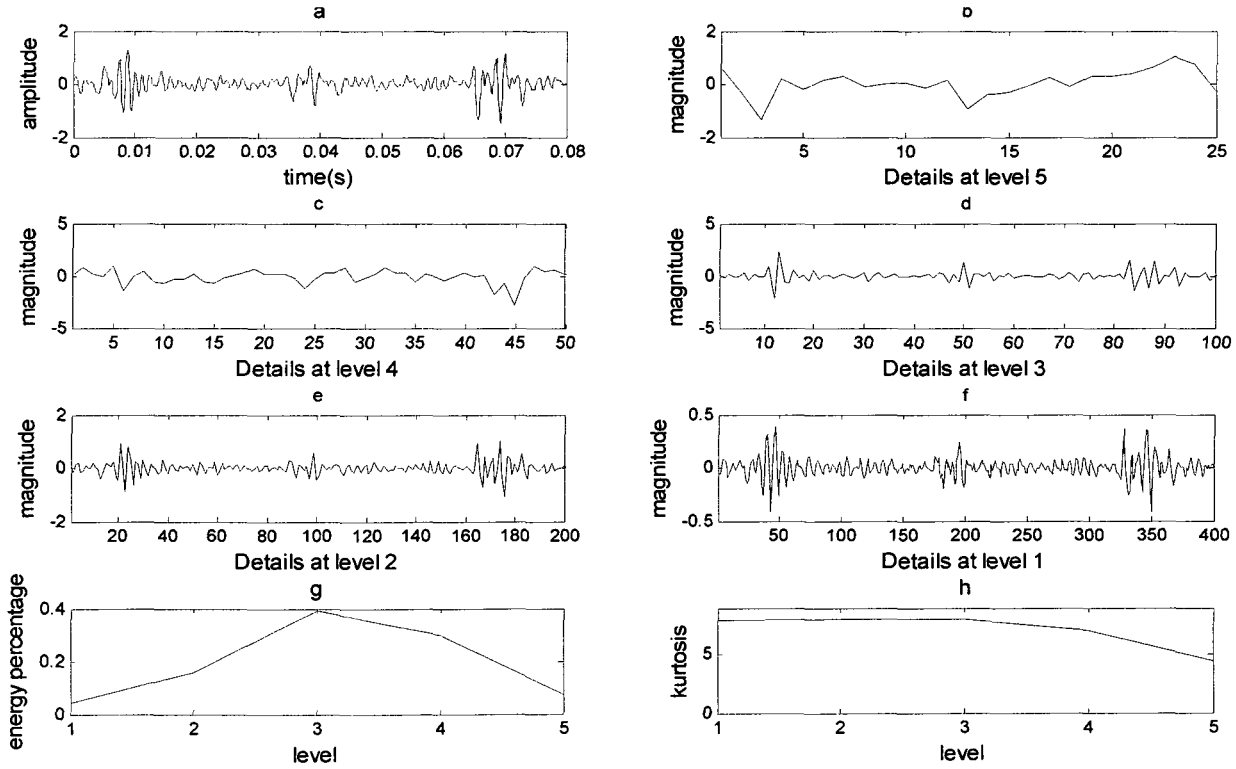


Figure 4.15 (a) Inner race faulty signal, (b)(c)(d)(e)(f) its discrete Wavelet transform details coefficients, (g) energy of Wavelet coefficients, (h) Kurtosis of Wavelet coefficients.

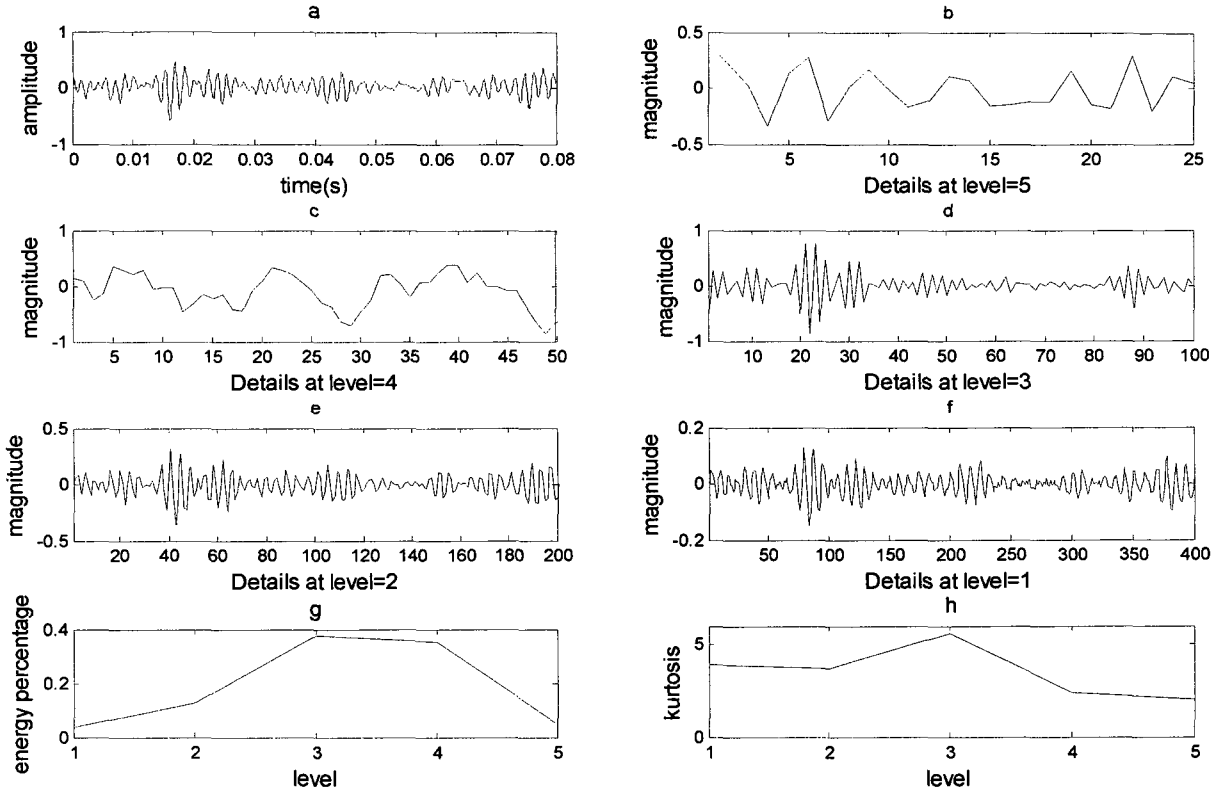


Figure 4.16: (a)Roller faulty signal, (b)(c)(d)(e)(f) its discrete Wavelet transform details coefficients, (g) energy of Wavelet coefficients, (h) Kurtosis of Wavelet coefficients

4.6 Conclusion

The above analysis suggests that Wavelet-ANN is a useful diagnostic system which can accurately detect different bearing defects under different operating conditions, if trained properly. Such a system is effective if the training data are readily available and the bearings operate under similar operating conditions. However, the training data may not always be available in reality and the operating conditions may vary substantially and quickly. In this case, it may not be realistic to acquire training data and re-train the ANN when the bearing experience substantially different working conditions due to the time pressure. This gives rise to a training-free method which will be presented in the next chapter.

Chapter 5. Fault Diagnosis Algorithm Using Continuous Wavelet Transform and Singular Value Decomposition

Discrete Wavelet transform (DWT) is computationally fast; however, because of scales and translation limitations the details of the signal may not be easily identified in the Wavelet plots. The second form of Wavelet transform, continuous Wavelet transform(CWT), can overcome the shortcoming of DWT by utilizing non-orthogonal bases and employing finer scales and translations at the expense of increased redundancy. Using finer scales and translations will amplify signal to noise ratio (SNR) resulting in the domination of impact impulses at some scales. Usually, at the early stage of fault development, impact impulses cannot be identified in the Wavelet plots by visual inspection because they are masked by noise. Selecting similar Wavelet to the transients and optimizing its parameters will improve the visibility of the transients to some extent; nevertheless, the interference of other components, noise, and existence of defect frequency in the normal bearing may still make the detection difficult. The method presented in this section, Wavelet-Energy Singular Value Ratio (Wavelet-ESVR), employs a mathematical decomposition technique and energy distribution of the Wavelet coefficients to measure the variations in the repeating pattern of Wavelet coefficients. If the coefficient's periodicities are found to vary at the rate of a certain characteristic defect frequency, it can be concluded that resonance due to the force impulses is present and the bearing is damaged.

In summary, in the Wavelet-ESVR method presented in this thesis, continuous Wavelet transform is used to convert a vibration signal to time-frequency plane and then the periodic behavior of the Wavelet coefficients is analyzed to diagnose faults. Figure 5.1 is a block diagram of Wavelet-ESVR fault detection scheme. The next few sub-sections will go through steps illustrated in figure 5.1. As the block diagram shows the first step in performing Wavelet-ESVR is to decompose the signal by CWT.

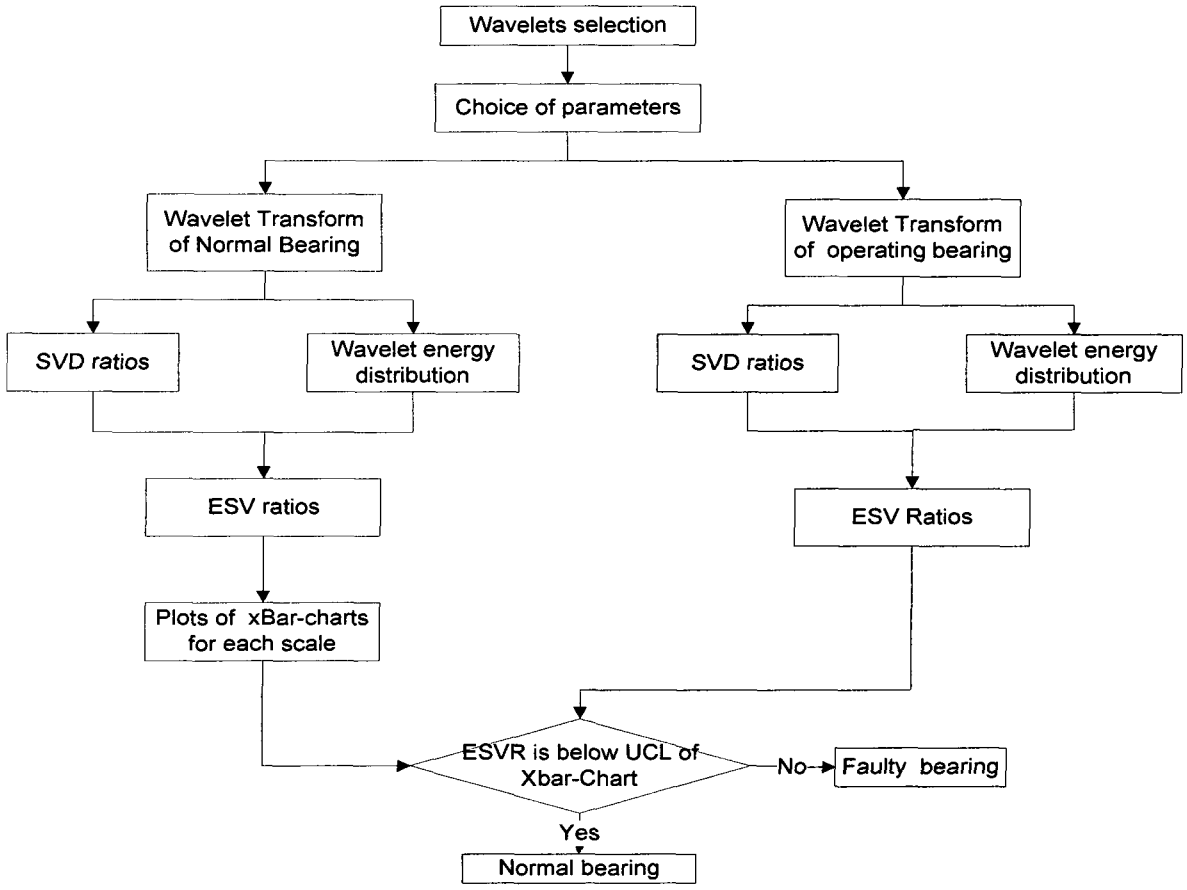


Figure 5.1 Block diagram of Wavelet-ESVR fault detection scheme.

5.1 Continuous Wavelet transforms (CWT)

The Wavelet analysis maps a time function into a two dimensional function of scale and time. By using a given function $\psi(t)$ as a basis and taking inner product of a shifted and scaled version $\psi(\frac{t-b}{a})$ by a continuous signal, the frequency components of the signal at specific intervals can be computed. The continuous Wavelet transform is defined by the following equation (Addison 2002):

$$W\ (a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t-b}{a})dt \tag{5.1}$$

where a is scale which determines the frequency content of the Wavelet, b is translation and determines the position of the Wavelet, and ψ^* is the complex conjugate of ψ, ψ is the basis function, and $W(a,b)$ is the frequency content of x(t) within a time interval and for a specific frequency band.

In order a function to be considered as a Wavelet, it should satisfy the following admissibility condition:

$$C_g = \int_{-\infty}^{\infty} \frac{|\psi(f)|^2}{|f|} df < \infty \tag{5.2}$$

Equation 5.2 shows that the Wavelet must have no zero frequency components or in other words it has zero mean. For detailed explanation refers to (Addison 2002).

There are numerous Wavelet bases to choose from. The best one depends on the nature of the signal and the type of the application. For example, for lossless compression of a signal or image an orthonormal bases is suitable. For lossy compression or noise reduction it is desirable to select a Wavelet similar to the component of interest (Castleman 1996). As will be discussed in the following section, in the case of bearing faults analysis the most appropriate Wavelet can be chosen from the Wavelets which have more similarity to the shape of impact impulses.

