

ROTATING MACHINE DIAGNOSIS USING SMART FEATURE SELECTION UNDER NON-STATIONARY OPERATING CONDITIONS

by

Robert George Vinson

Submitted in partial fulfillment of the requirement for the degree

Masters of Engineering (Mechanical)

In the

Faculty of Engineering, Built Environment and information Technology
Department of Mechanical and Aeronautical Engineering
UNIVERSITY OF PRETORIA

2014

ROTATING MACHINE DIAGNOSIS USING SMART FEATURE SELECTION UNDER NON-STATIONARY OPERATING CONDITIONS

by

Robert George Vinson

Supervisors: P. Stephan Heyns; Theo Heyns
Department: Mechanical and Aeronautical Engineering
University: University of Pretoria
Degree: Masters of Engineering (Mechanical)
Keywords: Condition based maintenance, Non-stationary operating conditions, Wavelet transform, Hidden Markov Model, Residual analysis

Summary

Condition based maintenance has received a substantial amount of interest and research in the last decade because of its ability in assisting asset management, which improves machine production, reliability and leads to notable financial and time savings. As both signal processing techniques and measuring technologies have improved, more complex systems have been undertaken resulting in the fault detection of rotating machines operating under non-stationary loads and speeds. Fluctuating load and speed conditions are found in many industrial applications and the standard condition monitoring techniques for stationary systems are inadequate at accurately detecting and diagnosing faults under such conditions. Gears are elements found in most machinery for transferring rotational power, and a common source of machine failure.

Discrepancy analysis is based on the similar concept of residual analysis where all healthy vibration components are removed from the signal, such as meshing frequency and its harmonics, thus only leaving damaged related vibration components. However discrepancy signal does not physically remove the vibration components, instead by statistically comparing a novel signal to a reference model the discrepancy signal is formed by its likelihood with respect to the reference model. Since the reference model is solely based on a healthy system, thus any discrepancy is assumed to be purely as a result of the presence of a fault. Also by allowing the empirical model to take the non-stationary operating conditions into account, the effects of fluctuating load and speed can be removed as well. Thus the discrepancy signal significantly reduces the complexity of fault detection and enables the type and severity of the fault to be determined using standard signal processing techniques.

To classify the operating conditions and detect the presence of a fault it is necessary to create statistical and classification models. The models are trained on data measured on the healthy systems under the full range of expected operating conditions. A novel signal can then piecewise be evaluated

against the reference models in order to detect instantaneous deviations. These deviations are indicative of gear faults. In order to account for fluctuating operating conditions a two-step approach is followed. First a hidden Markov model is used to detect the instantaneous operating conditions of the signal, by looking at the sequence of hidden states. Secondly a unique Gaussian mixture model for each of the identified operating conditions is used to represent the signal behavior under the respective operating condition. It is subsequently possible to detect instantaneous deviations in the signal in a manner that is robust to fluctuating operating conditions. The instantaneous negative log likelihood values of the Gaussian mixture models are used as the discrepancy signal.

For the two-step approach it is necessary to extract two types of features, namely operating condition sensitive features for hidden Markov model and fault sensitive features for the Gaussian mixture model. The features are chosen based on an understanding of the effect of operating condition fluctuation and fault mechanisms. For the hidden Markov model to accurately detect the instantaneous operating conditions, it is necessary to extract operating condition specific features from the vibration signal that capture both load and speed, but which is insensitive to possible faults in the signal. The low frequency region of the spectrogram, and in particular the meshing frequency and its harmonics, encapsulates both the speed and loads of the system. In the same manner it is necessary to extract signal features that are sensitive to faults, while being more robust to the fluctuating operating conditions, for the Gaussian mixture model to accurately detect the presence of a fault. The wavelet transform is ideal for identifying the presence of fault impulses because of the similarity between the shape of wavelets and exponentially decaying fault impulses. Also the frequencies at which the fault impulses are most evident is the natural frequencies and can be determined from either the experimental frequency response function or analysis of the high frequency region of the spectrogram.

The results of this dissertation prove that there is a cost and computationally effective method for detecting, diagnosing and determining the severity of faults in gearboxes under non-stationary operating conditions that is independent of historical fault data and not reliant on measures of either operating speed or load.

Acknowledgements

I would like to thank the following people and institutions that have helped me to complete this thesis:

- Professor Stephan Heyns my supervisor for his support and encouragement in this project
- Theo Heyns my co-supervisor for his guidance and constant assistance
- George Breitenbach and Herman Booysen for their assistance during the experimental tests
- My parents for their continuous love and support during my student career.

Table of Contents

1.	Introduction	1
1.1.	Background	1
1.2.	Literature	2
1.2.1.	Fault Mechanisms of Gears and Effects of Time-Varying Operating Conditions.....	3
1.2.2.	Tracking Shaft Rotational Displacement.....	6
1.2.3.	Data Processing and Feature Extraction	7
1.2.4.	Fault Detection & Classification	15
1.3.	Scope/Research Objectives.....	19
1.4.	Report Layout.....	21
2.	Discrepancy Analysis Using Smart Features	22
2.1.	Introduction	22
2.2.	Aims of the Methodology	23
2.3.	Identify the Instantaneous Operating Conditions	23
2.3.1.	Rough Windowing of Time-Domain Signal	24
2.3.2.	“Smart” Operating Condition Feature Extraction	25
2.3.3.	Classification of Instantaneous Operating Conditions using HMM	27
2.4.	Calculate Instantaneous Shaft Displacement	29
2.4.1.	IF Estimation – Maxima Tracking	30
2.4.2.	Vold-Kalman Filter and Phase Estimation.....	31
2.5.	Fault Detection.....	32
2.5.1.	“Smart” Fault Feature Extraction.....	32
2.5.2.	Multi-Resolution Representation.....	33
2.5.3.	COT and Windowing	35
2.5.4.	Extraction of Fault Sensitive Features for GMM.....	36
2.5.5.	Generating the Discrepancy Signal using the Fault Sensitive Features and the Operating Condition Specific GMM	36
2.6.	Account for Shaft Displacement Calculation Error	37
2.7.	Discrepancy Analysis.....	38
2.7.1.	Synchronous Averaging.....	38
2.7.2.	Spectral Analysis	38
2.7.3.	Cepstral Analysis	39
3.	Dynamic response of rotating machinery under non-stationary operating conditions	40

3.1.	Introduction	40
3.2.	Dynamic Gearbox Model	40
3.2.1.	Lumped Mass Model.....	40
3.2.2.	Time Domain Results	45
3.3.	Experiment Set-Up.....	46
3.3.1.	Instrumentation	47
3.3.2.	Data Acquisition	48
3.3.3.	Seeded Faults	48
3.3.4.	Load Cases.....	50
3.3.5.	Time Domain Results	51
3.3.6.	Frequency Response Function	52
4.	Results and Discussion	54
4.1.	Simulation Results.....	54
4.1.1.	Operating Condition Features and Classification.....	57
4.1.2.	Order Tracking Accuracy	61
4.1.3.	Feature Selection for Fault Detection.....	61
4.1.4.	White Noise in Signal	65
4.1.5.	Fault Magnitude.....	65
4.2.	Experimental Results	66
4.2.1.	Operating Condition Features and Classification.....	70
4.2.2.	Order Tracking Accuracy	74
4.2.3.	Feature Selection for Fault Detection.....	75
4.2.4.	Accounting for Shaft Displacement Calculation Error	80
4.2.5.	Fault Magnitude.....	81
4.2.6.	Spectra of Discrepancy.....	83
4.2.7.	Cepstra of Discrepancy	84
4.2.8.	Alternative Fault Types	85
4.3.	Conclusion.....	87
5.	Conclusion.....	88
	Bibliography	90
	Appendix A: Experimental Setup	95
A.1.	External Conditions.....	95
A.2.	List of Measurements	95
A.3.	File Name Convention.....	95

A.4.	Safety Precautions	95
A.5.	Additional Time Domain Results.....	95
A.5.1.	Shaft speed measurement.....	95
A.5.2.	Strain gauge measurement.....	96
A.5.3.	Alternator power measurement.....	96
A.5.4.	Load Measurement.....	97

Nomenclature

Symbols

a	State Transition Matrix
b	Observation Probability Distribution
C	Damping matrix
c	Contact ratio
f_a	Approximated frequency of wavelet
f_c	Wavelet center frequency
f_g, f_p	Gear/Pinion rotational frequency
f_m	Gear mesh frequency
f_r	Rotational frequency
f_s	Sampling frequency
I	Mass moment of inertia
K	Stiffness matrix
K_g	Gear meshing stiffness
K_x, K_y	Linear bearing stiffness
K_θ	Torsional Stiffness
M	Number of discrete HMM observations
m	Mass matrix
N	Number of hidden states in HMM
N_g, N_p	Number of gear/pinion teeth
n_r	Shaft rotational speed
O	HMM observation sequence
q	Estimated sequence of HMM hidden states w.r.t observation sequence
S	HMM hidden states
T_L, T_M	Torque of load and machine
T_m	Gear mesh period
T_r	Gear rotation period
t	Time
r_b	Base radius
V	HMM discrete observations
w	Weights
$x(t)$	Input time signal
Σ	Covariance matrix
α	Scale/dilation
β	Translation
δ	Displacement of line of action of gears
θ	Angular displacement
λ	HMM probability set
μ	Mean
π	Initial State Probability
ψ	Mother /Family wavelet

Abbreviations

ACFI	Auto-Correlation Function Indicator
ANC	Adaptive Noise Cancellation
ANN	Artificial Neural Network
CBM	Condition Based Maintenance

CHMM	Continuous Hidden Markov Model
COT	Computed Order Tracking
CWT	Continuous Wavelet Transform
DHMM	Discrete Hidden Markov Model
DOF	Degree of Freedom
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
FFT	Fast Fourier Transform
FGP	Fault Growth Parameter
FRF	Frequency Response Function
GMM	Gaussian Mixture Model
HFRT	High Frequency Resonance Technique
HMM	Hidden Markov Model
IF	Instantaneous Frequency
IFT	Inverse Fourier Transform
IMF	Intrinsic Mode Functions
LDN	Load Demodulation Normalisation
MRA	Multi-Resolution Analysis
MTI	Marginal Time Integration
NLL	Negative Log-Likelihood
ODE	Ordinary Differential Equation
OT	Order Tracking
PC	Principle Component
PCA	Principle Component Analysis
RES	Residual Signal
SA	Synchronous Averaging
SANC	Self-Adaptive Noise Cancellation
SNR	Signal to Noise Ratio
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
VKF	Vold-Kalman Filter
WPT	Wavelet Packet Transform
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

1. Introduction

1.1. Background

Condition Based Maintenance (CBM) is a maintenance strategy where maintenance decisions are based on the condition of the machine. It has been proved to be a generally superior maintenance strategy relative to older and more established maintenance strategies such as time based maintenance or operation till failure. With the demand for increased product reliability to decrease maintenance down time and reduce secondary damage, there has been a surge in research interest into CBM over the last two decades as recorded by Aherwar and Khalid (Aherwar & Khalid 2012). Research in CBM has also been stimulated by advances in other fields such as digital signal processing techniques, computational capacity and data acquisition devices. Due to the large advances in the field of CBM more complex problems and systems are being tackled. One problem commonly found in industry is machinery operating under non-stationary operating conditions. There is also a continual drive to implement more cost effective methods, which aims for reduced monitoring hardware, no historical fault data and simplified fault diagnosis.

CBM systems generally comprise of three stages: data acquisition, data processing and maintenance decision support. Data acquisition encompasses the collection, storage and transferring of measured data from the machine. The data can be in a variety of forms such as a vibration signal, acoustic emission, thermal signature and oil debris, etc. The data acquisition method is generally chosen with respect to the machine type and anticipated fault. However, possibly the most common form is the vibration signal because of its ability to operate on a wide variety of machine types, detect a wide range of fault types and intuitive nature to faults mechanisms.

The next stage of CBM is processing the data, which entails methods that improve the quality of the signal and extracting diagnostic information from the signal. The processing stage has a significant influence on the effectiveness of the CBM strategy, because of the dependence of machine learning techniques on fault related features (T. Heyns 2012). The aim of the signal processing techniques is to extract diagnostic information that is robust to operating conditions and environment, compact in size for reduced computational cost and intuitive for ease of fault detection and classification.

The final stage of CBM is to support maintenance decisions based on the interpretation of the extracted information, which results in detecting and classifying faults and aiding maintenance schedules and decisions. The aim of this stage is to cost effectively and accurately interpret the information on a continual real time basis. This stage can be implemented by trained and experience personnel, a rule based system, change detection alarms, automated classifiers, etc. Automated classifiers implementing supervised machine learning algorithm such as Artificial Neural Networks (ANN), Hidden Markov Models (HMM), fuzzy logic systems, etc. have gained popularity and success over the last decade in the field of CBM. However, one of their significant drawbacks is their dependence on training data, which can be expensive or impossible to attain in many industrial applications.

The foundation of CBM is based on a proper understanding of faults found in rotating machines and how they manifest themselves in the measured data. Thus to be able to detect faults it is critical to have a firm grasp of the fault mechanism expected. The expected faults found in rotating machines

range from gears faults such as broken teeth to bearing faults such as inner or outer race faults. Knowledge of faults mechanisms allows the efficient detection and classification of faults; thus, in many cases, reducing the need for historical fault data.

CBM has been implemented in countless industrial application with success. One such example is found at the Sasol power station in Secunda, where the motors, gearboxes and pulleys that transport the coal from the mines to the power station are all monitored at regular intervals and maintenance decisions are based on the vibration spectra of the measured signals. However, there is a need for CBM techniques to be able to operate on more complex machinery and under non-stationary operating conditions, such as the cutting head transmission of a continuous coal miner and drag line on opencast coal mines (C.J. Stander 2005). Other applications are wind turbine and helicopter gearboxes.

This dissertation addresses the challenge of monitoring the condition of a rotating machine under fluctuating operating conditions, especially in the scenario where fault representative data is limited and only vibration measurement possible.

1.2.Literature

This literature study is conducted in four sections:

The first section gives a background of the characteristics of vibration signals. This section explains how the fluctuating operating conditions in rotating machines (namely fluctuating speeds and loads) are manifested in the vibration signal in three ways. These are (1) amplitude modulation, (2) frequency modulation, (3) phase modulation. This section also discusses common faults found in gearboxes and their mechanisms. Understanding the vibration mechanism of rotating machines enables the user to choose 'smart' features that are relevant and sensitive to faults. In particular the interplay between the fluctuating operating conditions and the fault mechanisms are considered. This intuitive understanding of how the signal responds to different operating conditions and fault mechanisms makes it possible to specify rules which can be used to select features with desirable characteristics.

The second section reviews two methods for calculating the instantaneous speed of a gearbox by only analysing the vibration signal. This removes the need for a tachometer to track the shaft position, therefore significantly reducing the instrumentation requirements of CBM.

The third section looks at a few data processing methods for rotating machines under fluctuating operating conditions. These methods are implemented to extract information or features from a signal that is marred by noise and the effects of the fluctuating operating conditions. Each of the methods discussed have their own advantages and disadvantages, as well as their own bias towards a certain fault type or operating condition. Many of these methods are commonly found in the field of digital signal processing (DSP).

The fourth section is concerned with methods to detect, diagnose and quantify faults from the extracted features. The simple manual classification as well as a number of machine learning techniques are discussed. The concept of the discrepancy signal is also discussed as it is reliant on reference models that are representative of a healthy machine for all possible operating conditions.

1.2.1. Fault Mechanisms of Gears and Effects of Time-Varying Operating Conditions

The majority of machine condition monitoring research is focussed on gears and rolling element bearings as they are found in almost all rotating machines. They operate under dynamic cyclic contact forces which lead to high local stresses and fatigue. This either eventually results in uniform wear or in localized defects causing premature failure. There are many other faults found in rotating machines such as misalignment, lubrication contamination, etc. each with their own fault mechanisms. However, for simplicity sake, they are not analysed in this dissertation.

Non-defective gears each have their own characteristic vibration signatures determined by their design parameters and operating conditions. These vibration signatures can form the basis of machine condition monitoring, as any discrepancy from the characteristic signature may be indicative of a machine fault. Vibration signatures and fault frequencies of gears can be estimated from a basic understanding of the component and will be discussed below.

The basic assumption of CBM is that any change in the vibration signal of a machine is due to a fault. However, this is not a valid assumption under non-stationary operating conditions. Therefore it is necessary to understand the effects of time-varying operating conditions and take these into account when detecting and identifying faults. In general, fluctuating loads will tend to cause amplitude modulation and fluctuating speeds will tend to cause frequency modulation of the vibration signal.

1.2.1.1. Gears

The purpose of gears in rotating machines is to transmit and transform rotary motion from one shaft to another. This is accomplished by teeth on the input gear (pinion) meshing with the teeth on the output gear. The gear meshing period (T_m) is the interval of time that each tooth is the predominant meshed tooth. The gear meshing period of a gear with N_g teeth and rotational period of T_r can be calculated as follows.

$$T_m = T_r / N_g \quad 1.1.$$

The characteristic vibration signature of gears is generated by the cyclically varying gear meshing forces. Thus the vibration spectra of a healthy gear is characterised by the rotational frequency of the pinion (f_p) and gear (f_g), and the meshing frequency (f_m) of the pair as well as their harmonics and sidebands.

$$f_m = 1/T_m = N_g / T_r \quad 1.2.$$

In general the meshing frequency is the best indicator of any form of gear fault. However, under non-stationary operating conditions it can be difficult to identify the meshing frequency because of spectral smearing. Therefore another fault indicator has been identified which is commonly used to detect ball bearing faults but which can also detect gear faults. This is the presence of high frequency impulses. These faults impulses are caused each time a defect on a gear or bearing makes contact under load with another surface. These exponentially decaying impulses occur with a very short duration relative to the intervals at which they occur, thus the energy is distributed over a large frequency range. Fortunately these impulses excite the natural frequencies of the systems. Methods are detecting these impulses are discussed in Section 1.2.3.9.

The basic faults that occur in gears are listed below (Shiple 1967):

- **Wear:** is the progressive removal of layers of material from the contact surface of the tooth. This leads to thinning of the tooth, which ultimately reduces its strength and stiffness. Experimentally it is known that wear increases the amplitude of the sidebands of the gear meshing frequency with spacing equal to the shaft rotational frequency in the vibration spectrum. Also the wearing action excites the natural frequency of the gear. Gear wear does cause impulses from the uneven meshing surface of the gear tooth in contact.

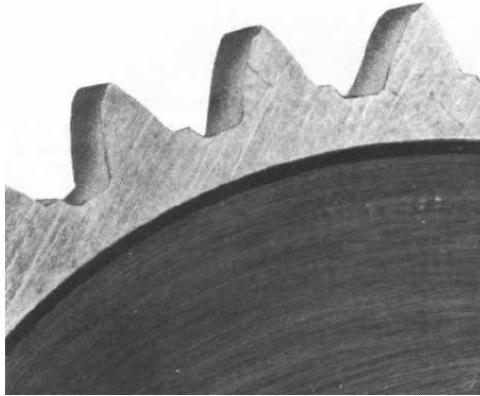


Figure 1.1: Excessive uniform wear from gear tooth surface(Shiplely 1967)

- **Surface Fatigue/Pitting:** is due to surface fatigue failure of the contact surface, which results in material removal from the surface forming pits. The mechanism of pitting does differ from that of wear; however, they both manifest themselves in a similar manner. Since they both result in the removal of material from the meshing surface, the effect of pitting manifests in the vibration signal in a similar manner to that of wear.

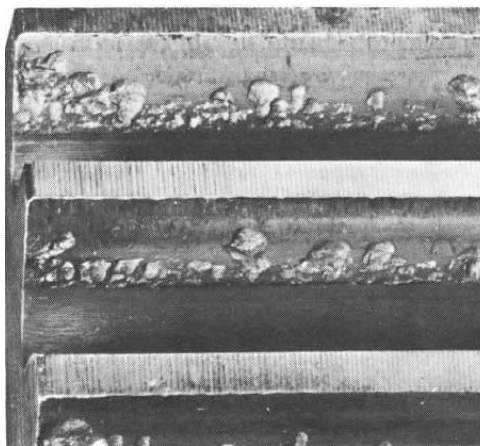


Figure 1.2: Destructive pitting in the dedendum region of the gear tooth(Shiplely 1967)

- **Tooth Crack/Breakage:** is when a substantial part of a tooth has either cracked or broken off from the gear due to overloading or fatigue. This causes an impulsive force once per gear revolution, thus it is characterised in the vibration spectrum by high amplitude at the gear rotational frequency and its harmonics. It is also known to excite the natural frequency of the gear. However, according to the author's experience, the simplest technique to detect a broken tooth is to identify the impulse in the time domain.

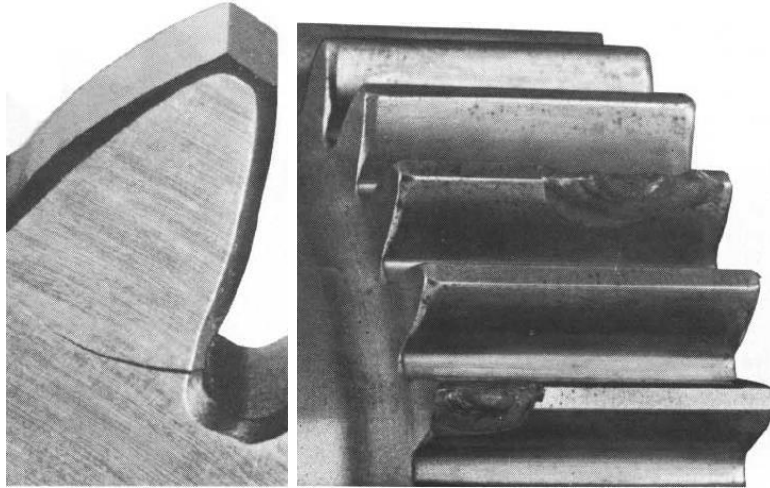


Figure 1.3: Tooth root crack just before complete tooth breakage and chipped tooth(Shiplely 1967)

- **Gear Misalignment:** occurs when there is an axial shift of the gear meshing surfaces within a gear meshing pair. This hinders the gear teeth from meshing, thus increasing the meshing force. This is characterised in the frequency spectrum by increased amplitude and harmonics of the gear meshing frequency. From experimental analysis it is known the most prominent amplitude change is that of the second harmonic of the gear meshing frequency.

1.2.1.2. Fluctuating Operating Conditions

CBM with respect to vibration analysis is based on the primary assumption that any change in the signal is due to the condition of the monitored system changing, i.e. any change is fault induced. However this is an invalid assumption when operating under fluctuating conditions, because changing load and speed have significant effects on the vibration signal. Randall determined that changing speed causes frequency modulation, while fluctuating load causes amplitude modulation (Randall 1989). Thus the many established and popular signal processing and fault detection techniques are generally inept when applied to fault detection under non-stationary operating conditions.

Amplitude modulation is essentially the multiplication of two time-domain signals. It is found in all gearboxes and is the cause of sidebands around the meshing frequency and its harmonics. These sidebands are commonly used to indicate the presence of wear. The mechanism behind the amplitude modulation is that the meshing frequency (carrier frequency) is modulated by the rotational speed of the gear. Fluctuating load will vary the amplitude of the meshing frequency. However, due to amplitude modulation this will also affect the amplitude of the sidebands, therefore eliminating the amplitude of the sidebands as a diagnostic feature.

Frequency modulation is similar to amplitude modulation, however instead of changing the amplitude of the carrier frequency it changes its frequency. Thus in the case of the vibration response of a gearbox, not only the frequency of the meshing frequency changes but also the frequency of the sidebands. This generally causes spectral smearing as the amplitude of the meshing frequency and its sidebands are distributed over a range of frequencies.

Stander et al. was one of the first groups to investigate vibration signals under fluctuating loading conditions (C.J. Stander et al. 2002). They measured vibration signals from a spur and helical signal stage gearbox operating at constant speed with varying loads and fault severities. They proposed a

novel technique of load demodulation normalisation (LDN) to detect gear faults. LDN implemented a combination of order tracking, synchronous averaging and the Hilbert transform. They were able to effectively detect faults; however, they were unable to quantify the severity of the faults.

Gaun et al. implemented ensemble empirical mode decomposition (EEMD) with order tracking to detect gear faults in a two stage helical gearbox (Guan et al. 2009). The order spectra of the intrinsic mode functions (IMF) were used as the diagnostic feature to detect gear faults. From their experimental data they concluded that their technique was effective for detecting faults and their severity, as well as being robust to non-stationary operating conditions and sensor location.

Bartelmus and Zimroz investigated the vibration signals from the planetary gearbox on a bucket wheel excavator that experienced a wide variation of load and speed fluctuations (W. Bartelmus & R. Zimroz 2009b). They understood that the interaction between the internal components of a gearbox and the external components in the system are crucial to fault diagnosis from vibration signals. Thus classical vibration monitoring methods are not suitable, because of their inability to take into account the influence of fluctuating speed and load. Therefore the identification of the external load variation is critical to efficient fault detection. Hence they proposed a novel method to identify external load variation which included filtration, enveloping and envelope frequency analysis. They also concluded that the vibration amplitude is not suitable to detect load change because it is known that vibration amplitude is dependent on the condition of the gear.

Bartelmus and Zimroz extended their research to develop a novel feature for condition monitoring of planetary gearboxes operating under fluctuating conditions (W. Bartelmus & R. Zimroz 2009a). The novel feature is the sum of 10 meshing amplitude components from STFT. The diagnostic may be simple, however when implemented in conjunction with an operating condition indicator (pinion speed) it proves to be very efficient, robust and intuitive. The proposed method determines the load susceptibility of the measured signal, from which the gearbox condition can be inferred.

Wang et al. knew that the residual signal of a gearbox, with respect to the raw signal and time synchronously average signal, is less sensitive to fluctuating load conditions, thus making it a more effective source for detecting gear faults (Xiyang Wang et al. 2010a). The residual signal represents the departure of the time synchronously averaged signal from the averaged tooth-meshing vibration signal. A novel fault growth parameter (FGP) was extracted from the amplitude of the CWT of the residual signal, because of the CWTs ability to effectively localise gear damage. The complex Morlet mother wavelet was implemented due to its similarity to the vibration signal caused by a mechanical fault. The FGP was able to quantify the severity of the gear damage because it used relative amplitude change of the CWT. They also investigated classical statistical indices such as kurtosis, mean, etc. to detect faults. However, they proved inadequate because of their sensitivity to load or insensitivity to early gear faults.

1.2.2. Tracking Shaft Rotational Displacement

Tachometers and shaft encoders track the location of a shaft and can provide invaluable information regarding the operating condition of a system and can significantly assist in the detection and classification of faults within a system. Common practices that are completely reliant on such information are TSA, resampling algorithms, etc. However installing tachometers does escalate the cost of CBM, and in many applications it is impossible to get access to the rotational shaft of interest. Thus without information regarding the shaft location fault identification can be severely hindered,

especially in the case of fluctuating operating speeds. Recently there has been a keen interest in estimation of the shaft speed from the measured vibration signal of the system operating under large speed fluctuations.

Urbanek et al. proposed a two-step procedure for estimation of the instantaneous rotational speed of a gearbox under large speed variations (Urbanek et al. 2013). The first step roughly estimates the instantaneous frequency (IF) of the system by applying a basic maxima tracking algorithm to the STFT of the vibration signal. The STFT transform of the vibration signal is useful for such an application because it is able to estimate the meshing frequency with respect to time, thus accounting for speed variations. The maxima tracking algorithm can be seen in the equation below. It is vital that the algorithm tracks a single harmonic and does not jump between harmonics because this severely reduces its overall accuracy. At this stage it is better to track a low harmonic that experiences less fluctuation, than a high harmonic that maybe more prominent.

$$f_{max}(t) = \text{Argmax}_f |X(t, f)|^2 \text{ for } f \in \Delta f_t \quad 1.3.$$

$$\Delta f_t \in \{f_{max}(t - d\tau) - \delta f, f_{max}(t - d\tau) + \delta f\}$$

Where δf is the given frequency tolerance and $X(t, f)$ is the STFT of the vibration signal. The next step is to resample the signal into the angular domain using the rough estimate of the IF. The resampled signal can then be bandpass filtered around the most prominent harmonic component of the IF of interest with a narrow angular band. The filtered signal is then resampled back into the time domain and the phase angle is calculated using the Hilbert transform of the new signal.

Another method proposed by Zhao et al. follows a very similar approach but is generally more efficient (Zhao et al. 2013). They began by first estimating the IF from the STFT, then they refined the IF estimation using the Chirplet transform. Next the most prominent harmonic was extracted by applying the Vold-Kalman filter (VKF), which is essentially a time varying band pass filter which uses the IF estimation as the dynamic center frequency of the filter. The bandwidth of the filter is generally half the distance between harmonics. The final step is to estimate the instantaneous phase using the Hilbert transform of the filtered signal and phase unwrapping. This proved to be a highly successful order tracking method as it achieved very accurate results and thus is highly suitable to be used in conjunction with other signal processing techniques such as SA and phase demodulation for fault detection and classification.

1.2.3. Data Processing and Feature Extraction

1.2.3.1. Synchronous Averaging (SA)

Synchronous averaging (SA) is an effective means of removing background noise from a measured signal and extracting the periodic components from the vibration signal. It is assumed that the noise is normally distributed. SA is the ensemble average of a time or frequency domain signal over multiple cycles. Bechhoefer and Kingsley proved that linear interpolation in the time domain is superior to Fourier domain based SA (Bechhoefer & Kingsley 2009). Thus SA is applied to gear vibration signals by synchronously averaging amplitudes at fixed angular increments for a number of shaft revolutions. Phase information for the angular resampling can either be gathered from a tachometer/shaft encoder which provides a certain number of pulses per revolution or through demodulation of the gear mesh signatures. Bechhoefer and Kingsley also calculated that non-synchronous noise is reduced by the inverse of the square-root of the number of revolutions. SA is ideal for gearbox vibration

analysis because of its ability to separate the vibration signal of the gear of interest from other gears and noise sources in the gearbox that are not synchronous with that gear.

1.2.3.2. Computed Order Tracking (COT)

Order tracking (OT) is an effective means of avoiding smearing of discrete frequency components due to speed fluctuations. OT samples are measured at fixed angular increments, not time intervals. This results in a frequency based on 'orders' of shaft speed (harmonics) and not Hertz (absolute frequencies (Hz)). It is crucial to OT that it has an accurate measurement or calculation of the instantaneous angular position. OT can be implemented mechanically with a shaft encoder triggering sample measurement or computationally with angular resampling (known as computed order tracking (COT)). With modern technology angular resampling can very easily be implemented on a signal, post measurement, with a myriad of methods; one of which is quadratic interpolation. Fyfe and Munck did extensive analysis on COT and determined its accuracy primarily depended on detecting the pulse of the tacho signal as early as possible, and concluded that a suitable high sampling frequency is necessary for accurate COT (Fyfe & Munck 1997). They also noticed that the number of pulses per revolution did not significantly affect the accuracy of the method. However, in cases where low rotational speeds are experienced they recommended using higher-order interpolation methods.

Eggers et al. used a variety of COT methods to detect gear faults on a dragline in the mining environment which operated under fluctuating operating conditions (Eggers et al. 2007). They discovered that COT can be implemented with insufficient speed data and in cases where there is bi-directional rotation. Features extracted from the COT signal using rotation domain averaging (angular version of SA) and statistical metrics are significantly more sensitive to faults compared to features extracted from the original signal.

1.2.3.3. Fourier/Spectra Analysis

Fourier analysis decomposes a signal into a summation of sinusoidal components, known as a Fourier series. Fourier analysis can efficiently be implemented on a digital signal using the Fast Fourier Transform (FFT). The FFT is one of the most common signal processing techniques for detecting gear faults.

The FFT is one of the primary techniques for detecting and classifying faults in rotating machinery, because of its ability to identify and isolate discrete frequency components that can be directly related to specific fault cases. This is possible because of the physical understanding of fault mechanisms and their characteristics. The Technical Associates of Charlotte publish a vibration diagnostics chart that gives a broad overview of vibration analysis and can be used to diagnose a wide variety of machine faults by only analysing the vibration spectrum.

However Fourier analysis is not sufficiently robust and sensitive to effectively detect faults in many industrial operating conditions, such as fluctuating operating speeds, low SNR, etc. The FFT cannot effectively deal with non-stationary signals, which are generated from machines operating under time-varying conditions such as a dragline gearbox (Eggers et al. 2007). The FFT is also unable to detect non-sinusoidal vibration components such as impulses which are found in ball bearing faults (Randall 2011).

1.2.3.4. *Cepstrum Analysis*

Cepstrum is also a standard signal processing technique for detecting faults in rotating machines. Cepstrum analysis is defined as the spectrum of the logarithmic spectrum. Cepstrum analysis is seen as a better alternative to autocorrelation function for detecting echo delay times (Randall 2011). Cepstra can either be calculated by the 'power cepstrum' or 'complex cepstrum'. The 'power cepstrum' is defined as the inverse Fourier Transform (IFT) of the log power spectrum while the 'complex cepstrum' is defined as the IFT of the log of the complex spectrum. The 'complex cepstrum' is fairly complex to compute because of the need to unwrap the phase into a continuous function of time. Hence the 'power cepstrum' is more commonly used technique because of its ease of implementation.

Cepstra is an effective tool to identify periodicity within the frequency spectrum, detect and remove echoes in audio signals and separate forcing functions. It is ideal to use cepstra to identify faults because it can collect families of sidebands and harmonics into a much more easily interpretable set of harmonics. Hence it is well suited to detect sidebands in the vibration spectra, which is known to be indicative of wear in gears (Dalpiaz et al. 1998). It is also able to estimate sideband spacing, thus enabling accurate measurement of sideband periodicity thus allowing fault diagnosis.

1.2.3.5. *Short-Time Fourier Transform Analysis (STFT)*

In many industrial applications rotating machines are subject to fluctuating operating conditions which generate non-stationary vibration signals. In these circumstances conventional spectra analysis is unable to effectively detect faults. Thus it is necessary to be able to detect change in frequencies with respect to time. Time-frequency analysis was seen as an effective method to monitor the transient and time-varying characteristics of a vibration signal (Safizadeh et al. 2000). One of the simplest time-frequency transforms is the Short-Time Fourier Transform (STFT). The STFT breaks down the signal into small segments that can be assumed to be locally stationary, and then applies the conventional FFT to these segments.

It is able to track the trend of the meshing frequency and its harmonics as they vary with time. Hence both Zhao et al. (Zhao et al. 2013) and Urbanek et al. (Urbanek et al. 2013) used the STFT to track the instantaneous meshing frequency of the gearbox in order to calculate the instantaneous shaft speed as discussed in Section 1.2.2.

Rajagopalan et al. implemented the STFT to detect faults in rotating machines under rapidly varying operating conditions (Rajagopalan et al. 2005). They noted that even though the STFT is simple and rugged, its basic flaws are that it makes the incorrect assumption that the signal is stationary within the time window. Therefore they recommended using quadratic time-frequency representations (such as Wigner-Ville Distribution (WVD), etc.), which are much more complex and computationally expensive.

Cocconcelli et al. used the STFT to detect faults in the bearing of an AC brushless motor (servomotor) that operates under non-stationary conditions (Cocconcelli et al. 2012). They implemented spectrum averaging (frequency version of SA) to improve the signal-to-noise ratio. Features were extracted from the averaged STFT using marginal time integration (MTI), which integrates along time and can be seen as the mean instantaneous power of the signal.

The primary disadvantage of the STFT is the trade-off between time and frequency. Resolution in the frequency domain is determined by window length; large windows produce good frequency resolution but result in poor time resolution (Nese et al. 2012). Therefore it is crucial to select the appropriate window size that will enable the frequencies of interest to be observable.

