Condition Monitoring of Industrial Machines Using Wavelet Packets and Intelligent Multisensor Fusion

by

Srinivas Raman

B.A.Sc., University of British Columbia 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Mechanical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2009

Abstract

Machine condition monitoring is an increasingly important area of research and plays an integral role in the economic competitiveness in many industries. Machine breakdown can lead to many adverse effects including increased operation and maintenance costs, reduced production output, decreased product quality and even human injury or death in the event of a catastrophic failure. As a way to overcome these problems, an automated machine diagnostics scheme may be implemented, which will continuously monitor machine health for the purpose of prediction, detection, and diagnosis of faults and malfunctions. In this work, a signal-based condition monitoring scheme is developed and tested on an industrial fish processing machine. A variety of faults are investigated including catastrophic on-off type failures, partial faults in gearbox components and sensor failures. The development of the condition monitoring scheme is divided into three distinct subtasks: signal acquisition and representation, feature reduction, and classifier design.

For signal acquisition, the machine is instrumented with multiple sensors to accommodate sensor failure and increase the reliability of diagnosis. Vibration and sound signals are continuously acquired from four accelerometers and four microphones placed at strategic locations on the machine. The signals are efficiently represented using the wavelet packet transform and node energies are used to generate a feature vector. A measure for feature discriminant ability is chosen and the effect of choosing different analyzing wavelets is investigated.

Since the dimensionality of the feature vector can become very large in multisensor applications, various means of feature reduction are investigated to reduce the computational cost and improve the classification accuracy. Local Discriminant Bases, a popular and complementary approach to wavelet-based feature selection is introduced and the drawbacks in the context of multisensor applications are highlighted. To address these issues, a genetic algorithm is proposed for feature selection in robust condition monitoring applications. The fitness function of the genetic algorithm consists of three criteria that are considered to be important in fault classification: feature set size, discriminant ability, and sensor diversity. A procedure to adjust the weights is presented. The feature selection scheme is validated using a data set consisting of one healthy machine condition and five faulty conditions.

For classifier design, the theoretical foundations of two popular non-linear classifiers are presented. The performance of Support Vector Machines (SVM) and Radial Basis Function (RBF) networks are compared using features obtained from a filter selection scheme and a wrapper selection scheme. The classifier accuracy is determined under conditions of complete sensor data and corrupted sensor data. Different kernel functions are applied in the SVM to determine the effect of kernel variability on the classifier performance.

Finally, key areas of improvement in instrumentation, signal processing, feature selection, and classifier design are highlighted and suggestions are made for future research directions.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

IAL	Industrial Automation Laboratory
LAN	Local Area Network
VFD	Variable Frequency Drive
CCD	Charge Coupled Device
OD	Outer Diameter
SIFT	Scale Invariant Feature Tracking
FFT	Fast Fourier Transform
DWT	Discrete Wavelet Transform
PCA	Principal Components Analysis
LDA	Linear Discriminant Analysis
WPT	Wavelet Packet Transform
LDB	Local Discriminant Bases
DPP	Dictionary Projection Pursuit
MI	Mutual Information
GA	Genetic Algorithm
ANN	Artificial Neural Networks
RBFN	Radial Basis Function Network
SVM	Support Vector Machine
FPGA	Field Programmable Gate Array
DAQ	Data Acquisition
VI	Virtual Instrument
STFT	Short Time Fourier Transform
CWT	Continuous Wavelet Transform
K-L	Karhunen-Loeve
KKT	Karush-Kuhn-Tucker

Acknowledgements

First, I would like thank Professor Clarence W. de Silva for supervising me these last two years. I am tremendously grateful to him for his skillful guidance and unwavering support of my academic and career goals. His academic accomplishments and humanitarian efforts are a true source of inspiration to me and I consider myself very lucky to have stumbled into his laboratory (Industrial Automation Laboratory—IAL) and benefited from his engineering contributions.

Other faculty I would like thank are Mr. Jon Mikkelsen, for all his support during my undergraduate and graduate studies, members of my research committee: Dr. Farrokh Sassani and Dr. Farbod Khoshnoud (SOFTEK), and Dr. Lalith Gamage—a former Ph.D. student of IAL and currently a Visiting Professor with us, for their helpful advice and suggestions.