5.1.1 Complex Morlet Wavelet

Signals containing impulsive components, such as bearing impact impulses, with some approximation can be modeled as a simple single freedom mass-spring-damper structure and formulated as [Junsheng et al. 2007, Yang et al. 2004].

$$S(t) = \frac{1}{M_a \omega_d} \sum_{k=0}^{\infty} F_k e^{[-\xi\omega_n (t-kT)]} \sin[\omega_d (t - kT)] + n(t) \qquad (5.3)$$

where F_k is the power of impact, $n(t)$ is noise, T is the fault period, M_a is mass, ω_n is the natural circular frequency defined by $\omega_n = \sqrt{Ks/M}$ (Ks is the stiffness of the structure), ξ is the damping ratio determined by $\xi = \frac{C}{2M\omega_n}$, C is the damping constant,

ω_d is the frequency of damped oscillation given by $\omega_d = \omega_n\sqrt{1-\xi^2}$, F_k is constant determines the intensity of the impact.

Equation 5.3 shows the oscillatory signal with decaying behaviors of an impulse impact. This transient oscillatory signal is repeated with time interval of T seconds. Among several Wavelet functions, the time-frequency structure of complex Morlet has more similarity to the transient impulses generated by a faulty bearing. The classical complex Morlet Wavelet function is simply a complex wave $e^{-j\omega_0 t}$ within a Gaussian window $e^{-t^2/2}$ given by

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-j\omega_0 t} e^{-t^2/2} \tag{5.4}$$

where ω_0 is the center of frequency which determines the number of significant oscillations in the Gaussian envelope. Coefficients produced by complex Morlet Wavelet are in a complex form and contains both amplitude and phase components of the signal (Addison 2002).

Figure 5.2 (a) shows the real and imaginary parts of Morlet Wavelet. The energy spectrum of Morlet Wavelet is shown in Figure 5.2 (b) and is defined by

$$\left|\psi(\omega)\right|^2 = e^{-(\omega-\omega_0)^2} \tag{5.5}$$

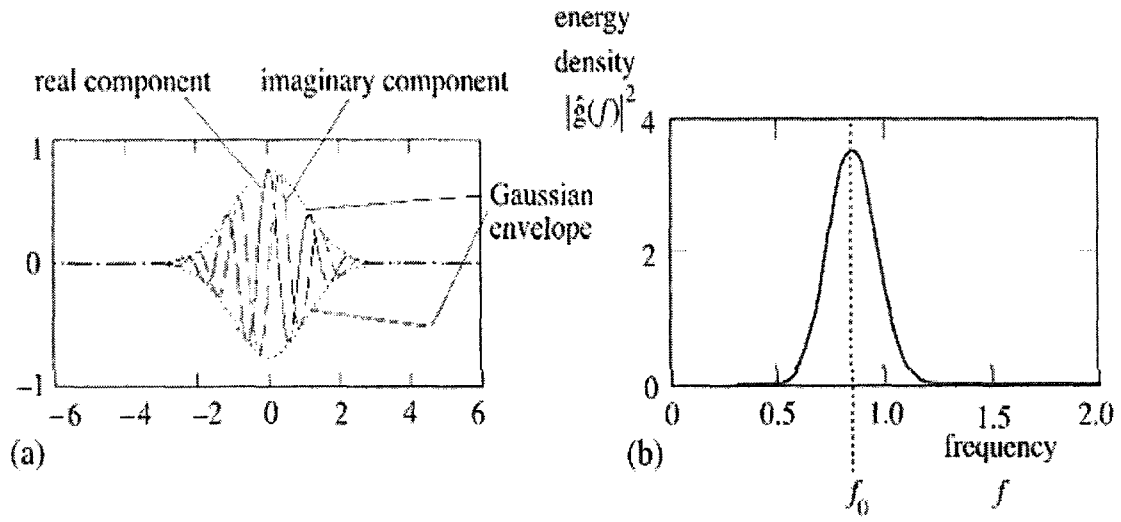


Figure 5.2: (a) Real and imaginary parts of the complex Morlet Wavelet, (b) Energy spectrum of the Morlet Wavelet (Adapted from Addison 2002, page 36)

As the plots show, the Morlet Wavelet has a compact support in both frequency and time domains. Morlet Wavelet is not really a Wavelet as its function has non-zero mean which makes it inadmissible. However, minimum error will be obtained by selecting large value for ω_0 or in practice within the range of $5 < \omega_0 < 6$ (Addison 2002).

Figure 5.3 shows impact impulses generated by real defective bearings with fault diameter FD = 0.007". In this case, both the outer race and roller signals have a strong resemblance to the Morlet Wavelet.

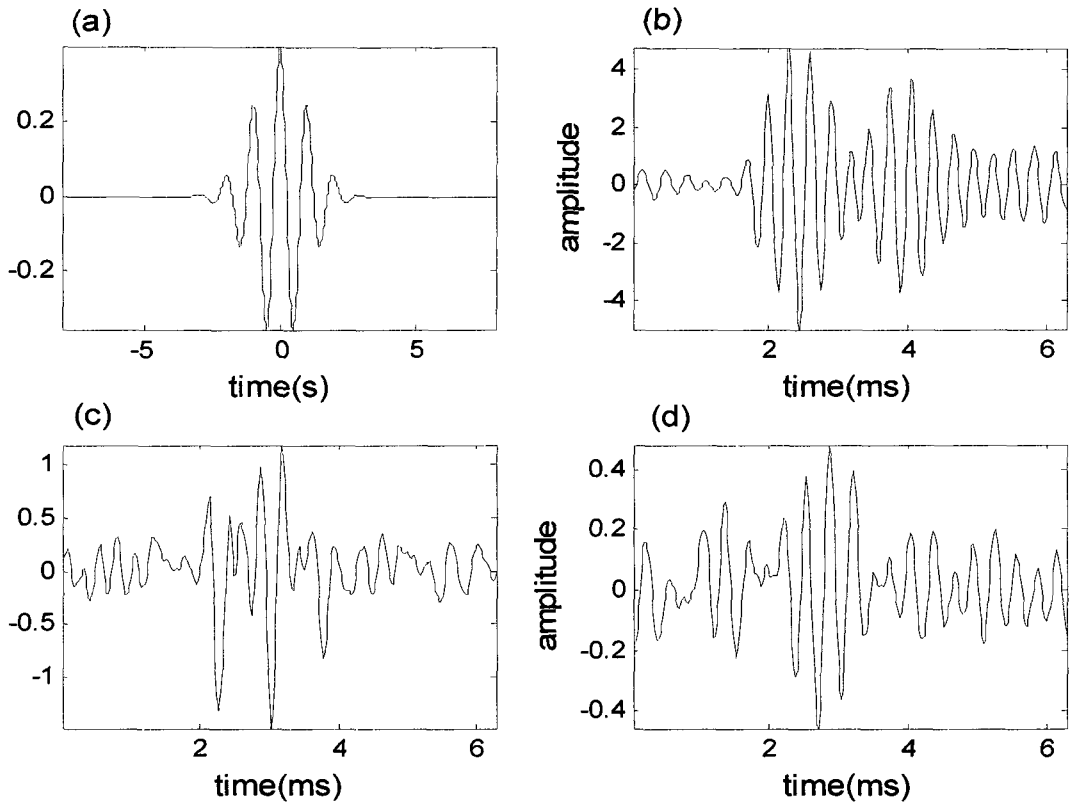


Figure 5.3: (a) Complex Morlet function, real part, central frequency $f_0 = 1Hz$, impact impulses, generated by local defect with FD=.007" on (b) outer race, (c) inner race, and (d) roller

5.1.2 Choice of scales

Based on the Heisenberg uncertainty principle, in the time-frequency plane the area of Heisenberg boxes must be greater than or equal to a constant value (Addison 2002). This implies that the bandwidth of Wavelet filters decreases as the scale increases. Therefore constant spacing with small distances between scales will cause redundancy, especially in high frequencies and when there are large distances between scales which will cause poor frequency resolution. As a result, in order to obtain a sufficient number of scales, the spacing between scales should be decreased as the values of the scales are increased. A common choice for orthogonal discrete Wavelet is dyadic grid scaling. Nevertheless,

dyadic discretization of scales is not appropriate for continuous Wavelet transform with non-orthogonal bases as it will make the time-frequency plane so sparse that a single Wavelet amplitude map cannot describe all details of the signal.

One way to overcome the shortcoming of constant and dyadic discretization is to define a degree of overlap between two adjacent filters (figure 5.4). This method generates an adequate number of filters which can cover the entire frequency range without redundancy in high frequencies. The mathematics of scale resolution of the Morlet Wavelet can be found in Jordan et al. (1997), Le (2007), Le et al. (2001) and can be presented in simpler form as follows.

If the peak frequency and the magnitude response of a Morlet filter at scale a^i are ω_p^i and $\psi(a^i\omega_p^i)$ respectively (see figure 5.4), then the magnitude response of the same filter for the peak frequency of next successive filter (scale a^{i+1} with peak frequency of ω_p^{i+1}) will be $\psi(a^i\omega_p^{i+1})$. The distance (D) between two adjacent band peak frequencies can be determined by specifying a decay constant (R) with a value in the range of $0 < R < 1$. The closer R is to one, the higher the scale resolution. Equation 5.6 shows the relationship between R and D:

$$\psi(a^i(\omega_p^i + D)) = R\psi(a^i\omega_p^i) \tag{5.6}$$

where R is a constant and its value is between [0 1], $D = \omega_p^{i+1} - \omega_p^i$ is distance between two peak frequencies, i is an integer index showing number of scales starting from 1 and , ω_p^i is the peak of frequency at the scale a^i, and a^i is scale and $a^1 = 1$ (Note: it should be

emphasized that i is an integer index and a^i is a scale. For example, $a^2 = 1.56$ means at index 2, scale is 1.56. However, we always have $a^1 = 1$).

Using equation 5.6 and the property of Morlet Wavelets $\omega^i = \frac{\omega_0}{a^i}$ can be rewritten as

$$R = \frac{\psi(a^i(D + \frac{\omega_0}{a^i}))}{\psi(\omega_0)} \tag{5.7}$$

The Fourier transform of the Morlet Wavelet is defined by (Chun et. al. 2000)

$$\psi(\omega) = e^{-\frac{1}{2}(\omega - \omega_0)^2} \tag{5.8}$$

Substituting the value of D, R in equation 5.7 results in

$$R = \frac{e^{\left[-\frac{1}{2}\left(a^i\left(D+\frac{\omega_0}{a^i}\right)-\omega_0\right)\right]^2}}{1} \tag{5.9}$$

$$R = \frac{e^{-\frac{1}{2}(a^i(D+\frac{\omega_0}{a^i})-\omega_0)^2}}{1} = e^{-\frac{1}{2}(a^i(\omega_p^{i+1}-\omega_p^i+\frac{\omega_0}{a^i})-\omega_0)^2} = e^{-\frac{1}{2}(a^i(\omega_p^{i+1}-\omega_p^i))^2} = e^{-\frac{1}{2}(a^i(\frac{\omega_0}{a^{i+1}}-\frac{\omega_0}{a^i}))^2} \tag{5.10}$$

Then

$$R = e^{-\frac{1}{2}(a^i(\frac{\omega_0}{a^{i+1}}-\frac{\omega_0}{a^i}))^2} \Rightarrow -2Ln(R) = (a^i(\frac{\omega_0}{a^{i+1}} - \frac{\omega_0}{a^i}))^2 \Rightarrow \pm\sqrt{-2Ln(R)} = (a^i(\frac{\omega_0}{a^{i+1}} - \frac{\omega_0}{a^i}))$$

Hence

$$a^{i+1} = \frac{\omega_0 a^i}{\omega_0 \pm \sqrt{-2\ln(R)}} \tag{5.11}$$

The plus sign produces results which is not acceptable as the scale a^{i+1} should have value higher than scale of a^i, therefore:

$$a^{i+1} = \frac{\omega_0 a^i}{\omega_0 - \sqrt{-2\ln(R)}} \tag{5.12}$$

This is a recursive equation and the value of each scale is determined by the value of the previous scale. The starting scale $a^1 = 1$.

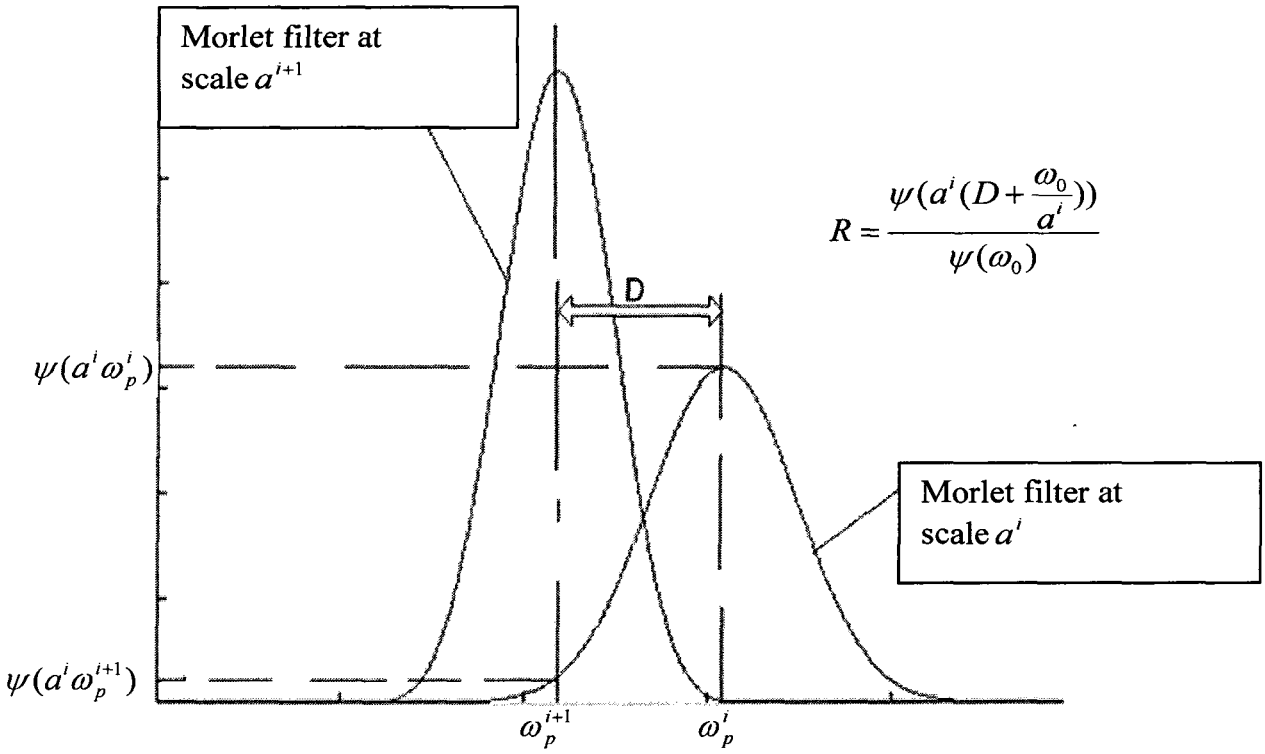


Figure 5.4: Frequency representation of complex Morlet at two scales

The value of *R*, scale resolution, highly depends on the application and the nature of the signal. Selecting high values will improve the chance of matching one of the Morlet filters with the transient signals; however at different segments of the signal it will increase the chance of matching at different scales. Performance of Wavelet-ESVR will

be improved if the resonance frequency of the signal remains mainly at one scale. In this thesis, R=0.8 is chosen. R=0.8 ensures that the resonance frequency is located within pass band of one of the Morlet filters with maximum 6.25% distance from the filter's peck. Figure 5.5 (a) depicts the selected scales and corresponding frequencies for R=0.8. The frequency representation of Morlet filters at the selected scales is shown in figure 5.5 (b).