1.2.3.6. Continuous Wavelet Transform Analysis (CWT)

A popular time-frequency representation that has been extensively researched over the last decade is the wavelet transform (WT). This has proved to have a number of advantages over the STFT. Morlet proposed the novel concept of wavelets and with the help of Grossmann was able to develop the Continuous Wavelet Transform (CWT) (Grossmann & Morlet 1984). The CWT uses wavelets instead of sinusoids as the basis function to decompose signals. Wavelets are defined by the parameters scale (otherwise known as dilation), α , and shift (also known as translation), β , and can be written as a function of the mother wavelet Ψ , as seen below.

$$\Psi_{\alpha,\beta}(t) = \frac{1}{\sqrt{\alpha}} \Psi\left(\frac{t-\beta}{\alpha}\right) \quad 1.4.$$

The CWT is defined by the following equation in terms of the time signal, $x(t)$.

$$CWT(\alpha, \beta) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} x(t) \Psi^*\left(\frac{t-\beta}{\alpha}\right) dt \quad 1.5.$$

Since the CWT operates in terms of scale and not frequency, the results are better known as a time-scale representation. The WT essentially projects a signal on to a complete set of translated and dilated versions of the mother wavelet, ψ .

Giurgiutiu and Yu discussed the advantages between the CWT and the STFT (Giurgiutiu & Yu 2003). The basic difference between the two methods is that the CWT allows adjustable window sizes, which resolves the time-frequency resolution problem of the STFT. Also the coefficients of the CWT are easier to extract in order to monitor critical frequency components within the signal. Even though the STFT is simpler to understand and apply, the CWT has a wider range of applications and can be used for both denoising and spectrum analysis.

Staszewski and Tomlinson were some of the first researchers to implement wavelet analysis to detect gear faults (Staszewski & Tomlinson 1994). They understood that gear faults cause impulses and discontinuities within the vibration signal and that the FFT is unable to effectively detect them because of their non-stationary nature. Thus they used CWT because of its ability to detect discontinuities in a signal and its derivatives. CWT reveals discontinuities by a localised increase in the amplitude at small scale values at the time instant of the discontinuity. In their experimentation they noted that the CWT detected the impulse in the signal due to the damaged tooth significantly earlier than the FFT.

Peng and Chu compiled a comprehensive review of the WT with respect to machine condition monitoring, where many practical aspects of the WT were discussed (Peng & Chu 2003). One of the primary disadvantages of the WT is the inability to directly relate frequency to scale, thus making the results of the CWT very difficult to manually interpret. Hence the CWT is always used to extract features from a signal and implemented in conjugation with a pattern recognition method. The most common feature extraction techniques listed by Peng and Chu are wavelet coefficients based, wavelet energy based, singularity based and wavelet function based.

Another solution to the problem of relating frequency to scale is discussed by both Staszewski and Tomlinson (Staszewski & Tomlinson 1994) and Giurgiutiu and Yu (Giurgiutiu & Yu 2003). They both introduced quasi-relationships between scale and frequency, with the simplest requiring the center-frequency (f_c) of the mother wavelet and the sampling frequency (f_s) of the signal. This allows wavelet coefficients to be extracted at specific fault frequencies, enabling some interpretation of the features. The equation below displays how the respective frequency (f_α) of the scale (α) for a mother wavelet can be approximated.

$$f_\alpha = \frac{f_c \cdot f_s}{\alpha} \quad 1.6.$$

The figure below displays a Daubechies 4 wavelet, with a sinusoid equal to its center frequency. It shows that they can be compared, but may not be perceived as identical.

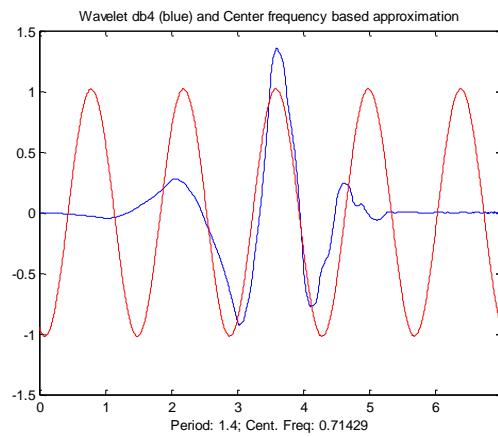


Figure 1.4: Wavelet and sinusoid comparison

Wang, et al. applied the CWT to the vibration signal of a gearbox operating under varying load conditions (Xiyang Wang et al. 2010b). They understood that the majority of vibration-monitoring techniques are unsuitable for fault diagnosis of gearboxes under fluctuating loads, because they are based on the assumption that the load is constant. It is known that the vibration amplitude of the gearbox casing, caused by the meshing gears, is modulated by the fluctuating loads on the gearbox. Thus the CWT is ideal for gearboxes under fluctuating loads and since scale and frequency can be related, Wang et al. were able to estimate fault frequencies. They concluded that the CWT effectively detected gear damage but was unable to evaluate the severity of the gear's fault or its rate of advancement.

There is a wide range of available wavelet families that can be implemented in the CWT. They either have their own properties and shapes that may be more or less sensitive to specific fault signatures. Kankar et al. investigated the optimum combination of wavelet family used in the CWT and machine learning techniques to detect faults in rolling element bearings (Kankar et al. 2011). They concluded that the Meyer wavelet family in conjunction with the Support Vector Machine (SVM) was most effective at detecting bearing faults among the wavelet families and machine learning techniques tested.

The CWT also generates a large amount of redundant data, which can be very computationally expensive when working with large amounts of data. Miao et al. implemented the CWT to the vibration signals from the accelerated life test of a cooling fan bearing (Miao et al. 2012). They wanted

to optimise the choice of scales to reduce computation time, thus they proposed a novel method, Autocorrelation function indicator (ACFI), to choose the optimum scale to detect faults. The ACFI correlates the values of different scales at the same time instant to determine the scale with the most fault related information. It was concluded that by optimising the scales, it significantly increases the performance of the CWT at detecting faults.

Elbarghathi et al. implemented the CWT to the SA vibration signal from a two stage helical gearbox (Elbarghathi et al. 2012). They used the wavelets to detect the fault impulses generated by the gears. The scales at which these fault impulses occurred where determined experimentally by comparing the spectrograms of the healthy and damaged gearboxes. They also experimented with different wavelet families to determine the optimum family for detecting fault impulses. They concluded that the most effective wavelet family produces the highest RMS value of the wavelet coefficients of the healthy signal, which in their case was the Daubechies 1.

1.2.3.7. Discrete Wavelet Transform Analysis (DWT)

The primary difference between CWT and discrete wave transform (DWT), is that instead of the scaling and translating factors being integers (1,2,3,etc.) they are powers of 2 (2,4,8,etc.), which results in the wavelets being orthogonal. This significantly reduces the amount of redundant data calculation, which is ideal for pattern recognition algorithms. Thus by selecting fixed values for $\alpha = \alpha_0^j$ and $\beta = k\beta_0\alpha_0^j$ for $j, k = 0, \pm 1, \pm 2, \dots$, the DWT can be defined as

$$DWT(j, k) = \alpha_0^{-j/2} \int_{-\infty}^{\infty} x(t) \Psi^* \left(\alpha_0^{-j/2} t - k\beta_0 \right) dt \quad 1.7.$$

However α and β were replaced by 2^j and $2^j k$ respectively, then the DWT can be written as.

$$DWT(j, k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x(t) \Psi^* \left(\frac{t - 2^j k}{s^j} \right) dt \quad 1.8.$$

An efficient manner to implement the DWT was proposed by Mallat (Mallat 1989), where two complementary filters spilt the time signal into a low frequency signal (approximation, A) and a high-frequency signal (detail, D). This can be repeated by further filtering the approximation into its own approximation and details. This decomposition process is represented in the figure below.

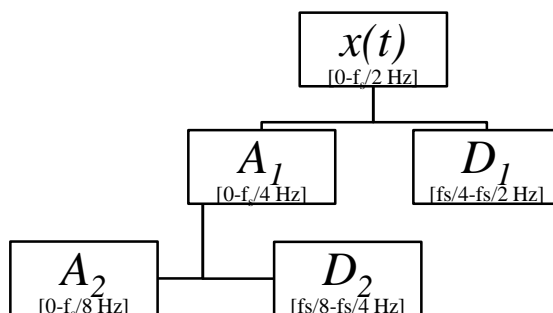


Figure 1.5: Discrete wavelet decomposition diagram

The DWT is an effective method for multi-resolution analysis (MRA), allowing the signal to be analysed at different frequencies with differing frequency resolution. The DWT does not require the scale to be related to frequency, because the frequency band of the decomposition level can be calculated from the sampling frequency and level of resolution. The frequency bands can be seen in the figure above.

Prabhakar et al. used DWT to detect inner and outer race faults on a rolling element bearing (Prabhakar et al. 2002). They noted that the time-domain and FFT signals do not detect impulses due to faults effectively, specifically inner race faults that are weakened by the transmission path and masked by noise. However using the DWT with Daubechies 4 mother wavelet and inspection, fault impulses were effectively detected by the second level decomposition and thus the faults were classified by comparing the frequency of the occurrence with the respective fault frequencies of the bearing.

Lou and Loparo implemented the DWT to detect faults from ball bearings (Lou & K. a Loparo 2004). Similarly to Prabhakar, they determined the frequency band of the fault impulses from the comparison of the healthy and damaged decomposed signals. To diagnose the faults they implemented a neuro-fuzzy classification system, which extracted the standard deviation of the wavelet coefficients from the frequency bands of interest as features.

Purushotham et al. also implemented the DWT to diagnose bearing faults (Purushotham et al. 2005). However instead of determining the fault impulse frequency band from inspection, they knew the frequency band of the bearing block and thus were able to select the appropriate frequency band without historical fault data. A HMM was used to classify faults from the complex cepstral coefficients extracted from the DWT.

Sanz et al. used the DWT to detect gears faults in rotating machines (Sanz et al. 2007). DWT has proven to be an effective method for detecting impulses caused by faulty bearings, however there has not been significant research of DWT with respect to gears. Similarly to the CWT, the wavelet families of the DWT vary in performance in detecting gear faults.

Wu and Liu implement DWT using a range of Daubechies wavelet functions to detect faults in an internal combustion engine from its sound emissions (Wu & C.-H. Liu 2008). Due to the high computational cost of the CWT, they opted to use the DWT which systematically decomposes a signal into sub-band levels which significantly reduces the computational cost. An advantage of the DWT that they used was its ability to characterise signals at different localization levels in time and frequency domains. They decomposed the original time signal into 8 levels, which results in 8 detail vectors and 1 approximation vector. The energy contained in the approximation and detail vectors at each decomposition level were calculated using the RMS value and used as features for an ANN.

Guo implemented the DWT to detect a wide range of gear faults found in gearboxes of wind turbines (Guo 2010). DWT was necessary due to the non-stationary components within the vibration signal and the loss of fault signals over the transmission path. They determined the scale band of gear faults by calculating and comparing the energy in each band, because faults generally cause an increase in energy in specific scale bands. Then the signal was reconstructed from the fault frequency band and the signal was analysed using the standard FFT, from which they were able to determine the location and type of the fault.

1.2.3.8. Wavelet Packet Transform (WPT)

Wavelet packet transform (WPT) is very similar to the DWT, as it is also a MRA tool. WPT decomposes both the approximation (low frequency) and detail (high frequency) coefficients while the DWT only decomposes the approximation coefficients. This overcomes the problem of the DWT, by decomposing the detail coefficients it allows analysis of high frequency components with high

resolution. This enhanced decomposition capability enables it to detect and differentiate transient components with high frequency characteristics (Yan et al. 2013).

Ocak et al. implemented the WPT in conjugation with a HMM to track fault severity in bearings (Ocak et al. 2007). They windowed the signal into epochs (2^{10} samples which equates to roughly 0.5s) and applied a 5th level WPT decomposition and calculated the energy of each scale band for each epoch. Thus they neglected none of the original signal. The energies of the epochs were used as features to train a HMM on healthy data. Bearing faults were detected and tracked by the probability of the HMM.

Li et al. calculated the kurtosis of the wavelet coefficients of a 3rd level WPT decomposition of ball bearing vibration signals (Li et al. 2008). The kurtosis values were plotted, thus enabling operators to trend the periodicity and severity of the faults. They understood the need for the WPT because of the need to detect the non-stationary or singular vibration components submerged within the vibration signal that are indicative of damage.

He et al. applied the WPT to the vibration signal from a gearbox because of its ability to extract the discriminatory information from the raw signal (He et al. 2010). Defect induced vibration signals are often corrupted by noise, distorted by coupling of machine components and the environment. Thus the WPT is the ideal tool to detect incipient faults generated under a variety of gear conditions.

Zhang et al. used the WPT to decompose the vibration signal of an unbalanced fan, then the decomposed signals are transformed with the FFT and the peak values were used as the feature vector for an ANN (Z. Zhang et al. 2012). They determined that the frequency resolution of the WPT must be close to the fundamental frequency of the bearings, thus dictating the level of decomposition required. This is contrary to the majority of methods which use simple visual inspection to determine the best wavelet decomposition level.

1.2.3.9. High Frequency Resonance Technique (HFRT) or Envelope Analysis

As mentioned in Section 1.2.1.1, pertaining to gear fault mechanisms, impulses are generated when defects are in contact with another surface while under load. These impulses are generally not visible in the spectra of the entire signal because of their high frequency and short duration resulting in a wide distribution of energy in the spectrum (McFadden & Smith 1984). One such technique to extract the fault impulses is envelope analysis or high frequency resonance technique (HFRT). The first step of the HFRT method is to band-pass filter the time domain signal around a high frequency where the fault impulses are expected to be amplified by the structural resonance of the machine. The filtered signal is then amplitude demodulated to generate the envelope signal, from which the periodicity of the impulses can be compared to the characteristic fault frequencies of the bearings and the fault can be diagnosed (Randall 2011).

McFadden and Smith commented that it is not worthwhile to calculate the resonance frequencies because calculations give no ranking of the relative magnitude of resonances when observed in practice (McFadden & Smith 1984). Thus the resonance frequencies are generally determined by comparing the high frequency bands of the healthy and damaged signals, and identifying the frequency band in which the fault impulses are most noticeable/amplified. This is commonly practised when using the DWT. Bozchalooi and Liang attempted to determine the high resonance frequency experimentally through an impact test and observing the frequency response function (FRF) (Bozchalooi & M. Liang 2007). They were able to identify the structure's resonance frequencies from

the FRF, which proved to be a valid estimation for the center frequency of the band-pass filter, which effectively removes the need for historic fault data to determine the relevant frequency band. It is however still necessary to approximate the width of the frequency band and it is vital to understand that the physical characteristics of the system do change with respect to operating conditions.

Rubini and Meneghetti implemented the CWT in the place of the standard band-pass filtering in the HFRT to increase its robustness (R. Rubini & Meneghetti 2001). The original signal is band-pass filtered by transforming it into the time-scale domain and then inverting it back into the time domain, however only using the coefficients at the scales which represent the frequency band. They noted that spectral analysis often proves inadequate at detecting faults as the magnitudes at the characteristic fault frequencies in the damaged signal are equal to those in the healthy signal. It was discovered that the reliability of the spectral analysis depended on not only the fault severity, but load intensity and other such parameters. It was concluded that that CWT proved to be an effective band-pass filtering method for a variety of operating conditions.

Since the WT is more effective at detecting the fault impulses, Rafiee et al. set out to determine the optimum wavelet family to detect fault impulses from both gears and bearings (J. Rafiee et al. 2010). After evaluating 324 wavelet families with experimental data from both faulty bearings and gears, the Daubechies 44 (db44) was determined to be the best overall wavelet family for determining impulses from both gear and bearing vibration signals.

1.2.4. Fault Detection & Classification

1.2.4.1. Manual Expert Classification

It is common practice for experienced/trained maintenance engineers to be able to detect and classify machine faults manually by applying basic rule-based interpretation. They primarily use indices and statistical features from the time domain such as RMS and kurtosis. Also the vibration spectrum is a popular tool for the diagnosis of rotating machines as discussed above. A common set of rules is the Vibration Diagnostic Chart produced by the Technical Associates of Charlotte, which can be used to diagnosis the most common faults under stationary operating conditions.

However the above mentioned features only accurately represent the machine condition if they are properly normalised with respect to operating conditions. Thus any vibration signal from a machine under fluctuating operating conditions needs to undergo appropriate processing (amplitude/frequency demodulation), to be able to be used in conjunction with manual interpretation.

The main disadvantages of manual classification are personnel related, as it is very labour intensive and requires personnel with specific condition monitoring training. Also data capturing that is not automated can be extremely labour intensive for large industries.

1.2.4.2. Machine Learning

Fault detection and diagnosis have been successfully implemented in many cases using pattern recognition models, because of its ability to handle significantly more complex features than humans can interpret. It has many advantages over manual classification as it is completely automatic and requires minimal labour. However, as will be discussed later, machine learning techniques must be trained based on historical data of faults. This is a disadvantage for many applications where there are no historical data or where it is too expensive to gather data for every fault type.

Miao and Makis identified the two most common approaches to pattern recognition as a knowledge-based approach or a statistical data-based approach (Miao & Makis 2007). An example of a knowledge-based approach is artificial neural networks (ANN) in conjunction with fuzzy logic models, which apply human knowledge in the form of logical rules applied to extracted features. ANNs have many advantages such as superior learning capability and noise suppression, etc.. However, its success strongly relies on the problem, the correct selection of model topology, good feature selection and sufficient training data. An example of a statistics-based approach is the HMM and GMM, which generally requires less human understanding of the system. Again however, it suffers from the same disadvantages, such as its need for training data. One of the advantageous features of the HMM is its ability to classify time-series data such as speech signals using the Markov assumption.

Gaussian Mixture Models (GMM)

The GMM has not seen extensive use in the field of CBM, as many researchers either opt for either a NN or a HMM to classify a fault. The GMM is a non-linear pattern recognition algorithm, which is based on a maximum likelihood model defined by weights (w), means (μ) and covariance (Σ). The GMM is trained using the Expectation Maximisation (EM) algorithm, which is comprised of two primary steps, namely expectation (E-step) and maximisation (M-step). The E-step guesses the probability distribution over the missing data with respect to the current model. Then the M-step re-estimates the model parameters using the missing data. This process is repeated until the model parameters reach the desired level of convergence.

Marwala et al. were the first to investigate the ability of a GMM at detecting bearing faults (Marwala et al. 2006). They compared the commonly used HMM to the new GMM. The models were investigated using features extracted by multi-scale fractal dimension estimated using box-counting dimension. They concluded that the HMM was a superior classifier, but that the GMM was much simpler to train and implement. Xinmin et al. also implemented a GMM to detect a variety of faults in rolling element bearings (Xinmin et al. 2007). The features extracted for the GMM were taken from the Lyapunov exponent spectrum. They agreed with Marwala that the GMM is a suitable tool for diagnosing bearing faults.

Hidden Markov Models (HMM)

The basic schematic for a continuous HMM can be seen in the figure below. The model below consists of 3 hidden states, with their respective transitional probabilities between them (a_{ij}). Then there is a single continuous observation, where each hidden state has its respective Gaussian probability parameters that it is observed ($b_i(V)$).

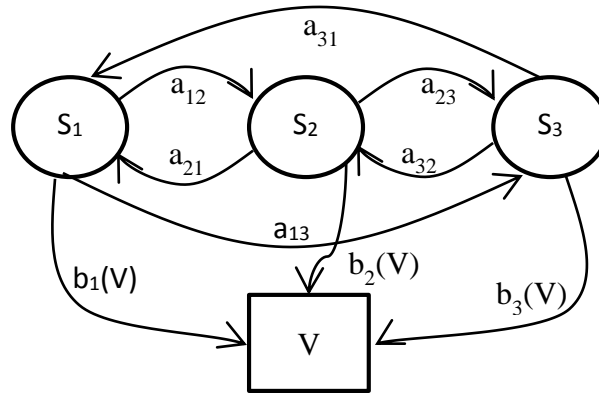


Figure 1.6: Schematic diagram of continuous HMM

There are two primary approaches to assigning physical significance to the hidden states within a HMM in the field of CBM. Firstly each of the hidden states represent the physical state of the machine (fault), while the second approach is where each of the hidden states represent an operating condition. The first approach has been implemented by Cartella et al., where the condition of the machine is predicted by decoding the observation sequence to reveal the hidden states using the Viterbi algorithm (Cartella et al. 2012). A left-right continuous HMM was used where the state transition matrix is upper triangular, which results in no transitions to states below the current state. This allows the model to track the changes of the severity of the bearing fault, while not allowing the model to reduce the severity of the fault.

The second approach was followed by Ocak and Loparo who used multiple HMMs to detect and classify rolling elements bearing faults in an induction motor (Ocak & K. A. Loparo 2001). Each HMM represents a machine condition and its hidden states represent possible operating conditions. Thus the machine condition is classified based on identifying the HMM with the highest probability. They also had to window the signal, to generate a time series set of measurements from which features could be extracted. However the length of the windows was determined experimentally by comparing the diagnosis accuracies of differing window length, thus giving no conclusion of how to decide the window length for novel systems.

Miao and Makis implemented a very similar HMM to Ocak and Loparo. They understood the need for proper selection of the observation sequence, which includes feature extraction and windowing (Miao & Makis 2007). Instead of simply dividing the vibration signal into equal arbitrary sized windows like Ocak and Loparo, they used the CWT which allows the wavelet coefficients at each time instance to form the time-series feature vectors. However the primary problem with both these approaches is the imperative need for large amounts of training data to train each of the individual models for the healthy condition and each of the expected fault conditions respectively.

Kang et al. compared different feature extraction techniques for HMMs, where the hidden states were representative of differing stages of gear wear. They compared the continuous HMM (CHMM) and discrete HMM (DHMM) which require vector quantisation (Kang et al. 2011). DHMMs may be simpler to implement, but they are more computationally expensive because they need to quantise each observation vector. They were also less accurate due to loss of information. The extracted features were a combination of 10 time domain features (mean, standard, deviation, etc.); 30 energy scales from 4 level wavelet package decomposition and 30 standard deviations of each package's

reconstruction coefficient. Kang et al. also used the amplitude demodulated vibration signal as a feature vector, but it proved to be significantly less effective than either the CHMM or DHMM with the above extracted features. They found that CHMMs were more effective and efficient than DHMMs with features extracted from continuous signals, because of the DHMM's need to quantise features.

1.2.4.3. Discrepancy Analysis

Discrepancy analysis attempts to combine machine learning with manual classification. Machine learning is used to transform a complex signal/high-dimensional feature vector into a much simpler signal that is simple to interpret manually.

The discrepancy signal concept originates from residual signal analysis (RES), which removes healthy vibration components, such as meshing and shaft components and their harmonics, from the time domain signal such that only fault related components remain. Discrepancy analysis is generated by comparing a novel signal to a reference model that is representative of the vibration signal of a healthy machine for the full range of expected operating conditions. This smooths the signal and enables faults to be significantly easier to detect. It is assumed that any discrepancy is the result of the presence of a fault. This has the huge advantage that the machine learning technique only requires training data for healthy conditions and not for every fault type. Heyns and Heyns determined the discrepancy signal of a gear with localised wear, with the use of 5 layer auto-encoder. Faults were then easily classified by analysing the spectra and cepstra of the discrepancy signal because of the signals smoothness and robustness to the fluctuating operation conditions (T. Heyns & P. S. Heyns 2012).

Heyns et al. implemented a novel discrepancy method to detect and trend gear faults (T. Heyns, P. S. Heyns, et al. 2012). They initially applied COT to the signal to remove any frequency modulation within the signal. Next they resampled the signal into windows, with the window length equalling the length of a single gear tooth pass. The sliding windows were so chosen as to represent a single gear meshing period. Each of the windows was thus expected to be approximately similar, which made it easier to train a representative model and subsequently to detect deviations. Avoiding overlapping windows also reduced the computational cost. They trained a Gaussian Mixture model (GMM) on purely healthy data; such data being readily available from any new machine. Using the GMM and windowed signal, they were able to produce a discrepancy signal in the form of Negative Log Likelihood (NLL) values. Gear faults were easily detected by analysing the spectra and cepstra of the synchronously averaged discrepancy signal.

Heyns et al. implemented an adaptive time series model based on Bayesian model selection to model the healthy waveform of a gearbox. When deducted from a novel vibration signal the fault induced waveform or residual signal can be determined (T. Heyns, Godsill, et al. 2012). The adaptive time series model is based on multiple auto-regressive models, each model representative of a specific operating condition. As such the combination of the models with their respective likelihoods to the novel vibration signal, allow the model to handle fluctuating operating conditions. Since the residual signal still contains a modelling error and white noise, the likelihood of the residual signal was used as a discrepancy signal instead of the residual signal. The fault location and severity was easily determined from the discrepancy signal, which is sensitive to gear damage and robust to fluctuating operating conditions.

1.3.Scope/Research Objectives

The focus of this dissertation is the data processing stage because of its predominant influence on the performance of CBM systems. The goal is to develop an accurate, robust and cost effective technique to represent the data in a simple and intuitive manner for classifying, locating and quantifying the severity of a variety of common faults found in rotating machines. Accuracy entails an ability to classify a given fault, without any misclassification of healthy data. It also encompasses the ability to locate the position and quantify the severity of the given fault. Robust techniques must be able to operate under a wide range of operating conditions, where both the load and speed fluctuates, as well as handle noise and transmission path interruption. Cost effective techniques often require minimal or no historic fault data for machine training and should avoid excessive computational needs. Also to improve the simplicity of the methodology it is assumed that only the vibration signal is available and that no measurements of speed or load are available.

The objective of this dissertation is to implement and investigate a novel discrepancy analysis technique, which generates a simple and intuitive discrepancy signal from which the type, location and severity of a variety of faults can be determined. From the literature and a basic understanding of statistics, it is proposed that the discrepancy signal can be generated from a series of steps which include the use of a HMM and GMM. In the first step a HMM is used to determine the sequence of instantaneous operating conditions. Then a GMM for each of the identified operating conditions trained solely on healthy data can determine the NLL value, which can be used as a measure of the discrepancy from the healthy signal.

Since the HMM is used to determine the sequence of instantaneous operating conditions it is crucial that it only receives operating condition sensitive features that are robust to the presence of a fault. Similarly the GMM, which detects the fault, be given only fault relevant features that are unaffected by the fluctuating load and speed. It is evident from the literature that effective features can be extracted from the time-frequency or time-scale domains because of their ability to handle the fluctuating operating speeds. The time-frequency domain is great for capturing the trend of the meshing frequency, therefore it is an ideal transform from which operating condition specific features can be extracted from. The time-scale domain is ideal for detecting discontinuities within a signal, thus making it well-suited for detecting fault related impulses. Also it is proposed that the frequencies or scales that the fault impulses excite can be predetermined from an in-depth knowledge of the expected fault mechanism, thus negating the need for historic fault data.

In order to validate the proposed discrepancy method, it is necessary for an experimental test rig to generate vibration signals of a gearbox under fluctuating operating conditions. The non-stationary operating conditions should have large fluctuations in both speed and load, thus be in a similar non-stationary form of the loads on the dragline gearbox investigated by Eggers (Eggers et al. 2007). The experiment needs to take the form of an accelerated life test of both the gears and bearings in the gearbox, to be able to track the severity of the faults.

The diagram below gives a basic run through of the proposed methodology, beginning with the unprocessed time domain vibration signal. The first step is to identify the operating conditions; by initially implementing a rough windowing scheme and extracting features that are representative of the instantaneous operating conditions from each window. Then the features are inputted into a HMM to detect the most probable sequence of underlying states. Since there is no measure of the

shaft displacement it is necessary to track the input shaft displacement. Thus the second step is to estimate the angular displacement of the shaft. Guided by the results of the HMM and a maxima tracking algorithm implemented on the STFT, the IF can be estimated. The IF is then used to center the Vold-Kalman filter, which in turn generates a filtered waveform from which the instantaneous phase can be calculated. The third step is to generate the discrepancy signal by detecting fault induced vibration components. Firstly the vibration signal is re-windowed with respect to the revised calculation of the shaft displacement, since the windowing scheme is critical to the accuracy of the fault diagnosis. Then fault sensitive features are extracted from each window and inputted into the GMM that correlates to the identified instantaneous operating condition. The GMM generates a discrepancy signal in the form of NLL values, from which basic techniques such as TSA, spectral and cepstral analysis can identify and locate faults.

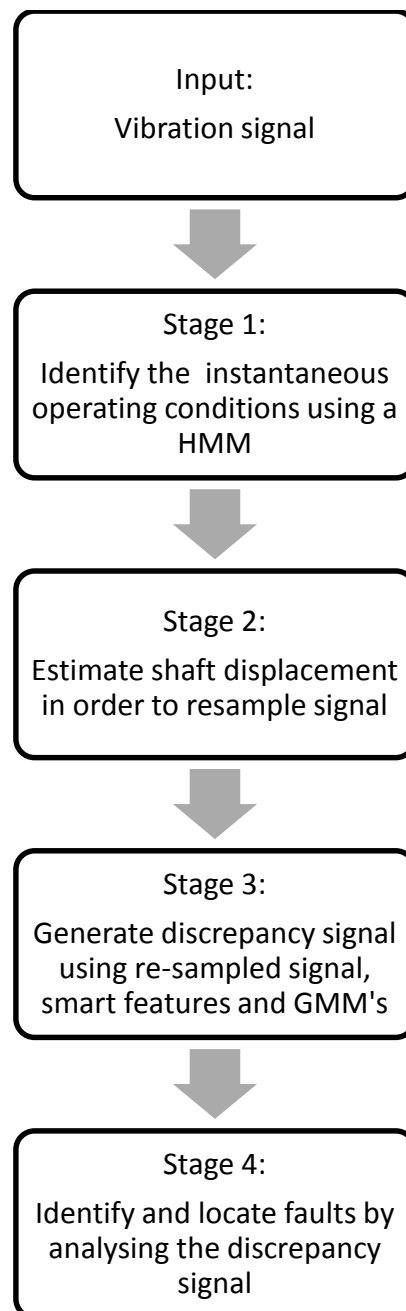


Figure 1.7: Flow diagram of proposed methodology

1.4. Report Layout

Chapter 2 presents a detailed description of the novel discrepancy methodology as well as its aims. The chapter follows the procedure displayed in Figure 1.7, firstly discussing the operating condition feature extraction and classification. Then the order tracking approach is presented, followed by the fault feature extraction and classification methodology. Finally some basic signal processing techniques that will be applied to the discrepancy signal are discussed.

Chapter 3 describes the lumped mass model used to generate simulated data and the physical experimental setup used to measure actual vibration responses of healthy and damaged gearboxes operating under non-stationary operating conditions.

Chapter 4 is the discussion of the results of the proposed methodology when applied to both the simulated and experimental data. The chapters begins by assessing the healthy signal to try extract as much useful information from the system as possible, such as region of meshing and resonance frequencies. Then it goes on to investigate the operating condition features and the HMM's ability to accurately determine the state that represents the instantaneous operating conditions. Next the order tracking approach is validated against the actual displacement of the shaft. Finally a range of fault features are investigated in order to validate which is the optimum feature set for fault detection.

2. Discrepancy Analysis Using Smart Features

2.1. Introduction

The proposed novel discrepancy analysis methodology proposed in this chapter is comprised of many of the signal processing, statistical modelling and fault detection techniques discussed in the previous chapter. This chapter begins with defining the scope and aims of the proposed algorithm. The method has two primary stages. The first stage involves extracting operating condition sensitive features and then identifying the instantaneous operating conditions. The second stage generates the discrepancy signal by extracting fault sensitive features and using a range of discrepancy models for each of the respective operating conditions.

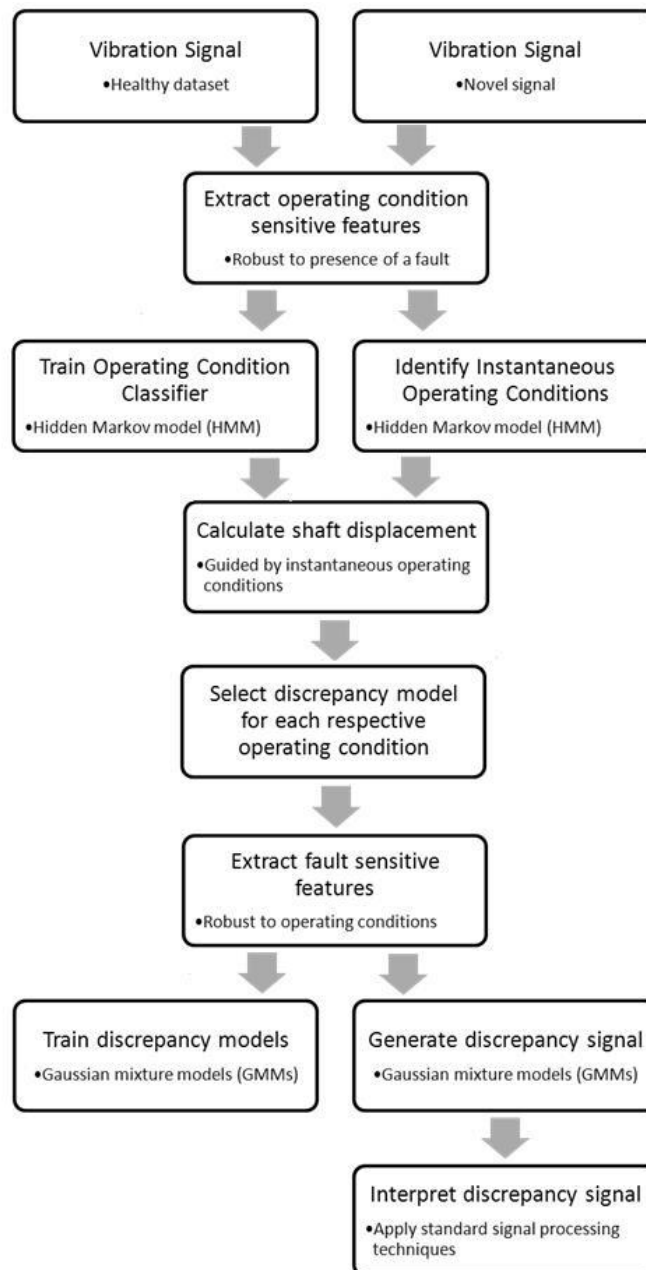


Figure 2.1: Flow diagram of proposed methodology

2.2. Aims of the Methodology

The primary aims of the proposed novel methodology are:

1. **Fluctuating operating conditions:** The algorithm must be able to detect faults from vibration signals generated from a gearbox functioning under non-stationary operating conditions, namely fluctuating speed and load. This is of importance due to the large number of industrial applications (mine drag lift, wind turbine, etc.) which operate under non-stationary conditions. Also the majority of research in the field of CBM is based on stationary operating conditions and most stationary process algorithms are inadequate at detecting faults under non-stationary operating conditions.
2. **Simple instrumentation:** It is assumed that only a single unidirectional accelerometer is available and that there are no tachometers or other transducers that are able to directly measure the operating conditions.
3. **Independent of historic fault data:** The algorithm must be able to diagnose faults in the absence of historical fault data. It is uncommon for mechanical components to have historically measured vibrations signals for a wide range of faults. Diagnosis of machine faults without fault data for comparison is a primary problem faced when installing condition monitoring and evaluation systems. Obtaining fault data is generally a costly process and thus significantly depreciates the cost effectiveness of CBM.
4. **Smart features:** Smart features are selected based on an understanding of the faults of interest and the dynamic characteristics of the machine under consideration. In the case of fault detection these smart features are selected to be sensitive to specific faults while also being more robust under different operating conditions. Smart features for the identification of operating conditions are sensitive to load and speed fluctuations, while also being robust to the presence of faults. The features are also selected so as to be computationally effective.
5. **Fault detection and diagnosis:** The faults must be easily located and classified by applying basic CBM techniques on the discrepancy signal produced by the algorithm. The location and classification of the fault are critical to CBM, as it enables the operator to make efficient maintenance preparations and decisions, such as part availability and scheduled down time.
6. **Time to failure:** The severity of the fault is also crucial to CBM. The severity of faults enables the operator to estimate the remaining useful life of the damaged components, as well as plan an efficient down time schedule.

2.3. Identify the Instantaneous Operating Conditions

One of the primary aims of the proposed method is to be able to detect faults under fluctuating operating conditions. To make the system more robust to fluctuating operating conditions, the instantaneous operating conditions are determined. This allows the presence of a fault to be detected with respect to the operating conditions, thereby reducing the effect the operating conditions on the detection scheme. The basic approach of the proposed method can be found in the schematic below. In general smart features that are specifically related to the operating conditions are extracted from the spectrogram, dimensionally reduced using PCA and sent to the HMM which identifies the most likely sequence of operating conditions.