I wish to thank my colleagues at the IAL for their friendship over the last two years; notably, Ramon, Guan-lu, Behnam, Gamini, Tahir, Arun, Ying, and Roland. In particular, Ramon has been extremely helpful during the final stages of the project and I am very thankful for his support.

Financial support for my research project has come from grants held by Prof. de Silva, particularly through the: Tier 1 Canada Research Chair (CRC), Canada Foundation for Innovation (CFI), British Columbia Knowledge Development Fund (BCKDF), and the Discovery Grant of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Words cannot describe how indebted I am to my parents. Short of writing my thesis, they have supported me in every way possible. From helping me change greasy gearboxes to encouraging me towards the finish line, they have done it all. I can only hope to be as dedicated, loving and caring a parent as they are. I dedicate this thesis to them.

Chapter 1

Introduction to Condition Monitoring

1.1. Rationale for Machine Condition Monitoring

In many industries, particularly those related to manufacturing and production, machine malfunction and failure is a cause for serious concern due to many reasons. Financially, it places a large burden on operation and maintenance costs. Production output and product quality can significantly degrade if the machine will directly affect the production process. In extreme situations, there can be human injury or death resulting from catastrophic machine failures. To prevent these serious problems, companies adopt various maintenance programs to monitor and service machinery and processes.

Maintenance strategies can be classified into several broad categories [1]:
1. Run to Failure: Machine maintenance is only performed when the machine has failed.
2. Scheduled Maintenance: Maintenance is performed at set time intervals.
3. Condition-Based Maintenance: Maintenance is performed according to the condition of the machine, as determined through a monitoring scheme.

It is estimated that maintenance costs contribute to approximately half of all operating costs in manufacturing and processing industries [1]. In view of this, the choice of an appropriate strategy of maintenance is rather important. The choice depends on many factors including the dependence of production level and quality on the machine condition, redundancy of machine and operations, lead-time for replacement of machine, safety of personnel and environment, and replacement costs of the machine. In particular, if the machine under consideration is crucial to the production process, has a high replacement cost, has a dangerous failure mode or is difficult to access (due to mobility or remote location), the use of a condition-based program is well justified.

Condition-based maintenance offers many potential advantages over other maintenance programs. These include increased availably and reliability of machines, higher operating

efficiencies, improved quality of products and services of the machine, lower downtimes, reduced maintenance costs, and improved safety. Traditionally, the disadvantages of condition-based maintenance programs included the high cost monitoring equipment, operational costs, and requirement of skilled personnel for operating and servicing the monitoring system. As electronics and software become more accessible and easy to use, engineers and technicians can afford to implement real-time monitoring systems at a fraction of what the cost used to be. As a result of the potential advantages, there has been a large effort to automate the maintenance process and move towards condition-based maintenance programs [2].

In implementing condition–based maintenance, certain machine parameters are monitored to determine if there a change in these parameters, which is indicative of machine failure. These parameters are measured by appropriate sensors and the condition diagnosis is performed by humans or by a computer-based system. If the machine is critical to the production process, continuous monitoring by a computer system may be necessary to provide an accurate and timely diagnosis about machine health. Once the diagnosis is determined, corrective action may be taken ranging from immediate machine shut-down to scheduling a maintenance procedure in the near future.

Certain complex machines may require the use of multiple sensors to capture machine health information. Wind turbines are examples of such machines because they contain numerous subsystems, each with their own modes of mechanical, electrical and structural failures (see Fig. 1.1). The use of more than one sensor is often necessary to make a robust and accurate diagnosis about the turbine condition. Also, wind turbines are challenging machines to maintain: they are frequently located in remote offshore locations and are subject to harsh environmental conditions. A faulty turbine may require a long period of time to access and repair due to the remote location of these machines. Therefore, sensor redundancy is also a useful feature, so that turbine performance can still be monitored if one or more sensors fail.