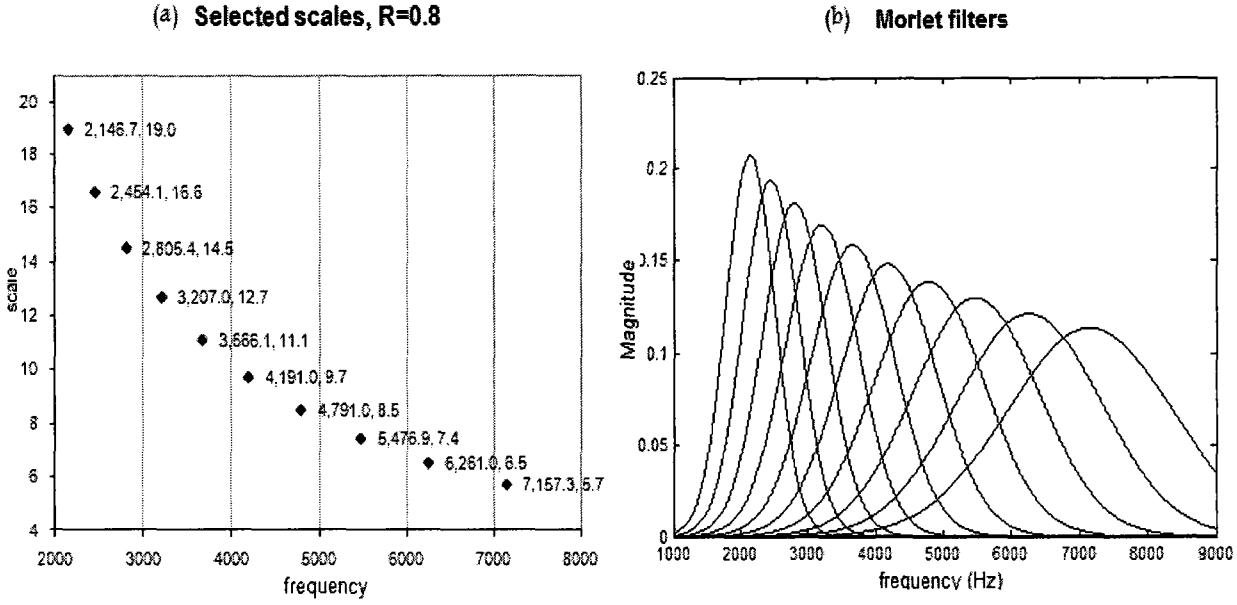


Figure 5.5: (a) Selected scales and corresponding frequencies (Note: in the figure, the first number is frequency and the second is scale), (b) the frequency domain representation of Morlet filters

5.1.3 Choice of central frequency

The central frequency determines the location of the Heisenberg box in the time frequency plane. Looking at the same location in the time frequency plane, a higher value of ω_0 results in a narrower frequency and wider time domain (figure 5.6). Thus Morlet Wavelets with lower ω_0's produce more temporal resolution than spectral resolution and those with higher ω_0's produce better spectral resolution (Addison 2002).

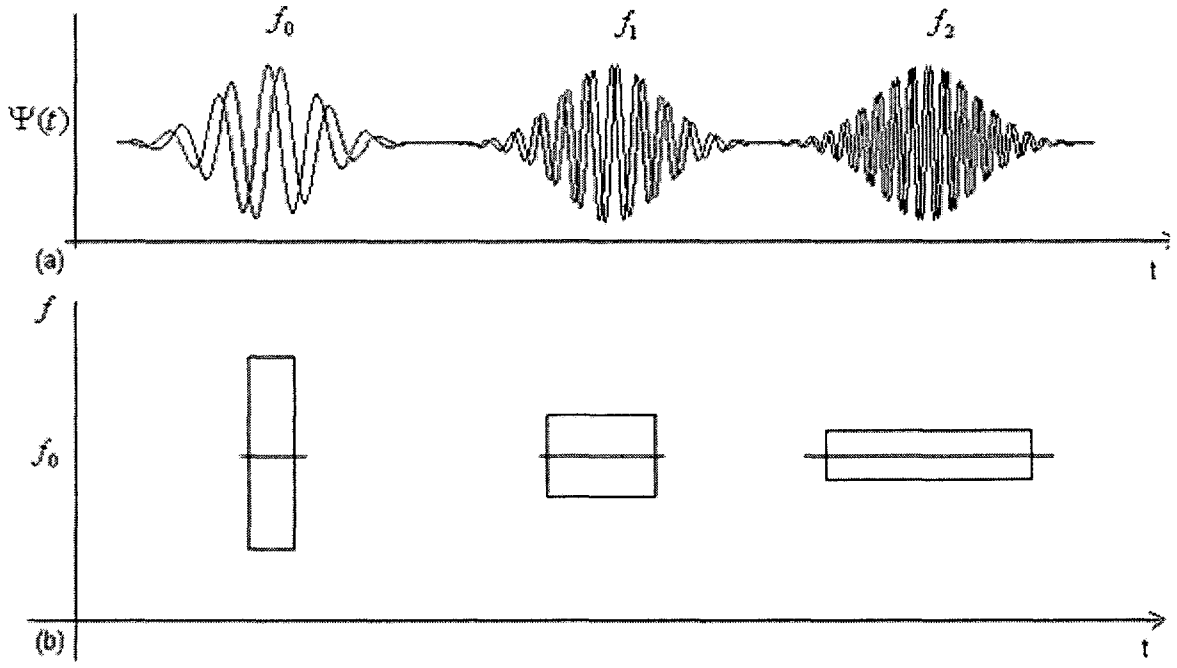


Figure 5.6 (a) Complex Morlet with 3 different central frequencies. (b)Heisenberg boxes in the time frequency plane.

The optimum value for ω_0 depends on the nature of the signal and its application. In practice, $5 < \omega_0 < 6$ is often used which can make the Mortlet Wavelet admissible. For example, the value of ω_0=5.336 is suggested b Lee et al.(1994). The ω_0=6 is recommended by Farye (1992). $\omega_0 = 1.75\pi$ was used to locate abrupt changes in a signal by Staszewski (1994). It was also suggested to use entropy as the rule for finding best ω_0 (Lin et al. 2000).

In section 5.3 an experiment is conducted to measure the effect of the central frequency on the periodicity of Wavelet coefficients. It will be shown that the value of $5 < \omega_0 < 6$ has the least effect on the performance of the system.

5.2 Periodicity detection by ESVR

5.2.1 Singular Value Decomposition (SVD)

Singular Value Decomposition is a powerful factorization technique in linear algebra and its application is extended to signal processing and statistics. This technique was applied on real square matrices in 1870 by Beltrami and Jordan (KLEMA, et al. 1980) and extended for complex matrices by Autonne (KLEMA, et al. 1980), and for rectangular matrices by Eckart and Young (KLEMA, et al. 1980). Since its introduction, SVD has become a standard decomposition and the best factorization technique.

The main idea of SVD is that to decompose a rectangular matrix A into two different orthogonal matrices, U and V, and a diagonal matrix Σ with non-negative elements arranged from high to low as shown in the following:.

$$A = U\Sigma V^T = [u_1 \quad u_2 \quad . \quad . \quad u_r \quad . \quad . \quad u_m] \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & . & & & & \\ & & & . & & & \\ & & & & \sigma_r & & \\ & & & & & 0 & \\ & & & & & & 0 \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ v_r \\ . \\ . \\ v_n \end{bmatrix} \quad (5.13)$$

where $\sigma_1 \geq \sigma_2 \geq \ldots\ldots \geq \sigma_r \ldots \geq 0$

The left singular vectors U are an orthonormal bases in the column space of matrix A and the right singular vectors V are an orthonormal bases in the row space of matrix A. The diagonal matrix (Σ) is the multiplying factor (singular values) of this transformation.

SVD of a complex matrix A is defined by

$$A = U\Sigma V^H \tag{5.14}$$

Where V^H is the Hermitian adjoint of matrix V. If A is a complex matrix, then there always exists such a decomposition with positive singular values (Hopkins 1996).

In numerical linear algebra, the singular values can be used to determine the effective rank of a matrix. If the matrix is not a full rank matrix, the singular values for the null space are zeros. In a rank deficient matrix, low level noise leads to small non-zero singular values. This property of SVD can be used to reveal the periodicity of the transient's signals hidden in the background noise. The number of non-zero singular values of a strictly periodic signal in a noiseless environment is one. In the presence of noise, a periodic matrix becomes full rank; however, for the high signal to noise ratio (SNR), the first singular value is much greater than other values. By monitoring the variations in the periodicities of the Wavelet coefficients, one can detect bearing localized defects.

5.2.2 SVD Matrix

In this work, in order to measure the periodicity of the Wavelet coefficients, the SVD of a $2\times n$ matrix at each scale is computed, see equation 5.15. The number of the rows determines the number of singular values.

$$Z = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} \tag{5.15}$$

$x_{11}, x_{12}, \cdots, x_{1n}, x_{21}, x_{22}, \cdots, x_{2n}$ are the Wavelet coefficients at a specific scale. The matrix must have at least two rows and in this thesis for computational speed the

minimum number of rows is used. As the number of rows suggests, two periods of fault frequency are required to form the matrix, therefore wavelet transform must be performed on the signal with the sample size twice as the fault period. Each row of the matrix needs half of the Wavelet coefficients at each scale. The number of columns (n) is determined by the fault characteristic frequencies and sampling rate as follows:

$n = Fs * Period$

where Fs is the sampling frequency and period is the time duration between two successive impulses of a bearing fault. For example, if the sampling rate is 48000Hz and the BPFO (ball passing frequency outer race) is 200Hz (fault period is 5 ms), the value of n is:

$n = 48000 * 0.005 = 240$

Numerical example:

Wavelet transform is performed on a signal with the sampling rate of 10Hz and fault period of 0.5s. The number of required columns to build the matrix is

$n = Fs * Period = 10 * .5 = 5 \ \ Columns$

Column 2 of the table 5.1 shows the simulated signal with sample size of 10. The signal is transformed to time-frequency domain by complex Morlet Wavelet for two scales. Column 3 and 4 of table 5.1 shows the Wavelet coefficients ('j' is imagery part) for two scales respectively.

n	x(n)	Wavelet coefficients at Scale a^i	Wavelet coefficients at scale a^{i+1}
1	0.025	-0.080 - 0.494j	-0.063 - 0.504j
2	0.443	0.059 - 0.511j	0.103 - 0.510j
3	0.359	0.225 - 0.472j	0.262 - 0.461j
4	0.722	0.382 - 0.385j	0.387 - 0.364j
5	0.902	0.471 - 0.245j	0.486 - 0.227j
6	0.757	0.529- 0.077j	0.541 - 0.090j
7	0.681	0.543 + 0.095j	0.540 + 0.087j
8	0.982	0.472+ 0.244j	0.482 + 0.251j
9	0.569	0.353+ 0.391j	0.379+ 0.373j
10	0.152	0.230 + 0.481j	0.241 + 0.465j

Table 5.1: Signal and Wavelet coefficients at two successive scales

Therefore Matrix Z for scale a^i is defined by :

$$z_{a^i} = \begin{bmatrix} -0.080-0.494j, 0.059-0.511j, 0.225-0.472j, 0.382-0.385j, 0.471-0.245j \\ 0.529-0.077j, 0.543+0.095j, 0.472+0.244j, 0.353+0.391j, 0.230+0.481j \end{bmatrix}$$

And for scale a^{i+1}

$$z_{a^{i+1}} = \begin{bmatrix} -0.063-0.504j, 0.103-0.510j, 0.262-0.461j, 0.387-0.364j, 0.486-0.227j \\ 0.541-0.090j, 0.540+0.087j, 0.482+0.251j, 0.379+0.373j, 0.241+0.465j \end{bmatrix}$$

As the matrix has only two rows, the computational time for the two singular values is reasonably short.

5.2.3 Measure of signal periodicity

Several alternatives have been considered as a measure of signal periodicity. Kanjilal et al (1995 and 1999) used the ratio of the first two biggest singular values $\frac{\sigma_1}{\sigma_2}$. Qiu et al (2006) used $\left(\frac{\sigma_1}{\sigma_2}\right)^2$ whereas Chetverikov et al. (2006) used $1-\frac{\sigma_1}{\sigma_2}$ and $\frac{\sigma_1-\sigma_2}{\sigma_1+\sigma_2}$.