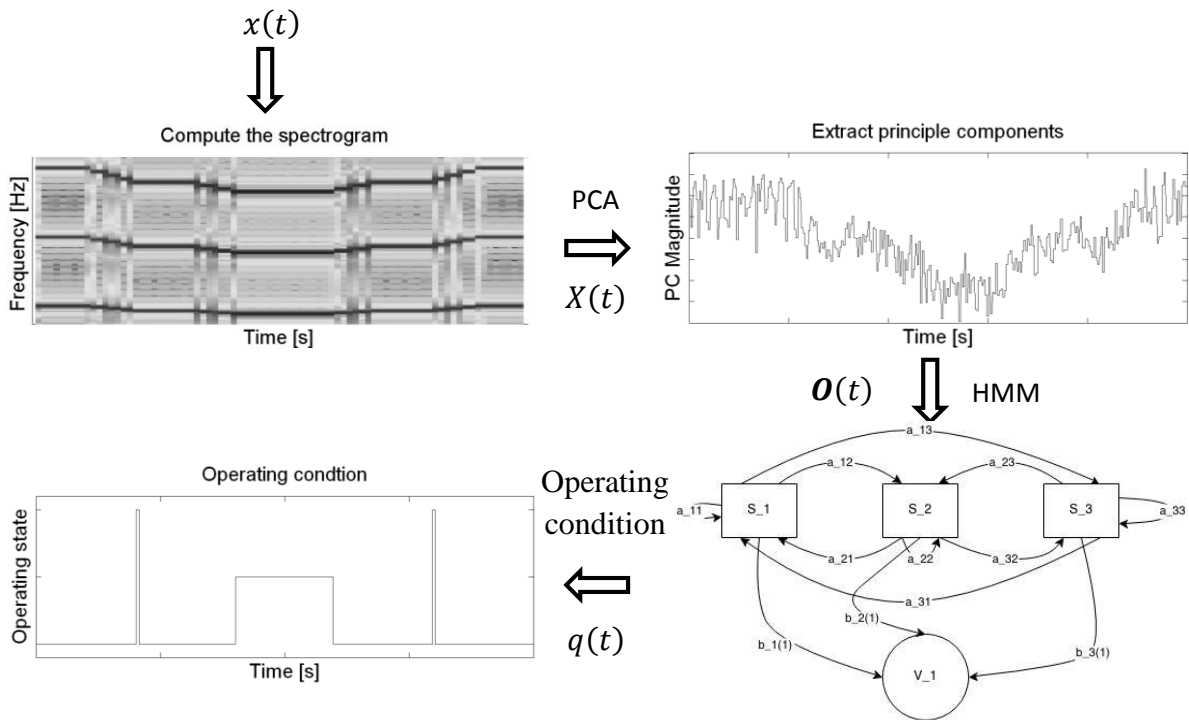


Figure 2.2: Schematic of operating condition classification approach

2.3.1. Rough Windowing of Time-Domain Signal

The initial step of estimating the instantaneous operating condition is to window the time-domain vibration signal into appropriately sized windows. The signal is windowed so that operating condition features can be extracted from each window. An appropriate window size is small enough to be able to accurately follow the change of operating conditions, however also long enough to contain enough information to classify the operating condition. An inappropriately sized window will lead to misclassification of operating conditions and thus reduce the overall accuracy of the methodology. Thus the appropriate window size covers a wide range of possible sizes and the selection of a size within that range has a negligible effect on the overall performance of the methodology.

Since the windowing size does not significantly affect the accuracy of the identification process, the window size used in this methodology is based on the time period that each tooth is meshed. The meshing period can be estimated by the number of teeth on the gear and the estimated average speed of the gear over the whole signal. The number of gear teeth is a known design value and the average shaft speed can be estimated from knowledge of the gearbox's application or simply measured using a hand held tachometer. The window takes the form of a rectangle with a 50% overlap with adjacent windows. The fixed angular windows can be seen in the figure below. Each of the windows will have the identical size or time period because the estimated average speed will be applied as constant over the entire length of the signal. The windows will not physically align with the gear teeth, however this does not affect the success of identifying the operating conditions.

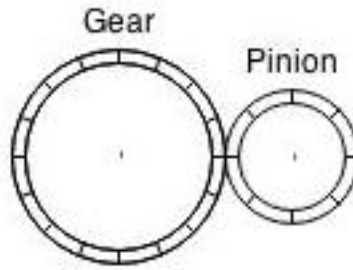


Figure 2.3: Fixed angular windowing scheme

2.3.2. “Smart” Operating Condition Feature Extraction

The next step is to extract operating condition specific features within each window to generate input vectors for the HMM. As discussed in the previous chapter the spectrogram of a vibration signal is able to track both the instantaneous frequency and amplitude of the meshing frequency with respect to time, thus making it an ideal tool to extract both speed and load variations. The figure below is an example of a spectrogram with varying meshing frequency; the dotted blue line follows the third harmonic and it is clear how the fourth harmonic also follows the same pattern. Thus the spectrogram is a highly suitable tool for the extraction of operating condition specific features.

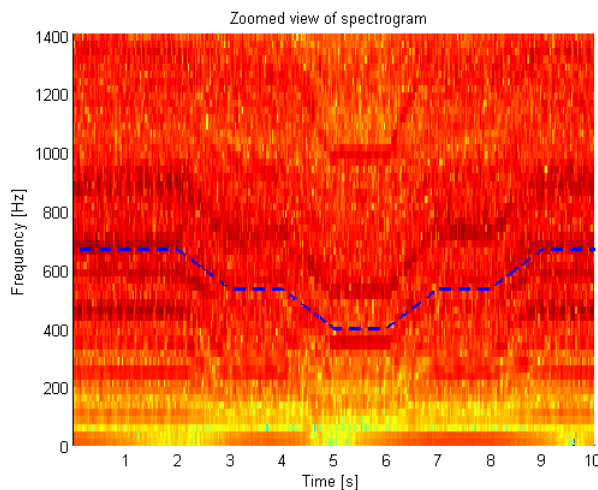


Figure 2.4: Example of spectrogram clearly displaying the time varying meshing frequency, highlighted by the blue dashed line.

The spectrogram of a time domain vibration signal, $x(t)$, can be generated by the squared magnitude of STFT, which is defined by the equation below.

$$X(t, f) = \left| \int_0^t x(t) \gamma(t - \tau) e^{-i2\pi f \tau} dt \right|^2 \quad 2.9.$$

Where $\gamma(\tau)$ is the analysis window function for the STFT.

The feature vector is comprised of each of the frequency components within the frequency range that contains the meshing frequency and its harmonics. For the spectrogram in the figure above that would be in the frequency range of 200 to 1000 Hz. The STFT has a trade-off between temporal and frequency resolution. Therefore the window size of the STFT is chosen such that it best represents the signal in both the time and frequency domain. It is expected that the frequency range of interest will be windowed into roughly 20-40 windows (this does depend on sampling frequency, period and STFT

window size). Therefore the observation vector of the HMM is composed of the value in each of the windows within the frequency range of interest. The observation vector is defined by the equation below. Next the observation vector must be windowed in the time domain according to the windowing scheme discussed in 2.3.1. It is highly probable that the windows will be smaller than the windows implemented by the STFT, therefore many consecutive features may have identical values. This is, however, advantageous for the HMM as will be discussed later. Thus an operating condition feature vector is generated that has a fairly high dimensionality but captures the necessary information.

$$\mathbf{O}_t = X(t, f_{interest}) \quad 2.10.$$

where $f_{interest} \in \{\min(f_m); 4 \times \max(f_m)\}$

The dimensionality of the vector is equal to the number of windows the frequency range of interest is divided into. Thus the feature vector does have a high dimensionality and there is a fair amount of unnecessary information within the feature vector. Therefore a popular technique for reducing the dimensionality is the Principle Component Analysis (PCA) which transforms the feature space and extracts the elements with the most variance, thereby reducing the total dimensionality and keeping the most prominent information. PCA does not only compress data, but identifies patterns in data and displays it in a manner that highlights their similarities and differences. In the same manner that eigen-faces are used in facial recognition, “eigen-spectra’s” will be used to characterise the variation in the STFT. Essentially “eigen-spectra’s” are a set of eigenvectors that capture the variation in a collection of spectra’s and can be used to compare spectra’s. Each “eigen-spectra” is associated with an eigenvalue which is a measure of its variability. Therefore, the higher the eigenvalue, the greater the ability of the “eigen-spectra” to capture the variation in the spectra. In Figure 2.5 is the first four “eigen-spectra’s” of the spectrogram plot in Figure 2.4. In the title of each plot the eigenvalue is given in brackets. Also the meshing frequency and its harmonics at the maximum speed is displayed by the red dotted line.

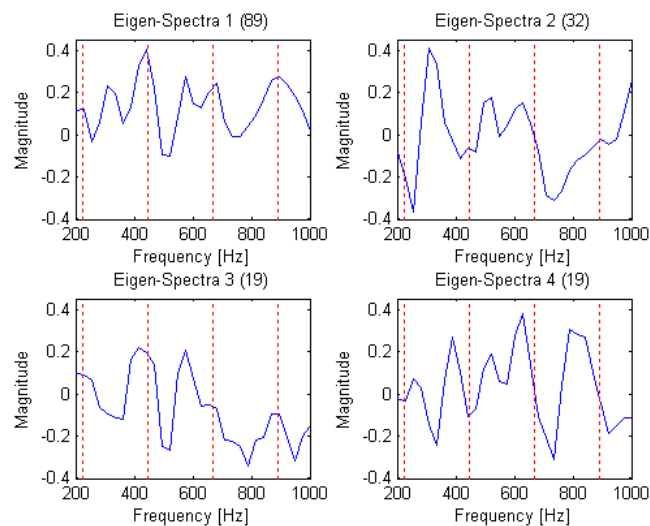


Figure 2.5: Example of "eigen-spectra's" of spectrogram in Figure 2.4. Dominating meshing frequency and its harmonics are displayed in red.

The first “eigen-spectra” has by far the highest eigenvalue of 89 and is able to capture the spectra with peaks at the meshing frequencies harmonics. The next three “eigen-spectra’s” have significantly reduced eigenvalues which is indicative that they contain less variation. Also they do not capture meshing frequency and its harmonics. However, it is possible they capture the meshing frequency and

harmonics of lower operating speeds. Therefore PCA is not simply a black-box approach to data compression, but a method of extracting a few spectra's that are representative of the variation in a large collection of spectra's (i.e. a spectrogram).

PCA is implemented by calculating the principle component coefficients ("eigen-spectra's") which are used to linearly transform the feature space and weights which are essentially the means of each of the elements within the vector. The new feature space lists the elements in descending order according to their variance. Thereby extracting the first n elements, the n most variable elements are drawn out. The new observation vector with the reduced dimensionality of $N_{reduced}$ is defined below with respect to the PC coefficients (C^{PCA}) and weights (μ^{PCA}).

$$\mathbf{O}_t^{PCA} = (\mathbf{O}_t - \mu^{PCA}) \times C_{1:N_{reduced}}^{PCA} \quad 2.11.$$

The figure below contains the first four transformed elements of the spectrogram in Figure 2.4. It is evident that the first PC contains a significant amount of variance and therefore captures the trend of the spectrogram well. The second PC which also has a relatively high variance still contains valuable information about the signal. Therefore, it is also able to capture the trend of the spectrogram, but its information is quite diminished relative to the first PC. This trend continues until you get to the fourth PC that is not at all able to capture the trend of the original spectrogram because it adds minimal valuable information.

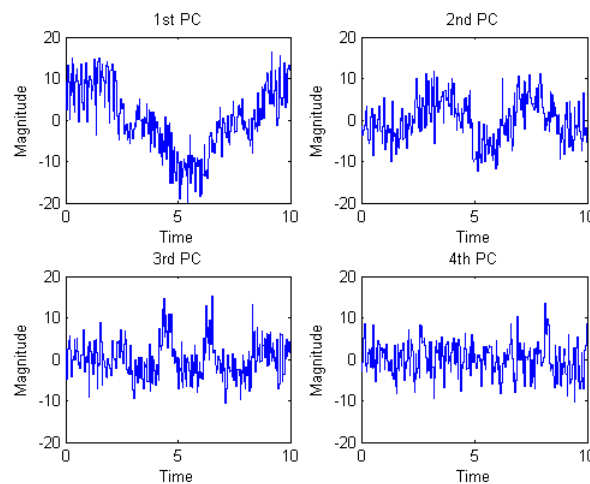


Figure 2.6: First four PCs of the spectrogram in Figure 2.4

2.3.3. Classification of Instantaneous Operating Conditions using HMM

Once the operating condition features have been extracted from the measured vibration signal, the next step is to send the features to the hidden Markov model (HMM) for classification. The reason that the feature extraction and statistical modelling of the operating conditions and faults are separated, is to ensure that the operating condition identification is not influenced by the presence of a fault. The HMM only operates with the operating condition sensitive features and solely determines the operating condition. The instantaneous operating condition is represented by the identified hidden state of the HMM.

The implemented HMM is based on the comprehensive report by Rabiner (Rabiner, 1989). There are two basic forms of HMMs, namely discrete and continuous probability distributions for the observations. The continuous HMM will be implemented in this algorithm because of its ability to

handle an infinite variety of observation vectors within the observation sample space, thus avoiding the need for quantising of the observation vectors.

The HMMs are comprised of N hidden states (S_1, S_2, \dots, S_N), M observation densities (V_1, V_2, \dots, V_M), state transition probability distribution ($a_{1:N;1:N}$), observation probability distribution ($b_{1:N}(\mathbf{O})$) and initial state probability distributions ($\pi_{1:N}$). A schematic view of a HMM displaying how everything interconnects can be seen in Figure 2.7. For convenience the HMM with its three probability measures can be written in a compact form.

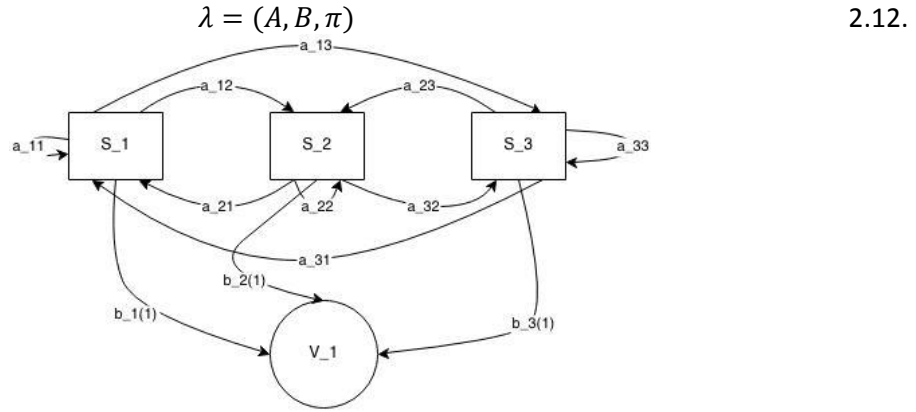


Figure 2.7: Schematic view of HMM with 3 hidden states, a single observation density and the interconnecting probabilities

With all the parameters and probability measures, the HMM is able to generate a sequence of T observations ($O_1 O_2 \dots O_T$), which each have a corresponding hidden state ($q_1 q_2 \dots q_T$). The state transition distribution (A) defines the probability of one hidden state transferring to another over time.

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N \quad 2.13.$$

The observation probability distribution (B) defines the probability of the observation with respect to the hidden state. For discrete observations, $b_j(k)$ is comprised of a matrix of probabilities.

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j] \quad 1 \leq j \leq N; 1 \leq k \leq M \quad 2.14.$$

However in this application the observations are from a continuous probability distribution (e.g. Gaussian), thus the probability $b_j(k)$ is approximated by a multivariate mixture model that consists of normal distributions with a set of means ($\mu_{1:N;1:M}$), covariance matrices ($\Sigma_{1:N;1:M}$) and mixture coefficients ($c_{1:N;1:M}$). This is a very flexible approach that can model complex density distributions if sufficiently many normal components (N) are used.

$$b_j(\mathbf{O}) = \sum_{m=1}^M c_{jm} P[\mathbf{O}, \mu_{jm}, \Sigma_{jm}] \quad 1 \leq j \leq N \quad 2.15.$$

The initial state probabilities (π) define the initial probabilities of the hidden states.

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N \quad 2.16.$$

The three basic challenges with respect to HMMs are evaluation, decoding and learning. Evaluation is the problem of computing the probability ($P(O|\lambda)$) of a given observation sequence (O_1, O_2, \dots, O_T) with respect to a given model. The second problem, decoding, is determining the optimum hidden

state sequence that corresponds to a given observation sequence. The third problem is training a HMM (adjusting λ) with a given set of observations sequences in order to maximise $P(O|\lambda)$.

The first obstacle of HMMs is evaluating the probability of a given sequence, which is overcome using the Forward-Backward Procedure. This procedure is comprised of three primary steps, which are as follows: initialization, induction and termination. This procedure operates with the forward variable ($\alpha_t(i)$), which is the probability of the partial observation sequence up until time t and for state S_i .

Unlike the solution for evaluating the probability of a given observation sequence, there are multiple methods of determining the optimum hidden state sequence that corresponds to a given observation sequence. There are so many approaches because of the variety of definitions and interpretations of what is optimum. Some approaches tend to select states which are individually most likely or sequence of states which are most correct. However, the most commonly used and computationally stable approach is the Viterbi Algorithm, which determines the hidden state sequence that maximises the probability of the given observation sequence. The Viterbi Algorithm is critical to the success of the proposed novel discrepancy method, because it determines the sequence of instantaneous operating conditions.

The most difficult challenge of HMMs is adjusting the model parameters (λ) to maximise the probability of a given observation sequence. There is no analytical solution to training the model, and, in fact, there is no optimal method of estimating the model parameters. However, using the Baum-Welch method (otherwise known as expectation-modification (EM)), the model parameters can be approximated by iteratively determining the localised maximum of the probability of the given observation sequence. This stage is equally crucial to the success of the proposed algorithm because the model needs to accurately portray the healthy system. The training process identifies the N (number of hidden states) distinct operating conditions and then clusters the operating condition features accordingly. The HMM is trained purely on data from a healthy system.

It is necessary to mention the importance of the transition probabilities of the HMM. The transition probabilities capture the dynamic properties of the operating conditions, which is why the HMM is used and not a clustering algorithm or just a set of GMMs. The transition probabilities contain the probabilities of the system transferring from one operating condition to another. Essentially the transition probability matrix should be diagonally dominant, which means that the system has a higher chance of remaining in its current operating conditions than transferring to another operating condition. This is to prevent unnecessary jumps between operating conditions. Also the transition probabilities capture the system's dynamics by favouring to transfer to one operating condition over another. This is desirable because in many industrial application machines operating under a constant sequence of operating conditions. Eggers, et al. who investigated a dragline gearbox mentioned the common operating cycle consisted of dropping the bucket and then dragging it through the ground (Eggers, Heyns, & Stander, 2007). Therefore the transition probabilities of the HMM play a crucial role in the capturing the operating condition dynamics.

2.4. Calculate Instantaneous Shaft Displacement

Once the sequence of the operating conditions has been determined by the HMM it is necessary to estimate the instantaneous shaft displacement. Unlike the operating condition classification, the windowing scheme for fault detection is critical to the performance of the methodology because it

determines the location of the detected fault. It is also necessary to calculate the shaft displacement, because of the assumption that there is no physical measure of the shaft displacement (i.e. no tachometer). The method implemented to track the shaft displacement is a combined version of the two approaches discussed in Section 1.2.2.10. It combines the advantages of both methods allowing for large speed fluctuations and a high level of accuracy. The basic approach of the order tracking scheme implemented can be seen in the schematic below. The instantaneous frequency is estimated from the STFT using maxima tracking and then the original vibration signal is passed through a Vold-Kalman filter (VKF) centred on the instantaneous frequency. The instantaneous phase can be calculated by applying the Hilbert transform to the filtered signal.

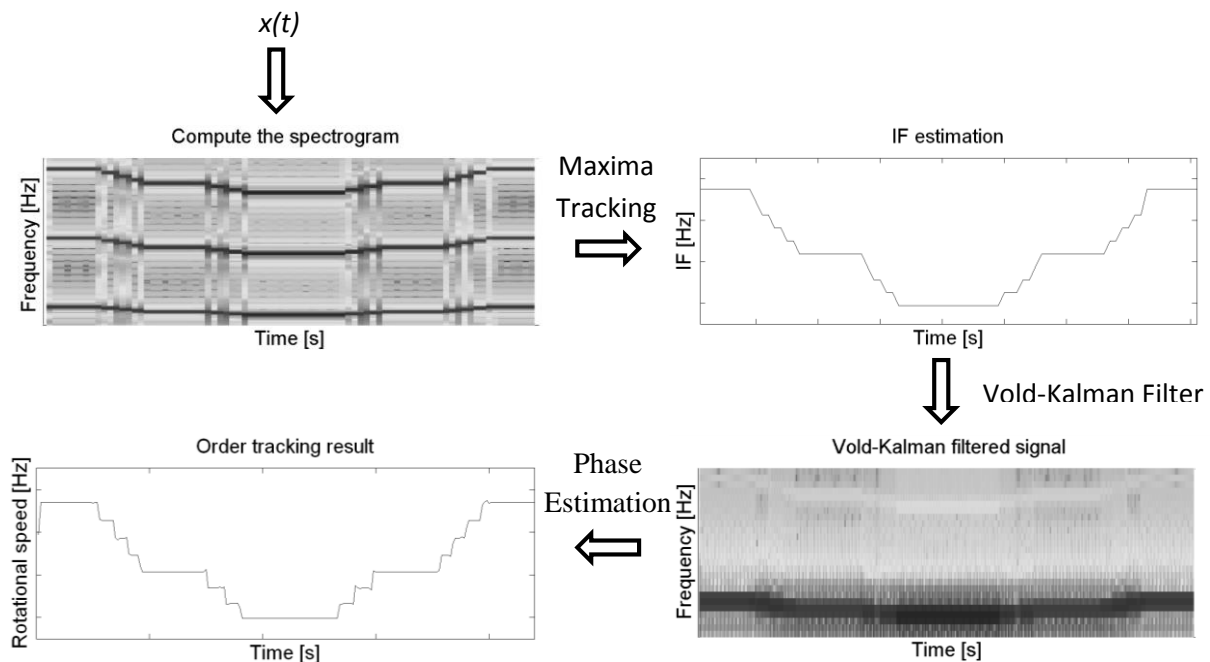


Figure 2.8: Schematic of order tracking approach

2.4.1. IF Estimation – Maxima Tracking

The first step is to estimate the IF from the spectrogram. This is accomplished by implementing the maxima tracking algorithm used by Urbanek et al. The algorithm can be found in the equation below. Urbanek only uses the previous location of the maximum to guide where the next maximum can be found. However, in our case, we also have the results of the HMM. It is possible to estimate the meshing frequency of each of the hidden states and therefore it is possible to use that frequency as a guide to where the maximum should be found. So the location of the maximum frequency in the STFT is guided not only by the previous frequency but also by the estimated frequency of the instantaneous operating condition. As mentioned in the literature study it is better to implement the maxima tracking algorithm on a low harmonic of the meshing frequency that experiences less fluctuation, even though it may be less prominent.

$$f_{max}(t) = \text{Argmax}_f |X(t, f)|^2 \text{ for } f \in \Delta f_t \quad 2.17.$$

$$\Delta f_t \in \{f_{max}(t - d\tau) - \delta f, f_{max}(t - d\tau) + \delta f\}$$

The results of the maxima tracking algorithm guided by the results of the HMM can be seen in the figure below, where the IF and its fourth harmonic have been superimposed (in blue) over the spectrogram.

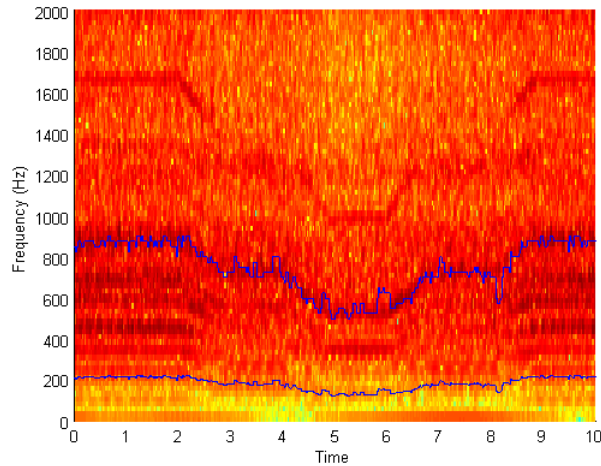


Figure 2.9: Maxima tracking results superimposed over spectrogram

2.4.2. Vold-Kalman Filter and Phase Estimation

The second step is based on Zhao’s et al. approach by applying the Vold-Kalman filter(VKF) to the original vibration signal. The VKF is a time-varying bandpass filter, where both the center frequency and bandwidth can be dynamically changed. Therefore the center frequency is defined by the IF and the bandwidth is defined by half the distance between harmonics. The spectrogram of the filtered signal can be seen in the figure below. It is clear how all frequency content outside the band has been removed and only the meshing frequency remains. Unlike the maxima tracking, it is preferred to implement the VKF on a high harmonic that experiences high fluctuation.

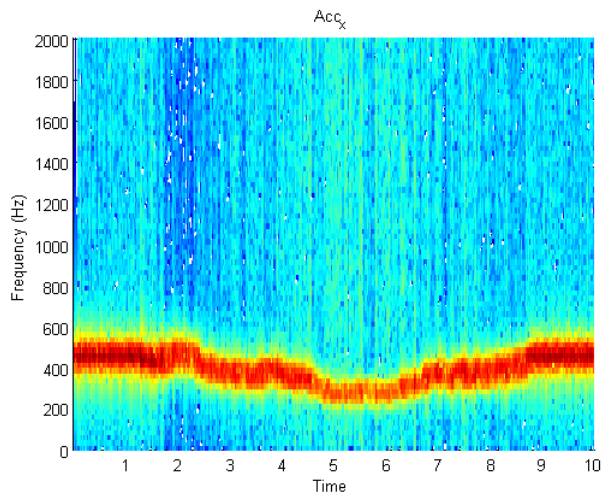


Figure 2.10: Spectrogram of signal after VKF

The instantaneous phase of the harmonic about which the signal was filtered can be calculated using the Hilbert transform and phase unwrapping. Simply dividing the phase by the number of the harmonic the instantaneous phase of the shaft can be calculated. The equation below displays the instantaneous phase calculation.

$$\phi(t) = \text{unwrap} \left[\arctan \left(\frac{\tilde{x}_{VKF}(t)}{x_{VKF}(t)} \right) \right] / k \quad 2.18.$$

Where $x_{VKF}(t)$ is the filtered signal, $\tilde{x}_{VKF}(t)$ is the Hilbert transform of the filtered signal and k is the number of the harmonic that the signal was filtered around.

It is speculated that the first estimation of the IF using maxima tracking, will not have sufficient accuracy to filter the signal and hence will result in poor phase estimation. Thus it is possible to re-estimate the IF from the calculated instantaneous phase and repeat the whole process using the improved version of IF and have improved accuracy.

2.5. Fault Detection

2.5.1. “Smart” Fault Feature Extraction

This dissertation defines the concept of “smart features” as features that are developed based on an engineering understanding of the data. In many machine learning applications (e.g. hand writing recognition) it is possible to iteratively try out numerous features and models and experimentally determine which feature and model configurations allow for the best classification of a test set. In this context it is however assumed that it is not possible to select optimum features for each specific rotating machine based on trial and error, because fault data is often not available. The smart features investigated in this paper aim to define rules that can be used to develop features for anomaly/change detection without first being able to test it. Ideally these features should be sensitive to rotating machine faults, fairly robust to fluctuating operating conditions, and computationally cost effective.

Smart frequencies or scales fulfil the aim of being able to efficiently and effectively detect both gear and bearing faults within a vibration signal. The popular characteristic fault frequency of gear’s is the meshing frequency since it is fairly simple to detect. However, it is easily affected by operating speed and is unable to capture bearing faults. Bearings on the other hand have three basic characteristic fault frequencies (BPFI, BPFO, BSF) that are not simple to detect. Thus a popular technique for detecting bearing faults is the high frequency resonance technique (HFRT), which detects high frequency impulses of the fault moving through the load zone. Therefore the smart frequencies can relate to the high frequency impulses caused by defects in gears or bearings while respectively meshed or under load.

The high frequency bands at which the impulses resonate are generally determined experimentally by comparing the high frequency spectra of the healthy and damaged gearbox. This approach is commonly found in most papers that implement the DWT to detect fault impulses such as the proposed methodology by Lou and Loparo (Lou & Loparo, 2004). However since one aim of the algorithm is to operate in the absence of damaged gearbox signals, it is necessary to determine an alternative approach to determine the frequencies at which these fault impulses resonate.

It is known that the fault impulses excite the natural resonance frequencies of the gearbox and the system it is in. Bozchalooi and Liang determined the fault impulse frequencies through a simple impact test (Bozchalooi & Liang, 2007). The fault impulse frequencies can be estimated from the resonance frequencies extracted from the FRF of the gearbox, while connected in the system. The FRF can be simply determined through an impact test, requiring the accelerometer that would be used to measure the vibration, modal hammer and spectrum analyser.

It was discovered during testing from the STFT plot of the healthy gearbox that there were high frequencies that dominated the vibration spectra of the gearbox which were independent of the

operating speed. An example of the spectrogram for the healthy gearbox from the experimental set up discussed in the next chapter can be seen in the figure below. It is proposed that these dominating frequencies bands can be used as an estimation of the natural frequencies of the system. The high frequency resonance technique can then extract the impulses generated by gear and bearing fault in these natural frequency bands.

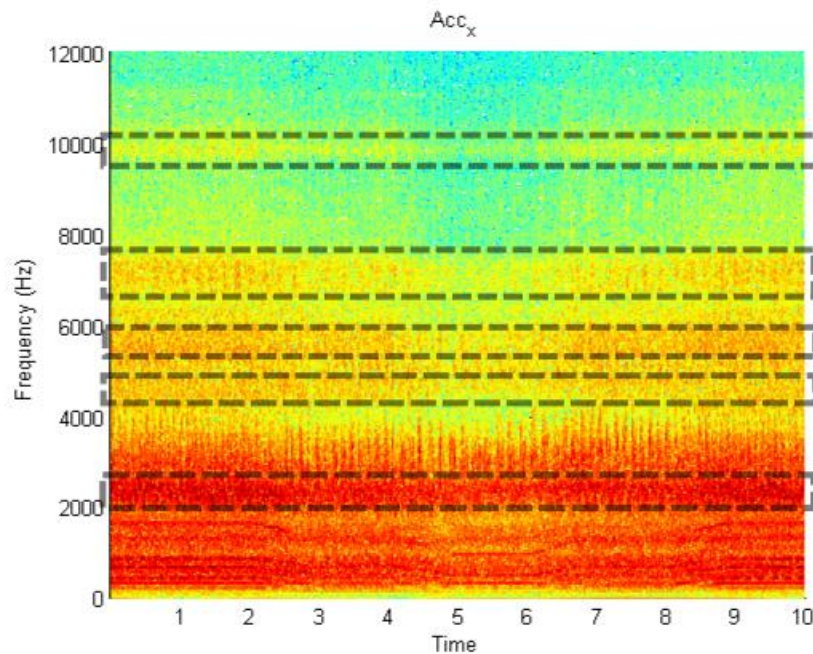


Figure 2.11: Example of STFT, with dominating frequencies bands highlighted

It is not recommended to determine the resonance frequencies of the gearbox mathematically by modelling the gearbox. The gearbox, as well as the system it is connected in, is an extremely complex system with many non-linearities, making the modelling process difficult. Thus it is not worthwhile modelling the gearbox since it is not guaranteed that the fault impulses will excite the calculated frequencies.

A final note with regard to smart frequencies, their respective scales can be easily estimated using the central frequency of the mother wavelet, as mentioned in the previous chapter. However, it is more advantageous to estimate a frequency band rather than a distinct frequency as it leaves room for error within the estimation process.

2.5.2. Multi-Resolution Representation

The time-frequency representation was implemented in the algorithm to satisfy the aim to be able to operate under non-stationary operating conditions. This representation allows the spectra of the signal to be evaluated with respect to time while also taking into account that the spectra does change with respect to operating conditions. Also, since only the high frequency region is being analysed, the coefficients should be unaffected by the meshing frequency and its harmonics. Finally, multi-resolution analysis is an appropriate way to remove noise because of the removal of certain frequency bands.

The discussion below goes through the evolution and development of the time-frequency representation to determine the optimum representation for the algorithm. The general progression is from the STFT to the CWT to the DWT and finally ending at the WPT.

The first time-frequency representation of interest is the short-time Fourier transform (STFT). It is extremely simple to implement and to evaluate because the results are similar to the results found from the common spectrum analysis. However the primary disadvantage of the STFT is the fixed window size, which controls the resolution along both the time and frequency axes. This disadvantage combined with the inability of the STFT to detect transient impulses, highlighted the fact that the STFT was not ideally suited to be implemented in this algorithm.

The next representation to be evaluated was the continuous wavelet transform (CWT). It is very similar to the STFT, except with the use of wavelets instead of sinusoids as the basis function. This allows it to overcome the problem of fixed window lengths. It also makes the CWT ideal for detecting discontinuities such as transient impulses. However the CWT evaluates the signal at discrete scales, which makes it challenging to extract the fault impulses at the resonant frequencies because their exact frequencies are unknown. Also the CWT evaluates a large range of scales which can be extremely computationally expensive and produce a large amount of redundant information. This is not ideal, specifically in the case of an accelerated life test where there is a large amount of data to be processed. Also since the CWT decomposes the signal in terms of scale and time, it is challenging to visually interpret the results.

Next the discrete wavelet transform (DWT) representation was evaluated. It requires significantly less computation time compared to the CWT, as well as not producing the large amount of redundant information. Also the DWT decomposes a signal within frequency bands, which is perfectly suited towards the smart frequencies because generally they are only experimental approximations. Thus the smart frequencies can be evaluated within a specified frequency resolution. However the DWT only decomposes the low approximation coefficients, thus there is poor frequency resolution at the high frequencies where the fault impulses are expected.

The final representation to be evaluated was the wavelet packet transform (WPT). While the above time-frequency representations all have their disadvantages, the WPT appears to be the perfect time-frequency representation for the algorithm. It uses wavelets which allow variable window lengths, which result in it effectively detecting impulses and it has good resolution in both the time and frequency domains. The WPT is similar to the DWT as it is not as computationally expensive as the CWT, as well as decomposing the signal into frequency bands and not distinct frequencies. This is ideal in a real world situation where exact natural frequencies are unknown. The WPT is an improvement of the DWT because it allows the details coefficients to be decomposed allowing improved frequency resolution within the high frequency sub-bands. Therefore the WPT is ideally suited for the algorithm because it can analyse all possible frequencies within a specific frequency bandwidths.

The primary design parameters of the WPT are the level of decomposition and the wavelet family. The level of decomposition is determined by the impulses frequencies and desired frequency resolution. As discussed above the fault frequencies can either be determined using a modal impact test or analysis of the STFT plot. The wavelet family is chosen with respect to their similarity to the impulses generated by both gear and bearing faults. Meyer wavelets are commonly used to detect bearing faults while Daubechies are generally used to detect gear faults. Rafiee et al. determined that the

Daubechies 44 was the perfect wavelet family for detecting both gears and bearing fault impulses (Rafiee, Rafiee, & Tse, 2010). The Morlet wavelet family is also commonly used to detect local damage but, the wavelet is not orthogonal so it can not be used in conjunction with the DWT or WPT. However from studying the literature pertaining to the wavelet transform, the most commonly used wavelet is the Daubechies 4. Therefore in this dissertation the Daubechies 4 wavelet family will be used.

The WPT is defined by the same equation as the DWT, the only difference being that the details coefficients are also further decomposed. The basic equation can be seen below.

$$WPT(j, k) = \frac{1}{\sqrt{2^j}} \int x(t) \psi^* \left(\frac{t - k2^j}{2^j} \right) dt \quad 2.19.$$

Where $x(t)$ is the time domain signal, ψ^* is the complex conjugate of the mother wavelet and j and k are the scale (level of decomposition) and time indices respectively. However Mallat proposed the pair of low-pass and high-pass filters denoted by $h(t)$ and $g(k) = (-1)^k h(1 - k)$ respectively. These filters are constructed from the mother wavelet and its corresponding scaling function ($\phi(t)$), they are expressed as.

$$u_{2n}(t) = \sqrt{2} \sum_k h(k) u_n(2t - k) \quad 2.20.$$

$$u_{2n+1}(t) = \sqrt{2} \sum_k g(k) u_n(2t - k) \quad 2.21.$$

Where $u_0(t) = \phi(t)$, $u_1(t) = \psi(t)$ and n is the sub-band number. Therefore using the above two wavelet filters, the signal is decomposed into a set of low and high frequency coefficients.

$$d_{j+1,2n} = \sum_m h(m - 2k) d_{j,n} \quad 2.22.$$

$$d_{j+1,2n+1} = \sum_m g(m - 2k) d_{j,n} \quad 2.23.$$

Where $d_{j,n}$ denotes the wavelet coefficients at the j^{th} level and n^{th} sub-band.

2.5.3. COT and Windowing

For the algorithm to be able to diagnose the type and location of the fault, the algorithm must be able to locate the angular position of the features with respect to either a bearing or gear meshing pair of interest. Thus a form of COT will be implemented to produce features vectors in the angular domain and not in the time domain. This section is critical to be able to identify the type of fault detected. Also it is necessary to have an accurate calculation of the instantaneous shaft displacement to correctly identify the fault, hence Section 2.4.

The first step is to decompose the time domain signal into epochs or windows of equal angular distance, using the calculated instantaneous phase of the shaft in Section 2.4. This results in windows that are of unequal size in the time domain. The angle size of the windows may vary depending on whether gears or bearings are being investigated, but it is possible for the window lengths to be equal for both gears and bearings. However, it must be ensured that the number of epochs can capture individual high frequency fault impulses.

The angle size of the window for gears will be the size of a single tooth pass, for a 23 tooth gear the window size will be $360^{\circ}/23 = 15.65^{\circ}$. This windowing scheme which is identical to one implemented for the operating condition and is illustrated in the figure below. Therefore each window

will relate to an individual gear tooth. For bearings there is no definite angle size. It is recommended to have sufficient epochs per revolution to be able to detect the characteristic fault frequencies but not too many such that the windows are too small to pick up an entire fault impulse as it decays over time.

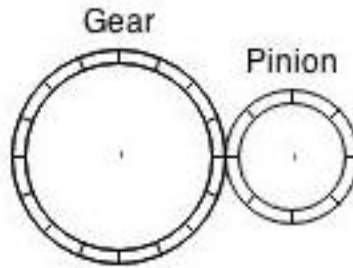


Figure 2.12: Fixed angular windowing scheme

2.5.4. Extraction of Fault Sensitive Features for GMM

So the question remains as to what fault features are extracted from the wavelet coefficients in the different frequency sub-bands and in each window. These are the features that will be interpreted by the GMM in order to determine the discrepancy signal. It must be noted that while the windows are of equal angular rotation distance, they are of different time lengths. Features that do not take this into account will produce inaccurate results. The instantaneous feature/s extracted from the coefficients must quantify the presence and magnitude of a fault impulse, independent of the length of the window.

There are many examples of features that can be extracted from the wavelet coefficients. Fairly often the coefficients themselves can be used as features. It is also popular to implement statistical measures such as kurtosis. However to take into account the varying window length and the need to quantify the presence of a fault impulse, the simple RMS will be extracted from the coefficients. The RMS is also known as the energy of the coefficients and is ideally suited to capture the energy increase due to the presence of a fault impulse. The RMS of a set of values can be calculated by the following equation.

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad 2.24.$$

2.5.5. Generating the Discrepancy Signal using the Fault Sensitive Features and the Operating Condition Specific GMM

A Gaussian mixture model (GMM) takes as input fault features and determines the discrepancy of features from the healthy case. The GMMs are implemented subsequent to the HMM. The HMM determines the state which best represent the combination of instantaneous operating conditions. The most likely state from the HMM is then used to select the GMM which best represent the baseline healthy behaviour of the vibration signal for the specific operating conditions. In other words there are many GMMs. Each GMM is matched to one operating state. The GMM are trained to represent the expected fault feature values for when the machine is in good condition. The GMM can now be used to detect when the fault sensitive features deviate from the expected norm in a way that is robust

to fluctuating operating conditions. The basic relationship between the HMM and set of GMMs can be seen in Figure 2.13 .

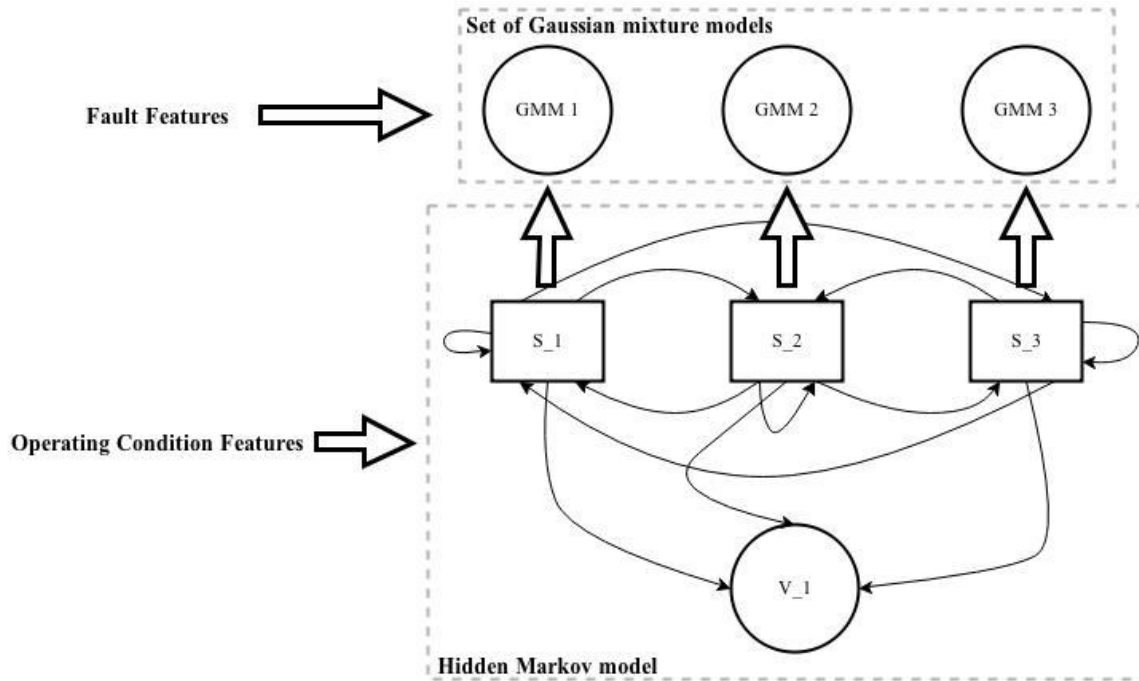


Figure 2.13: Relationship between HMM and set of GMMs

The GMM is trained on purely healthy data for each of the identified operating conditions using the Expectation Maximisation (EM) algorithm. The algorithm is comprised of two primary steps, namely expectation (E-step) and maximisation (M-step). The E-step guesses the probability distribution over the missing data with respect to the current model. Then the M-step then re-estimates the model parameters using the missing data. This process is repeated until the model parameters reach the desired level of convergence. As a Gaussian model the parameters that define it are means (μ), covariance's (Σ) and mixture coefficients.

The discrepancy of the extracted fault features is calculated in the form of the Negative Log-Likelihood (NLL). The NLL is simply the negative logarithm of the likelihood(probability) of the extracted features generated by the GMM. The NLL is ideal as the measure of discrepancy because of its ability to capture the deviation of the measured signals from the model trained with purely healthy data. The NLL for a Gaussian distribution with a single mixture is calculated with the following equation:

$$P(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad 2.25.$$

Where n is the number of dimensions of the extracted feature (x) and T is the transpose of the vector.

2.6.Account for Shaft Displacement Calculation Error

It is expected that the order tracking approach is not completely accurate because the phase demodulation approach results in cumulative error over time. Thus it is highly possible that the same fault can be detected in different locations with each successive rotation. This is visible in the figure below where the location of the maximum NLL value changes for many successive revolutions. However, it can be assumed that if the location of the fault has moved only a single gear tooth in

either the forward or backward direction it is, in fact, the same fault and the location change is due to an order tracking error. The error can easily be erased by rotating the discrepancy signal for that revolution and all subsequent revolutions, such that the fault locations align. In the case of a 42 tooth gear, looking at a single gear tooth forward and backwards accounts for only a 4.75% error in the order tracking method, which is within reason. Thus the realigning does not misrepresent the information, but does account for the possible order tracking error.

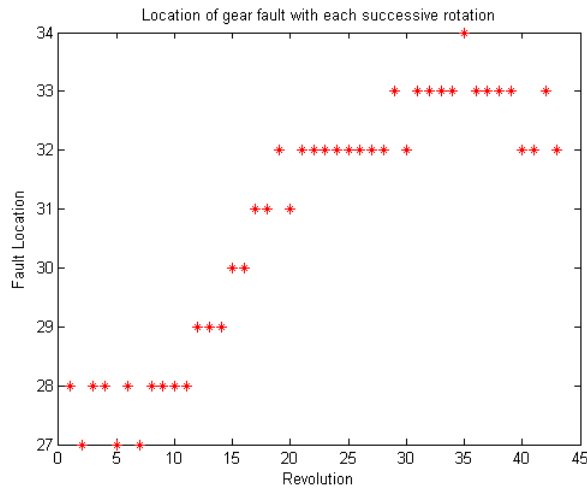


Figure 2.14: Location of maximum NLL for successive revolutions

2.7. Discrepancy Analysis

2.7.1. Synchronous Averaging

The first and simplest signal processing applied to the discrepancy signal to detect and classify faults is synchronous averaging (SA). Since the discrepancy signal is already in the angular domain, it is not necessary to again implement COT. Thus the SA is simply calculated by averaging the discrepancy signal over a single revolution. This effectively removes noise in the signal, while at the same time highlighting the repetitive discrepancies. It is expected that the simple SA will be able to effectively detect general wear as well as damaged gear teeth. This is because it will detect the uniform deviation from the healthy model over the full revolution; however, it is not able to locate which component is suffering from wear. The SA is defined by the equation below.

$$\bar{x}_i^g = \frac{1}{r} \sum_{n=1}^r x_{n \times N_g} \quad \text{for } i = 1:N_g \quad 2.26.$$

Where r is the number of revolution of the gear, x is in our case the discrepancy signal and N_g is the number of gear teeth.

2.7.2. Spectral Analysis

The second step for processing the discrepancy signal is applying the Fast Fourier Transform (FFT), which is defined by the equation below. The FFT is commonly used to detect faults in rotating machines, but as mentioned above, it is unable to handle non-stationary operating conditions. However, it is fully capable of detecting faults in the discrepancy signal which is unaffected by the change in operating conditions. The FFT is able to detect periodicity within the discrepancy signal,

which may not be synchronous with the shaft's rotational speed (e.g. bearing faults), thus is undetectable using the SA.

$$X(f) = \int_0^T x(t)e^{-2\pi itf} dt \quad 2.27.$$

2.7.3. Cepstral Analysis

The final analysis technique is cepstral analysis, which essentially looks for periodicity within the spectrum and is defined by the equation below. This is ideal for our situation where the spectrum is dominated by not only the fault frequency but also many of its harmonics and sidebands, thus making it difficult to interpret especially in the early stages of the fault. Therefore by grouping the sidebands and harmonics into families the early stages of the fault are significantly easier to detect.

$$C_p = |F^{-1}\{\log(|F\{x(t)\}|^2)\}|^2 \quad 2.28.$$

Where $F\{\}$ if the Fourier transform as defined in equation 2.19.

3. Dynamic response of rotating machinery under non-stationary operating conditions

3.1. Introduction

In order to validate the novel discrepancy method proposed in the previous chapter it is necessary to attain data from both healthy and damaged rotating machines that operated under non-stationary conditions. The two possible options for generating data are from a simulated model or a physical experiment. Simulated dynamic models are a cost effective way of generating data, however there is a measure of uncertainty that the models are able to accurately replicate physical systems. Physical experiments are ideal for generating data because of their close similarity to the industrial applications, however they can be costly and time consuming to undertake.

The simulated dynamic model was used during the development of the proposed discrepancy method, because of its ease of implementation and its ability to represent, in a general sense, the expected vibration signal. The model takes the form of a lumped mass system, and the dynamic response of the system can be calculated by solving the set of differential equations of the system.

The experimental test setup is the final verification step of the propose discrepancy method. It is designed to measure the dynamic response of a rotating machine under fluctuating operating conditions. The rotating machine is a single stage helical gears box with single row ball bearings supporting the shafts. The fluctuating operating conditions are a cyclic repetition of a finitely long period of controlled fluctuating speed and load. The fluctuation period comprises of 3 primary operating conditions defined by speed and load and the transfers between them. The experimentation procedure will take the form of an accelerated life test, from initial fault occurrence to final failure.

3.2. Dynamic Gearbox Model

There have been many dynamic gearbox models developed in the last decade to generate expected vibration signals from gearboxes in both healthy and damaged states. The model implemented to verify the effectiveness of the proposed method is the lumped mass model proposed by Chaari et al (Chaari, Bartelmus, Zimroz, Fakhfakh, & Haddar, 2012). This model was selected because of its ability to accurately simulate the effect of large fluctuations in load and speed on a gearbox. It is able to do this because the model takes into account the influence of varying loads on the gear mesh stiffness.

3.2.1. Lumped Mass Model

The model is of a single stage spur gear, incorporating a synchronous induction motor (11), pinion (12), gear (21) and machine load (22) as can be seen in Figure 3.1. The model has a total of 8 degrees of freedom (DOFs), the pinion and gear each have two translation DOFs and one rotational DOF while the motor and machine each have a single rotational DOF. It is assumed that the pinion and gear and their teeth are rigid, thus no tooth deflection occurs during meshing. The shafts between the motor and pinion as well as the gear and machine each have a torsional stiffness of K_θ . The rolling element bearings supporting the pinion and gears, which allow the two translational DOFs, both have a linear stiffness in the x- (K_x) and y-direction (K_y).

The meshing of the pinion and gear is modelled as a time varying stiffness $K_g(t)$, and is a function of the position of the pinion position, θ_{12} , and the displacement of the line of action, δ . The displacement of the line of action can be expressed as follows.

$$\delta(t) = (x_1 - x_2) \sin(\alpha) + (y_1 - y_2) \cos(\alpha) + \theta_{12}r_{b12} + \theta_{21}r_{b21} \quad 3.29.$$

Where x_i and y_i are the translations of the bearing i ($i=1,2$). θ_{ij} and r_{bij} is the angular displacement and base radius of the pinion and gear respectively, while α is the pressure angle.

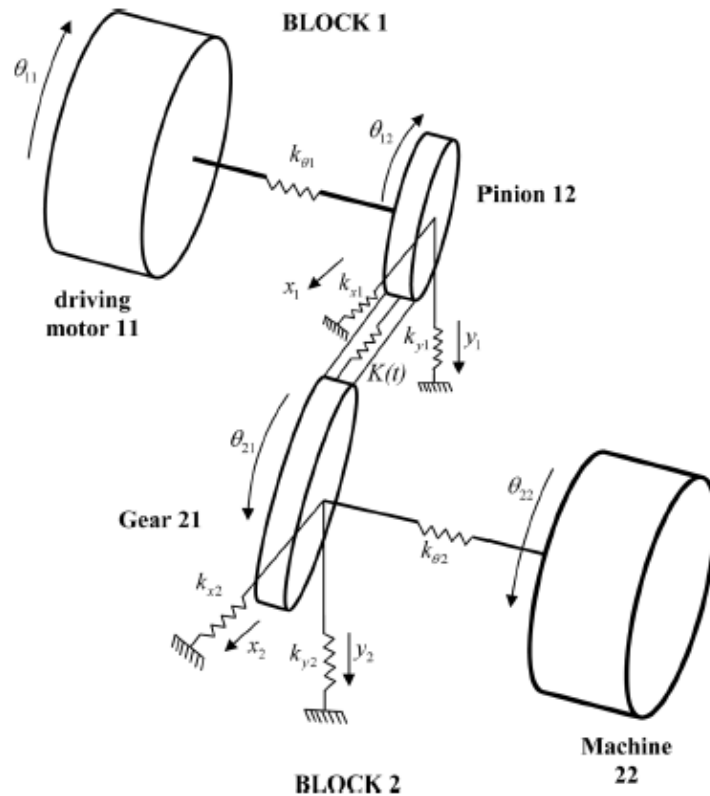


Figure 3.1: Single stage spur gear lumped mass model (Chaari et al., 2012)

The load on the system is controlled by the load created by the machine T_L , while the rotational speed of the gears is also dictated by the machine load because the torque required from the synchronous induction motor, T_M , varies with its speed. The speed torque characteristics of the motor are defined by the physical characteristics of the motor and create a relationship that can be seen in Figure 3.2. The motor has an inversely proportional relationship between load and speed, thus when the machine load increases the shaft speed decreases and vice versa.

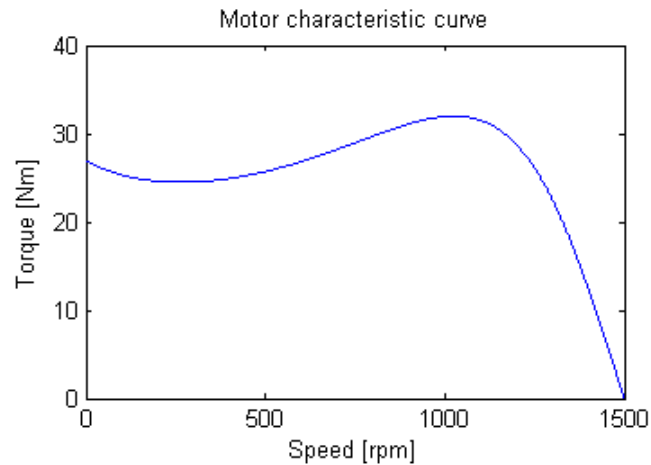


Figure 3.2: Speed-Torque characteristics

The time varying gear mesh stiffness is defined by a maximum stiffness, K_{gmax} , when two teeth pairs are in contact and a minimum stiffness, K_{gmin} , when only a single pair of teeth are in contact. The time period of a single set of teeth in a contact and a pair is defined by the contact ratio of the gears c , and gear mesh period T_g which is determined by the following equation.

$$T_g = \frac{60}{n_{r1}Z_1} \quad 3.30.$$

Where n_{r1} and Z_1 is the rotational speed and number of teeth of the pinion gear respectively. A gear fault is simply modelled as a reduction of the meshing stiffness of a single tooth. Chaari et al. experimented modelling various gear faults by altering the gear mesh stiffness and as a result gives appropriate advice for simulating gear faults (Chaari, Baccar, Abbas, & Haddar, 2008). They calculated the stiffness of an individual tooth based on deflection due to bending, contact and fillet foundation deflections. Then the gearmesh frequency can be generated as a combination of the individual teeth stiffness as they mesh with one another. A gear tooth crack was modelled by removing material, which results in decreasing the contact surface between teeth. They calculated that small cracks can cause up to a 5% decrease in the gearmesh stiffness of the damaged gear tooth. The gear mesh stiffness for a healthy and damaged gear can be seen in the figure below.

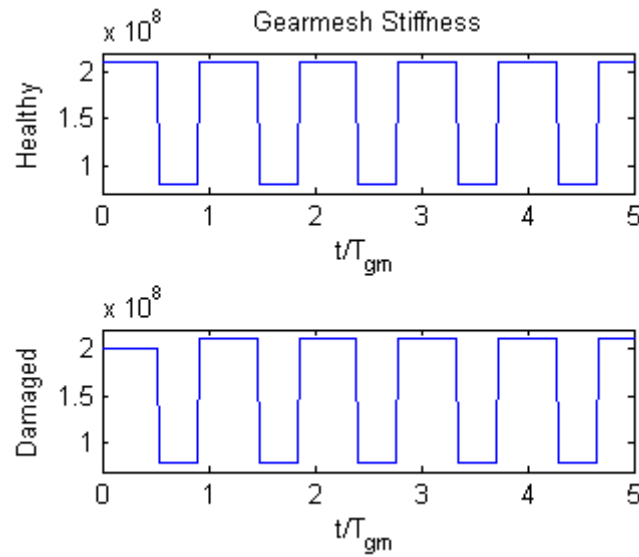


Figure 3.3: Time varying gear mesh stiffness for a healthy and damaged gear

The equation of motion of the entire system is controlled by the time varying machine load and gear mesh stiffness and can be developed by applying the Lagrangian formulation and can be seen in the equation below.

$$M\ddot{q} + C\dot{q} + K(t)q = T \quad 3.31.$$

where $q = \{x_1; y_1; \theta_{11}; \theta_{12}; x_2; y_2; \theta_{21}; \theta_{22}\}$ is the vector containing the 8 DOFs of the system and $T = \{0; 0; T_M; 0; 0; 0; 0; T_L\}$ is the applied torques on the model. M , K and C are the mass, stiffness and damping matrices of the systems and will be defined below. The system is an unstable set of ordinary differential equations (ODEs), and as a result regular ODE solvers such as Runge-Kutta prove incapable of solving them. It is necessary, therefore, to solve it using the implicit Newmark algorithm because by altering the simulation parameters it can be unconditionally stable and able to solve the system with sufficient accuracy. m is defined by the diagonal matrix below.

$$m = \begin{bmatrix} m_1 & & & & & & & 0 \\ & m_1 & & & & & & \\ & & I_{11} & & & & & \\ & & & I_{12} & & & & \\ & & & & m_2 & & & \\ & & & & & m_2 & & \\ & & & & & & I_{21} & \\ 0 & & & & & & & I_{22} \end{bmatrix} \quad 3.32.$$

where m_i is the mass of the bearing. I_{ij} is the mass moment of inertia of the respective components computed by considering them as solid disks.

The K matrix for the entire model is expressed below. It incorporates the constant stiffness's of the bearings and shaft as well as the time varying gear mesh stiffness.

$$\begin{aligned}
 & K(t) \\
 = & \begin{bmatrix}
 s_3 K_g(t) + k_{x1} & s_5 K_g(t) & 0 & s_7 K_g(t) & -s_3 K_g(t) & -s_5 K_g(t) & 0 & s_9 K_g(t) \\
 s_5 K_g(t) & s_4 K_g(t) + k_{y1} & 0 & s_6 K_g(t) & -s_5 K_g(t) & -s_4 K_g(t) & 0 & s_8 K_g(t) \\
 0 & 0 & k_{\theta 1} & -k_{\theta 1} & 0 & 0 & 0 & 0 \\
 s_7 K_g(t) & s_6 K_g(t) & -k_{\theta 1} & k_{\theta 1} + s_{10} K_g(t) & -s_7 K_g(t) & -s_6 K_g(t) & 0 & s_{12} K_g(t) \\
 -s_3 K_g(t) & -s_5 K_g(t) & 0 & -s_7 K_g(t) & k_{x2} + s_3 K_g(t) & s_5 K_g(t) & 0 & -s_9 K_g(t) \\
 -s_5 K_g(t) & -s_4 K_g(t) & 0 & -s_6 K_g(t) & s_5 K_g(t) & k_{y2} + s_4 K_g(t) & 0 & s_8 K_g(t) \\
 0 & 0 & 0 & 0 & 0 & 0 & -k_{\theta 2} & -k_{\theta 2} \\
 s_9 K_g(t) & s_8 K_g(t) & 0 & s_{12} K_g(t) & -s_9 K_g(t) & s_8 K_g(t) & -k_{\theta 2} & k_{\theta 2} + s_{11} K_g(t)
 \end{bmatrix} \quad 3.33.
 \end{aligned}$$

The values for the coefficients s can be found in the table below. r_{b12} and r_{b21} is the base radius of the pinion and gear respectively.

Table 3.1: Coefficients of $K_g(t)$

s_1	$\sin(\alpha)$	s_7	$r_{b12} \sin(\alpha)$
s_2	$\cos(\alpha)$	s_8	$r_{b21} \cos(\alpha)$
s_3	$\sin^2(\alpha)$	s_9	$r_{b21} \sin(\alpha)$
s_4	$\cos^2(\alpha)$	s_{10}	r_{b12}^2
s_5	$\sin(\alpha) \cos(\alpha)$	s_{11}	r_{b21}^2
s_6	$r_{b12} \cos(\alpha)$	s_{12}	$r_{b12} r_{b21}$

The damping in the system is proportional and is therefore defined by the equation below.

$$C = 0.05m + 10^{-6}K_g \quad 3.34.$$

The motor, pinion and gear parameters that define the lumped mass model can be found in the table below

Table 3.2: Motor, pinion and gear characteristics

	Motor	
Electrical characteristics	4 pole, 50Hz, 3 phase, 415V	
Synchronous speed N_s [rpm]	1500	
Full load torque T_f [Nm]	10	
Ratio breakdown T_b/T_f	3.2	
Slip s_b	0.315	
Motor constant a	1.7111	
Motor constant b	1.316	
	Pinion	Gear
Teeth numbers Z	20	40
Mass m [kg]	0.6	1.5
Mass moment of inertia [kgm ²]	2.7×10^{-4}	0.0027
Diameter d [mm]	60	120
Pressure angle α [deg]	20°	
Contact ratio c	1.6	
Bearing Stiffness [N/m]	$k_{x1} = k_{y1} = k_{x2} = k_{y2} = 10^8$	
Torsional Stiffness [N rd/m]	$k_{\theta 1} = k_{\theta 2} = 10^5$	
Max and Min gear mesh stiffness [N/m]	$K_{gmin} = 0.8e^8; K_{gmax} = 2.1e^8$	

The fluctuating machine load can be modelled in a simpler form to what the experimental step up will experience. The primary difference, as will be seen later, is that the load and speed are inversely proportional to each other, while the experimental load and speed are directly proportional to one

another. The maximum and minimum loads experienced are 25 N.m and 10 N.m respectively. The resulting motor speed can be calculated from the load on the motor and torque speed characteristics, thus the corresponding speeds to the maximum and minimum load are 1267 rpm and 1420 rpm respectively. The machine load and motor speed can be seen in the figure below. These load cases may have a fairly slow rate of change and do not contain any impulsive fluctuations. This means they are not representative of all possible load fluctuations found in industry but they do capture the basic nature of varying operating conditions.

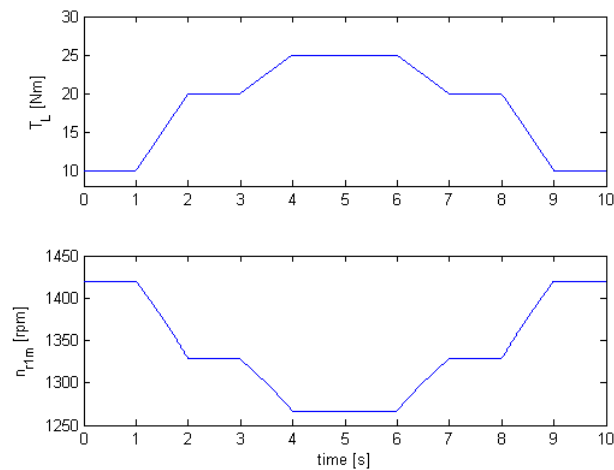


Figure 3.4: Machine load and respective motor speed

3.2.2. Time Domain Results

The time domain acceleration of the bearing near the pinion in the x-direction is used as the vibration signal of the gearbox. The healthy gear waveform as well as the damaged gear waveform can be seen in Figure 3.5. The fluctuating operating conditions have an obvious effect on the magnitude of the vibration signal. From simple observations of the waveform below it is evident that standard vibration monitoring techniques are inadequate to handle such large amplitude fluctuations created by fluctuating operating conditions as the fault conditions may be hidden and undetected. The damaged gear tooth fault is also evident from the once per revolution peak in the vibration signal. The amplitude of the impulse due to the gear fault varies significantly with respect to the operating conditions, thus increasing the difficulty to quantify the severity of the fault.

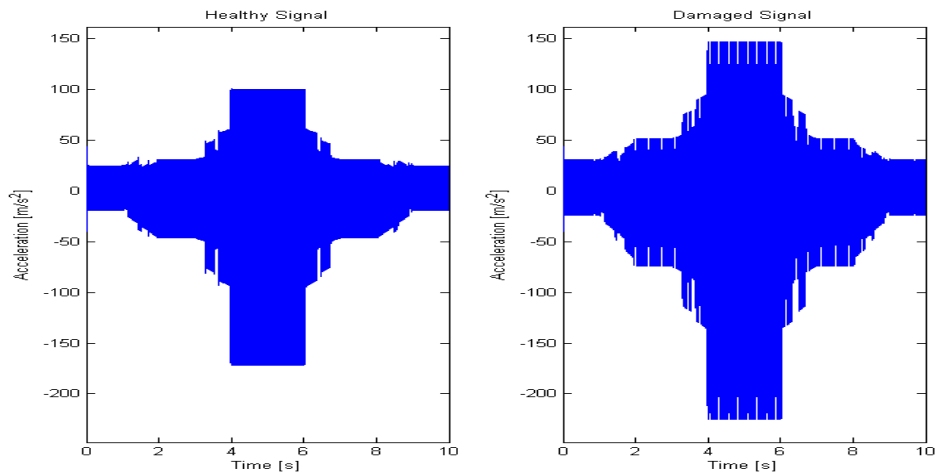


Figure 3.5: Time domain acceleration of healthy and damaged gearbox

3.3. Experiment Set-Up

The experimental setup used to gather experimental data to validate the proposed method was based on the setup created by Dr C.J. Stander (Standar, 2005). The experimental setup is comprised of three Flender helical gearboxes, a 5.5kW Weg electric motor and a 5.5kVA Mecc Alte alternator. The induction motor speed is controlled by an analog speed controller that receives an input signal from the user. In the same manner the alternator load is controlled by manipulating the electromagnetic field strength using a circuit diagram developed by C.J. Stander and a signal generated by the user. A photo and schematic of the experimental setup can be seen in Figure 3.6 and Figure 3.7 respectively.

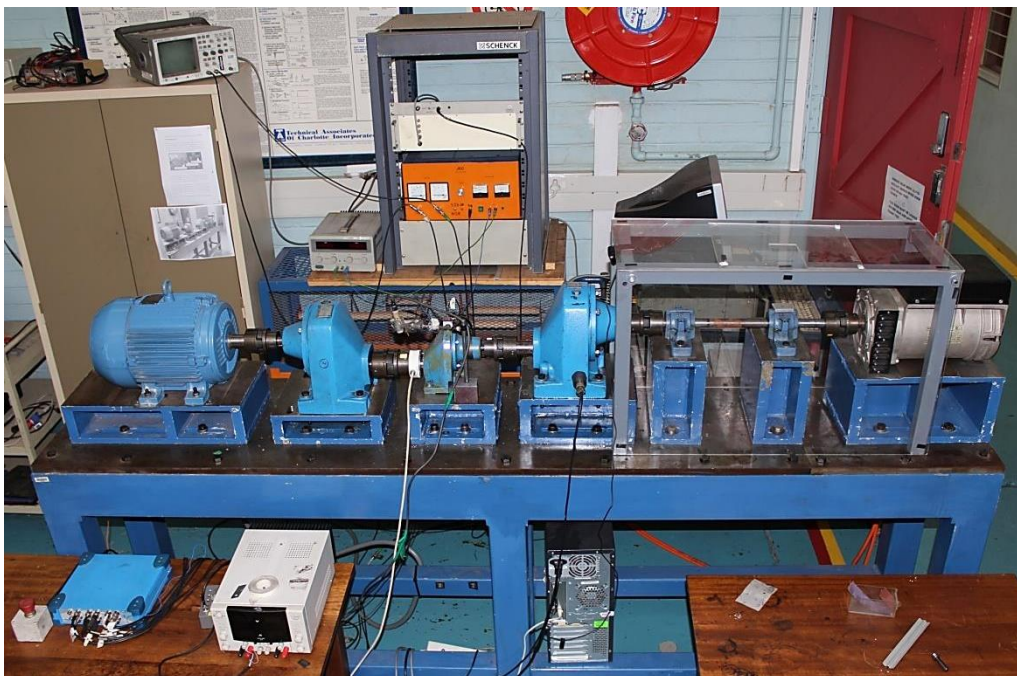


Figure 3.6: Experimental setup

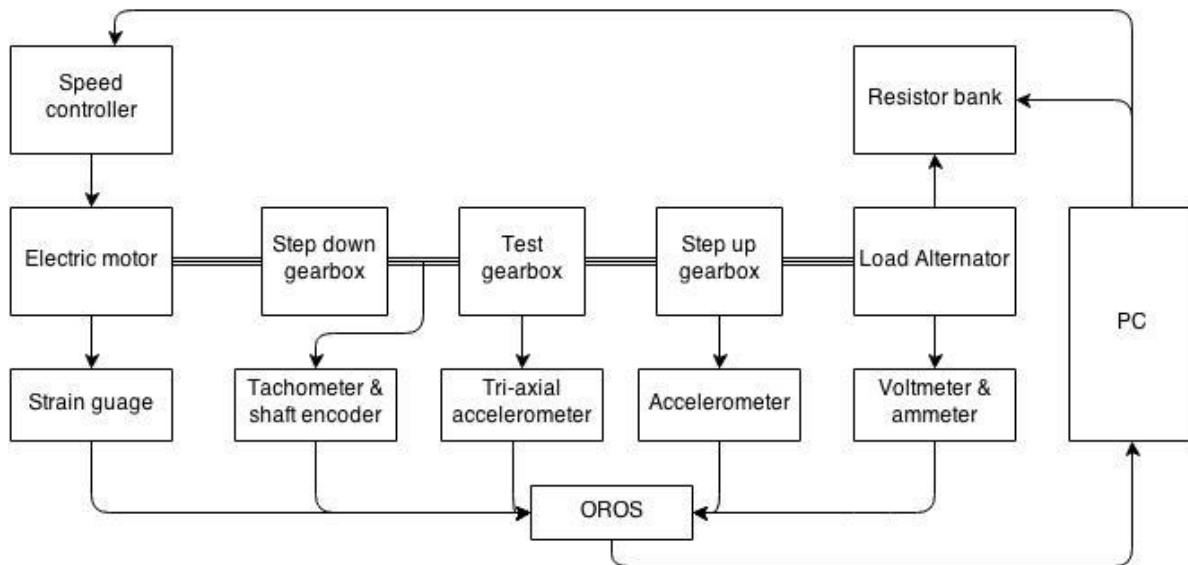


Figure 3.7: Schematic of experimental setup

3.3.1. Instrumentation

A comprehensive instrumentation setup is implemented to gain maximum information from the system. Not all data recorded will be used for this project, however, it will be available for future projects. All the measurements were recorded with an OROS data acquisition system in conjunction with a Personal Computer (PC).

Integrated Circuit Piezo (ICP) accelerometers are used to measure the vibration response of the gearbox casings. Accelerometers are placed on the test gearbox as well as the step-up gearbox on the output side of the test gearbox. The accelerometer on the test gearbox is a tri-axial accelerometer to determine its vibration along all three axes to have the best understanding of its response. All of the results are taken from the single accelerometer that is measuring acceleration in the vertical direction on the test gearbox. There is also an accelerometer on the step-up gearbox, which measures the response as if the gearbox is a multi-stage gearbox, which is more relevant to industrial applications. However it must be noted that there was a rubber coupling between the gearboxes, which would have a significant effect on the vibration transfer between the gearboxes. Hence only a uni-directional accelerometer was placed on the step-up gearboxes and only measures the vibration in the horizontal plane in a radial direction. It is critical for the accelerometers to have sufficient sensitivity to be able to accurately measure low amplitude response, which helical gearboxes are known for.

The proposed methodology is based on the assumption that there is no measure of the rotational displacement of the shaft. However, to validate the shaft displacement calculation it is necessary to measure the shaft displacement. Therefore a simple digital tachometer and accurate shaft encoder are used. The shaft encoder gives an accurate angular position of the rotating shaft due to it measuring 1024 pulses per revolution, while the tachometer simply measures a single pulse per revolution. Both devices measure on the identical shaft, thus allowing the instantaneous speed difference of the two devices to be a measure of the load on the shaft, again for validation purposes.