In this work, the following measure is used to detect bearing characteristic frequencies:

$$ESVR = Ln\left(E\left(\frac{\sigma_1}{\sigma_2}\right)^2 \right) \tag{5.16}$$

In this equation E is the energy factor and its value is determined by the Wavelets' energy distribution. The use of E in the above equation is to emphasize the importance of energy level in the detection process. The reason is that a high $\frac{\sigma_1}{\sigma_2}$ does not always indicate the presence of a fault, particularly when this occurs at a low energy scale. Therefore, using $\frac{\sigma_1}{\sigma_2}$ alone as a measure of periodicity may yield in misleading results. The rationale for using E in Equation (5.16) will be further elaborated in the next sub-section. Though $E\left(\frac{\sigma_1}{\sigma_2}\right)^2$ could be directly used as the indicator, its value could be very high. For this reason, its logarithm value is adopted in this work.

Theoretically the value of ESVR can be in the range of $-\infty \le ESVR \le \infty$. The positive infinity occurs when the signal is strictly periodic as the number of non-zero singular values of a strictly periodic signal in a noiseless environment is one, $ESVR = Ln\left(E\left(\frac{\sigma_1}{0}\right)^2 \right) = +\infty$. The negative infinity occurs when the signal has a very low energy level, i.e., $ESVR = Ln\left(E\left(\frac{\sigma_1}{\sigma_2}\right)^2 \right)_{E->0} = -\infty$. As it will be shown in the section on statistical analysis, the values of ESVRs roughly fit the normal distribution.

5.2.4 SVD and Wavelet energy, parameter E

The following section shows the effect of signal energy on the singular values.

The SVD of a full rank m × 2 matrix A is defined by

$$A = USV^H = U\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} V^H \tag{5.17}$$

Therefore the singular value ratio of matrix A is defined by $\frac{\sigma_1}{\sigma_2}$. Now consider the matrix $B = \varepsilon * A$ where ε is an infinity small scalar value. Matrix B can be decomposed by:

$$B = \varepsilon A = U\begin{bmatrix} \varepsilon\sigma_1 & 0 \\ 0 & \varepsilon\sigma_2 \end{bmatrix} V^H \tag{5.18}$$

Then the SVD ratio of Matrix B is defined by.

$$\left(\frac{\varepsilon\sigma_1}{\varepsilon\sigma_2}\right) = \left(\frac{\sigma_1}{\sigma_2}\right) \tag{5.19}$$

As the SVRs show, matrix A and B have the same singular value ratios. This property implies that scales with extremely low energy can produce very large SVRs.

It should be noted that high SVRs at some low energy scales are not always a symptom of a faulty bearing, as the characteristic frequencies of the bearing can be observed at any scale even in a normal condition. For instance, part 'a' of figure 5.7 illustrates a signal obtained from a normal bearing. Part c of the figure 5.7 shows the SVR plot which shows roller defect periodicity (high SVR values at low scales). These high SVRs occur at scales with very low energy level. Therefore, one way to tackle this problem is to incorporate a percentage of the scale's energy into the value of the SVRs, in other words

normalizing SVRs based on the energy of scales. In equation 5.16 this normalization is done by multiplying the energy of the scales, E, by the SVR.

Figures 5.7 (c) and (d) respectively show the SVR and energy distributions for all scales. The energy distribution in both time and scale domain is shown in figure 5.7 (b). As it can be seen from figure 5.7 (c), the maximum SVR has occurred at the scale with low energy. Part 'e' depicts the contribution of energy to the SVRs. The SVRs of low-energy scales are considerably reduced.

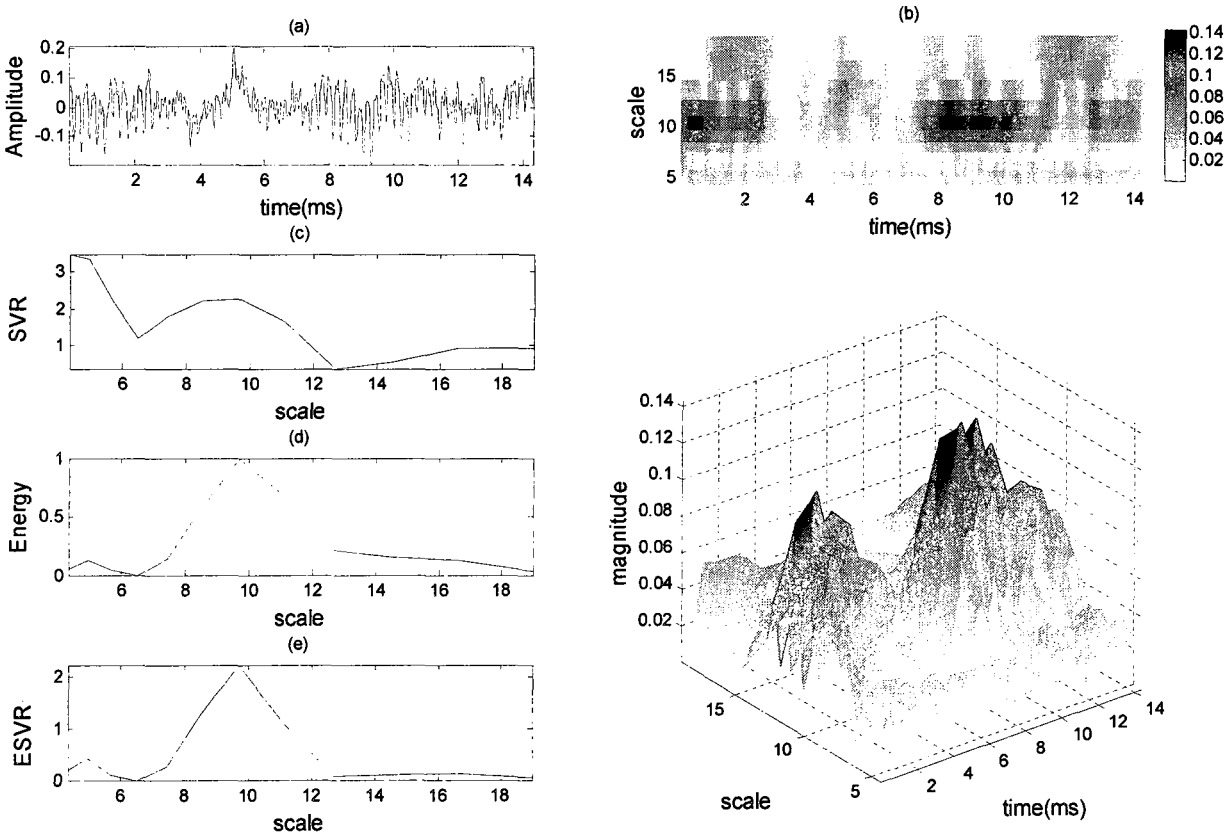


Figure 5.7 (a) Normal bearing signal, (b) Wavelet plot, (c) SVRs based on roller frequency, (d) Energy distribution of scales,(e) ESVRs

5.3 Effect of central frequency on the periodicity detection by measuring SVRs

In this section, an experiment is conducted to see the effect of central frequency on the periodicity of Wavelet coefficients. The central frequencies are selected in the range of $5 < \omega_0 < 6$ with step size of $2\pi*0.02=0.125664$ which result in roughly $8\omega_0$'s. The ω_0 with higher average SVD ratios will be the best possible choice. Computer-simulated data and real bearing vibration data are used to perform initial testing and validation of the results.

In the first experiment, 241 Computer simulated signals are generated with one second duration each. The sampling rate is 48000Hz. Generated signals contain transient impulses similar to impact impulse of a faulty bearing. The transient impulses of all the signals have the same periodicity and duration; however, their resonance frequencies are different. White Gaussian noise is added to all the signals with SNR 10db per sample. Table 5.2 shows the details of the simulated signals. For visual presentation, one of the test signals with resonance frequency of 6000Hz has been shown in figure 5.8.

	Signal duration(Sec)	Transient frequency (Hz)	Transient length(ms)	Periodicity (ms)	White noise	Sampling Rate(Hz)
Test Signal #1	1	2000	2.083	6.25	10db	48000
Test Signal #2	1	2025	2.083	6.25	10db	48000
Test Signal #3	1	2050	2.083	6.25	10db	48000
..	...	...	...	...		...
Test Signal #241	1	8000	2.083	6.25	10db	48000

Table 5.2 Test signals specifications

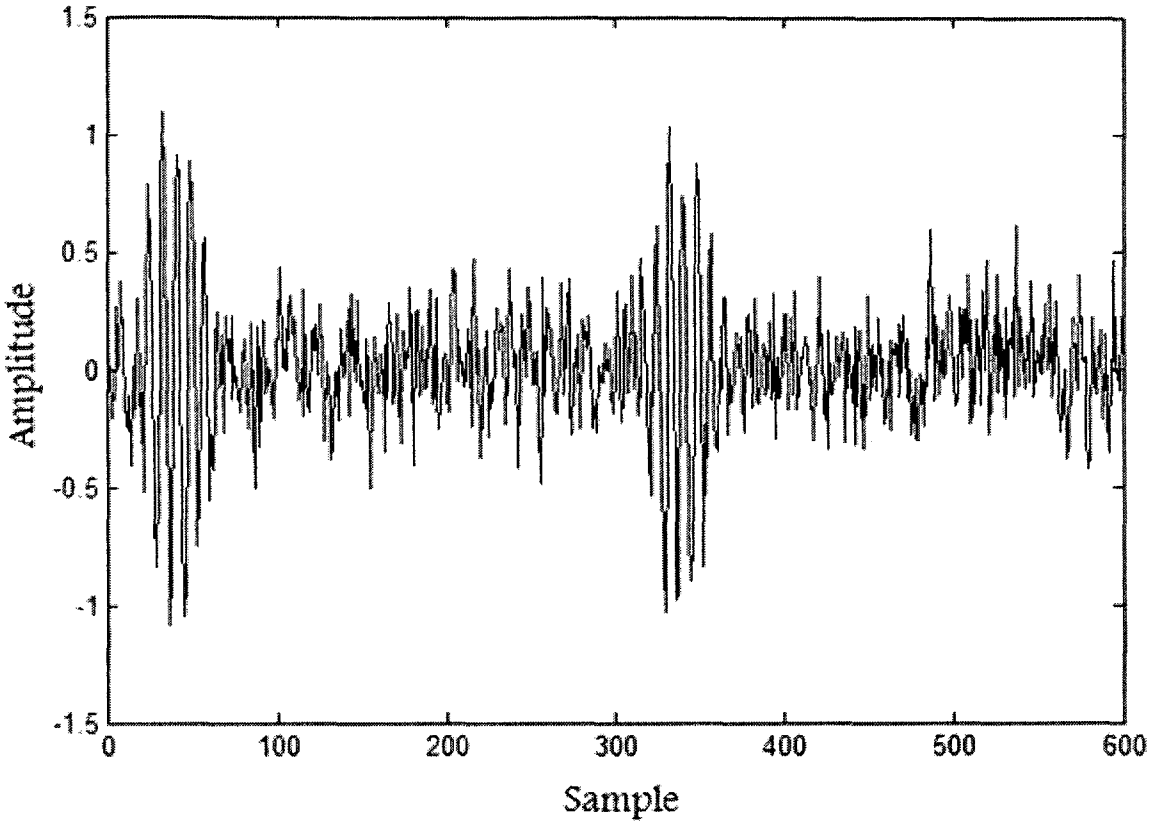


Figure 5.8 Simulated signal with resonance frequency of 6000Hz

Wavelet transforms are performed on the signals with analyzing Morlet Wavelet. The SVD matrix is formed based on the period of 6.25. The number of rows is two and the number of columns is computed as follows:

$n = Fs * Period = 48000 * 6.25/1000 = 300$

Each signal is broken up into 80 segments of 12.5ms (80*.0125=1Sec) which result in 80 matrices. SVRs of all segments are computed and then averaged to minimize the randomness. In real vibration signal the exact value of resonance frequency is unknown. At $\omega_0 = 2\pi * .849 = 5.334$ the maximum distance from peak of one of the Morlet filters to the resonance frequency can be derived from equation 5.12, i.e.:

$$a^{i+1} = \frac{\omega_0 a^i}{\omega_0 - \sqrt{-2\ln(R)}}$$

If R=80% and $\omega_0 = 5.334$, then

$$a^{i+1} = \frac{5.334a^i}{5.334 - \sqrt{-2\ln(.8)}} = 1.143a^i \tag{5.20}$$

If Fs=48000Hz then $n = Fs * Period = 48000 * 6.25/1000 = 300$

Using the property of Morlet Wavelet

$$a = \frac{\omega_0}{\omega} \tag{5.20a}$$

where 'a' is the scale corresponding to the frequency 'f'.

From equation 5.20, the relationship between two successive frequencies is defined by

$$\frac{5.334}{(2 * \pi * f_p^{(i+1)})} = \frac{1.143 * 5.334}{(2 * \pi * f_p^{(i)})}$$

Then

$$f_p^{i+1} = .8748 * f_p^{(i)} \tag{5.21}$$

Therefore the distance between two successive frequencies can be computed from equation 5.21. Therefore any frequency will locate within 6.26% of the peak of one of the Morlet filters as shown below

$$\frac{1}{(1 - .8748)/2} * 100 = 6.26\% \tag{5.22}$$

For example, if the peak frequency of the second Morlet filter is located at $f_p^{(2)} = 1000Hz$ then for R=0.8 and from equation 5.21 we will have $f_p^{(3)} = 874.8Hz$,

therefore any resonance frequency between these two peak frequencies is within 6.26% of the peak of one of these filters, see figure 5.9.