Other instruments installed to detect load change on the system are a strain gauge on the foot of the motor. Also the power output of the alternator was measured. The exact value of the torque (in N.m.)

on the system is easily calculated from the power output of the alternator and its instantaneous shaft speed. These instruments were used to measure the instantaneous load on the system and help to validate the results of the HMM, which identified the instantaneous operating conditions.

Table 3.3: Experiment instrumentation

Instrument	Specification	Location
Personal Computer	Acer TravelMate 6592	Adjacent to test bed
Data Acquisition System	OROS OR35	Adjacent to test bed
Tri-Axial Accelerometer	PCB 103mV/g Model 354B22	Test gearbox
Uni-directional Accelerometer	IMI 502mV/g Model 626A02	Step-up gearbox
Shaft Encoder	Hengstler Model RI 58-0	Input side of test gearbox
Tachometer	PROVA digital Model RM-1000	Input side of test gearbox
Strain Gauge	PCB 740B02	Motor foot
Thermal Camera	FLIR ThermoCam Model E65	Test gearbox
Laser Vibrometer	Polytec OFV056	Test gearbox

3.3.2. Data Acquisition

It is essential that the sampling frequency is able to capture the vibration components that are of interest to determine faults. The first frequency of interest is the meshing frequency of the test gearbox at maximum speed which is in the vicinity of 200Hz, but it is desirable to measure up to the fourth harmonic at 800Hz. However due to the Nyquist criteria and to prevent aliasing, it is necessary to measure at 3 times the desired frequency, which only amounts to a sampling frequency of 2.4 kHz. However, it is desirable to measure the high frequency resonance response of the bearing/gear housing due to impulses from defective gears and bearing which are generally in the range of 10-50kHz (McFadden & Smith, 1984). Therefore a significantly higher sampling frequency is required. A suitable sampling frequency and the maximum that is achievable and realistic is 25.6 kHz. Thus allowing a maximum observable frequency of 10 kHz, which should be able to capture some of the high resonance frequencies.





The period of measurement determines the resolution in the frequency domain for spectral analysis, thus it is necessary to use a period of 10s, allowing a minimum measured frequency of 100 mHz. Also the length of measurement determines the number of revolutions seen, which in turn determines the effectiveness of the SA. The measurement period is equal to the period of fluctuating operating conditions which is cyclically repeated. Measurements will be taken every half hour, which will give sufficient resolution in the time domain to track the severity of the fault as the components trends towards failure.

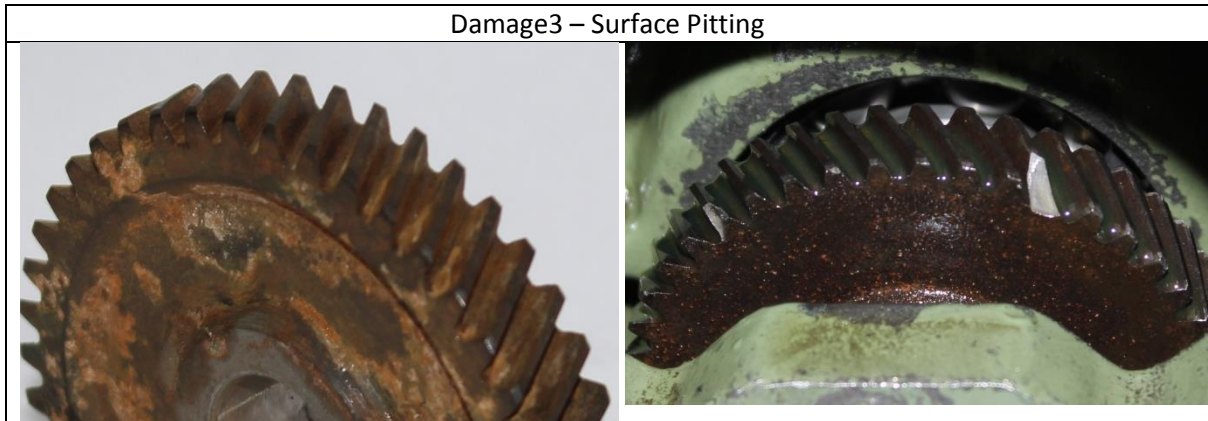
3.3.3. Seeded Faults

Once the experiment was set up, various faults were seeded into the machinery. The primary goal of the seeded faults was to best represent the faults that occur in industry. However, due to time constraints, only gear faults were investigated.

- **Tooth Breakage:** Tooth breakage is fairly self-explanatory. Two forms of tooth breakage will be investigated. The first one is when a crack develops at the root of the tooth and generally results in the entire tooth breaking off. The second one is when a crack forms on the meshing surface of the gear tooth and results in the gear tooth chipping. Both these faults were seeded by cutting minute lines in the respective positions on the gear tooth using wire erosion.
- **Wear and surface fatigue:** Wear is a common fault of gears, where material is gradually removed from the meshing surface due to contact between meshing gears. Surface fatigue occurs in similar manner due to the cyclic loading, but is generally not visible as it affects the molecular integrity of the surface. Wear and fatigue are found to be distributed over all the teeth on the gear, although wear is localised with respect to its position on the meshing surface of the gear. Both wear and surface fatigue manifest themselves in the form of material removal and pitting. This type of surface was created by leaving a gear in salt water for a month to degrade the meshing surface.

Table 3.4: Before and after photos of damaged gears

Before	After
Damage1 - Tooth root crack	
	
Damage2 - Tooth chip	
	



3.3.4. Load Cases

A critical component of the research of this thesis is the non-stationary operating conditions of the gearbox because fluctuating speed and loading conditions causes frequency and amplitude modulation in the vibration signal. Non-stationary operating conditions can be sub-divided into two classes; namely smooth and impulsive transitions between load cases. The smooth transitions allows the system to gradually reach the load cases and remain there briefly before changing to the next load case (temporarily stationary). The impulsive transitions consist of very rapid transitions between load cases and does not allow the system to reach stationary conditions. Due to the physical limitations of the induction motor, only smooth transitions were implemented in this experiment.

For the purpose of this project, it is necessary to measure the vibration response of the test gearbox over its entire lifespan, from inception to failure. This can be a very time consuming process, thus an accelerated life test can be used. Accelerated life tests do not replace long-term testing, but are good approximations. By simply operating the gearbox at loads above its design load, the design life of the gear can be dramatically reduced. The seeded faults discussed above also significantly reduce the operating life of the gears, however, accelerated life test methods are required to ensure ultimate failure of the component occurs within the allocated time.

The loads cases are defined by percentages of the rated load of the gearbox, 20 N.m, and the maximum speed of the electric motor, 1470 rpm. The loads cases, their percentages and their duration (in form of percentage) can be seen in the table below. Note that these operating conditions are similar in shape, but are quite different to those applied to the simulated model.

Table 3.5: Loads cases

Load Case No.	Gearbox Load		Motor Speed		Duration
	[%]	[N.m]	[%]	[rpm]	[%]
1	200	40	60	186.8	10
2	250	50	80	249.0	20
3	300	60	100	311.3	30
Trans. 1	Load case 1 → 2				20
Trans. 2	Load case 2 → 3				20

The above values are used to determine the speed of the motor and load of the alternator. The smooth transition can be seen in the figure below and the sequence of load cases are shown in Table 3.6. It

must be noted that only over the last 10s of the operating condition sequence will the measurement be taken.

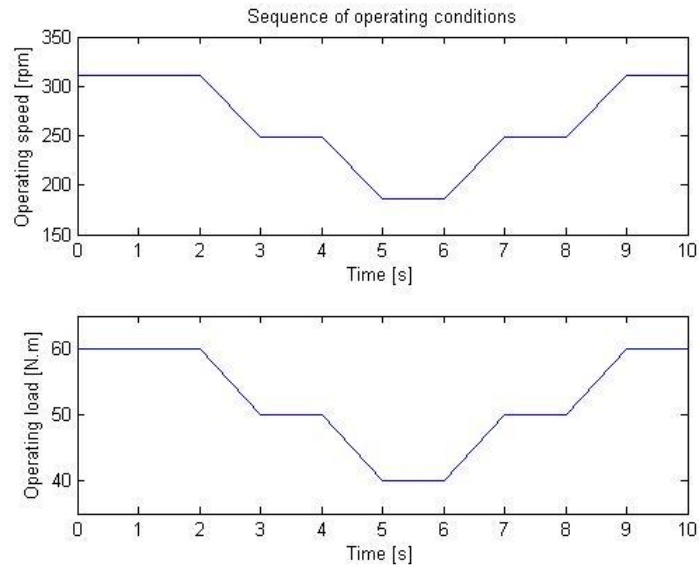


Figure 3.8: Illustrative view of the fluctuating operating conditions

Table 3.6: Sequence of load cases seen in Figure 3.8

Order	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th
Load Cases No.	3	3→2	2	2→1	1	1→2	2	2→3	3

3.3.5. Time Domain Results

The figure below is the vibration signal of the test gearbox along the 3 primary axes. The greatest vibration magnitude is measured along the x-axis, which corresponds to the axial direction of the gearbox. The fluctuating operating conditions have a clear effect on the vibration magnitude of the gearbox. From literature and the time domain results below, the x-axis (axial) vibration signal will be used because it is known to contain the most information.

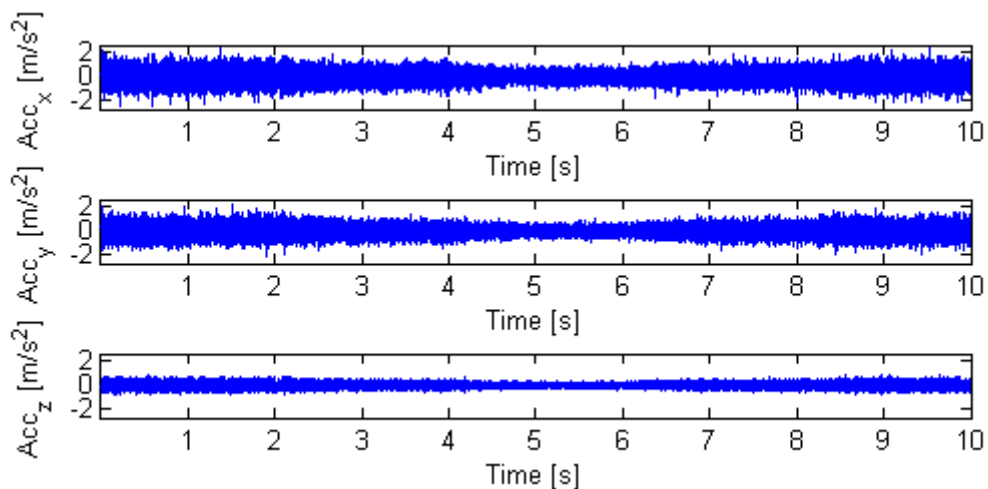


Figure 3.9: Time domain vibration of the test gearbox in all 3 directions

The figure below is the vibration signal of the step-up gearbox on the output side of the test gearbox. The vibration is measured in the y-direction, which is radial with respect to the gearbox. It is also affected by the fluctuating loads.

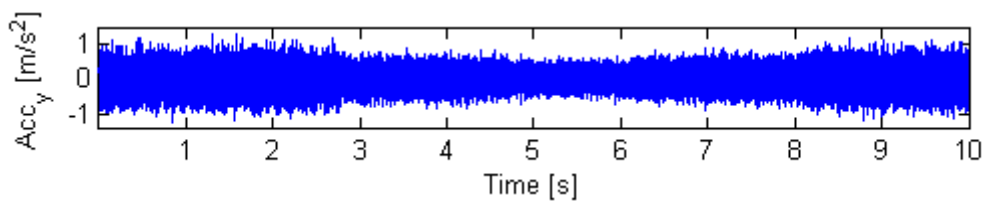


Figure 3.10: Time domain vibration of the step-up gearbox on output side

3.3.6. Frequency Response Function

As discussed in section 2.5.1. a possible approach to determining the resonance frequencies of a system is by experimentally determining the frequency response function (FRF). The FRF of the gearbox within the system was experimentally determined using the impact hammer modal test. The gearbox was struck, at multiple locations, with a modal hammer that theoretically generates a perfect impulse with an infinitely small time period, which excites a wide range of frequencies in the system. An accelerometer measures the dynamic response of the gearbox at the position where the accelerometers will be placed for condition monitoring. A spectrum analyser takes the measured input and output and generates the experimental FRF. The FRF at point 3, can be seen in the figure below.

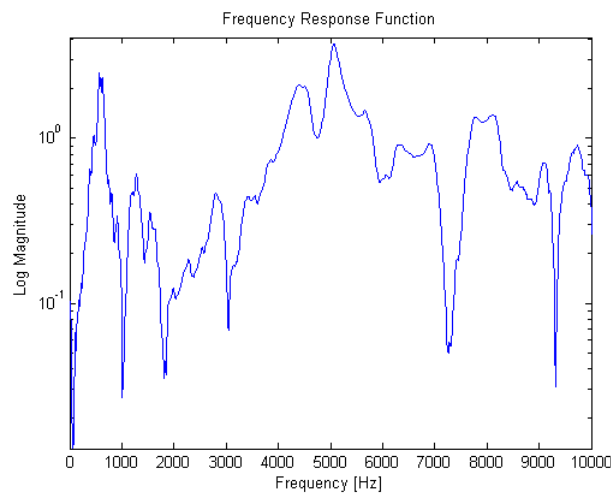


Figure 3.11: FRF of experimental gearbox

Two tests were conducted on the 8 individual points, generating 16 unique FRF plots. The natural frequencies that were common to all the plots were extracted. It is important to know the natural frequencies of the gearbox within the whole system, as it is expected that any fault will cause a minor impulse, which will be most noticeable at the natural frequencies. Thus we can learn a lot about a system with a simple experimental test. The extracted natural frequencies can be seen in the table below. The highest resonance frequency extracted is 8.16 kHz, which agrees that a sampling frequency of 25.6 kHz is suitable for detecting such a frequency.

Table 3.7: Natural frequencies of gearbox within experimental setup

Mode Number	Frequency [Hz]
-------------	----------------

1	605
2	1267
3	2909
4	5038
5	7475
6	8160

4. Results and Discussion

The proposed algorithm was tested using data from both the simulated lumped mass model and the accelerated life test on the physical experimental setup. From both experiments, healthy data was used to train the HMM and set of GMMs, and damaged data was used to verify the ability of the algorithm to detect the fault and determine its severity. The features extracted from the signals were based on an understanding of the expected fault mechanisms, modal properties of the system (determined either mathematically or experimentally) and characteristics of the healthy signal. This enabled the selection of fault features without the need for historic fault related data.

4.1. Simulation Results

The simulated model is a cost effective method to determine the feasibility of the proposed algorithm. The time domain vibration signals for a healthy and various degrees of damaged gearboxes are generated using the lumped mass model as discussed in the previous chapter. Before the proposed algorithm can be implemented it is necessary to extract as much useful information as possible from the system and its vibration signals.

It is not possible to conduct an impact test on the simulated model to generate the FRF, from which the expected fault impulse frequencies can be deducted. However, it is possible to use the eigenfrequencies of the system as the expected fault impulse frequencies. The eigenfrequencies for the undamped model can be calculated as follows.

$$(-\omega^2 M + \bar{K})\phi = 0 \quad 4.35.$$

The eigenfrequencies of the simulated model are in Table 4.1. If these frequencies contain any diagnostic information it should be evident in the comparison of spectrograms of the healthy and damaged signals.

Table 4.1: Eigen-frequencies of lumped mass model

Eigen Number	Eigen- frequency [Hz]
1	696
2	795
3	1146
4	1186
5	1399
6	1756
7	5203

As well as using the eigenfrequencies more useful information can be gained by analysing the measured vibration signals in the system's healthy state. Below are some of the standard signal processing transforms that have been applied to the healthy signal in order to extract some of the system's dynamic properties. It is possible that the transform that best represents the signal and its properties, may be the best transform to extract diagnostic features for the HMM and GMM.

Spectra of healthy and damaged vibration signal

Spectra analysis is the most common approach to fault detection in vibration signals and is presented here as a baseline for comparison. The fluctuating operational speed and load, results in frequency

and amplitude modulation, which limits the efficiency of spectral analysis. The model experiences drastic speed fluctuations, as the speed varies between 1200 and 1450 rpm over a time period of 10s. The speed fluctuations cause serious blurring of the meshing frequencies, known as spectral smearing, since only the gear meshing frequency at the maximum speed of 436Hz and its harmonics are clearly visible. There is not a significant difference between the spectra of the healthy and damaged gears because the spectra is dominated by the meshing frequency peaks. There is no apparent increase in amplitude of the meshing frequencies, which is generally indicative of gear damage. Also there seems to be no increase in amplitude of the sidebands around the meshing frequencies which is an early indicator of wear. The visible impulses seen in the time domain in the previous chapter have been lost in the spectra of the signal. The most visible effect of the damage is the increase of the amplitude of the high harmonics of the meshing frequency.

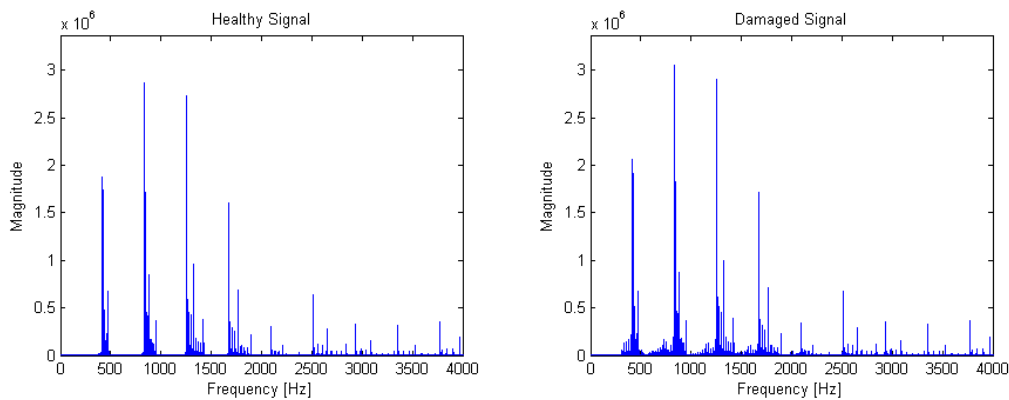


Figure 4.1: Spectra of healthy and damaged vibration signal

Spectrogram of healthy and damaged vibration signal

The gear meshing frequency and its harmonics dominate the spectrogram in the figure below. They are the horizontal lines that follow the operating speed of the system. It has this form because the gear meshing frequency and its harmonics vary with time. It is an error of the lumped mass model that the high frequency region (6 to 12 kHz) is still dominated by the meshing frequency harmonics and not resonance frequencies. The effect of the gear fault is clear by the increase in amplitude of the area between meshing frequency and its harmonics. Also the fault impulses instantaneously excite a wide range of frequencies as can be seen in the lower frequency region of the spectrogram. The concept of the spectrogram, that is investigating the spectral content of a signal with respect to its position in time, is ideal for signals undergoing fluctuating operating conditions; however as seen below it lacks clarity to be able to effectively detect gear damage.

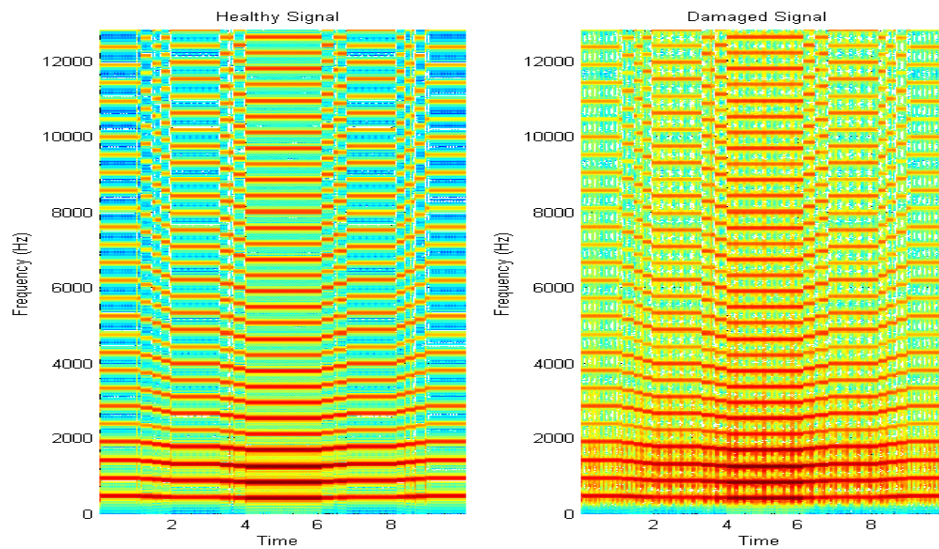


Figure 4.2: Spectrogram of simulated model vibration signals

Continuous wavelet transform of healthy and damaged vibration signal

The CWT of the healthy and damaged signals can be seen in Figure 4.3. The plots cover the identical frequency range (converted to scale) as the spectrogram plots above. Therefore the main difference between Figure 4.2 and Figure 4.3 is that the spectrogram decomposes using sinusoids while the CWT uses wavelets. Wavelets relate better to impulses, thus in the figure below it is not dominated by the meshing frequencies and its harmonics but fault impulses. Unfortunately due to the scaling nature of the plot, the response of the system under the highest load (middle of the plot) dominates the figure. Also by comparing the CWT of the healthy and damaged signals, there is no visible significant difference between the two plots. Therefore the fault may not be visible in the CWT plot below. However, it is expected that the CWT will be able to significantly better detect the presence of a fault compared to the spectrogram in the figure above.

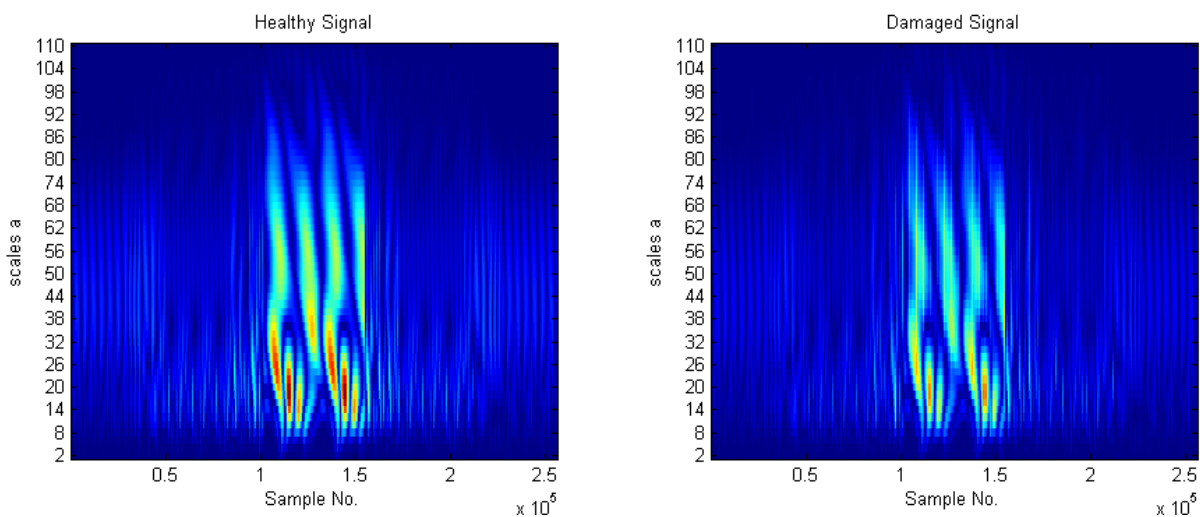


Figure 4.3: Continuous wavelet transform of healthy and damaged signal from simulated model

Wavelet packet transform of healthy and damaged vibration signal

The figure below is the wavelet coefficients from the WPT of the healthy and damaged signals. The 8 plots are the total number of decomposed frequency bands found at the third level of decomposition. Since the bandwidth of the signal is 12,8kHz, each decomposition set of coefficients have a frequency

width of 1,6 kHz. It was unexpected that each frequency band would follow the similar trend of the original signal. The figure is in a sense similar to the spectrogram in Figure 4.2, which displays that the meshing frequencies and its harmonics dominate the entire spectra of the system. From a visual comparison of the figures below, it is difficult to identify the scale band that best captures the presence of the fault.

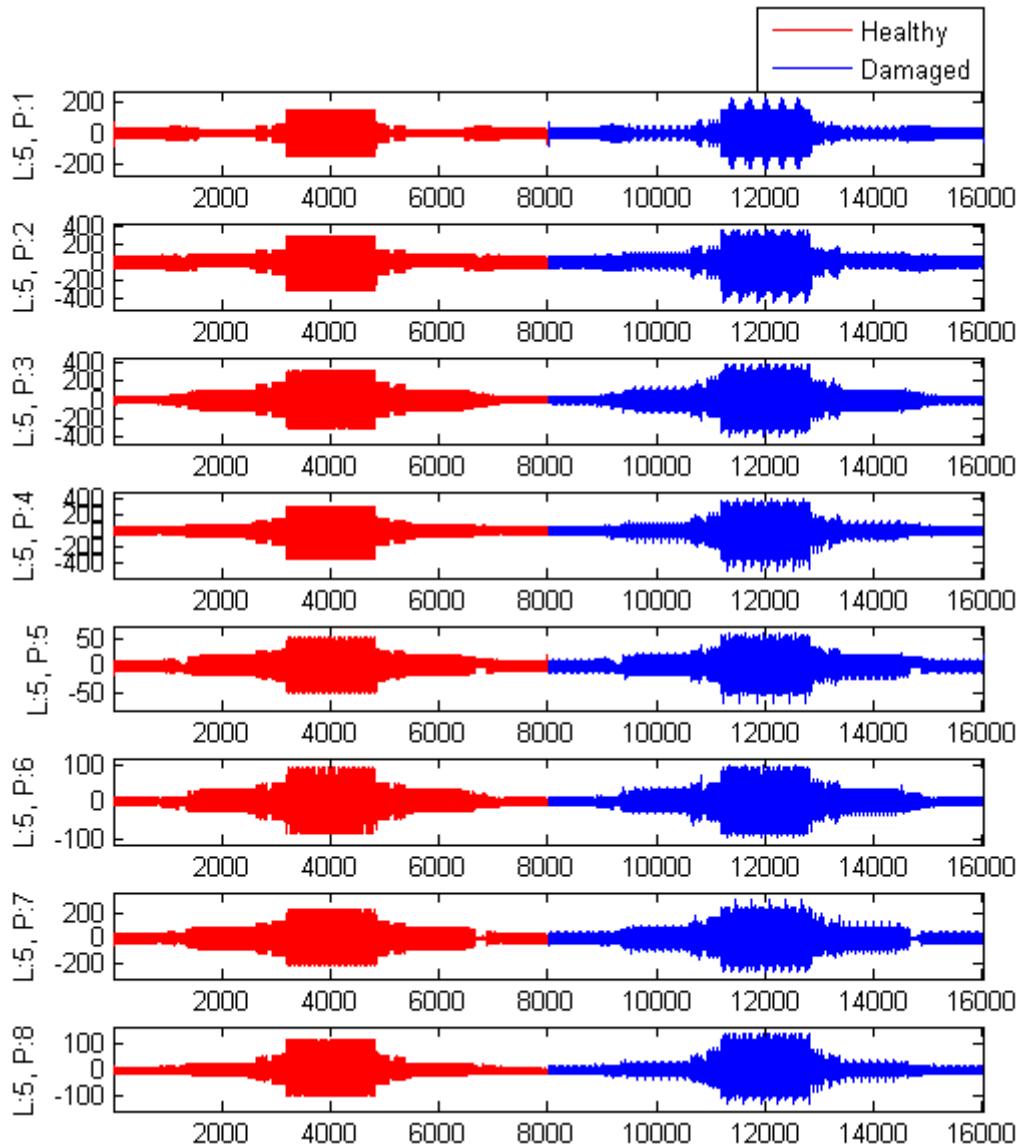


Figure 4.4: WPT of simulated signal, coefficients of packets 1 to 8 at a decomposition level 3

So far it is difficult to discern which transform best captures the presence of the fault, the CWT or WPT. By analysing the signals above, it can be concluded that practical information regarding the system can be deduced from the spectrogram, but not from the CWT and WPT. But it is the CWT and WPT that can best detect the presence of the fault because they are less affected by the fluctuating operating conditions.

4.1.1. Operating Condition Features and Classification

A critical component of the proposed algorithm is its ability to determine the instantaneous operating condition of each measurement. The two critical elements to operating condition classification are extracting operating condition relevant features and modelling the data in such a way to most

accurately classify the condition. Each load case is defined by both load and speed, thus it is necessary to be able to extract features that capture both load and speed. The model must be able to identify N distinct operating conditions and cluster the extracted features accordingly. The HMM does not only use the extracted feature for each window to classify the instantaneous operating condition of that window, but also the operating condition of the previous window. Thus the HMM is able to capture the dynamics of the operating conditions because it understands the relationships between the N distinct operating conditions.

Operating Condition Features

As was discussed in Chapter 2 the smart features for identifying the instantaneous operating conditions must be sensitive to load and speed and robust to the presence of a fault. The STFT was seen as the best transform for capturing the operating conditions. The coefficients between the frequencies of 200Hz and 1000Hz (31 in total) form the operating condition feature for the HMM. Initially the number of hidden states in the HMM was set to 3 for simplicity, however the complexity of the HMM will be investigated later. The sequence of operating conditions for a healthy signal can be seen in the Figure 4.5. The first deduction from the figure is that the HMM and features are only able to identify two of the three expected operating conditions. The model and features are unable to detect differences between the first two primary operating conditions, however it is able to identify the third primary operating condition. Also there are some unnecessary transfers between operating conditions. The figure below is from a healthy signal, so there is no fault present, therefore the poor performance of the operating condition classifier is the result of insensitive features and not the presence of a fault.

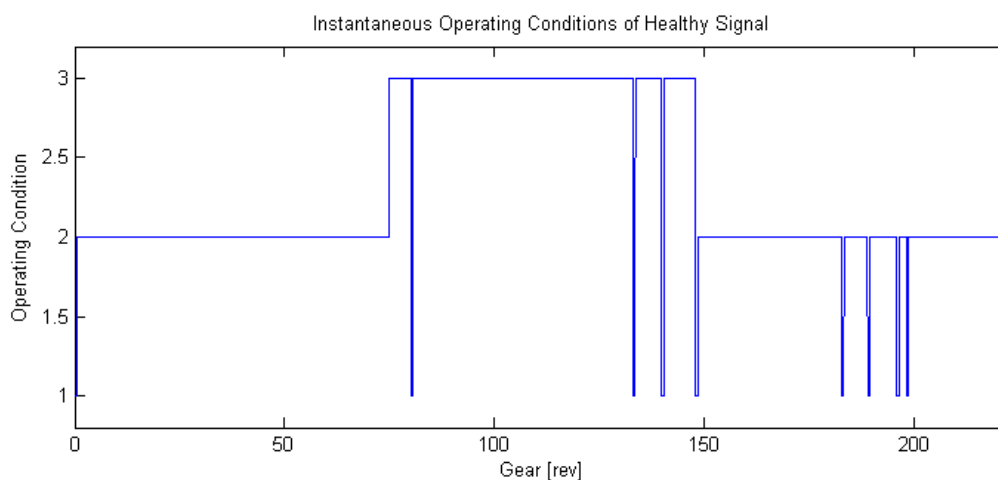


Figure 4.5: Instantaneous operations conditions for healthy signal using all STFT coefficients are features

In line with the conclusion from the figure above it is clearly necessary to refine the operating condition features to be more sensitive to load and speed. Therefore as discussed in Section 2.3.2 , principle component analysis (PCA) can be performed not only to reduce the dimensionality of the features but also extract the components with the most variance. By extracting only the components that capture the speed and load, this significantly increases the sensitivity of the features. The first four “eigen-spectra” of the simulated healthy signal can be seen in the figure below. It can be clearly seen that only the first “eigen-spectra” captures the two distinct peaks that align with the average meshing frequency and its first harmonic (indicated in red). Thus using the only first principle

component of the extracted features will make the feature more sensitive to speed and load as well as reduce the dimensionality of the feature.

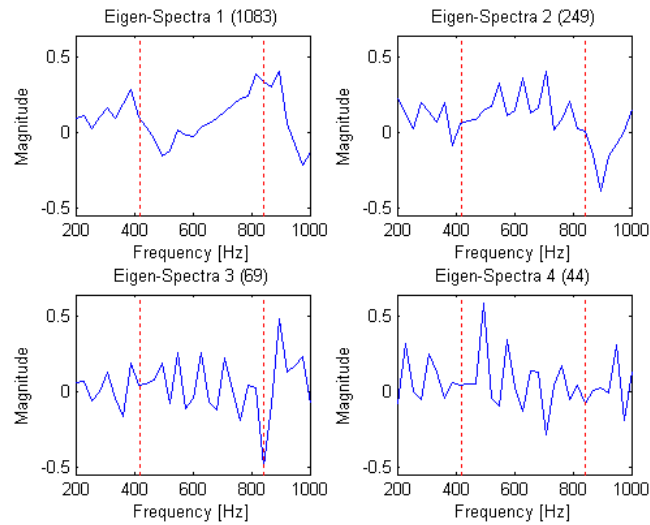


Figure 4.6: First four "eigen-spectra" of simulated model

Since the first eigenvalue contain most of the variability within the features and best captures the trend of the operating conditions, the new operating condition feature is comprised of the first eigenvalues. The new sequence of instantaneous operating conditions can be seen in the figure below. It is a significant improvement to the previous result in Figure 4.5. Firstly the HMM is able to identify the three distinct primary operating conditions. The basic trend of the operating conditions in the figure below is very similar to the trend defined by the user in Chapter 3 when the simulated model was run. There is still a couple of unnecessary transfers between states, however this has a minimal effect on the overall performance of the technique. Therefore it can be concluded that the PCA reduced features from the spectrogram are sensitive to both the speed and load fluctuations in the signal. Also the HMM is able to model the dynamics of the operating conditions and understand the relationships between operating conditions.

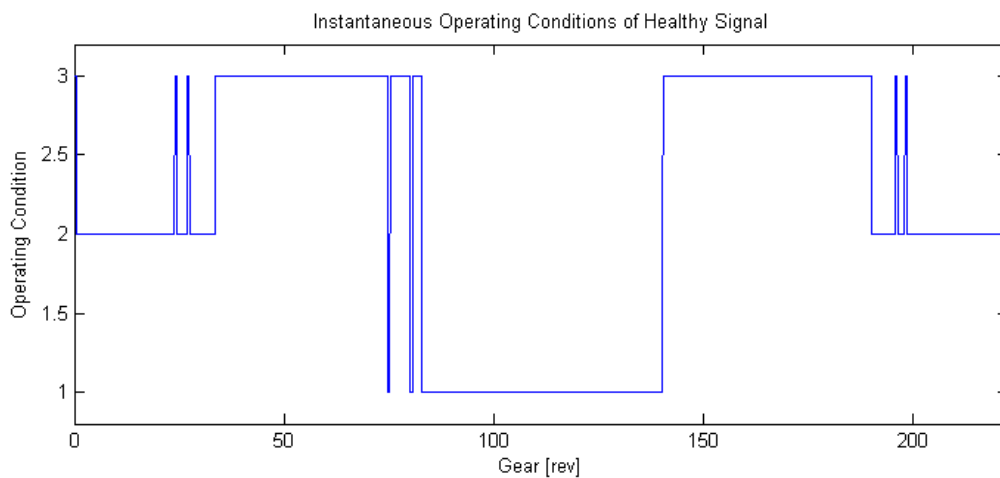


Figure 4.7: Instantaneous operations conditions for healthy signal using PCA reduced STFT coefficients are features

The next test is to see if the operating condition features are robust to the presence of a fault. This is investigated by determining the instantaneous operating conditions of a damaged signal undergoing

the exact same operating condition sequence as the healthy signal in the figure above. Then the instantaneous operating conditions of the healthy and damaged signals are compared using the confusion matrix in the table below. The overall accuracy of the method on the damaged signal was only 58%, which is low and the reason for this is evident in the confusion matrix. The operating conditions features are unable to distinguish between the 2nd and 3rd operating conditions in the damaged signal, because it classified the majority of the 2nd operating conditions as the 3rd. However it was very good at classifying the 1st and 3rd operating conditions. In general the operating condition classification methodology did not perform well, mainly due to lack of robustness of the features to the presence of a fault.

Table 4. 2: Confusion matrix of operating conditions for simulated data

		Predicted Operating Condition (Damaged signal)		
		1	2	3
Actual Operating Condition (Healthy signal)	1	1170	0	0
	2	2	98	1174
	3	676	52	1281

Operating Condition Classification

Up to now the HMM has only consisted of 3 hidden states because there are 3 stationary load cases applied to the system. However it is necessary to ask what is the optimum number of hidden states that best represents the data. Thus a simple overfitting test is implemented and can be seen in Figure 4.8. The test is in the form of the average NLL value of the healthy and damaged measurements. The HMM was trained on the healthy signal, thus the NLL continues on a downward trend. However the NLL values of damaged signal, which is novel to the HMM, initially decreases but begins to increase around 8 hidden states which is indicative of overfitting. Thus it can be concluded that any number of hidden states below 8 is suitable for accurately fitting the data. It is interesting to see that 3 hidden states causes a spike in the NLL of the damaged signal, however it is concluded that this is purely due to noise in the signal and would be different if re-tested.

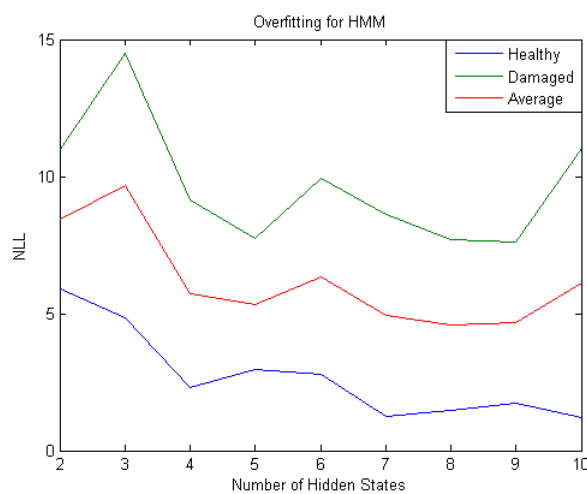


Figure 4.8: Overfitting test for HMM

Therefore in general it can be concluded that a HMM trained on the first PC of the extracted features and comprised of 8 hidden states is the optimum solution for classifying the instantaneous operating conditions.

4.1.2. Order Tracking Accuracy

The accuracy of the implemented order tracking methodology is evaluated by comparing it to the actual displacement of the gear which is contained in the lumped mass model parameters. Three basic methods of order tracking are investigated in Figure 4.9. The first method is the initial windowing scheme based on the constant average shaft speed implemented for the operating conditions classification. The next method revises the original estimation by estimating the average speed of each hidden state of the HMM and then recalculating the shaft displacement based on the path of the hidden states (this is used to guide the maxima tracking algorithm to track the IF). The final approach to estimate the shaft displacement is the one explained in Chapter 2 with the Vold-Kalman filter (VKF). The figure below plots the absolute error between the actual displacement and estimated displacement in degrees. It is clear that the first two methods are not at all suitable with a maximum deviation from the actual displacement of over 7 revolutions, which over a period of 220 revolutions is an error of 3.18%. The order tracking method using the VKF is a significant improvement with a maximum deviation of 120° , which over the entire period is only an error of 0.15%. It is extremely important to accurately calculate the instantaneous phase of the shaft because it is necessary to correctly identify the location and periodicity of detected faults.

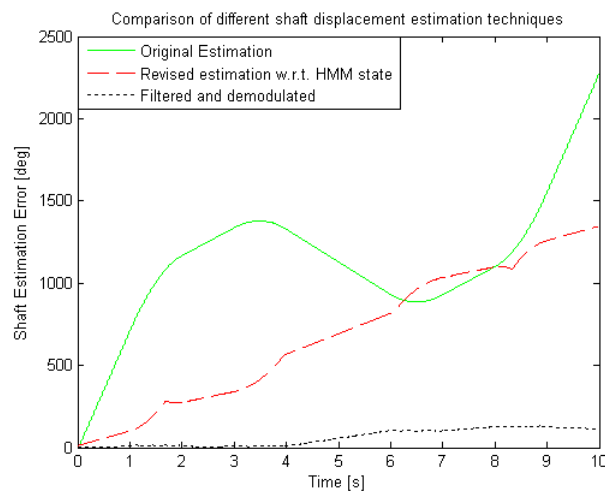


Figure 4.9: Absolute error of order tracking methods

4.1.3. Feature Selection for Fault Detection

This section displays and discusses the results when using a variety of different fault features. All of the fault features are linked to the plots discussed earlier in the chapter. The basic progression is from time domain metrics, to STFT coefficients, to CWT and WPT coefficients. This is to illustrate the power of smart fault features, that as the features get 'smarter' the results should improve.

Time Domain Features

The time domain features used are those listed in section 1.2.3.5. There is nothing smart about the time domain metrics used, they contain minimal fault information. The discrepancy signals of the healthy and damaged gearbox are compared in Figure 4.10, it is clear that the GMM is able to detect

the presence of the fault due to the large magnitude difference between the signals. However it is also noticeable that the discrepancy signal is still affected by the operating conditions. The SA plot of the discrepancy signal displays that the fault occurs at a single point, and the smearing of the fault can be attributed to the error in the order tracking.

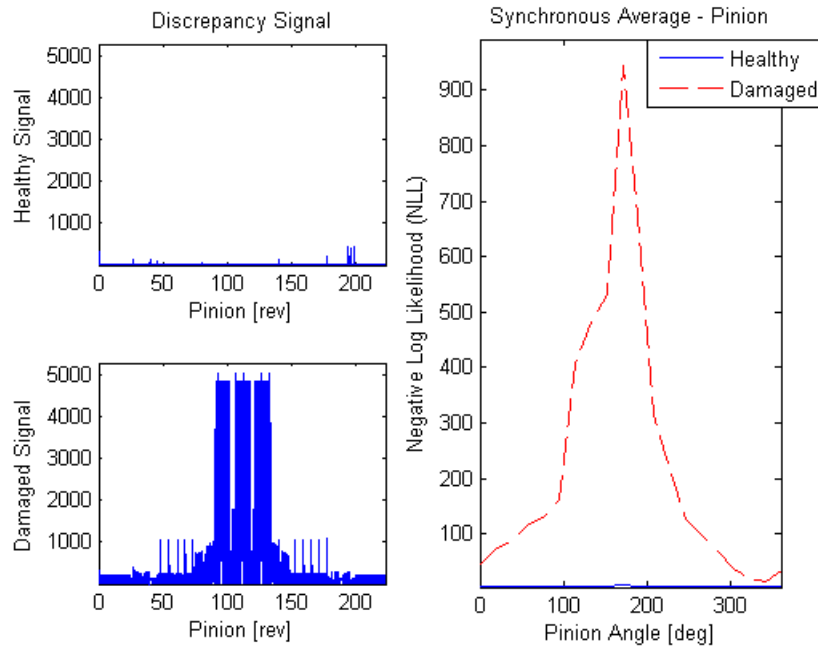


Figure 4.10: Discrepancy signals of the simulated model using time domain features

Frequency Domain Features

Frequency domain features start to take into account the basic mechanisms of expected fault, by looking at the frequencies where the fault impulses are expected. From analysis of the system and the spectra plots of the signal, it is expected that the eigenfrequencies at 1399, 1756 and 5203 Hz are the optimum frequencies to detect the gear fault impulses. Using the frequency domain features, there is a improvement to the discrepancy signal compared to the time domain features. However the discrepancy signal still suffers from a similar problem as that in the time domain features, that it is still heavily affected by the operating conditions, because the fault is so pronounced during only one of load cases. The SA plot is so accurate because it is dominated by the fault detected in the middle load case, where under stationary operating speed the order tracking method is very accurate and therefore does not smear the location of the fault.

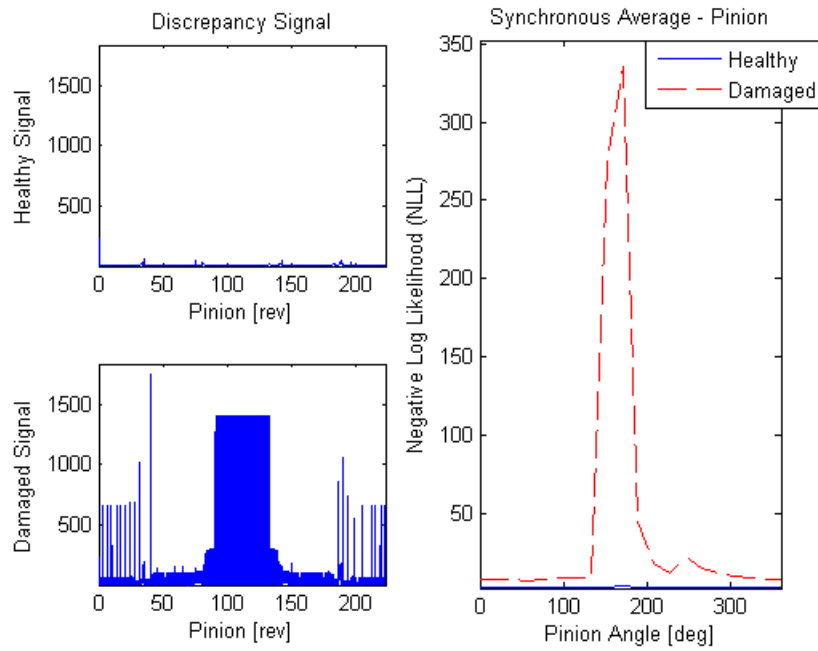


Figure 4. 11: Discrepancy signals of the simulated model using frequency domain features

Continuous Wavelet Features

For the continuous wavelet features, a Daubechies 4 wavelet was used with scales that were representative of the frequencies used for the frequency domain features. This step enables the features to be able to better detect impulses and discontinuities compared to the sinusoids used in the frequency domain features. The scales used were 13.1, 10.4 and 3.5 respectively. The discrepancy signal is an improvement to the signal generated used frequency domain features because it is less affected by the operating conditions. The SA plot again it very clear in locating the position of the fault because of it being dominated by the impulse in a single load case. However there is minor smearing which is due to the fault impulses from the other load cases.

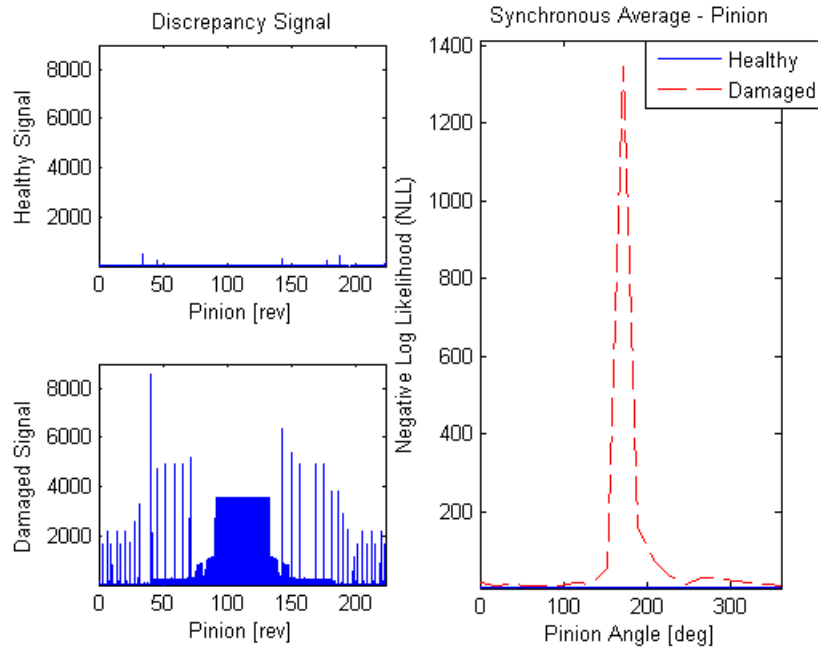


Figure 4.12: Discrepancy signals of the simulated model using CWT features

Wavelet Packet Features

The final step is the WPT features, which uses wavelets and frequency bands to detect the fault impulses. However the results do not improve relative to the results from the CWT. This is due to the nature of the simulated model. The eigenfrequencies where the fault impulses are amplified are exact values because they have been mathematically calculated. Hence using the discrete frequencies is better than the frequency bands used by the WPT. Therefore in the case of the simulated model the CWT produces better results, but in the experimental data where the exact natural frequencies are unknown it is expected that the WPT will generate better results.

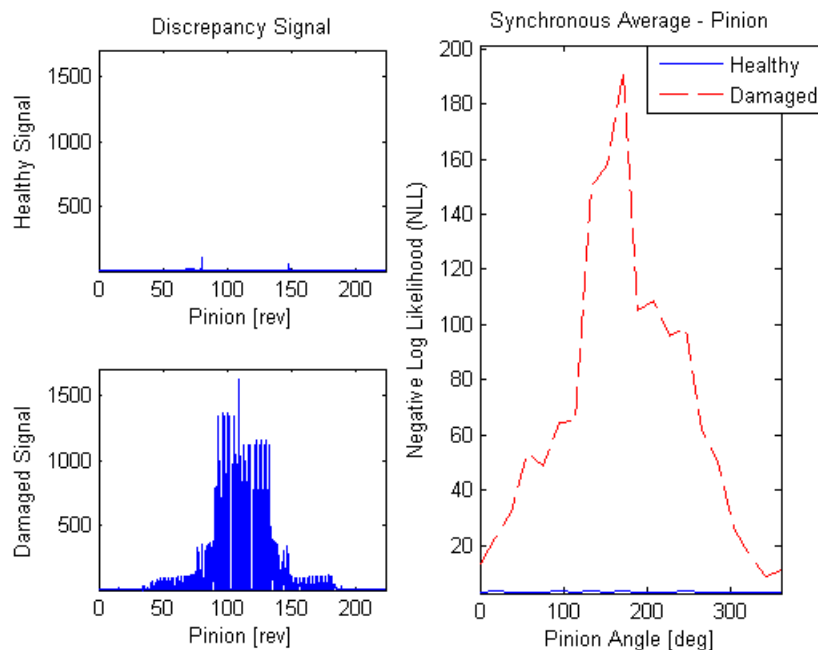


Figure 4.13: Discrepancy signals of the simulated model using WPT features

In general all the features discussed above were all able to accurately detect the gear fault in the system. It is not expected to be so easy in physical experiment. Thus the simulated signal proves that the proposed methodology does operated in the intended manner and is able to detect faults. However it is expected the results from the experimental test will give a more realistic idea of the effectiveness and robustness of the method.

4.1.4. White Noise in Signal

For the proposed algorithm to be effective in practical applications it must be able to handle a certain level of white noise in the signal. So far all of the investigation has been completed with no added white noise. The discrepancy signals below are generated using CWT features and discrepancy plots are compared to those in Figure 4.12.

The figure below is the discrepancy signals where the SNR has been increased to 2. The effect of the added noise is noticeable in the discrepancy signal of both the healthy and damaged system, however identical to the previous investigation this level noise has a small effect on the ability of the algorithm to detect and classify the fault. Therefore it can be concluded that the proposed algorithm is robust in the area of noise as it is not negatively affected by noise in the original vibration signal.

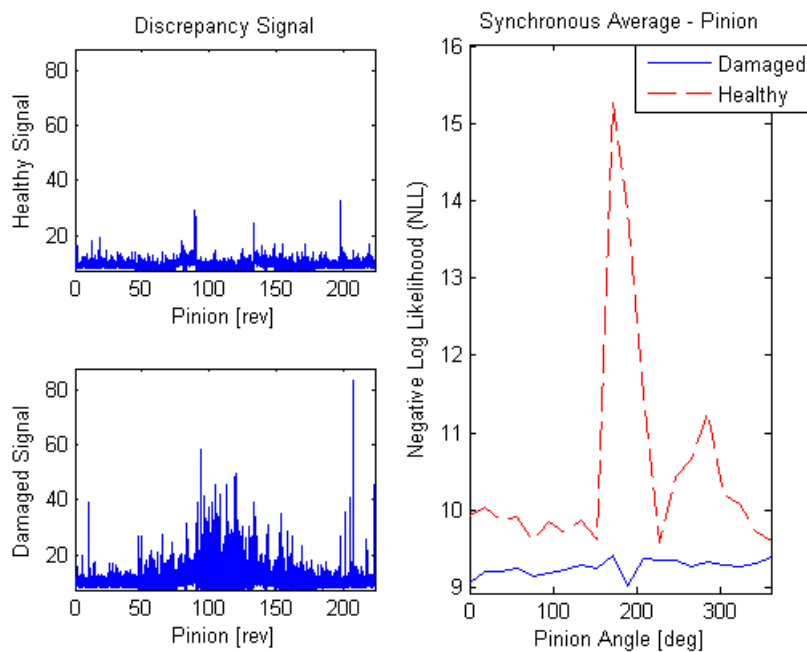


Figure 4.14: Discrepancy signals of the simulated model with an SNR of 2

4.1.5. Fault Magnitude

One of the primary aims of the proposed algorithm is that it is able to quantify the severity of the detected fault, and thus track the progression of the fault until failure. The concept is that the amplitude of the discrepancy is a valid measure of the severity of a fault. The figure below is a waterfall plot of the logarithm of the discrepancy signal of the simulated model, over a range of increasing fault severity. The logarithm is plotted because the fault magnitude of the less severe faults are overshadowed by the fault magnitude of the greatest fault severity. It is clear that the algorithm is more than capable to quantify the severity of the tooth damage and track the increasing damage as

the damage progresses. Thus the amplitude of the discrepancy signal is a valid metric of the severity of the fault.

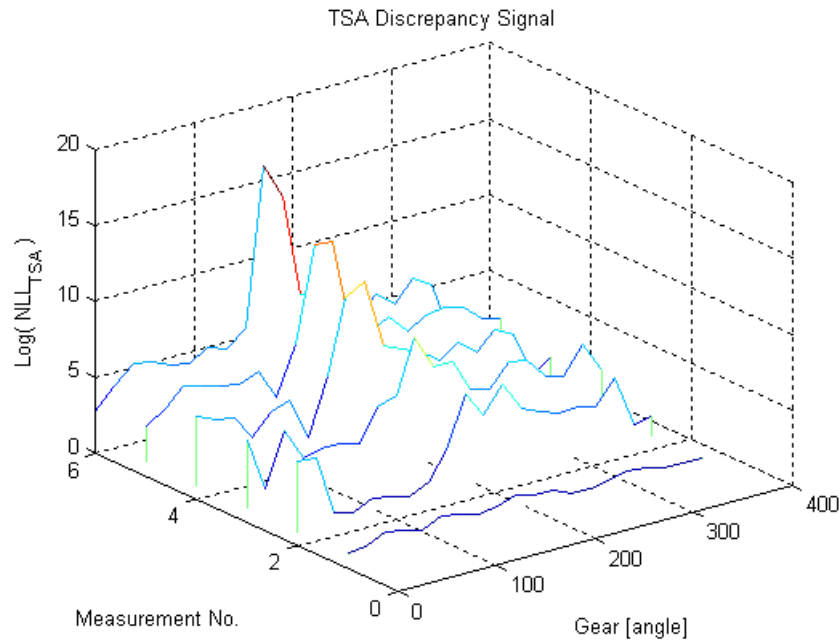


Figure 4.15: Waterfall plot of discrepancy signal of the simulated model for different fault severity levels

4.2. Experimental Results

The results of the experimental setup is the final stage of the validation of the proposed algorithm. The experimental measurements is comprised for four different datasets, namely a healthy dataset and then three datasets with seeded faults. The seeded faults in order of execution were a crack at the root of an individual tooth, a crack diagonally along the meshing face of an individual tooth (simulated a tooth chip) and surface pitting which was induced by leaving the gear in salt water for a period of a month. For the majority of the result analysis only the healthy and tooth crack datasets will be used

Similarly to the simulated model, it is desirable to extract as much information of possible from the set up. Hence the modal impact test was conducted to determine the natural frequencies of the gearbox within the setup. The natural frequencies can be found in the previous chapter and it is at the natural frequencies where the effect of impulses due to faults is expected to be the most visible. Also the healthy signals of the experiment are analysed using a range of basic transforms, as can be seen in the section above, to extract properties of the gearbox and its dynamic response characteristics.

Spectra of health vibration signal

The spectra of the measured signal, contains a fair amount of information that can be visually identified and extracted. The expected dominating frequencies are the meshing frequency and its harmonics, however the meshing frequency for the varying speeds range from 133 to 223 Hz, but it is not visible in the spectra. However the first, second and third harmonic of the meshing frequency at maximum speed is identifiable from the three large peaks. It is interesting to see how the maximum speed operating condition dominates the spectrum, this is partially due to it have the greater magnitude and that it is the operating condition with the longest percentage duration. Other

frequencies of interest are the resonance impact frequencies of the gearbox which were experimentally determined to lie at 605, 1267, 2909, 5038, 7475 and 8190 Hz, but there seems to be no significant information nears these frequencies. It is interesting to note the high energy in the frequency range or 2 to 2.5 kHz, as well as the two smaller peaks at 1.3 and 1.6 kHz. Otherwise these is almost no other information that can be extracted from the spectra of the signal.

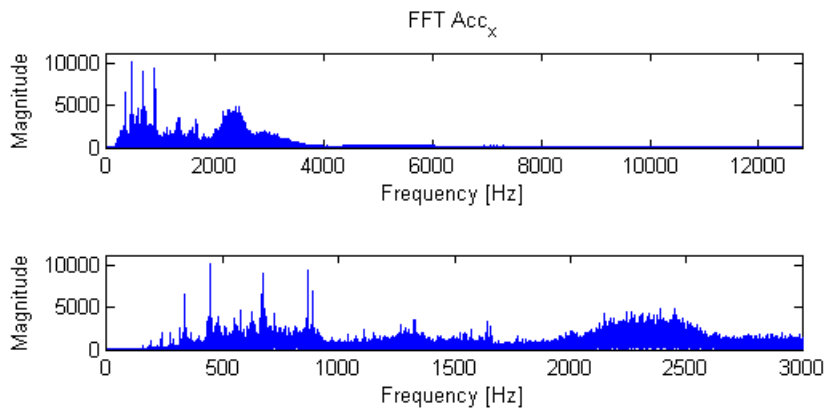


Figure 4.16: Spectra (full bandwidth and zoomed view) of healthy experimental signal

Spectrogram of healthy vibration signal

The spectrogram in the figure below, gives a good indication of how the spectra of the signal changes with respect to operating conditions (or time). The meshing frequencies that change with operating speed can be seen as the faint horizontal lines below 2 kHz that change in a step-wise manner. As discussed in Chapter 2, it is possible to estimate the natural frequencies of the system from the STFT plot, by looking at the frequencies with high amplitude that do not vary with speed. From analysing the STFT plot below, the extracted dominating frequencies can be estimated as follows; 2275, 4525, 5700, 7200 and 9900 Hz. These estimates do not correlate well with the frequencies extracted from the FRF, however they are another set of frequencies of interest that are worth investigating.

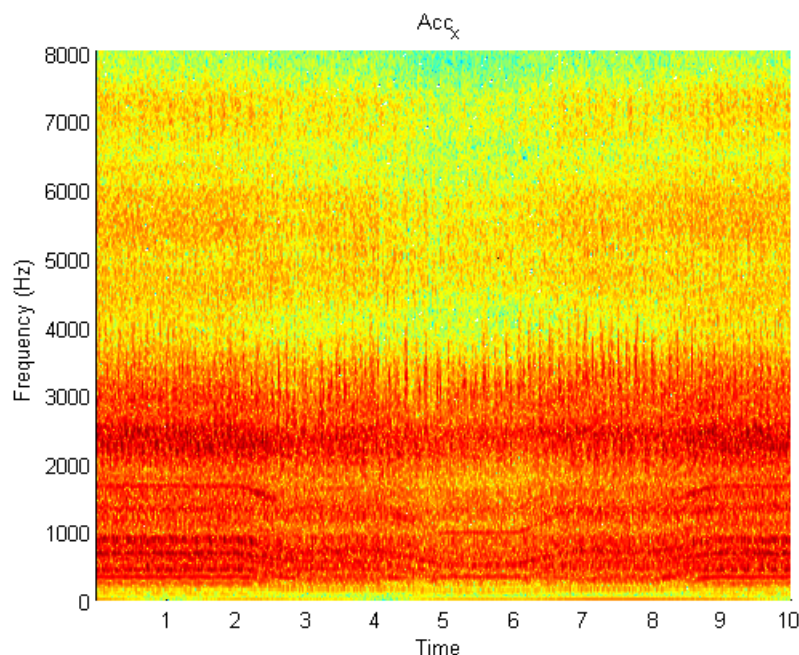


Figure 4.17: Spectrogram of healthy experimental signal

Continuous wavelet transform of healthy vibration signal

The CWT plot of the healthy signal can be found in Figure 4.18. No physical information regarding the system can be inferred from the figure. Because the CWT uses wavelets instead of sinusoids it detects impulses rather than continuous waves. Thus the CWT is less affected by the operating conditions, because from the plot there is only a minor amplitude change between operating conditions. Also the scale band (12 to 44) in which the high amplitude content is found remains constant. Therefore even though no useful information can be extracted from the plot, it is ideal for fault features because of its insensitivity to operating conditions.

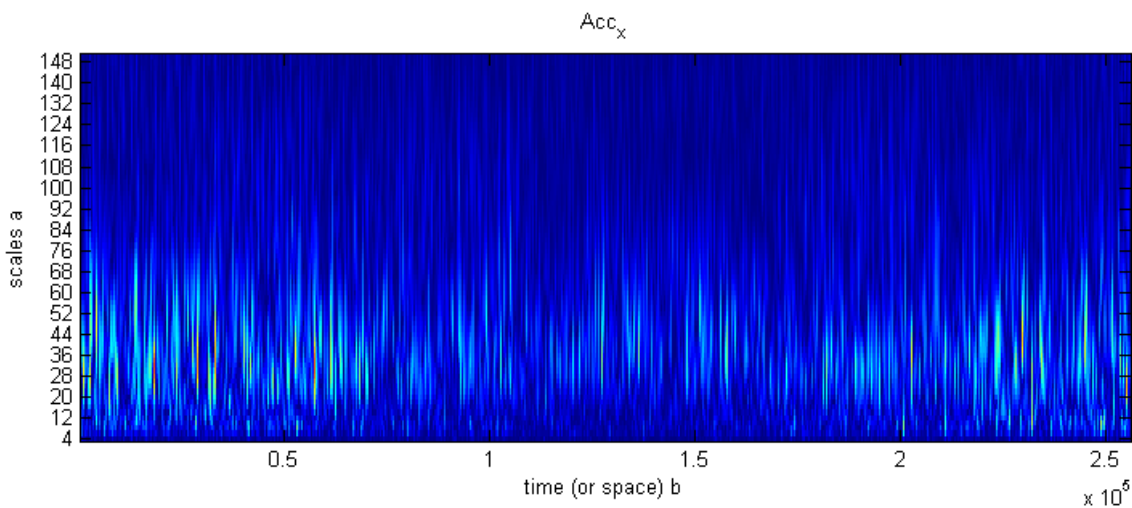


Figure 4.18: CWT plot of healthy experimental signal

Wavelet packet transform of healthy vibration signal

The WPT is fairly similar to the CWT, however it analyses the signal into scale bands and not at distinct scales. The plots below are all the wavelet coefficients from a 3rd level decomposition of the healthy signal. Each plot covers a frequency band of 1.6 kHz, it is possible to decompose the signal down to the 5th level for greater resolution, however for simplicity and ease of plotting the 3rd level decomposition has been plotted. The basic trend of the WPT is that as the central frequency increases the coefficient magnitude increases. The 4th set of coefficients do however not follow the trend, as they have greater magnitude than its previous set of coefficients and at the same time they have a very impulsive nature. The improved frequency resolution of the WPT coefficients at the higher frequencies, enables a more accurate estimation of the fault impulse frequency band.

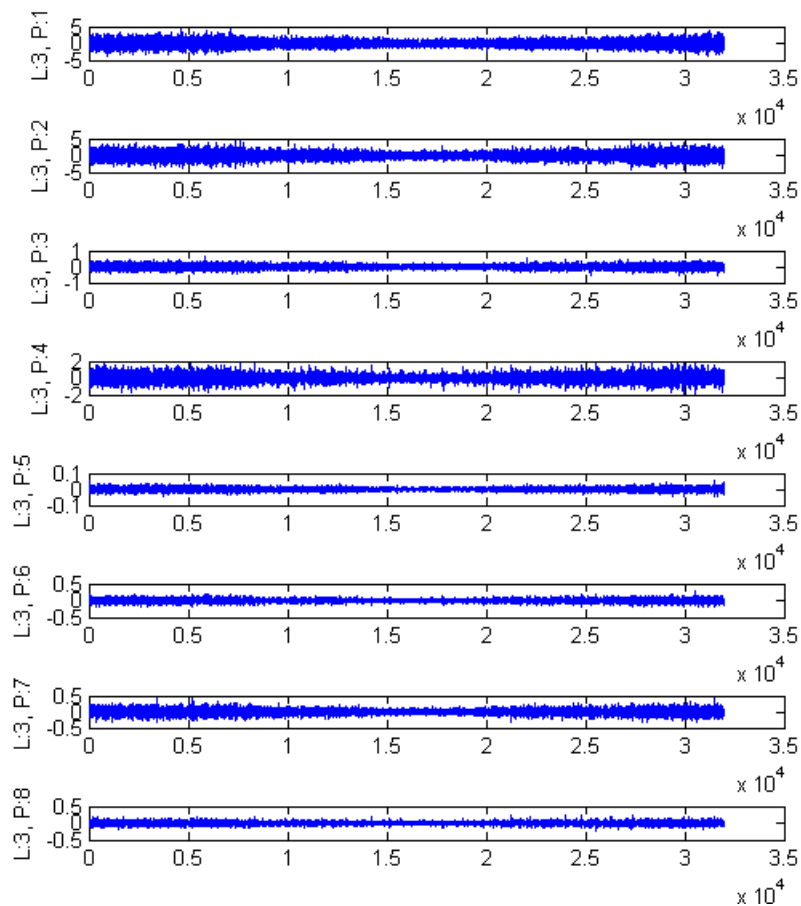


Figure 4.19: 3rd level WPT coefficients of the healthy experimental signal

Figure 4.20 is the packets that only relate to the natural frequencies of the system determine by the experimental FRF. Also it using the 4th level of decomposition, which has a frequency bandwidth of 800 Hz. The 4th and 7th packets covering the frequency bands of 2.4 to 3.2 kHz and 4.8 to 5.6 kHz respectively have a much greater amplitude than that other two frequency bands, because they are closer to the meshing frequency. The coefficients in the 7th and 11th packets seems to be less affected by the operating conditions, since their amplitude does not vary significantly over the time period relative to the coefficients in packets 4 and 10. Therefore it is expected that the 7th and 11th packets would form good fault features because of their independence of the operating conditions. However it is still to be seen if they can also detect the presence of a fault and not simply be robust to operating conditions.

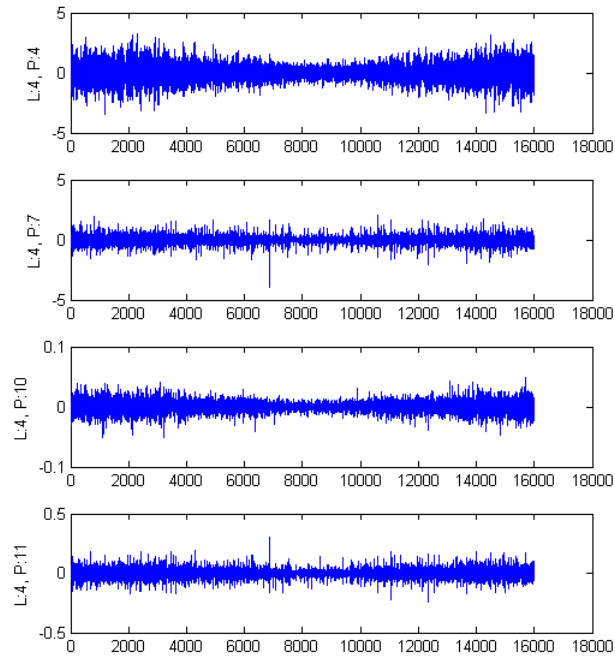


Figure 4.20: 4th level WPT coefficients of healthy signal, displaying only the packets relating to natural frequencies determine by FRF

4.2.1. Operating Condition Features and Classification

A vital element of the proposed methodology is its ability to classify the instantaneous operating conditions during the measurement. To identify the operating conditions it is necessary to extract operating condition relevant features and create a data-based model that takes into account the effect of sequential features. The fundamental aims of the operating condition classifier is to be able to accurately detect operating conditions and be robust to the presence of faults.

Operating Condition Features

The first set of operating conditions features are identical to those used on the simulated data. They are simply the STFT coefficients between the frequencies of 200 Hz and 1000Hz (31 coefficients in total), which capture the variable meshing frequency and its first harmonic. Also for the sake of simplicity the number of hidden states of the HMM was set to 3. The sequential path of the operating conditions for a healthy signal, with which the HMM was trained can be seen in Figure 4.21. The HMM and operating condition features are able to identify 3 distinct, however there is an excessive amount of transfers between the states. The reason for the excessive transfers can be related to noise, other vibration component independent of the operating conditions or a lack of sensitivity to the load and speed. Therefore the STFT coefficients are able to capture the fluctuating load and speed, but they are also being negatively affected by other element.

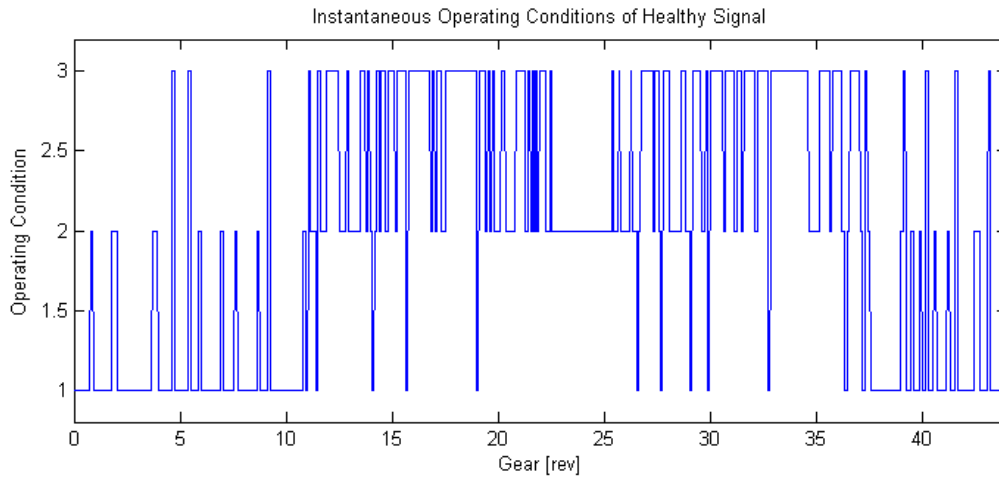


Figure 4.21: Instantaneous operating conditions of experimental healthy signal

It is necessary for the operating condition features to be more focussed on the speed and load. Therefore in the identical manner as the simulated model, the “eigen-spectra’s” of the STFT coefficients need to be evaluated. The first four “eigen-spectra’s” and their eigenvalues can be seen in the figure below. Identically to the simulated model, the first “eigen-spectra” captures the peaks at the average meshing frequency and its harmonics (indicated in red). Also it has by far the highest eigenvalue indicating its high level of variation in the signal. Therefore by extracting only the first “eigen-spectra” of the data, it focuses the feature onto the speed and load.