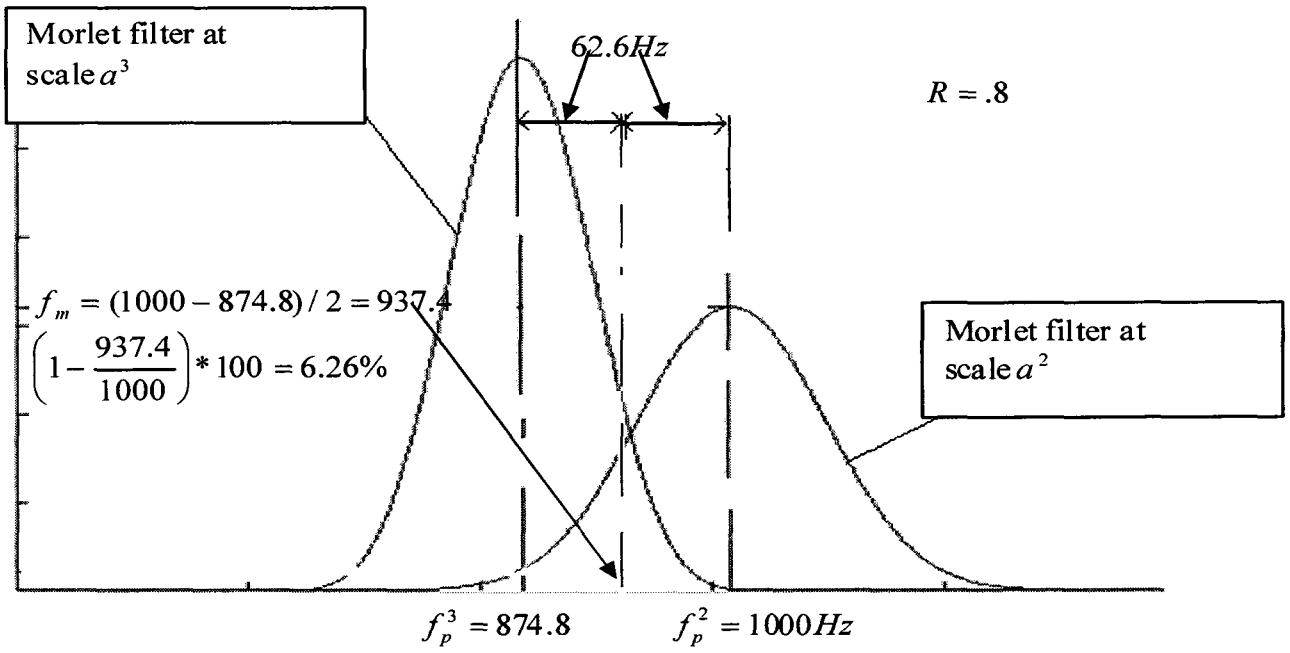


Figure 5.9: Complex Morlet filters at scale indices 2 and 3

As in the real vibration signals the value of the resonance frequency is unknown, in this experiment the maximum deviation error or in other words the maximum distance that the resonance frequency can have from one of the filter peaks, 6.26%, is added to the value of scales. Figure 5.10 illustrates 8 plots each corresponds to one of 8 central frequencies. In practical situation the resonance frequency is unknown and its value depends on the bearing and machine structure. In this experiment ESVRs of frequencies ranging from 2000Hz to 8000hz are computed. The x-axis of the plot is the resonance frequency. The minimum distance between two peak frequencies is about 300Hz, see figure 5.5. Therefore to reduce the computational time the spacing between two successive resonance frequencies is fixed at 25HZ which ensures at least 6 resonance

frequencies are within the narrow band of Morlet filters (lower frequencies, see figure 5.5). This produces 241 points (=(8000-2000)/25+1). The ESVR for each point (Note: Each point is corresponding to one resonance frequency is then computed. The y-axis is the natural logarithm of ESVRs which is computed for each resonance frequency. All the plots show the same pattern and almost the same ESVRs, in other words changing central frequency will not change the pattern of ESVRs. In this case moving from lower to higher resonance frequencies will cause reduction in the ESVRs. The average of ESVR's approximately remains the same too.

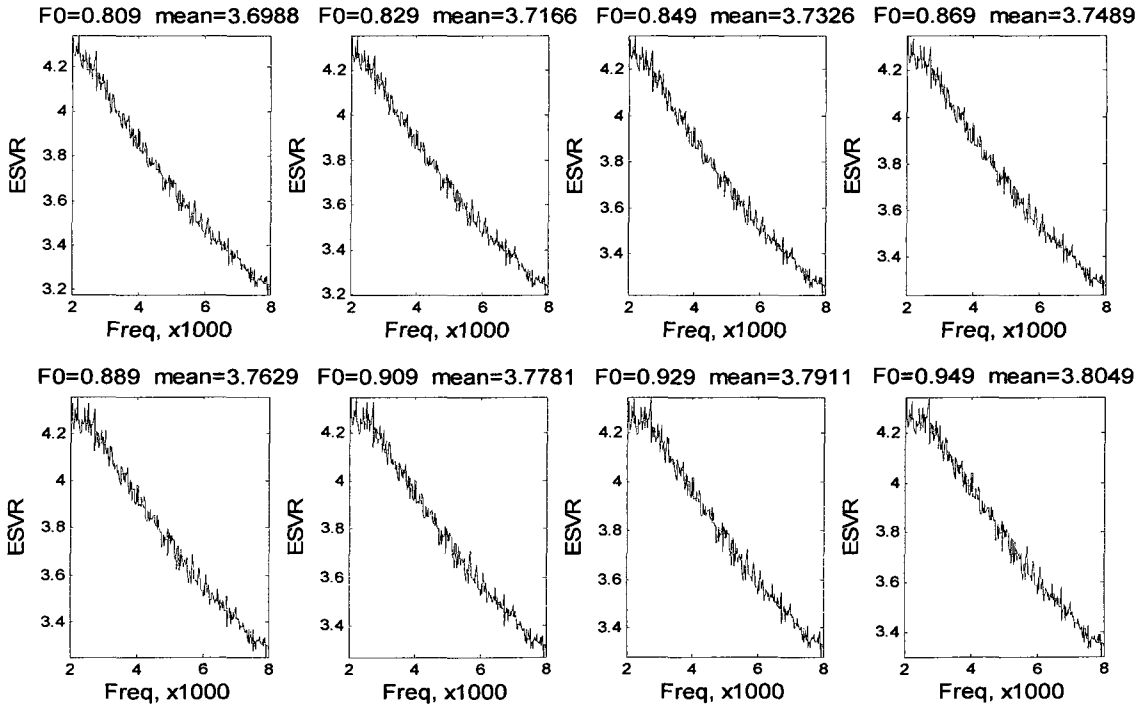


Figure 5.10 effect of w0 at different frequency on the ESVR ratios, simulated signal

The same experiment is performed on the real vibration signal collected from a bearing with the faulty outer race, see figure 5.11. As the plots show, the maximum ESVRs are located around 2600Hz which indicates the resonance frequency of impact impulses.

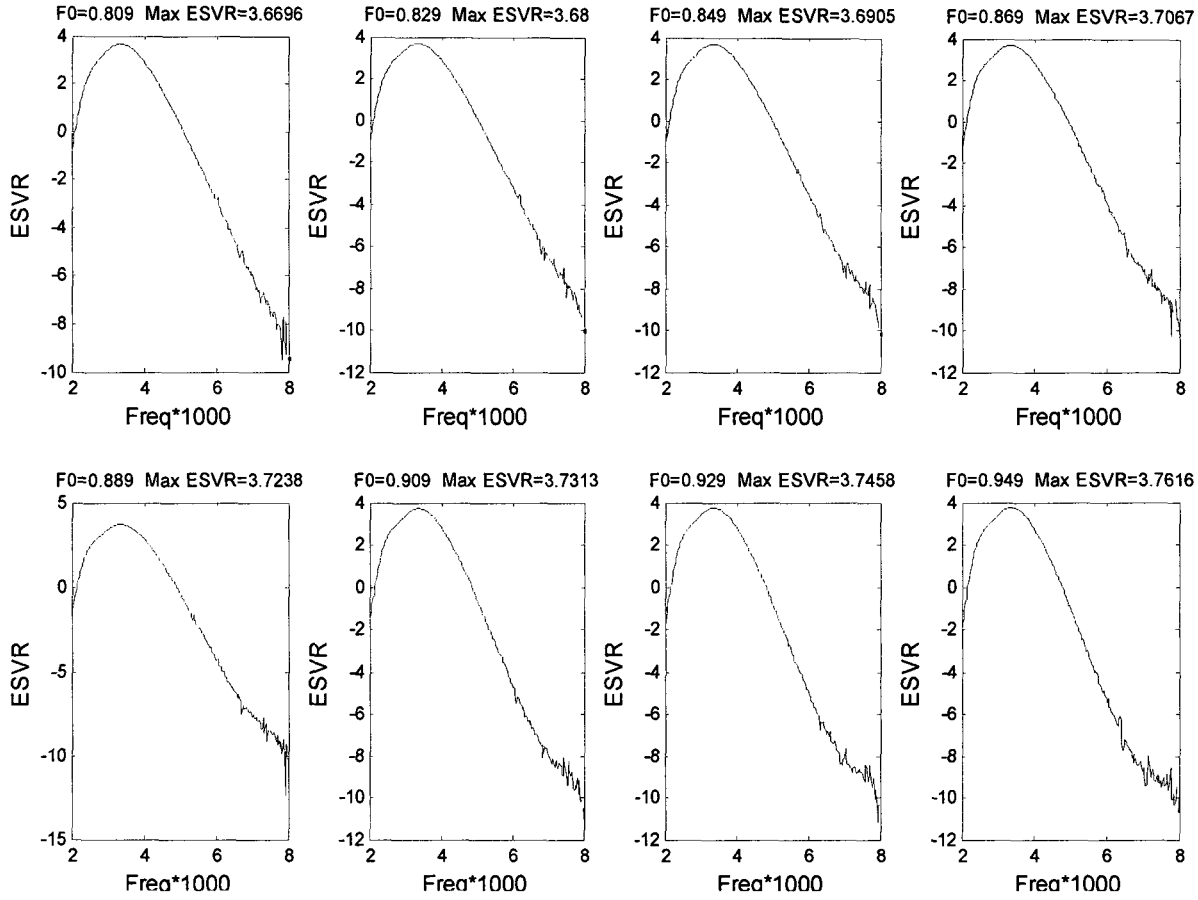


Figure 5.11 effect of ω_0 at different frequency on the ESVR ratios, Faulty outer race signal

Table 5.3 shows the relationship between ω_0 and the number of the Morlet filters for the frequency range of 2000-8000Hz.

	ESVR	Number of Morlet filters (Sampling rate=48000Hz Freq range: 2000-8000Hz)
$f_0 = .809$	3.6635	10
$f_0 = .829$	3.6761	10
$f_0 = .849$	3.6874	10
$f_0 = .869$	3.6994	11
$f_0 = .889$	3.7148	11
$f_0 = .909$	3.7295	11
$f_0 = .929$	3.7436	11
$f_0 = .949$	3.7592	12

Table 5.3 Number of Morlet filters required to cover frequency range of 2000-8000Hz, computed for different central frequencies

Both experiments show that though the higher value of ω_0 can improve the performance of the system, this improvement is so small that hardly worth the computational cost due to the increase in the number of required Morlet filters. In practice the value of $f_0 = 0.849(\sqrt{1/2\ln 2})$ is often used for ω_0 which produces decay where the magnitude of the two peaks in the real waveform adjacent to central peak are half of its amplitude (Addison 2002).

5.4 Experimental results

This section shows our experimental results of the Wavelet-ESVR fault detection approach. It contains three parts, one for the experiment on the simulated signal, one on the real vibration data, and one for statistical analysis of the data.

5.4.1 Simulated signal

In this experiment, two sets of signals are generated, one simulating normal conditions and the other simulating faulty conditions. Both signals contain transients with the same resonance frequencies but different periodicities. White Gaussian noise of the same strength is also added to the two signals. The simulated signals are sampled at 1 kHz and contain frequency components as shown in Table 5.4.

Test signal 1	*Test signal 2*
Resonance frequency=290Hz, Transient period = 0.2s, Length of each impulse =0.08s	Resonance frequency= 290Hz Transient period 0.25s, Length of each impulse =0.08s

Note: (a) Both signals contain white Gaussian noise of the same strength, (b) for the test signal 2, the frequency of transient signals are multiplied by a random number with a mean of 1 and variance of 0.01, and the maximum amplitude of transient signals are multiplied by a random number with a mean of 1 and variance of 0.01. (c) Transient period is the time duration between two successive impulses, length of each impulse is the time span of each single impulse.

Table 5.4 Components of the simulated signal.