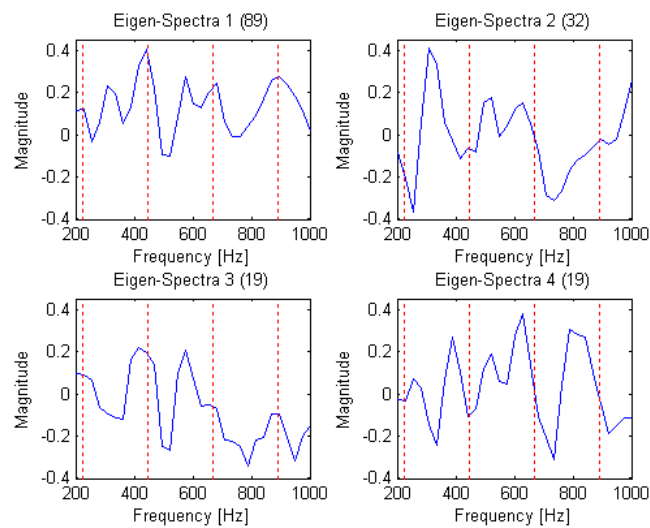


Figure 4.22: "Eigen-spectra's" of experimental signal

The sequence of operating conditions identified using the focussed operating condition features can be seen in Figure 4.23. It is not huge improvement compared to the previous results in Figure 4.21, as the HMM is still able to identify 3 distinct operating conditions and the basic trend of the operating conditions is in line with what is expected. The small improvement is that there is much less misleading transfers between states. Again it has been proved that by identifying the relevant “eigen-spectra” increases the sensitive of the features to the load and speed fluctuations. With all the unnecessary transfers, it means the HMM has not managed to fully capture the dynamics of the systems. To reduce

the transfers the diagonal of the transition probability matrix should be increase to reduce the chances of the model transferring between hidden states.

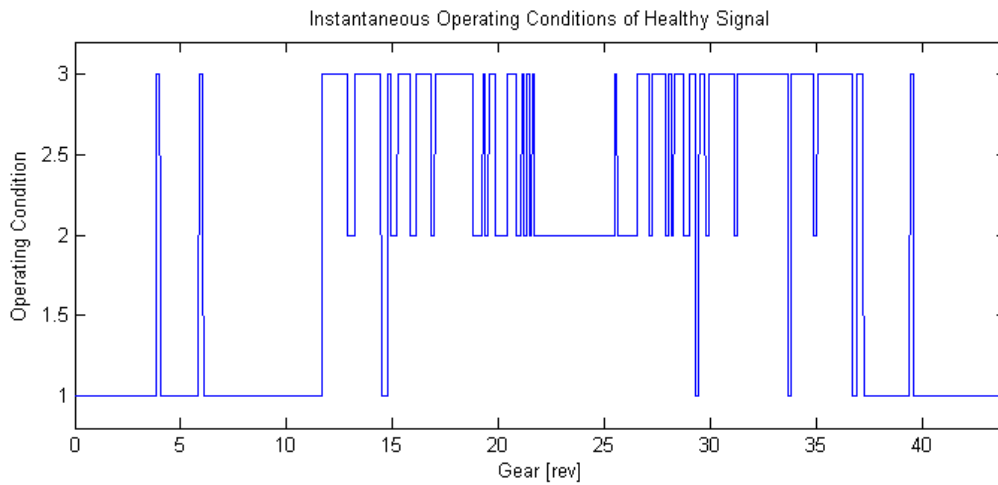


Figure 4.23: Instantaneous operating conditions of experimental signal using reduced features

The final test is to determine whether the operating condition features are robust to the presence of a fault. This is conducted by determining the sequence of operating conditions of a damaged signal under the identical sequence of operating conditions as the healthy signal in Figure 4.23. The two sequences of operating conditions are then compared using a confusion matrix. The results can be seen in the table below. The total equivalence between the two sequences is only 53% which is quite low. The HMM is very good at identifying operating conditions 1 and 2, however it misidentifies the 1st hidden state with the 3rd too many times. Therefore in general the operating conditions features and the HMM are not very good at classifying the instantaneous operating conditions. This is because of the features insensitivity to speed and load and lack of robustness to the presence of a fault.

Table 4. 3: Confusion matrix of instantaneous operating condition of experimental signals

		Predicted Operating Condition (Damaged signal)		
		1	2	3
Actual Operating Condition (Healthy signal)	1	315	169	288
	2	244	460	55
	3	94	28	198

It is crucial to ensure that the PCA method is stable for the whole range of measurements it will encounter. Thus the “eigen-spectra’s” were computed for a random number of healthy files and then a random set of novel healthy and damaged files were transformed using the first PC. The results can be seen in Figure 4.24, and it is visible that the transforms are all consistent and that the PCA is stable for the full range of expected signals.

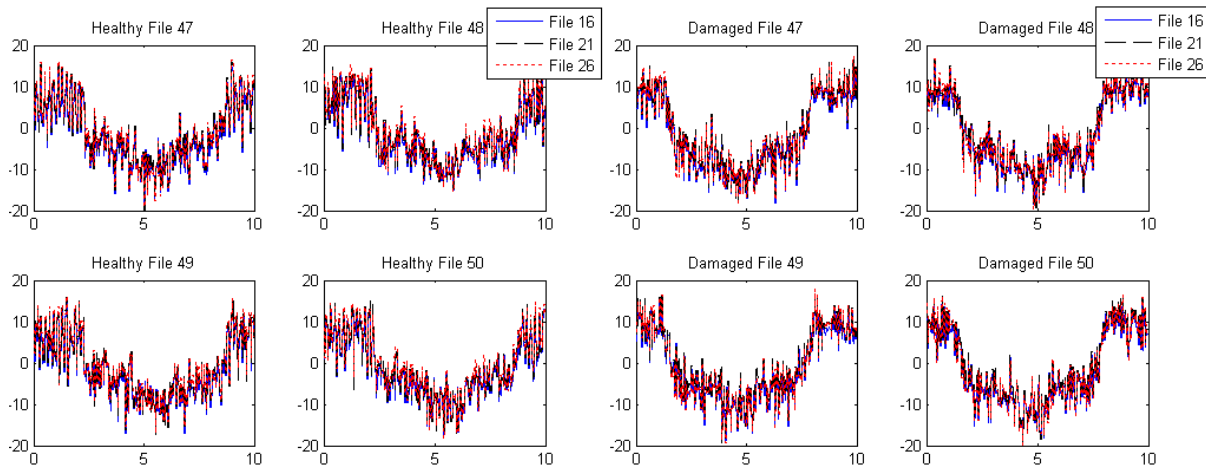


Figure 4.24: Transformed operating conditions features from healthy and damaged signals respectively.

Operating Condition Classification

Up to now the HMM has only consisted of 3 hidden states because there are 3 stationary load cases applied to the experiment. However it is necessary to determine the optimum HMM that best fits the data. Thus the results of a set of held back healthy measurements will be compared for varying number of training files and hidden states. Also it is not desirable to over fit the HMM to the data, thus reducing its accuracy. The results of the overfitting test can be found in Figure 4.25. The first conclusion of the figure is that the number of training files has a minimal effect on the overall accuracy of the HMM, and therefore it will take an excessive amount to training files to over fit the HMM. The second conclusion from the figure is that the number of hidden states does progressively improve the results of the HMM. It was expected to increase the number of hidden states to the point where the results begin to appreciate, which signifies the optimum model complexity and passing that point leads to overfitting. However it was not possible to train the HMM for an excessive amount of hidden states because the training algorithm (expectation maximisation) does not converge, thus not producing any results. However it can be concluded that any number of hidden states below 15 does not over fit the HMM.

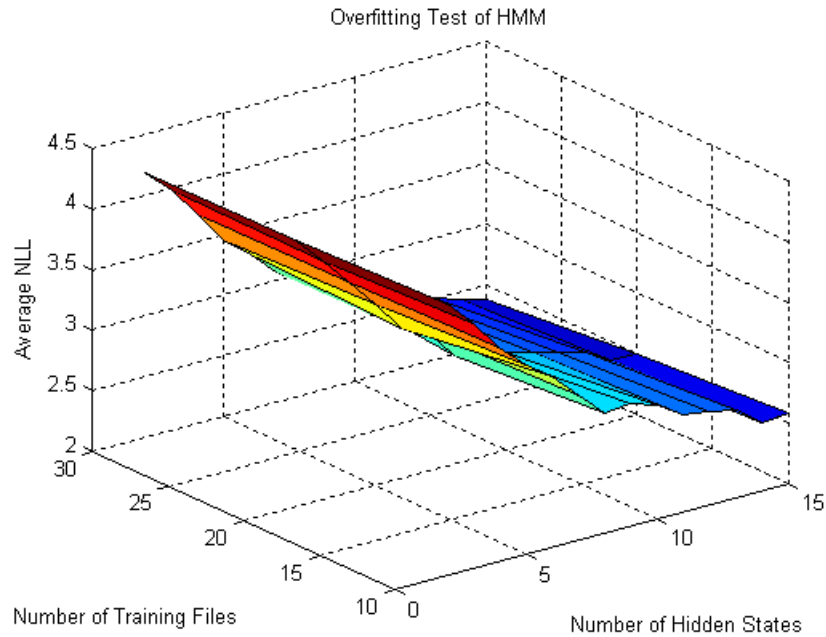


Figure 4.25: Test for overfitting of HMM

In general the classification of the instantaneous operating conditions is not very accurate, however it is interesting to note that this accuracy is not critical to the overall performance of the proposed methodology.

4.2.2. Order Tracking Accuracy

The next stage of analysing the effectiveness of the proposed methodology is determining the accuracy of the proposed order tracking algorithm. It is known that it works within reasonable accuracy for the simulated data, but does it generate reasonably accurate results for actual measurements. The identical three order tracking approaches applied to the simulated data were applied to the experimental data here. The accuracies of the approaches were compared to the actual shaft displacement measured by either the tachometer or shaft encoder. The absolute errors relative to actual angular displacement can be seen in Figure 4.26. The results are very similar to those of the simulated data in Figure 4.9. The original shaft displacement estimation used for the operating condition feature windows has a maximum error 750° , which is just over 2 revolutions which is an error of 4%. It is only by chance that the final error is small and should not be misleading of the approaches accuracy. The revised estimation based on the path of HMM is more indicative of the poor accuracy of these two approach, it has a total error of 5.6 revolutions which over a period of 50 revolutions is a total error 11.11% (this approach is used to guide the maxima tracking algorithm). Thus the two approaches based on constant speed are a good estimation but not sufficiently accurate to identify exact positions of faults. The approach based on the VKF is a significant improvement with a maximum error of 120° , which is only an error 0.67%. This huge improvement makes this approach highly suitable for accurately detecting the location of faults on the gear.

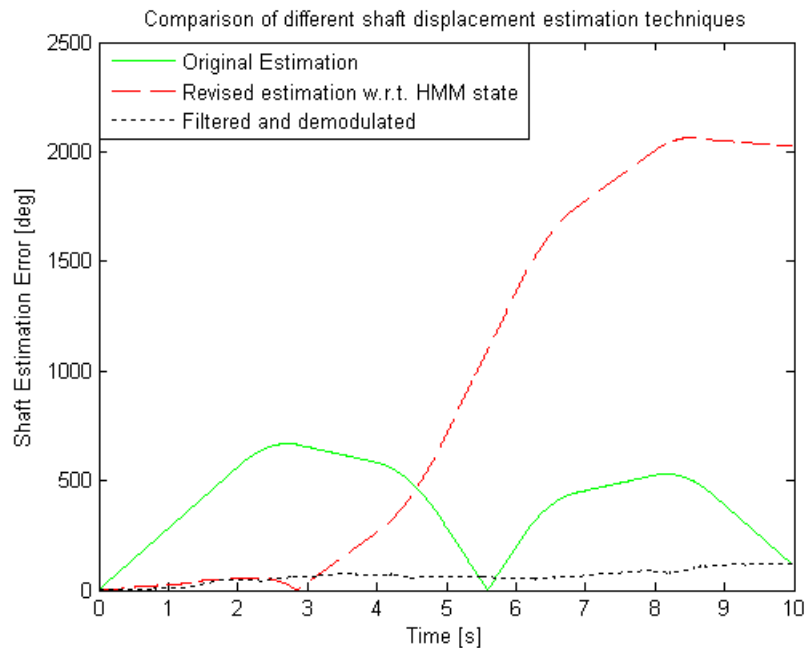


Figure 4.26: Accuracy of order tracking algorithms on experimental signals

4.2.3. Feature Selection for Fault Detection

Now that there is a better knowledge of the system's dynamic properties and response and that the HMM is able to account for some of the effect of fluctuating operating conditions, it is necessary to select fault features for the set of GMMs that are accurate in both location and severity of the fault. In the identical manner to the simulated data, a range of fault sensitive features will be investigated. These features progress from the time domain, to the frequency domain, then the CWT and finally the WPT. The progression should display that as the features become more sensitive to fault impulses and robust to the operating conditions, the performance should improve. It is expected that the time domain features will have the worst performance and it should improve until the WPT features which should have the best performance. The results in the figures below were generated for the first measurement of the first damaged dataset.

Time Domain Features

The time domain features used are very simplistic, as they require minimal insight into the system's dynamic properties and is purely based on previous research. It was fairly unexpected that the time domain features produced such good results for the simulated model. However, as can be seen in the figure below, the time domain features are completely unable to capture the presence of the fault within the damaged signal. There is minimal contrast between the discrepancy signals of the healthy and damaged gearbox, as they both seem to contain the same amount of impulses with similar magnitudes. The SA is able to align the peaks, however the healthy signal has a greater magnitude than that of the damaged signal, which is incorrect. It is supposed that the peak is defined by the outlying impulses (excessively higher than all other impulses) which are known to be misclassifications, thus the single peak in the SA is misleading. Thus the time domain features are completely inadequate to detect gears faults in both location and severity, which was expected within such a complex system.

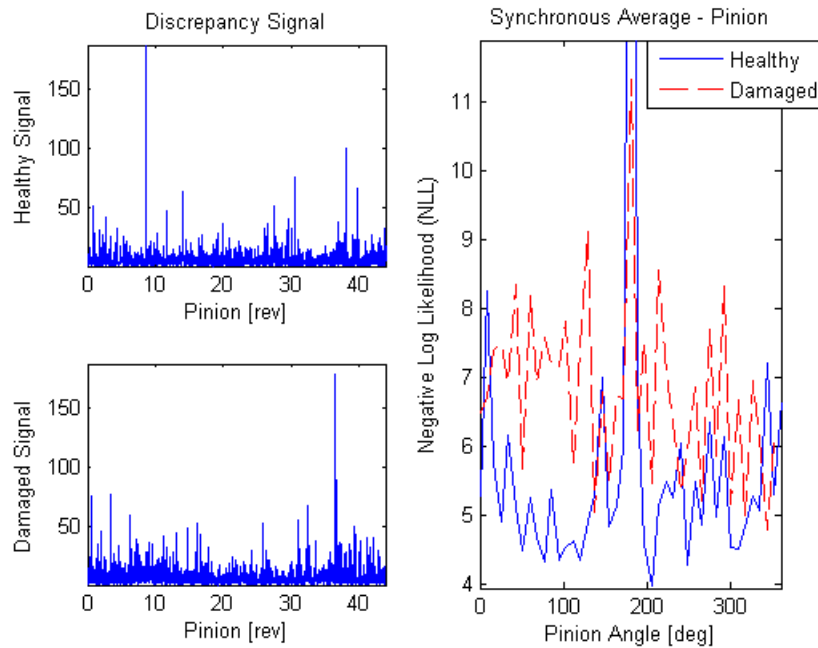


Figure 4.27: Discrepancy signals of the experimental setup using time domain features

Frequency Domain Features

From the figure below it is clearly evident that by simply focussing on the natural frequencies of the gearbox, there is a significant improvement in the quality of the results relative to the results of the time domain features. The discrepancy signal of the healthy gearbox has a significantly lower magnitude than that of the damaged signal. The discrepancy signal of the damaged gearbox contains peaks remarkably higher than those in the healthy signal, indicating the definite presence of a fault. The amplitudes of the discrepancy peaks of the damaged signal do seem to vary with respect to the operating conditions, which is indicative that the features are not completely robust to operating conditions. The SA is it able to align the impulses and extract useful information out of the discrepancy plot. The average of the damaged discrepancy signal is much higher than that of the healthy signal, indicating the clear presence of a fault. However it is the rough peak in the SA signal that indicates the presence of a localised gear fault. It Therefore by choosing features are that more sensitive to fault impulses, there is a evident in improvement in the models ability to detect discrepancies. It is anticipated that as 'smarter' features are investigated the discrepancies detected by the GMMs will become more pronounced. The frequency domain features display how simple foreknowledge of the system's properties and characteristics are able to significantly improve the ability of the GMM to detect faults.

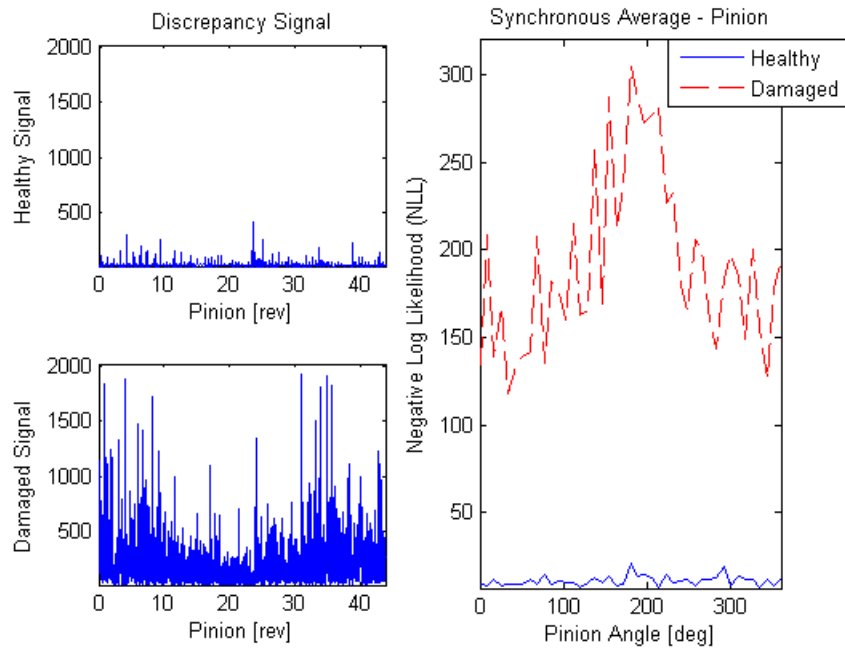


Figure 4.28: Discrepancy signals of the experimental setup using frequency domain features

Continuous Wavelet Features

The continuous wavelet features are simply the approximated scales for the Daubechies 4 wavelet family of the natural frequencies of the gearbox. It is very similar to the features from the frequency domain, it just decomposes the signal using wavelets instead of sinusoid which are more sensitive to discontinuities. The primary observation between the discrepancy signals using the CWT features and the previously discussed features is the pronounced distinction between the discrepancy signal of the healthy and damaged gearbox, however the peak in the SA is not very defined. The discrepancy signal of the healthy gearbox is almost completely smooth relative to the damaged gearbox. This is because of wavelets ability to better detect impulses than sinusoidal waveforms. This is ideal, because it is indicative of the features ability for earlier fault detection. Also the amplitude of the peaks in the discrepancy signal seem unaffected by the changing load and speed, since they seem constant over the entire signal. This displays that wavelets are less affected by operating conditions compared to sinusoids. The CWT features are unable to accurately detect the exact location of the fault in the SA, since peak is fairly wide. Thus the CWT features are a substantial improvement with respect to the previously discussed features, however there is still a need for refinement.

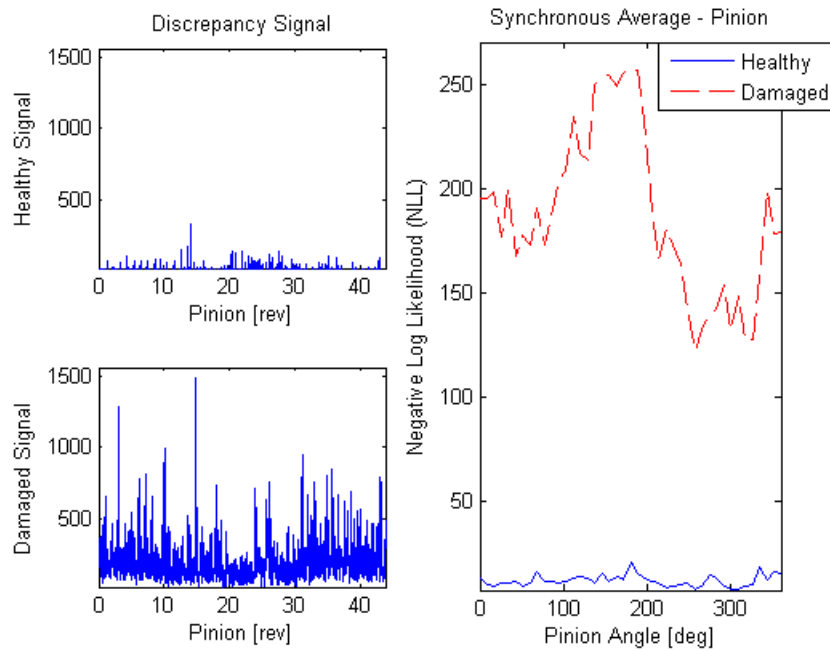


Figure 4.29: Discrepancy signals of the experimental setup using CWT features

Wavelet Packet Features – FRF frequencies

The WPT features are similar to CWT features, however instead of decomposing the signal at distinct scales the signals is decomposed into frequency bands. This has a significantly effect as it allows the features to take into account that the natural frequencies are only estimates and not exact (as in the case of the simulated model). This has a positive effect as the discrepancy signal of the health signal is much smoother and lower in magnitude than that of the damaged signal. It is also very good to see a much more defined fault peak in the SA plot of the damaged signal. Thus using frequency bands around the natural frequencies, taking into account that the natural frequencies are more of an estimation than exact, is able to capture the presence of the fault remarkably better. Since the WPT uses wavelets the discrepancy signal is robust to fluctuating operating conditions like the CWT features in the previous set of results. From the results in the figure below it is clear that WPT features are the ‘smartest’ features. They are sensitive to fault impulses and robust to fluctuating operating conditions.

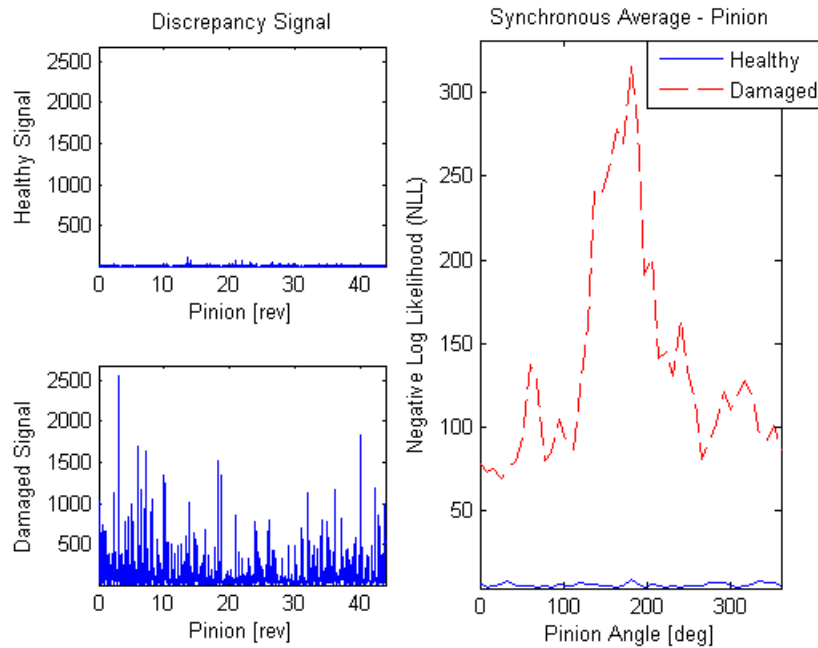


Figure 4.30: Discrepancy signals of the experimental setup using WPT features

Wavelet Packet Transform – STFT frequencies

Up until now all the frequencies of interest that determine the selected frequency and scale bands have been based on the experimental modal analysis, that generated the FRF in Chapter 3. However from analysing the spectrogram of the healthy signal earlier in the chapter, a variety of dominant high frequency bands were observable that did not correlate to the natural frequencies from the FRF. Thus using the WPT with the frequencies of interest from the spectrogram determining the packets of interest and not the FRF, the discrepancy signal for the damaged gear can be seen in Figure 4.31. The discrepancy signal is fairly similar to the previous one, however the fault peak in the SA is less defined than that in Figure 4.30. However the spectrogram frequencies are more than capable of detecting the fault, because the discrepancy peaks have a higher amplitude. Thus it can be concluded that the spectrogram is a viable option for determining the dominant frequencies of the system if it is not possible to perform an impact modal test to generate the FRF of the system.

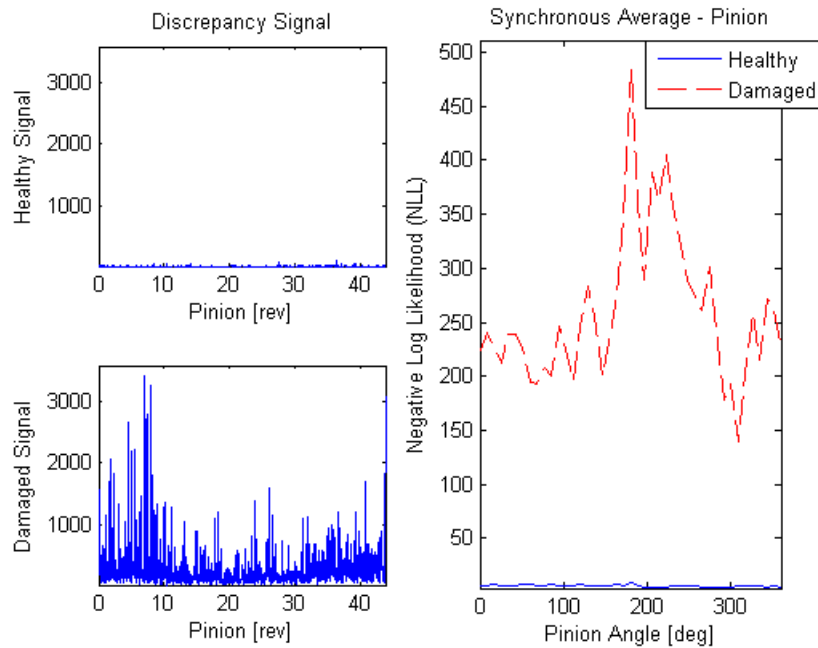


Figure 4.31: Discrepancy signals of the experimental setup using WPT features

From all the results above it is firstly concluded that the proposed algorithm is able to detect and locate a localised fault on a gear. Also from analysing the discrepancy signals, they do not seem to be influenced at all by the time varying operating conditions. By comparing the results of the time domain features and the features dependent of the system's natural frequencies (i.e. frequency domain), there is clear positive influence of knowing what frequencies to extract as features to detect the presence of a fault. Also by comparing the results generated by frequency based features and scale based features, the wavelets are more robust to operating conditions compared to sinusoids. In general the proposed fault detection and identification methodology using WPT features performed very well for a localised gear fault.

4.2.4. Accounting for Shaft Displacement Calculation Error

As seen in section 4.2.2, the calculation of the instantaneous shaft displacement is not perfect and does have a minor error. So it was proposed in section 2.6 to adjust the discrepancy signal to align the teeth within reason. Therefore the discrepancy signal of each gear revolution was compared to one another and if the maximum discrepancy was found to be either one gear tooth to the left or right, the discrepancy signal for that gear revolution was adjusted so that the maximum discrepancy occurred at the same gear tooth. The figure below displays the location of the maximum discrepancy for each successive gear revolution. As can be seen each maximum is only a signal gear tooth away from the previous tooth. Therefore all the maximums can be aligned to a single gear tooth, as it assumes the same tooth generates the maximum discrepancy each revolution and it is a result of the shaft displacement calculation error that it is detected at different teeth.

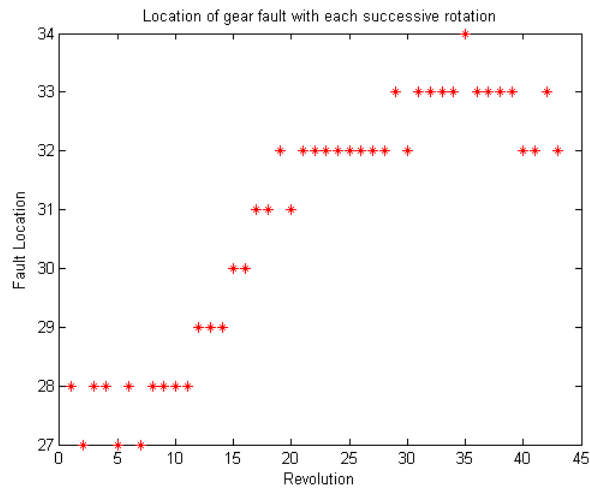


Figure 4. 32: Location of maximum discrepancy for each successive revolution

It is by adjusting the discrepancy signal to account for the shaft displacement error, that it is possible to generate the results in Figure 4.30 and Figure 4.31 where the discrepancies align well to indicate clearly the location and type of fault.

4.2.5. Fault Magnitude

The second primary aim of the proposed method is to track the severity of the fault. The ability to track fault severity also depends highly on the selected features, however based on the previous discussion on feature selection the WPT features will be implemented in this discussion. The frequencies of interest are based on the frequencies extracted from the FRF and not the STFT, since it was concluded that there was not a significant difference between the results of the two frequency sets.

The fault severity can be tracked by implementing a waterfall style plot of the SA of discrepancy signals of all the measurements of the accelerated life test of the gear. The waterfall plot that tracks the discrepancy severity of the first damaged data set can be seen in the figure below. From the figure, it can be concluded that there was a critical increase in the magnitude of the fault and it was not long until ultimate failure of the gear occurred. However due to the high magnitude of the fault, its initial stages of the fault progression has been lost in this plot due to the scale of the plot. However it can be concluded that the method is able to detect the magnitude of severity of the fault, even if the severity increases in a step-wise manner and not the expected small incremental increases.

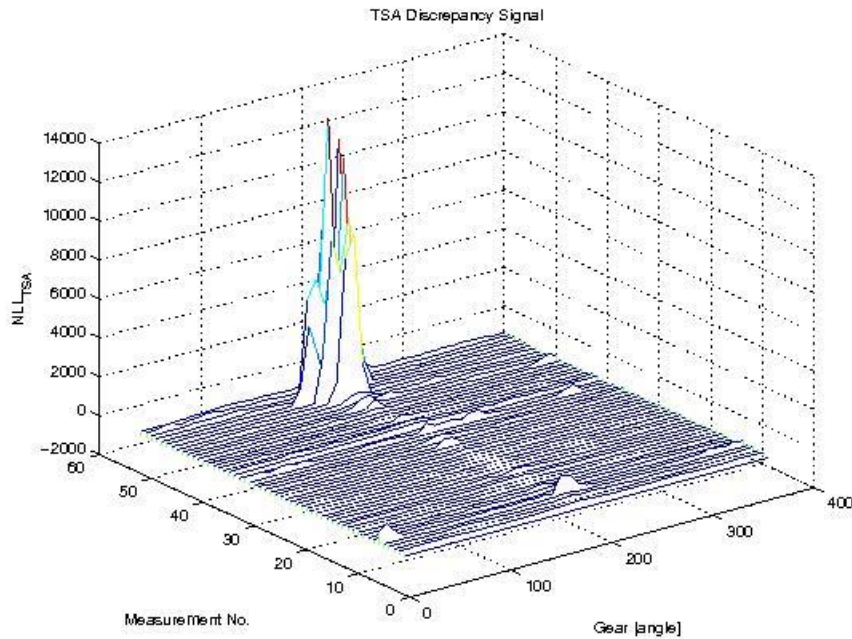


Figure 4.33: Waterfall plot of the SA of the discrepancy signal using WPT features

As mentioned in the previous paragraph, the above waterfall plot does not adequately display the early discrepancy signals because of the scale difference of the fault magnitude. A zoomed-in view of the initial measurements can be seen in the figure below. There is an evident ridge pattern that aligns itself with the location of the single fault, thus the fault is detectable in the earlier measurements. However, it is not completely smooth on either side of the ridge, due to the presence of noise and other irregularities within the signal as seen in the previous section where only a single damaged measurement was analysed. Nevertheless, these outlying discrepancy peaks can be disregarded because there is no consistency to them over time. Only the peaks that appear continuously are indicative of the presence of a fault. From this figure, it can be concluded that the method is able to detect the early stages of faults because of the visible ridge in the plot that aligns itself with the fault location.

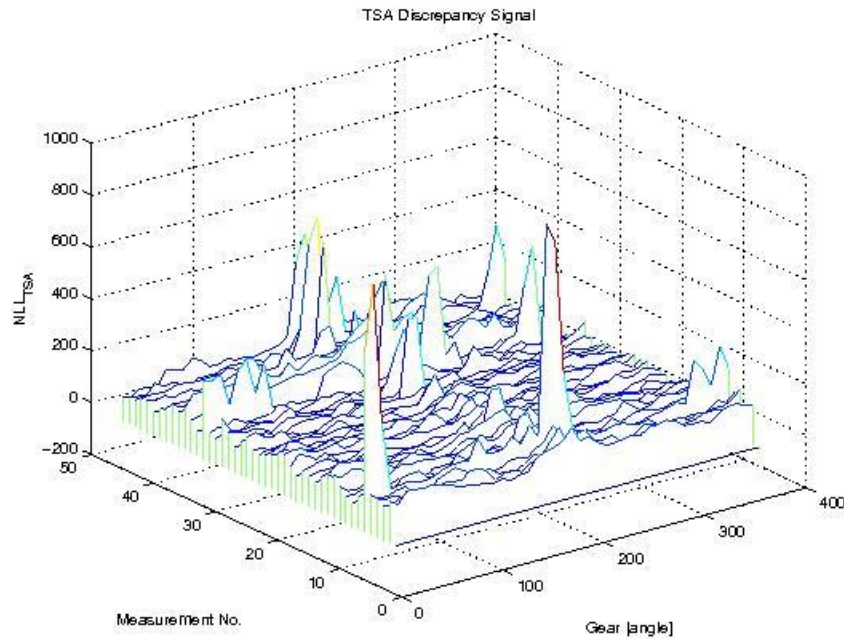


Figure 4.34: Zoomed in view of the initial measurements of the waterfall plot

4.2.6. Spectra of Discrepancy

The SA method is ideal for events that occur in time with the period that the signal is averaged over, however for fault impulses with periods that do not coincide with the averaging period, they are lost using the SA. Thus by analysing the spectra of the discrepancy signal, non-synchronous fault impulses and their periodicity are detectable. Figure 4.35 is the spectra of the discrepancy signals found in Figure 4.30 generated by the WPT features. The fault order of once per revolution and its harmonics are clearly visible in the spectra of the damaged signal. There is a large deviation between the spectra of the healthy and damaged signals, thus a clear indication of a presence of a fault, however the large number of harmonics can make the identification of the fault difficult. However it is obvious from the figure below it is more than capable at detecting the single impulse per revolution in the spectra and it is a viable option for fault detection of non-synchronous fault impulses.