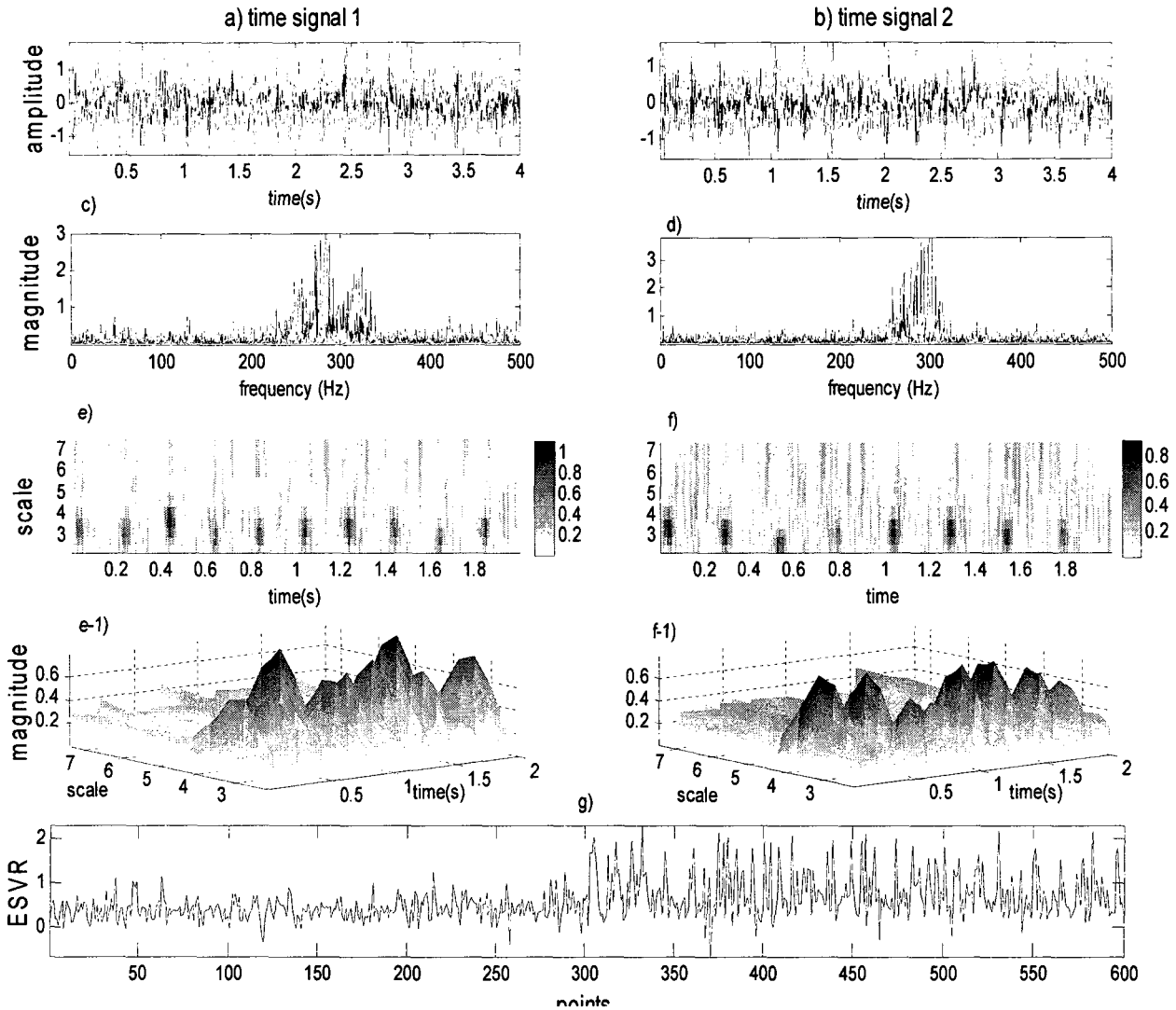


Figure 5.12 (a) Plot of simulated signal 1, periodicity 200ms, (b) Plot of simulated signal 2 , periodicity 250ms, (c)(d) Power spectra of the two signals, (e,e1)(f,f1) Wavelet Transforms of the two signals, (g) ESVDRs graph, points 0-300 computed for signal 1 , points 301-600 for signal 2.

Parts (a) and (b) of figure 5.12 show the plots of simulated signals. The goal of the experiment is to detect transient impulses with a resonance frequency of 290Hz and a period of 250ms. Wavelet transform is performed on the simulated signals with analytical complex Morlet function with central frequency $f_0 = 0.849$ Hz and scale resolution R=0.8. The scales are chosen for the frequency range of 100-400Hz. Parts (c) and (d)

illustrate the power spectrum of the two signals. Power spectrum indicates that the energy of the signal at the resonance frequency has been decreased for the second signal (which contains the 250ms transients). Parts (e)(e1)(f) and (f1) illustrate Wavelet plots of two signals. Part (g) is the ESVRs of the two signals. The first 300 points were computed for the signal containing transients with 200ms periodicity and the last 300 points were computed for the signal containing transient impulses with period of 250ms. As the figure shows the average of ESVR ratios is considerably increased and the detection can be done easily by monitoring a few points. The results clearly demonstrate the detection power of the proposed Wavelet-ESVR method.

5.4.2 Real vibration signals

In this section, actual bearing vibration data generated by CWRU were used to verify the feasibility of applying Wavelet-ESVR to diagnose bearing health. Table 5.5 illustrates the geometry and operating conditions of the bearings that are used in this experiment.

Bearing condition	Load	Fault Diameters (Inches)	Speed, RPM	Fault depth (Inches)	Sampling Frequency (Hz)
Normal	1hp	-	1772	-	48000
Defect on outer race	1hp	.007"	1772	0.01	48000
Defect on outer race	1hp	.014"	1772	0.01	48000
Defect on outer race	1hp	.021"	1772	0.01	48000
Defect on inner race	1hp	.007"	1772	0.01	48000
Defect on inner race	1hp	.014"	1772	0.01	48000
Defect on inner race	1hp	.021"	1772	0.01	48000
Defect on roller	1hp	.007"	1772	0.01	48000
Defect on roller	1hp	.014"	1772	0.01	48000
Defect on roller	1hp	.021"	1772	0.01	48000

Table.5.5 Fault geometry and operating condition if bearings, Taken from CWRS

Parts (a), (b), (c) and (d) of figure 5.13 depict the normal condition signal and faulty signals generated by the outer race defect at different fault sizes. Power spectrums of the signals are shown in parts (e),(f),(g) and (h). For all fault sizes, high energy levels are observed in the frequency range of 2500-3500Hz. Wavelet transform is performed on the vibration signals with analytical complex Morlet function ($f_0 = .849Hz$, R=.8). The scales are chosen within the range of 2000-8000Hz which will cover the high energy frequency range of 2500-3500 Hz. As expected, the Wavelet plot at the fault diameter FD=.007" (figure 5.13 (j)), demonstrates the strongest repeating pattern. As the fault progresses, the repeating pattern of defect frequency diminishes (figures 5.13 (k) and (l)). Therefore, the existence of faults of FD =0.014" and 0.021" cannot be detected by observing the Wavelet plots. For this reason, the ESVR ratios are plotted for the fault conditions of FD=0.007", 0.014" and 0.021" as shown in figures 5.13 (m),(n), and (o), respectively. The X-axis of ESVR plots consists of 400 points. Each point represents a segment of the vibration signal containing 908 samples. The first 200 points belong to the normal bearing and the last 200 points belong to the defective bearing.

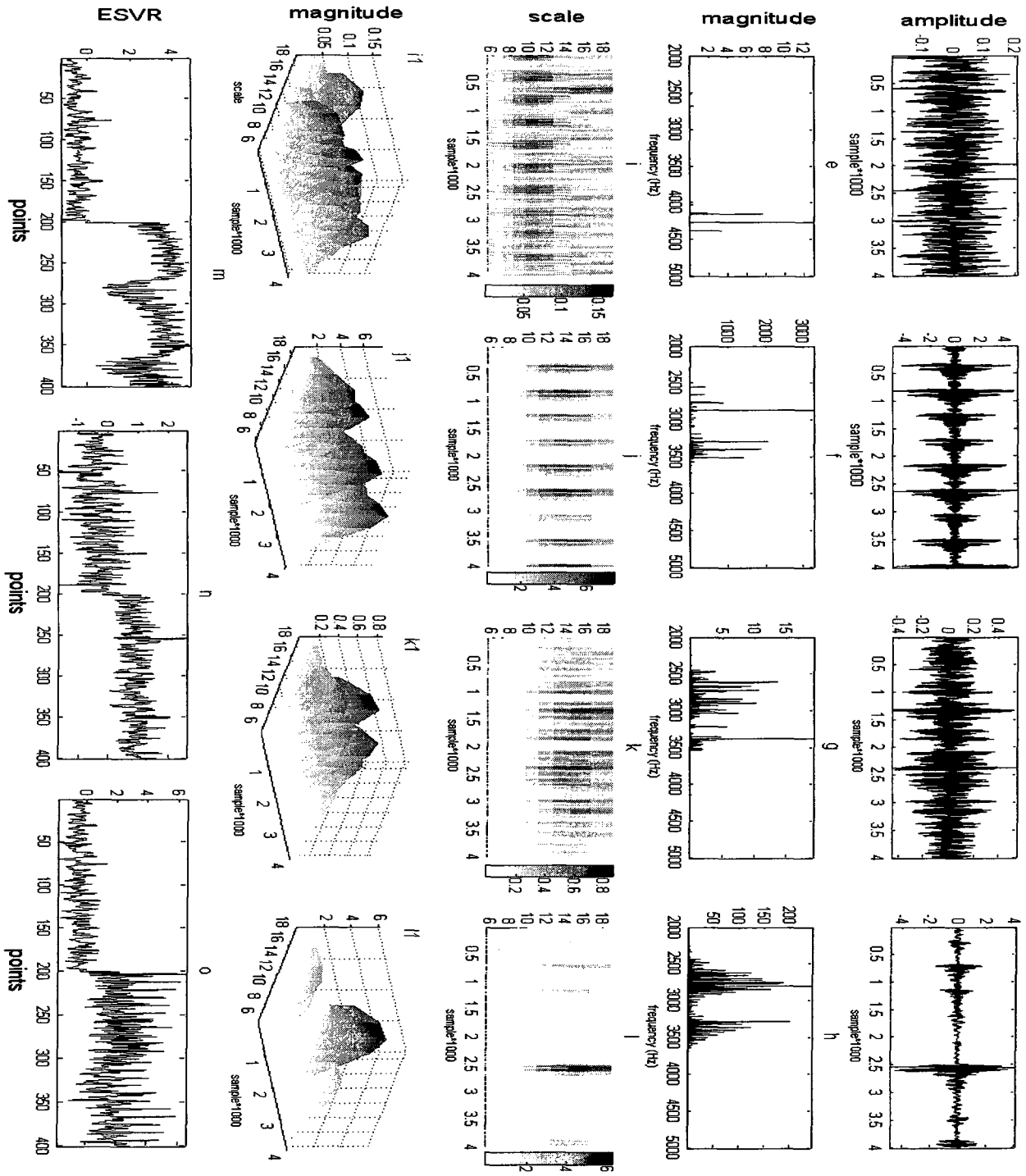


Figure 5.13 Faulty outer race signals with fault diameters of (a) normal, (b).007",(c) 014",(d) .021"; (e)(f)(g)(h) Power spectrum of the signals; (i,i1)(j,j1)(k,k1)(l,l1) Wavelet plots; (m)(n)(o) ESVRs graphs, points 0-200 computed for the normal bearing, points 201-400 computed for the faulty outer race signal. .(m) Mean(1:200)= -0.42684 Mean(201:400)= 3.2926 Scale = 12.7071(3207.0233Hz), (n) Mean(1:200)= -0.3069 Mean(201:400)= 0.92226 Scale = 14.5263(2805.3974Hz),(o) Mean(1:200)= -0.40074 Mean(201:400)= 2.1087 Scale = 16.6059(2454.0684Hz)

In all three plots, i.e., figures 5.13 (m),(n), and (o), an increase in the average of ESVR ratios for the faulty signals can be easily seen. At the FD=.007" the frequency of transient impulses located at the scale 12.42(3280Hz). As the fault develops the impact impulses are moved to the higher scales, 14.87(2740Hz) and 16.27(2504Hz). This may be caused by the flattened fault edge which leads to a lower resonance impact frequency.

The same procedure was applied to the signals obtained from faulty inner race (The results are shown in figure 5.14), and faulty roller (figure 5.15). Observations similar to those for the outer race can be made. This indicates that the Wavelet-ESVRs method can successfully detect bearing conditions in all fault diameters within the scope of this study.

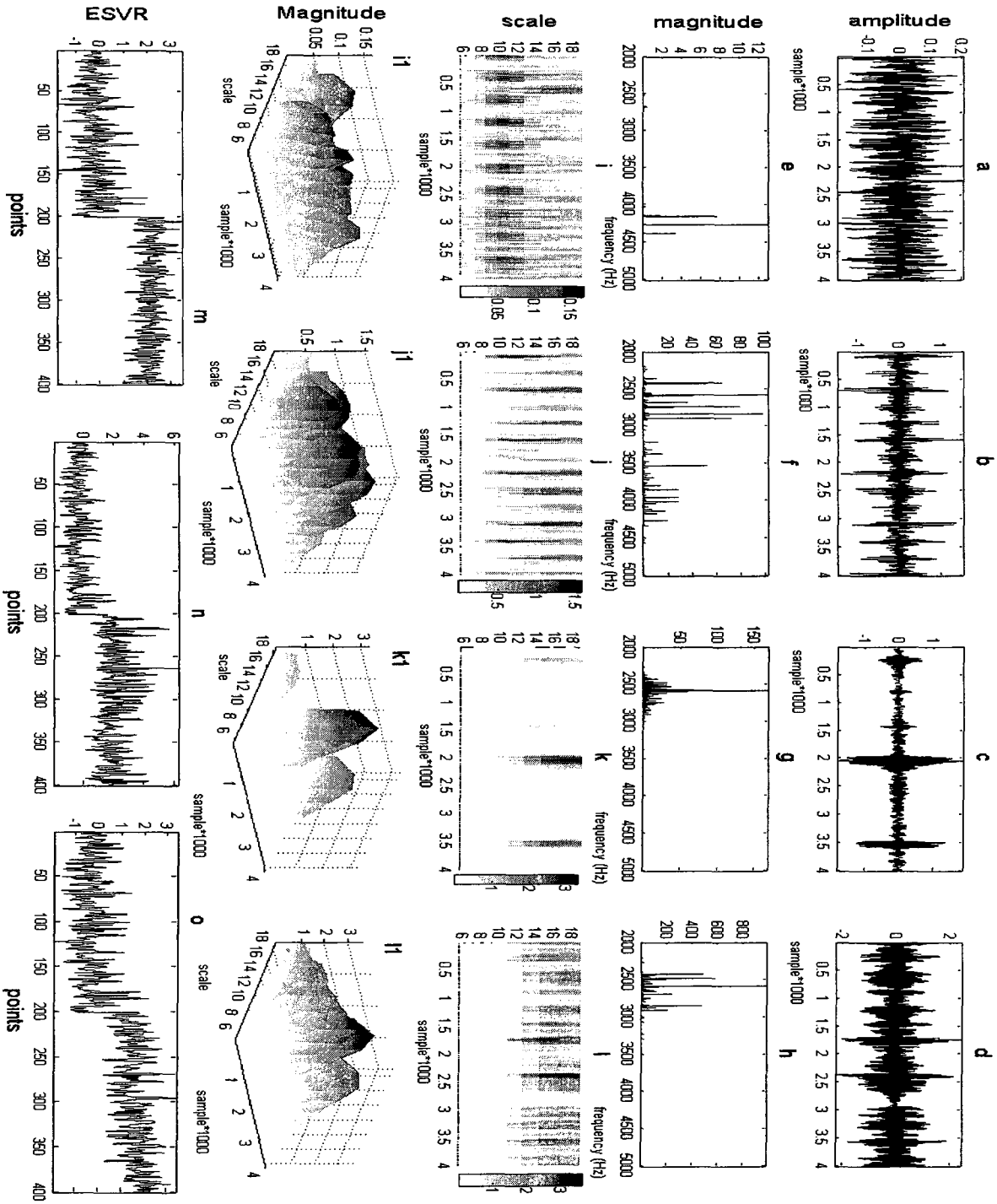


Figure 5.14 Faulty inner race signals with FDs of (a) normal, (b).007",(c) 014",(d) .021". (e)(f)(g)(h) Power spectrum of the signals.(i,i1)(j,j1)(k,k1)(l,l1) Wavelet plots. (m)(n)(o) ESVRs graphs, points 0-200 computed for the normal bearing, points 201-400 computed for the faulty inner race signal,(m) Mean(1:200)= -0.14206 Mean(201:400)= 2.1864 Scale = 14.5263(2805.3974Hz),(n) Mean(1:200)= -0.26495 Mean(201:400)= 2.0496 Scale = 16.6059(2454.0684Hz),(o) Mean(1:200)= -0.26495 Mean(201:400)= 1.6392 Scale = 16.6059(2454.0684Hz)

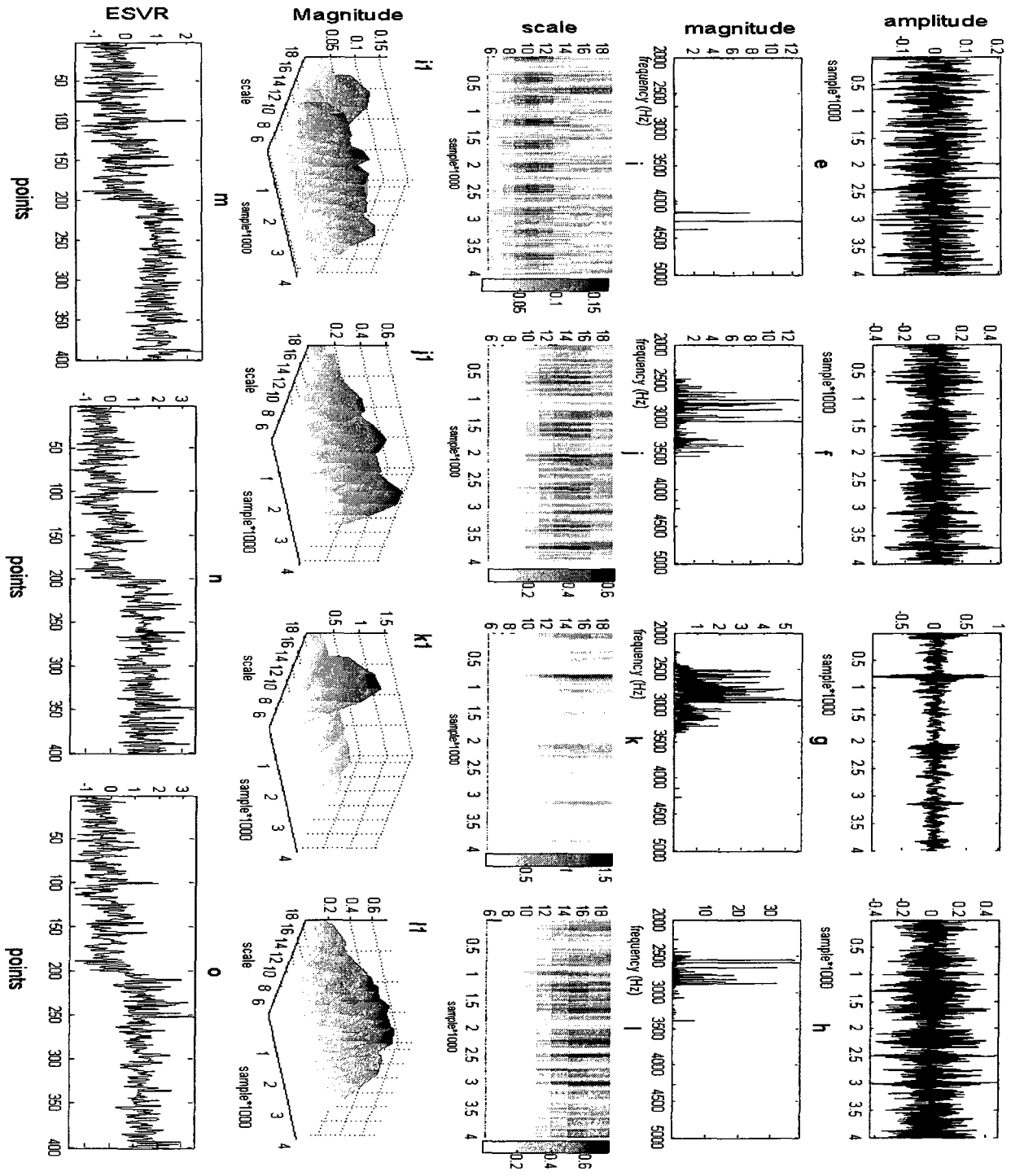


Figure 5.15 Faulty roller signals with FDs of (a) normal, (b).007",(c) 014",(d) .021". (e)(f)(g)(h) Power spectrum of the signals.(i,i1)(j,j1)(k,k1)(l,l1) Wavelet plots. (m)(n)(o) ESVRs graphs, points 0-200 computed for the normal bearing, points 201-400 computed for the faulty roller signal.(m) Mean(1:200)= -0.18518 Mean(201:400)= 1.0369 Scale = 14.5263(2805.3974Hz), (n) Mean(1:200)= -0.18518 Mean(201:400)= 1.4446 Scale = 14.5263(2805.3974Hz), (o) Mean(1:200)= -0.18518 Mean(201:400)= 1.3493 Scale = 14.5263(2805.3974Hz)

Discussion

Looking at the ESVR ratio plots for a given defect fault size, it is notice that the mean amplitude of ESVRs of the defective outer race is higher than the mean amplitude of ESVRs of the defective inner race and roller. In other words, the outer race defect generates stronger impulses. This can be explained by the fact that when the ball hits a defect in the inner race, the vibration should pass through oil film between inner-race and rollers, oil film between rollers and outer race, outer race, and then mechanical interference between outer race and sensors, while signal due to outer race defect passes through only mechanical interference. Thus the energy of the inner race faulty signals is attenuated more than the energy of outer race defect (Taylor et al. 2004). Another reason is that when a fault appears in the non-rotating outer race, all the rollers hit the defect at the same angular position, generating the same transient signals shapes. For an inner race defect or a defective rolling element, the property of transient signal may changes as the angular position of impact is changing [Liu et al. 2008]

5.4.3 Statistical analysis of ESVRs

Statistical analysis can help to summarize a collection of data in a clear and understandable way. Statistical analyses are often based on the probability distribution function (PDF) of data. The distribution does not need to be the best-fitting distribution, but an adequate model will result in valid conclusions. To evaluate the fit of a normal distribution of data, the empirical cumulative density function (CDF) of 200 computed ESVRs of the roller, inner race, and outer race defect frequencies are illustrated in figure

5.16. Minitab is used for drawing the plots. As the plots show, the normal distribution roughly fits the data.

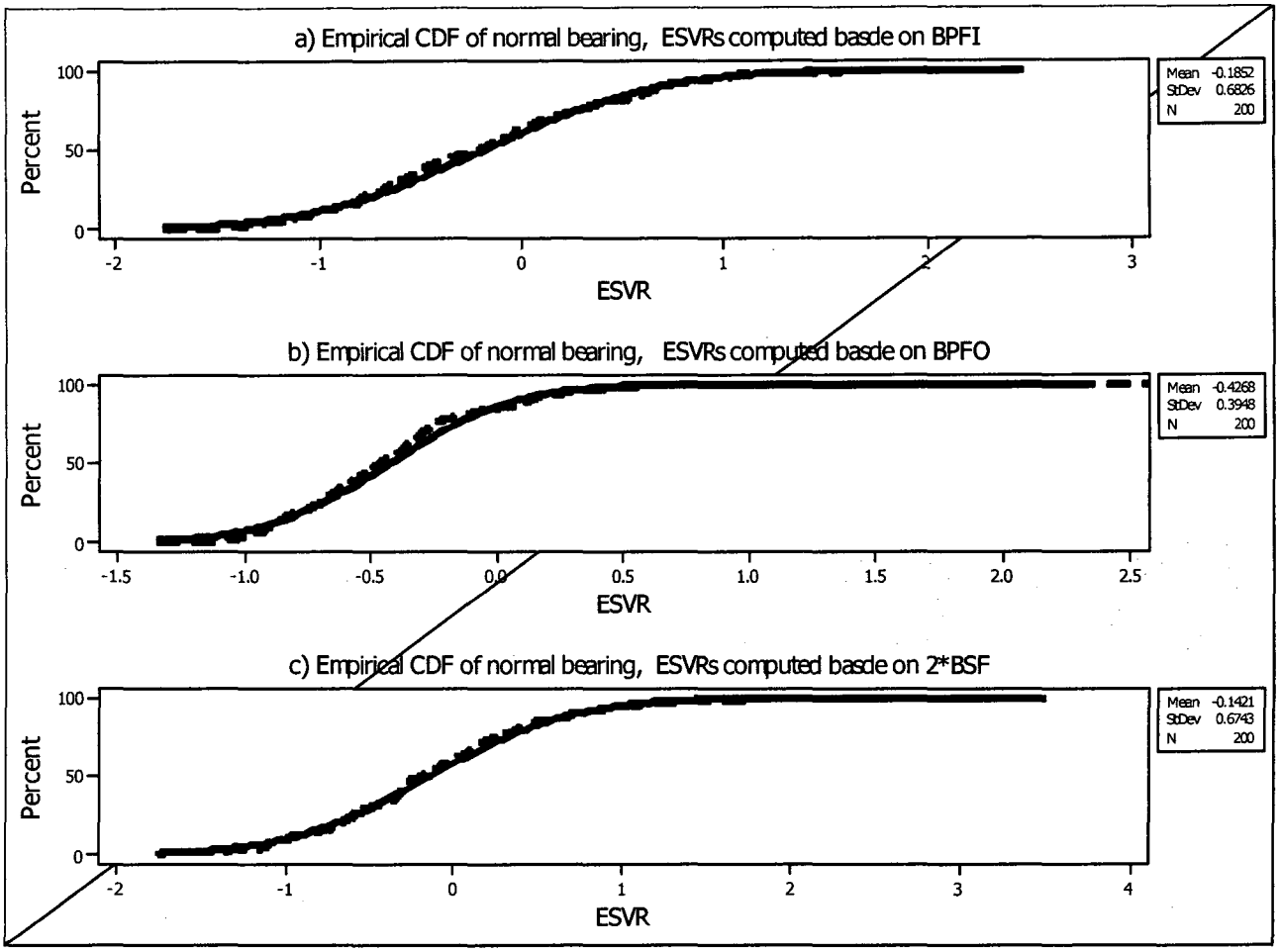


Figure 5.16 Empirical CDF graphs of (a) Normal bearing-BPFI (b) Normal bearing-BPFO (c) Normal bearing-2BSF. (The solid line in parts (a), (b) and (c) is the CDF curve of normal distribution. The dash line is the plot based on the 200 ESVRs.

Control charts (X-Bar charts) of ESVRs are used to analyze and monitor data. Control charts are a graphical representation of certain statistics of a process. These descriptive statistics are displayed in the control chart in comparison to their "in-control" sampling distributions. The comparison detects any unusual variation in the process that may be the result of a defect in the system. If the process is under control, all points will be

within the control limits. Any observations outside the control limits suggest possible faults in the bearing.

For each bearing condition and fault diameter, 200 ESVRs are used to plot X-bar charts. The first 100 ESVR ratios of normal bearing are used to establish mean, upper control limit (UCL), and lower control limit (LCL). The UCL and LCL lines are drawn at a distance of plus and minus three standard deviations from the mean. Figure 5.17 depicts X-bar charts for three fault diameters. If the ESVRs fall within the control limits in a random manner, it can be concluded that the bearing is in the normal condition. The plots show that a bearing in the normal condition generates the ESVR ratios which fall within the UCL and LCL (points 1 to 100). Contrastingly, in faulty conditions all the points have been located above the central line, with a huge majority above the UCL.

In conclusion a normal bearing's ESVRs should fall within the UCLs and LCLs of X-Bar charts.