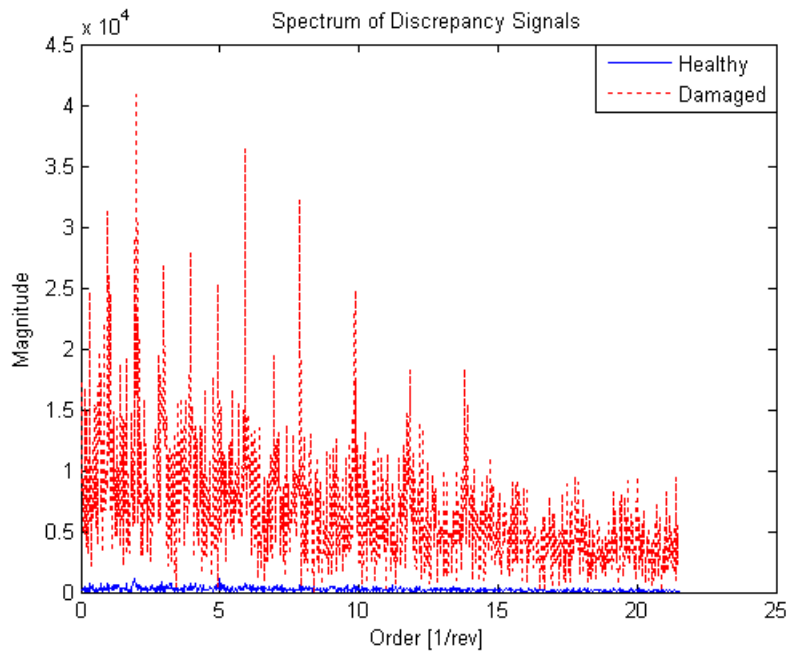


Figure 4.35: Spectra of discrepancy signal for both healthy and damaged signals

4.2.7. Cepstra of Discrepancy

So far the gear has been identified from the discrepancy signal analysing both the SA and the spectra, however each have their own drawbacks. The SA can be misled by outlying impulses and the spectra is overwhelmed with harmonics of the fault frequency. Therefore an ideal analysis approach to find periodicity within the spectra is cepstrum analysis. Thus the cepstra of the healthy and damaged discrepancy signals can be seen in Figure 4.36. It must be noted that the cepstra has been zoomed in because there is an extremely large peak at a quefreny of zero because of the highly periodic nature of the spectra. This partially explains why the healthy and damaged cepstra's have identical peaks at a quefreny of 1, because all of the energy in the damaged signal has gone into the DC offset. The damaged cepstra also has a peak at quefreny of 0.5 revolutions which is indicative of a fault twice per revolution. From the figure it can be concluded that the cepstra is not an ideal tool to identifying localised gear of this severity. It is proposed that less severe gear faults that have a much less periodic spectra will be must more easily identifiable in the cepstra, thus the cepstra will be the ideal tool for identifying the very initial stages of fault progression.

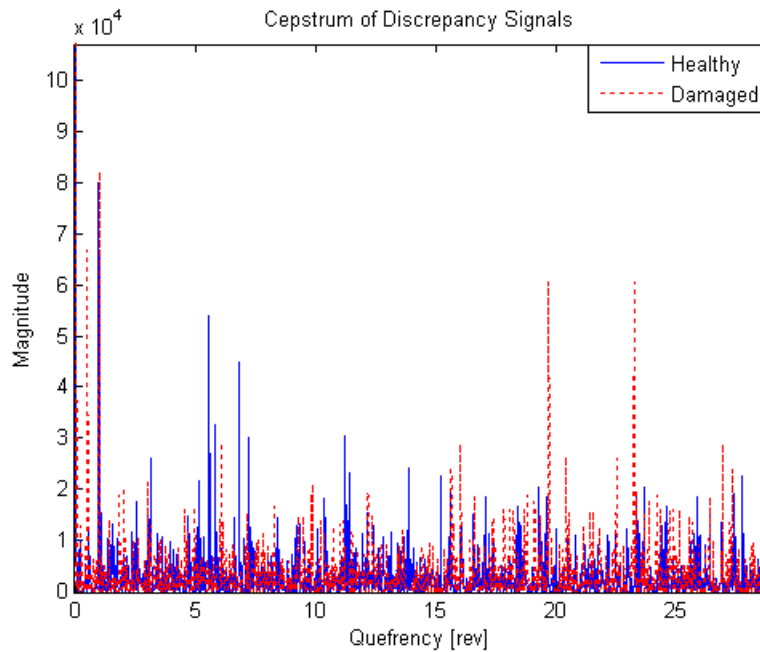


Figure 4.36: Cepstra of discrepancy signal for both healthy and damaged signals

4.2.8. Alternative Fault Types

Up until now only the dataset with the seeded tooth root rack has been analysed, but the question must be asked if the proposed methodology is also able to detect and identify other forms of faults. Thus the datasets based on the seed tooth chip and surface pitting will now be briefly analysed. All results will be based on WPT features using the frequencies from the FRF.

Tooth chip

A waterfall plot tracking the SA of the discrepancy up until failure for the dataset with the seeded tooth chip can be seen in Figure 4.37. It must be noted that the failure mechanism of this dataset was uncertain because the gear shaft sheared as well as the tooth breaking, hence it was unable to clarify which occurred first. The SA contains a lot of discrepancy peaks, however the general trend is one large peak, which is assumed to correspond to the seeded local fault. Also there is not the sudden step up in severity as seen in the previous waterfall plot in Figure 4.33, which lends to the view that the shaft sheared before the tooth chipped since there was no sudden tooth breakage. Thus the tooth breakage was as a result of the shaft shearing. The weakness is the shaft is therefore probably responsible for the uncorrelated peaks in the discrepancy signal. Therefore it can be concluded that the algorithm is able to detect the discrepancy caused by faults other than localised gear tooth faults, but not necessarily classify the nature of the fault. This also highlights a weakness in the proposed methodology namely that it can detect faults, but not necessarily identify the true fault when there are multiple faults occurring simultaneously.

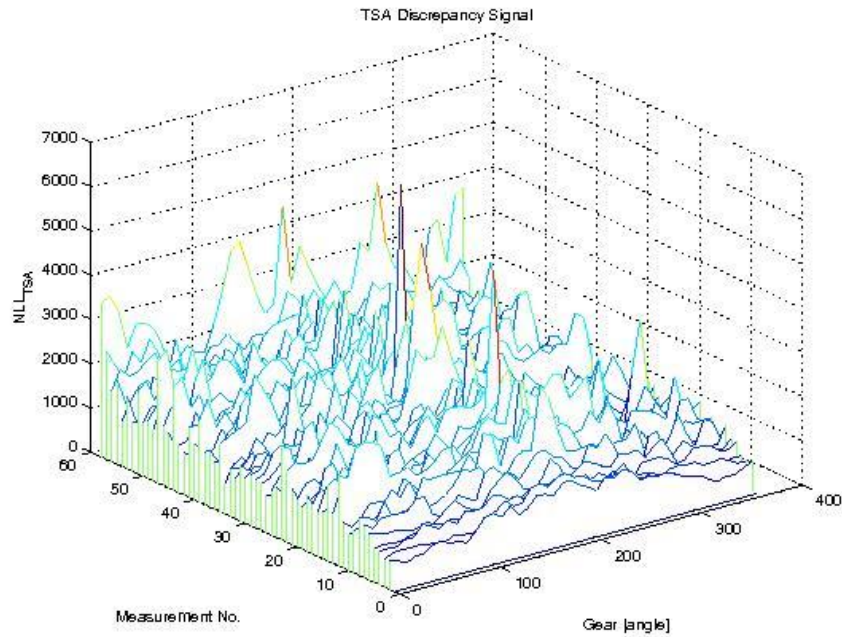


Figure 4.37: Waterfall plot of SA of discrepancy of tooth chip dataset

Surface pitting

The next seeded fault was seeded surface pitting induced by corrosion. Sadly again the shaft sheared as well as a tooth chip making the exact failure mechanism unknown, but it is reckoned that the shaft sheared before the gear tooth chipped because it was the second time it occurred and the time to failure was fairly fast. The waterfall plot of the SA of the discrepancy signal can be seen in Figure 4.38. The first observation is the large increase in discrepancy of the final measurement indicating imminent failure. It is good to see that the signal is fairly smooth over the full gear revolution, thus not indicating a localised fault but more of an uniform fault. There is a slight peak, but that is primarily due to the signal reshuffling to account for order tracking error. Therefore the proposed methodology is able to detect both local and uniform gear faults. However similar to the previous results in Figure 4.37, it is challenging to identify the exact type of fault from the SA of the discrepancy signal.

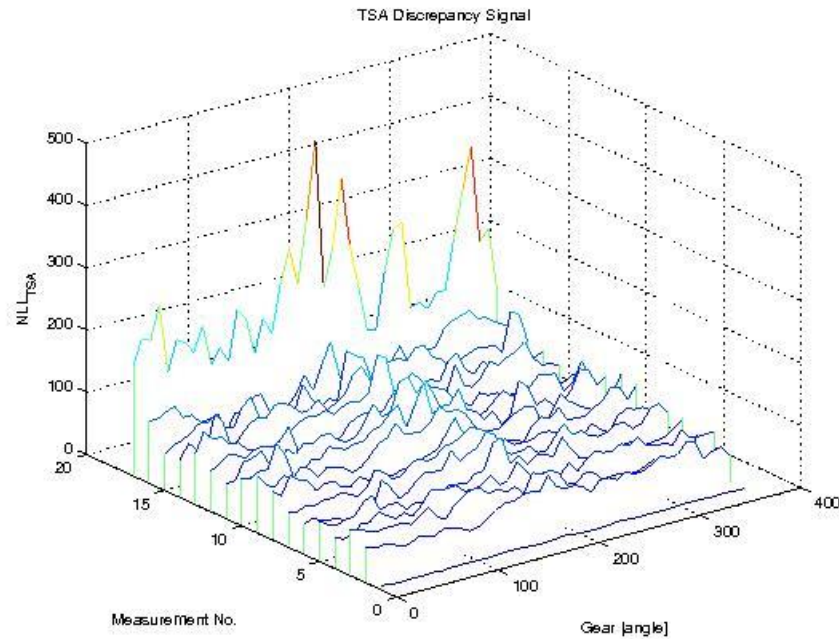


Figure 4.38: Waterfall plot of SA of discrepancy of surface pitting dataset

4.3. Conclusion

This concludes the evaluation of the results of the proposed method for the first damage set. In general the algorithm fulfilled all the specified aims set out in chapter 2. It was able to identify the operating conditions and classify signal accordingly. The smart operating condition features were sensitive to load and speed, however they were not sufficiently robust to the presence of a fault. However it is noticed that accurate operating condition classification is not crucial to the accurate performance of the methodology. It was able to detect and diagnose the fault in a rotating machine operating under non-stationary conditions. The smart fault features were both sensitive to the presence of a fault and robust to fluctuating operating conditions. There was no historical data of faults for the experimental setup, and the features selection was based on the understanding of fault mechanisms, system properties and characteristics of the healthy dynamic response. The algorithm was able to locate the faults to some extent, however the detection and diagnosis of the initial fault stages requires more work. Finally, the algorithm was capable to quantify the relative severity of faults, thus enabling it to track fault progression. From the tooth chip and surface pitting faults, it was concluded that the methodology was able to detect the discrepancy cause by other fault mechanism, however it was an intricate process to be able to identify the type of fault from the discrepancy signal. However in general the proposed novel discrepancy algorithm was successful in fulfilling the pre-set targets.

5. Conclusion

This dissertation proposes a novel fault detection and diagnosis methodology for rotating machinery that operate under non-stationary operating conditions. The methodology produces a discrepancy signal that is conditionally independent of the operating conditions. This is accomplished by identifying the instantaneous operating conditions using operating condition sensitive features and a HMM. The discrepancy signal is then generated from a set of GMMs, one for each of the respective identified operating conditions. A variety of fault sensitive features for the GMMs were investigated that were based on the physical understanding of the fault mechanisms and robust to operating conditions. It was concluded that the RMS values of the WPT coefficients at the natural frequencies of the systems proved to be the most successful at detecting faults. The natural frequencies were determined using both a modal impact test to generate the FRF of the system and by analysing the dominant frequency bands in the spectrogram that did not change with respect to operating conditions. The discrepancy signal produced by the GMMs allows faults to be detected, diagnosed and quantified through simple signal processing techniques such as SA and frequency analysis.

The methodology was verified using results from both a simulated lumped mass model of a gearbox and an experimental accelerated life test of a gearbox. The lumped mass model was based on a model proposed by Chaari et al (Chaari, Bartelmus, Zimroz, Fakhfakh, & Haddar, 2012), and incorporates an 8 DOF model that operates under fluctuating operating conditions. Gear damaged was modelled by reducing the stiffness of an individual gear tooth. The accelerate life test set up was based on the design by Stander (Stander, 2005). The experiment consisted of a gearbox with seeded faults operating under fluctuating load and speed conditions. Also the load was above the rated load of the gear in order to accelerate the remaining useful life of the gear. The data gathered from the experiments was fairly comprehensive and is suitable for other projects. Both datasets were imperative to the verification of the methodology to detect, diagnose and quantify gear faults under fluctuating operating conditions.

The discrepancy signals produced by the methodology when implemented on the signals created by the lumped mass model were very encouraging. The method was able to successfully classify the instantaneous operating conditions of the system. Also independent of the selected features, the method was able to detect and locate the fault in the system. The method was also able to handle the fluctuating operating conditions extremely well, as there was minimal effect of the operating conditions in the discrepancy signal. It also managed to operate effectively under a range of SNR, thus displaying its robustness and ability to operate in real world environments. The simulated results clearly exhibited its ability to quantify the severity of faults as seen in the waterfall plot. The only negative result of the simulated test was that it produced good discrepancy signals independent of the features and it is known that, in practice, feature selection is critical to the success of fault detection. In general, the discrepancy signals generated from the simulated model clearly displayed the feasibility and success of the methodology, however actual data was still needed to fully verify the methodology.

The results from the experimental accelerated life test were much more realistic and offered greater insight into the verification of the methodology. Firstly, the method was able to classify the instantaneous operating conditions with good accuracy. However it was concluded that accurate identification of the instantaneous operating conditions was not critical to the overall performance of

the methodology. Secondly, the importance of correct feature selection was clearly demonstrated by the different discrepancy signals generated using time domain and WPT features. Similarly to the results of the simulated model, the methodology was more than capable of handling the non-stationary operating conditions and noise within the measured signals. The SA and spectra of the discrepancy signal were able to detect, diagnose and quantify the severity of the fault within the gearbox. The method was not perfect as often it detected unexpected discrepancy peaks, which is attributed to unknown faults within the system and the general running process of the new gear. However, when analysed over time, the incorrect discrepancy peaks can be easily disregarded and the peaks that are detected regularly clearly reveal the presence of a fault. In conclusion the methodology fulfils all of its aims to detect, diagnose and quantify gear faults. There is, however, room for improvement.

The results produced by the methodology for the experimental measurements display the need for improvement to reduce the amount of unexpected peaks in the discrepancy signal. The first area of improvement is implementing a graphical model to replace the GMM and HMM, so that the operating condition and discrepancy estimation occur in a single model. Another area that can be investigated is the use of other wavelet families to detect fault impulses. There is an obvious need to broaden the scope of the methodology to be able to detect and diagnose a greater variety of faults found in rotating machines, such as bearing faults, misalignment, etc. Also further investigation is required into the discrepancy signal to be able to extract more diagnostic information from it. There is also significant research needed to be conducted on more complex machinery and measurement methods.

In final conclusion, the proposed methodology of this dissertation was able to fulfil all the initial aims set forward. It was able to clearly detect faults from the vibration signal of a gearbox operating under non-stationary conditions. It used only a single axis accelerometer and no measure of the shaft location. It was successfully able to diagnose faults with no historical fault data, but only using readily available healthy data to train the model. Feature selection plays a significant part in the accuracy of the method, and it was found that using either the FRF or spectrogram were useful in gaining a significant amount of understanding of the physical properties of the system. Using separate features for the operating condition classification and fault detection was critical to the overall performance of the method. The faults were easily detectable and diagnosable using the simple signal processing techniques of SA and frequency analysis. Finally the method was more than capable to quantifying the severity of the faults, thus allowing reasonable estimations of the remaining useful life.

Bibliography

- Aherwar, A., & Khalid, S. (2012). Vibration Analysis Techniques for Gearbox Diagnostic: A Review. *International Journal of Advanced Engineering Technology*, III(II), 04–12.
- Bartelmus, W., & Zimroz, R. (2009a). Vibration condition monitoring of planetary gearbox under varying external load. *Mechanical Systems and Signal Processing*, 23(1), 246–257. doi:10.1016/j.ymssp.2008.03.016
- Bartelmus, W., & Zimroz, R. (2009b). A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing*, 23(5), 1528–1534. doi:10.1016/j.ymssp.2009.01.014
- Bechhoefer, E., & Kingsley, M. (2009). A Review of Time Synchronous Average Algorithms. *Annual Conference of the Prognostics and Health Management Society* (pp. 1–10).
- Bozchalooi, S. I., & Liang, M. (2007). Identification of the high SNR frequency band for bearing fault signature enhancement. *2007 14th International Conference on Mechatronics and Machine Vision in Practice*, 39–43. doi:10.1109/MMVIP.2007.4430711
- Cartella, F., Liu, T., Meganck, S., Lemeire, J., & Sahli, H. (2012). Online adaptive learning of Left-Right Continuous HMM for bearings condition assessment. *Journal of Physics: Conference Series*, 364, 012031. doi:10.1088/1742-6596/364/1/012031
- Chaari, F., Baccar, W., Abbes, M. S., & Haddar, M. (2008). Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. *European Journal of Mechanics - A/Solids*, 27(4), 691–705. doi:10.1016/j.euromechsol.2007.11.005
- Chaari, F., Bartelmus, W., Zimroz, R., Fakhfakh, T., & Haddar, M. (2012). Gearbox vibration signal amplitude and frequency modulation. *Shock and Vibration*, 19, 635–652. doi:10.3233/SAV-2011-0656
- Cocconcelli, M., Zimroz, R., Rubini, R., & Bartelmus, W. (2012). STFT based approach for ball bearing fault detection in a varying speed motor. In T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, & M. Haddar (Eds.), *Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 41–50). Springer Berlin Heidelberg. doi:10.1007/978-3-642-28768-8_5
- Dalpiaz, G., Rivola, A., & Rubini, R. (1998). Gear Fault Monitoring : Comparison of Vibration Analysis Techniques. *3rd International Conference on Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques* (pp. 623–632).
- Eggers, B. L., Heyns, P. S., & Stander, C. J. (2007). Using computed order tracking to detect gear condition aboard a dragline. *The Journal of The South African Institute of Mining and Metallurgy*, 107, 1–8.
- Elbarghathi, F., Wang, T., Zhen, D., Gu, F., & Ball, A. (2012). Two Stage Helical Gearbox Fault Detection and Diagnosis based on Continuous Wavelet Transformation of Time Synchronous Averaged Vibration Signals. *Journal of Physics: Conference Series*.

- Fyfe, K. R., & Munck, E. D. S. (1997). Analysis of Computed Order Tracking. *Mechanical Systems and Signal Processing*, 11(2), 187–205. doi:10.1006/mssp.1996.0056
- Giurgiutiu, V., & Yu, L. (2003). Comparison of Short-time Fourier Transform and Wavelet Transform of Transient and Tone Burst Wave Propagation Signals For Structural Health Monitoring. *4th International Workshop on Structural Health Monitoring* (pp. 1–9). Stanford, CA.
- Grossmann, A., & Morlet, J. (1984). Decomposition of Functions into Wavelets of Constant Shape and Related Transform. In L. Streit (Ed.), *Mathematics + Physics, Lectures on Recent Results* (1st ed.). Singapore: World Scientific Publishing Co.
- Guan, L., Shao, Y., Gu, F., Fazenda, B., & Ball, A. (2009). Gearbox Fault Diagnosis under Different Operating Conditions Based on Time Synchronous Average and Ensemble Empirical Mode Decomposition. *SICE Annual Conference* (pp. 1–6).
- Guo, Y. (2010). Gear fault diagnosis of wind turbine based on discrete wavelet transform. *8th World Congress on Intelligent Control and Automation* (pp. 5804–5808). Jinan, China: Ieee. doi:10.1109/WCICA.2010.5554606
- He, Q., Yan, R., & Gao, R. X. (2010). Wavelet Packet Base Selection for Gearbox Defect Severity Classification. *IEEE Prognostics & System Health Management Conference*. Macau, China.
- Heyns, T. (2012). *Low Cost Condition Monitoring under Time-Varying Operating Conditions*. University of Pretoria.
- Heyns, T., Godsill, S. J., De Villiers, J. P., & Heyns, P. S. (2012). Statistical gear health analysis which is robust to fluctuating loads and operating speeds. *Mechanical Systems and Signal Processing*, 27, 651–666. doi:10.1016/j.ymssp.2011.09.007
- Heyns, T., & Heyns, P. S. (2012). Gear fault detection under fluctuating operating conditions by means of discrepancy analysis. In T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, & M. Hadder (Eds.), *Condition Monitoring of Machinery in Non-Stationary Operations* (First Edit., pp. 81–88). Springer.
- Heyns, T., Heyns, P. S., & De Villiers, J. P. (2012). Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. *Mechanical Systems and Signal Processing*, 32, 200–215. doi:10.1016/j.ymssp.2012.05.008
- Kang, J.-S., Zhang, X.-H., & Wang, Y.-J. (2011). Continuous hidden Markov model based gear fault diagnosis and incipient fault detection. *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering* (pp. 486–491). Ieee. doi:10.1109/ICQR2MSE.2011.5976659
- Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 11(2), 2300–2312. doi:10.1016/j.asoc.2010.08.011
- Li, F., Meng, G., Ye, L., & Chen, P. (2008). Wavelet Transform-based Higher-order Statistics for Fault Diagnosis in Rolling Element Bearings. *Journal of Vibration and Control*, 14(11), 1691–1709. doi:10.1177/1077546308091214

- Lou, X., & Loparo, K. a. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*, 18(5), 1077–1095. doi:10.1016/S0888-3270(03)00077-3
- Mallat, S. G. (1989). A Theory for Multiresolution Signal Decomposition : The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693.
- Marwala, T., Mahola, U., & Nelwamondo, F. V. (2006). Hidden Markov Models and Gaussian Mixture Models for Bearing Fault Detection Using Fractals. *International Joint Conference on Neural Networks* (pp. 5876–5881). Vancouver, Canada.
- McFadden, P. D., & Smith, J. D. (1984). Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review. *Tribology International*, 17(1), 3–10. doi:10.1016/0301-679X(84)90076-8
- Miao, Q., & Makis, V. (2007). Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models. *Mechanical Systems and Signal Processing*, 21(2), 840–855. doi:10.1016/j.ymssp.2006.01.009
- Miao, Q., Xie, L., Chen, Y., Liang, W., & Pecht, M. (2012). Fan bearing fault diagnosis based on continuous wavelet transform and autocorrelation. *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference* (pp. 1–6). Beijing: IEEE. doi:10.1109/PHM.2012.6228837
- Nese, S. V., Kilic, O., & Akinci, T. C. (2012). Analysis of wind turbine blade deformation with STFT method. *Energy Education Science and Technology Part A: Energy Science and Research*, 29(1), 679–686.
- Ocak, H., & Loparo, K. A. (2001). A New Bearing Fault Detection and Diagnosis Scheme based on Hidden Markov Modeling of Vibration Signals. *IEEE, ICASSP* (pp. 3141–3144, vol. 5).
- Ocak, H., Loparo, K. a., & Discenzo, F. M. (2007). Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics. *Journal of Sound and Vibration*, 302(4-5), 951–961. doi:10.1016/j.jsv.2007.01.001
- Peng, Z. K., & Chu, F. L. (2003). Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing*, 18(2), 199–221. doi:10.1016/S0888-3270(03)00075-X
- Prabhakar, S., Mohanty, A. R., & Sekhar, A. . (2002). Application of Discrete Wavelet Transform for Detection of Ball Bearing Race Faults. *Tribology International*, 35(12), 793–800. doi:10.1016/S0301-679X(02)00063-4
- Purushotham, V., Narayanan, S., & Prasad, S. a. N. (2005). Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT & E International*, 38(8), 654–664. doi:10.1016/j.ndteint.2005.04.003
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE*, 77(2).

- Rafiee, J., Rafiee, M. a., & Tse, P. W. (2010). Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, 37(6), 4568–4579. doi:10.1016/j.eswa.2009.12.051
- Rajagopalan, S., Restrepo, J. a., Aller, J. M., Habetler, T. G., & Harley, R. G. (2005). Selecting time-frequency representations for detecting rotor faults in BLDC motors operating under rapidly varying operating conditions. *31st Annual Conference of IEEE Industrial Electronics Society, IECON* (pp. 2585–2590). Ieee. doi:10.1109/IECON.2005.1569314
- Randall, R. B. (1989). *Frequency Analysis* (3rd Editio.). Bruel & Kjaer.
- Randall, R. B. (2011). *Vibration-based Condition Monitoring* (First Edit., pp. 1–309). Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/9780470977668
- Rubini, R., & Meneghetti, U. (2001). Application of the Envelope and Wavelet Transform Analyses for the Diagnosis of Incipient Faults in Ball Bearings. *Mechanical Systems and Signal Processing*, 15(2), 287–302. doi:10.1006/mssp.2000.1330
- Safizadeh, M. S., Lakis, A. A., & Thomas, M. (2000). Using Short-Time Fourier Transform in Machinery Diagnosis. *International Journal of COMADE*, 3(1), 1–14.
- Sanz, J., Perera, R., & Huerta, C. (2007). Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms. *Journal of Sound and Vibration*, 302(4-5), 981–999. doi:10.1016/j.jsv.2007.01.006
- Shiple, E. E. (1967). Gear Failures.
- Stander, C. J. (2005). *Condition Monitoring of Gearboxes Operating Under Fluctuating Load Conditions*. University of Pretoria.
- Stander, C. J., Heyns, P. S., & Schoombie, W. (2002). Using Vibration Monitoring for Local Fault Detection on Gears Operating Under Fluctuating Load Conditions. *Mechanical Systems and Signal Processing*, 16(6), 1005–1024. doi:10.1006/mssp.2002.1479
- Staszewski, W. J., & Tomlinson, G. R. (1994). Application of the Wavelet Transform to Fault Detection in a Spur Gear. *Mechanical Systems and Signal Processing*, 8(3), 289–307.
- Urbanek, J., Barszcz, T., & Antoni, J. (2013). A two-step procedure for estimation of instantaneous rotational speed with large fluctuations. *Mechanical Systems and Signal Processing*, 38(1), 96–102. doi:10.1016/j.ymssp.2012.05.009
- Wang, X., Makis, V., & Yang, M. (2010a). A wavelet approach to fault diagnosis of a gearbox under varying load conditions. *Journal of Sound and Vibration*, 329(9), 1570–1585. doi:10.1016/j.jsv.2009.11.010
- Wang, X., Makis, V., & Yang, M. (2010b). A wavelet approach to fault diagnosis of a gearbox under varying load conditions. *Journal of Sound and Vibration*, 329(9), 1570–1585. doi:10.1016/j.jsv.2009.11.010

- Wu, J.-D., & Liu, C.-H. (2008). Investigation of engine fault diagnosis using discrete wavelet transform and neural network. *Expert Systems with Applications*, 35(3), 1200–1213. doi:10.1016/j.eswa.2007.08.021
- Xinmin, T., Baoxiang, D., & Yong, X. (2007). Bearings Fault Diagnosis based on GMM Model using Lyapunov Exponent Spectrum. *The 33rd Annual Conference of the IEEE Industrial Electronics Society* (pp. 2666–2671). Taipei, Taiwan.
- Yan, R., Gao, R. X., & Chen, X. (2013). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, 1–15. doi:10.1016/j.sigpro.2013.04.015
- Zhang, Z., Wang, Y., & Wang, K. (2012). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-012-0657-2
- Zhao, M., Lin, J., Wang, X., Lei, Y., & Cao, J. (2013). A tacho-less order tracking technique for large speed variations. *Mechanical Systems and Signal Processing*, 40(1), 76–90. doi:10.1016/j.ymsp.2013.03.024

Appendix A: Experimental Setup

A.1. External Conditions

The experiment will be run in the Sasol Labs at the University of Pretoria. The ambient temperature of the laboratory is in the region of 24°C and only fluctuates by a couple of degrees throughout the day, which will have insignificant effects on the system because the lubrication and motor are rated to operate in temperatures in excess of 40°C.

A.2. List of Measurements

The time schedule for all the experiments can be seen in the table below

Table 1: Beginning and end dates of measurement sets

Start Date	End Date	Experiment
2013-11-21	2013-11-22	Healthy gearbox
2013-11-25	2013-11-28	Tooth root crack
2013-12-06	2013-12-07	Tooth chip
2014-02-27	2014-02-27	Surface pitting

A.3. File Name Convention

The file name convention used to keep track of data files will contain the fault number, date and time of the measurement. The format of the file name will be *Measurement yyyy-mm-dd HH mm ss.mat* and an example is *Healthy 2013-09-18 15 35 10*.

A.4. Safety Precautions

The experiment will be operating 24 hours a day to ensure optimum amount of testing and failure of the component, however this entails that the experiment will not be continuously under human supervision. Thus the need for automatic monitoring arises, to shut down the system in case of an emergency. Since the entire system is controllable and observed from the PC, it can easily be remotely monitored using the software package TeamViewer, thus allowing the measured vibration signal to be observed, control of the motor and alternator to be remotely accessible through the internet. The PC also was equipped with a webcam allowing a visual view of the gearbox as well. Finally there was an emergency stop button next to the system connected directly to the analog speed controller to stop the system immediately in the case of emergency.

Students and personal in the labs also required protection from the experiment, specific from noise generation and moving parts. Thus a frame with Perspex glass was place over all the rotating parts which prevented access to the parts while in operation and suppress noise, while not reducing the visibility of the system.

A.5. Additional Time Domain Results

A.5.1. Shaft speed measurement

The shaft speed profile of the input shaft to the test gearbox can be seen in the figure below. Both the results from the tachometer and shaft encoder can be seen in the figure below. It is noticed that there is a lot of fluctuation in the speed of the shaft that is detected by the shaft encoder because of its

greater number of pulses per revolution, while the tachometer gives a nice smooth version of the speed profile of the shaft.

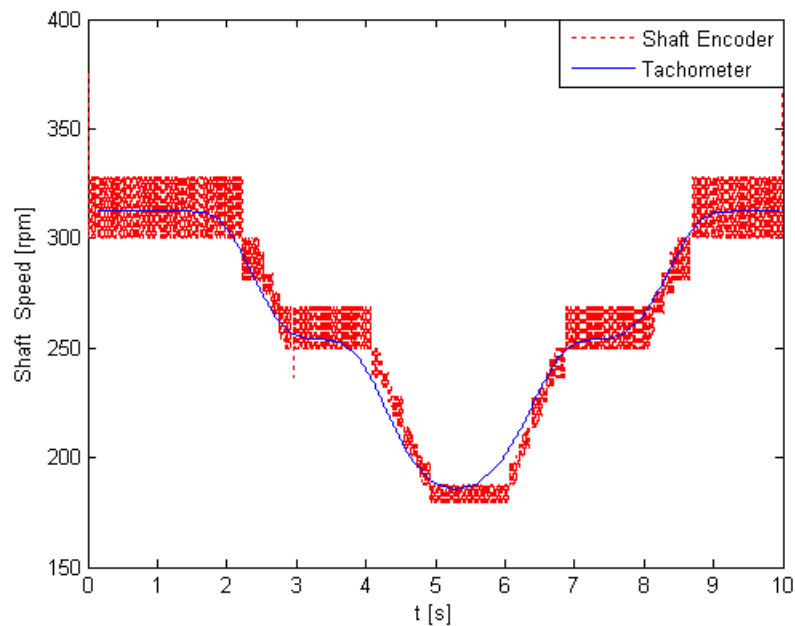


Figure 1: Speed profile of the system from both the tachometer and shaft encoder

A.5.2. Strain gauge measurement

The time signal for the strain measured on the motor foot can be seen in the figure below. The minor variation in stress amplitude is clear evidence of the fluctuating conditions.

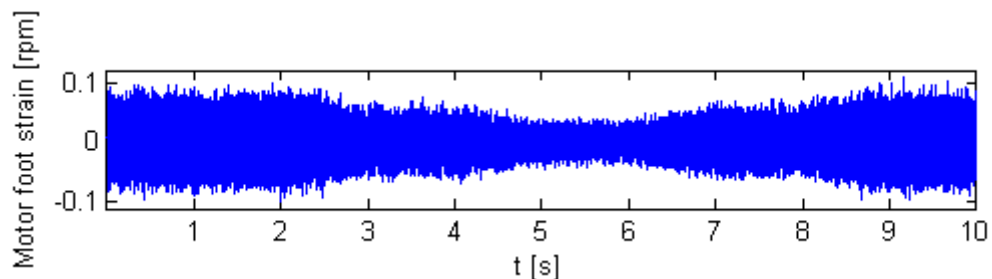


Figure 2: Time domain strain measurement of motor foot

A.5.3. Alternator power measurement

The figure below is the current, voltage and power output of the alternator. The load on the gearbox can be estimated from the instantaneous speed and power output the system. The power output of the alternator is the simplest and most reliable method to determine the load on the system.

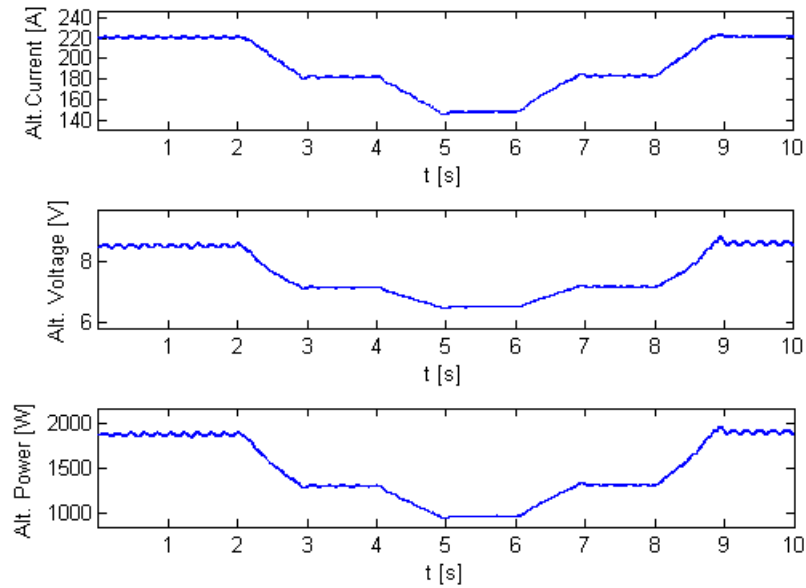


Figure 3: Current, voltage and power output of alternator

A.5.4. Load Measurement

There are many ways to measure the load on the gearbox. Firstly it can be deduced that the load is directly proportional to the strain experience by the motor foot. Secondly the load can be estimated from the output power of the alternator. Thirdly the load can be estimated to be proportional to the instantaneous speed difference between the tachometer and shaft encoder. The above load estimation techniques can be seen in the figure below. None of the figures below are the clear display of the expect load, however they all give good indications of the load profile. The strain plot may seem noisy, but the absolute amplitude gives a good measure of the load acting on the system. The measure of the load generated by the alternator is the best feature to quantify load, however it might be better to use the power output of the alternator, as seen in the figure above, than the load in the figure below. Sadly the instantaneous speed difference does not give a clear indication of the load, this is primarily due to the fluctuating speed profile that the shaft encoder picks up but the tachometer misses.

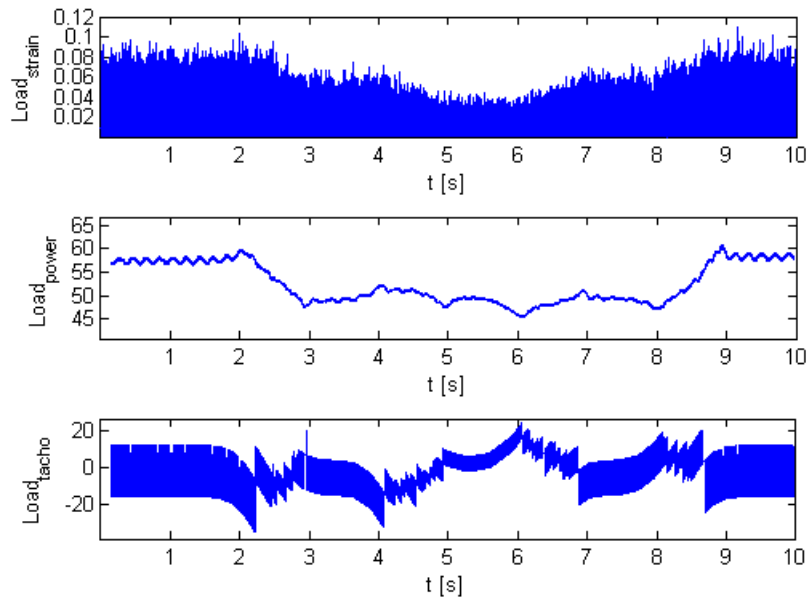


Figure 4: Different measures of load on the system