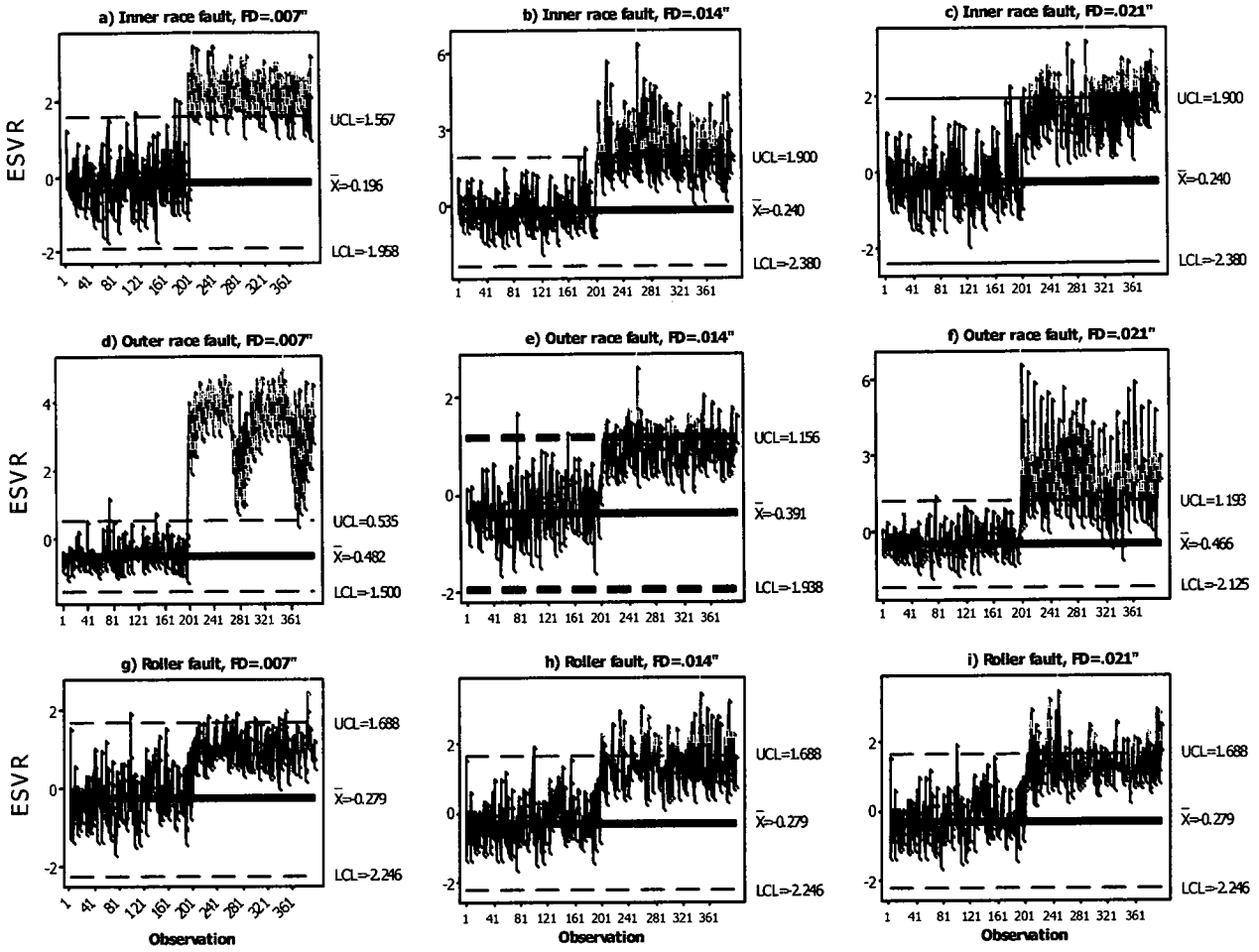


Figure 5.17 X-Bar charts of faulty inner race with FDs (a).007" (b) 014",(c) .021". X-Bar charts of faulty outer race with FDs (d).007" (e) 014",(f) .021" X-Bar charts of faulty roller with FDs (g).007" (h) 014",(i) .021".

Chapter 6. Conclusion and Future Work

In this thesis, two approaches for bearing fault detection and classification are developed. In the first approach, discrete Wavelet transform with three different analyzing bases is used for signal decomposition. Features, energy and kurtosis are extracted from Wavelet coefficients. The extracted features are then used as the input vectors to train a feed-forward neural network with 15 and 20 hidden neurons. ANN is trained by using the back-propagation algorithm. The test results show that the extracted features from the Haar Wavelet transform can classify different kinds of faults with a very high accuracy rate. Using D8 as an analyzing Wavelet was found to have relatively better performance than Haar analyzing Wavelet but the amount of improvement may not justify the computational cost of D8 Wavelets. A properly constructed and trained Wavelet-ANN classifier provides adequate performance in classification for the localized faults in the mechanical vibration signals of the bearing.

The second approach is a Wavelet-ESVR method. This method monitors variation in defect frequencies inside the Wavelet coefficients. Continuous Wavelet transform is used to convert time domain signal into time-frequency domain. Complex Morlet function was used as the analyzing Wavelets. Scales are selected in the range of resonance frequency with the adjustable overlap parameter between adjacent Morlet filter banks. Matrices were formed based on the defect frequencies for the complex Wavelet coefficients. Normalized singular value ratios were used to reveal any changes in the defect periodicity of the complex Wavelet coefficients. X-bar charts were made to find

inconsistencies in the behavior of the system. The results of the experiments show that the combination of Wavelet transform and ESVR can be used to successfully detect variation in impulses hidden in noise and distinguish between different kinds of the faults in the rotary machines.

The present work is done based on vibration data provided by CWRS. The available data may not allow more comprehensive analysis. If additional data are available in future, further studies in the following directions could be worth exploring:

1. With respect to Wavelet-ANN
    a. Generate real vibration signals with various fault diameters and with a different number of local faults. Repeat training and testing of the neural network to examine the performance and versatility of the system.
    b. Investigate the influence of fault depth and diameter on the classifier performance.
    c. Investigate the Wavelet packet transform as a preprocessor for the neural networks.
    d. Investigate using continuous Wavelet as the preprocessor
2. With respect to Wavelet-ESVR
    a. Generate real vibration signals with various fault diameters and with a different number of local faults to verify the accuracy of the system.
    b. Optimize Morlet parameters by using ESVR as the criterion for the measuring performance.

c. Build a new basis which is more similar to impact impulses and hence higher accuracy rate could be achieved. ESVRs may be used as the criterion for measuring similarity.

d. Use ESVRs as inputs of a neural network for automatic fault detection and classification

e. Develop a new approach by transforming the signal to time-frequency domain, de-noising the transformed data, going back to time domain, and then performing SVD on the time signal.

References

Abbasiona, S., _, Rafsanjania,A., Farshidianfarb, A., Irani, N., 2007, Rolling element bearings multi-fault classification based on the Wavelet denoising and support vector machine, Mechanical Systems and Signal Processing, Vol. 21, Issue: 7, 2933-2945

Addison, P. S., 2002, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, Publisher: Taylor & Francis; 1st edition

Al-Raheem, K.F., Roy, A., Ramachandran, K. P., Harrison, D. K., Grainger, S., 2007 Rolling Element Bearing Fault Diagnosis Using Laplace-Wavelet Envelope Power Spectrum, EURASIP Journal on Advances in Signal Processing Volume 2007, Issue 1, 14 pages

Antoni, J., 2006, The spectral kurtosis: a useful tool for characterizing non-stationary signals, Mechanical Systems and Signal Processing , Vol. 20, Issue: 2, 282-307

Antoni, J., Randall, R.B., 2006, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mechanical Systems and Signal Processing , Vol. 20, Issue: 2, 308-331

Case Western Reserve University (CWRS), http://www.eecs.case.edu/laboratory/bearing/

Castleman, K. R., 1996, Digital Image Processing (New Jersey: Prentice-Hall, inc)

Cavacece, M., Introini, A., 2002, Analysis of Damage of Ball Bearings of Aeronautical Transmissions by Auto-Power Spectrum and Cross-Power Spectrum, Journal of Vibration and Acoustics, Vol. 124, Issue 2, 180-185

Chen, D., Wang, W. J., 2002, Classification of Wavelet Map Patterns Using Multi-Layer Neural Networks for Gear Fault Detection, Mechanical Systems and Signal Processing , Vol. 16, Issue 4, 695-704.

Chetverikov, D., Fazekas, S., 2006, On Motion Periodicity of Dynamic Textures, Computer and Automation Research Institute Budapest, Hungary,

http://www.macs.hw.ac.uk/bmvc2006/papers/032.pdf

ChongChun, L., Zhengding Q., 2000, A Method Based on Morlet Wavelet for Extracting Vibration Signal Envelope, Signal Processing Proceedings, Volume 1, 337 - 340

Chow, M.Y., 1998, World Scientific Publishing Company, Methodologies of Using Neural Network and Fuzzy Logic Technologies for Motor Incipient Fault Detection

Devaney, M.J., Eren, L., 2004, Detecting motor bearing faults, Instrumentation & Measurement Magazine, IEEE, Vol. 7, Issue: 4, 30 - 50

Daubechies, I., 1992, Ten Lectures on Wavelets, 1 edition (Philadelphia: Society for Industrial and Applied Mathematics)

Farye, M., 1992, Wavelet transforms and their applications to turbulence, Annual Review of Fluid Mechanics, Vol. 24, 395-457

Hong, H., Liang, M., 2005, De-noising Mechanical Signals by Hybrid Thresholding, International Workshop on Robotic and Sensors Environments Ottawa, ON, Canada, Sept. 30 – Oct. 1, 2005

Hopkins, j., 1996, Matrix Computations (Studies in Mathematical Sciences), 3rd Edition, (Baltimore: The Johns Hopkins University Press)

Gabor, D. (1946), Theory of communication, Proceedings of IEEE, Vol 93(III), 429-457

Jordan, D., Miksad, R. W., 1997, Implementation of the continuous Wavelet transform for digital time series analysis, Review of Scientific Instruments, Vol. 68, No. 3, 1484-1494

Junsheng, C., Dejie, Y., Yu, Y., 2007, Application of an impulse response Wavelet to fault diagnosis of rolling bearings, Mechanical Systems and Signal Processing, Vol. 21, Issue: 2, 920-929

Kanjilal, P. P., Bhattacharya, J., Saha, G., 1999, Robust method for periodicity detection and characterization of irregular cyclical series in terms of embedded periodic components, Vol.59, No.4, 4013-4025

Kanjilal, P. P., Palit, S., 1995, On multiple pattern extraction using singular value decomposition, IEEE Transactions on Signal Processing , Vol. 43, No 6, 1536 - 1540

Karray, F.O., Silva, W.D.S., 2004, Soft Computing and Intelligent Systems design, (New York: Pearson/Addison Wesley)

Kessissoglou, N. J., Peng, Z, 2003, Integrating Vibration and Oil Analysis for Machine Condition Monitoring, School of Engineering, James Cook University, Practicing Oil Analysis Magazine.

Klema, V. C., Laub, A. J., 1980, The Singular Value Decomposition: Its Computation and Some Applications, IEEE Transactions on Automatic Control, Vol. AC-25, No. 2, 164-176

Le, K. N., 2007, High-order symmetrical hyperbolic Wavelets, Journal of Sound and Vibration, Vol. 304, Issue: 1-2, 297-325

Le, K. N., Dabke K. P., Egan, G. K., 2001, The Hyperbolic Wavelet Function, Proceedings of SPIE - The International Society for Optical Engineering, Vol. 4391, 411-422

Lee, D.T.L., Yamamoto, A., 1994, Wavelet Analysis: Theory and Applications, Hewlett-Packard Journal, Vol.45, No.6, 44-54

Li, B., Chow, M. Y., Tipsuwan, Y., Hung, J. C., 2000, Neural-Network-Based Motor Rolling Bearing Fault Diagnosis, IEEE Transactions on Industrial Electronics, Vol. 47, Issue 5, 1060 - 1069

Lin, J., Qu, L., 2000, Feature Extraction Based on Morlet Wavelet and its Application for Mechanical Fault Diagnosis , Journal of Sound and Vibration, Vol. 234, Issue 1, 135-148

Lin, J., Zuo, M. J., 2003, Gearbox Fault Diagnosis Using Adaptive Wavelet Filter, Mechanical Systems and Signal Processing, Vol. 17, Issue 6, 1259-1269

Liu, J., Wang, W., Golnaraghi, F., Liu, k., 2008, Wavelet spectrum analysis for bearing fault diagnostics, Measurement Science and Technology. 19 (2008) 015105 (9pp)

Love, A.E.H., 1944, Treatise on the Mathematical Theory of Elasticity, Dover Publication, Inc., New York , 284-286 , 451-453 (Note: The two sets of page numbers are associated with two the book sections that are referred to)

Lou, X., Loparo, K. A., 2004, Bearing fault diagnosis based on Wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing , Vol. 18, Issue 5, 1077-1095

Martin, R. L., 1970, Detection of ball bearing malfunctions, Instrumentation and Control Systems. Vol. 43, 79-82

Norton, M., Karczub, D.,2003, Fundamentals of Noise and vibration Analysis for Engineers, Second edition (New York : Cambridge University Press)

Paya, B. A., Esat, I. I., Badi, M. N. M., 1997, , Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Preprocessor, Mechanical Systems and Signal Processing , Vol. 11, Issue 5, 751-765

Qiu, H., Lee, J., Lin, J., Yu, G., 2006, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration, Vol. 289, 1066–1090

Rao, B. K. N., 1996, Handbook of Condition Monitoring, (Oxford: Elsevier Science Publishing Corporation)

Sawalhi, N., Randall, R.B., 2005, Spectral kurtosis optimization for rolling element bearing, Proceedings - 8th International Symposium on Signal Processing and its Applications, ISSPA, Vol. 2, 839-842,

Strang, G., Nguyen, T., 1997, Wavelets and Filter Banks, (Wellesley: Wellesley-Cambridge Press)

Staszewski, W. J.; Tomlinson, G. R., 1994, Application of the Wavelet transform to fault detection in a spur gear, Mechanical Systems and Signal Processing, Vol. 8, No. 3, 289-307

Tandon, N., Nakra, B.C., 1992, Vibration and Acoustic Monitoring Techniques for the Detection of Defects in Rolling Element Bearings , The Shock and Vibration Digest, Vol. 24, No. 3, 3-11

Taylor, J. I., Kirkland, D. W., 2004, The Bearing Analysis Handbook: A Practical Guide for Solving Vibration Problems in Bearings, (Vibration Consultants)

Wang, W. J., 1993, Mcfadden, P. D., Early detection of gear failure by vibration analysis, I: calculation of the time-frequency distribution. Mechanical Systems and Signal Processing, Vol. 7, Issue 3, 193-203

Wang, W. J., Mcfadden, P. D., 1995, Application of orthogonal Wavelets to early gear damage detection, Mechanical Systems and Signal Processing , Vol. 9, Issue: 5, 497-507

Wang, W. J., Mcfadden, P. D., 1996, Application of Wavelets to gear box vibration signals for fault detection, Journal of Sound and Vibration, Vol. 192, Issue 5, 927-939

Wang, W. J., 2001, Wavelets for Detecting Mechanical Faults, with High Sensitivity, Mechanical Systems and Signal Processing, Volume: 15, Issue: 4, July, 2001, 685-696

Yang, W. X., Ren, X. M., 2004, Detecting Impulses in Mechanical Signals by Wavelets, EURASIP Journal on Applied Signal Processing , Vol. 8, 1156–1162