The European Commission and the Institut fur Solare Energieversorgungstechunik at the University of Kassel, Germany [3] have collaborated to develop a conditioning monitoring system to monitor the health of wind turbine systems. The system uses multiple accelerometers, strain gauges, speed sensors, and current sensors to continuously assess the performance and condition of the wind turbine. The signals are acquired by an onboard data acquisition system

and communicated to a central server using Ethernet LAN technology, where the data is analyzed using time series and frequency domain techniques. The condition monitoring systems have been implemented and tested in three wind farms in Germany. As a result of the project's success, more wind farms in Europe are upgrading or planning to upgrade their systems to include advanced fault diagnostics capabilities.
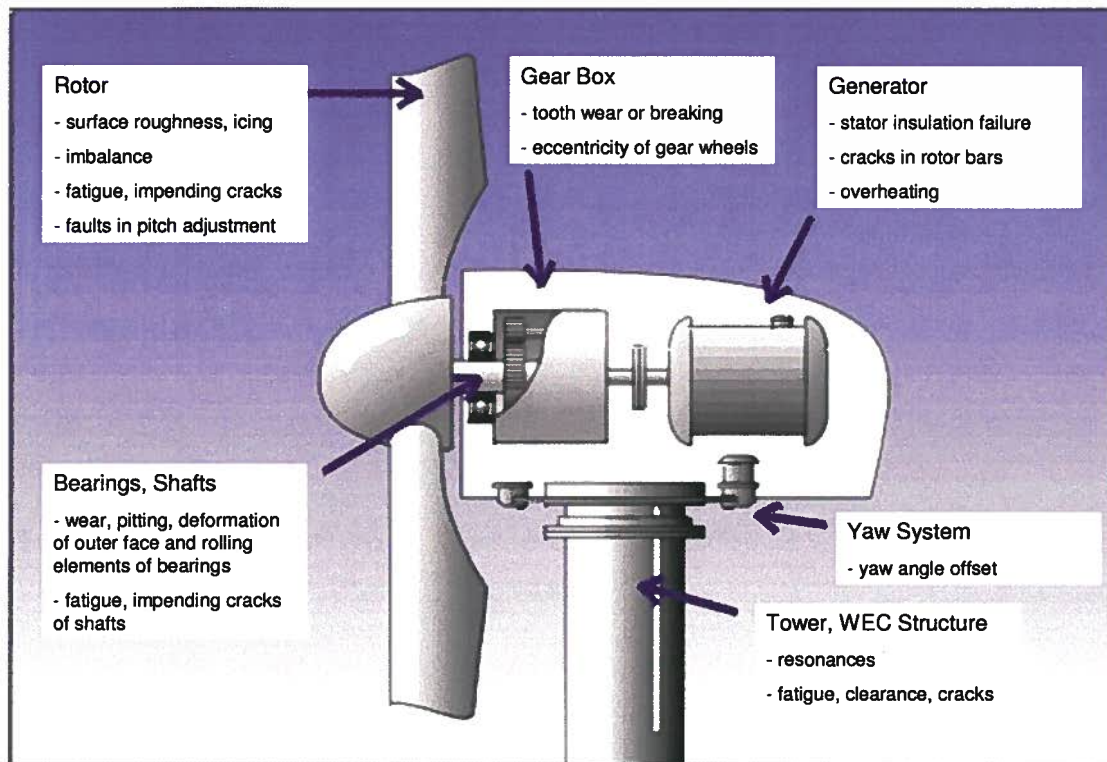
Figure 1.1: Potential Wind Turbine Faults. © P. Caselitz and J. Geibhardt, 2002, adapted by permission.

1.2. Objectives

As the previous example illustrated, multisensor-based condition monitoring may be required in applications where a single sensor may not be able to give a complete diagnosis of the machine condition and where sensors are prone to failure. The goal of the present thesis is to develop a multisensor condition monitoring scheme capable to:

- Acquire machine signals from multiple, heterogeneous sensors
- Represent the signals in a form that requires minimal computational resources and preserves faulty features
- Require minimal knowledge about physical characteristics of the machine
- Require minimal knowledge about physical characteristics of the fault

- Be implemented in real-time to minimize delay between failure and corrective action
- Classify the machine faults with high accuracy
- Be robust to disturbances, signal noise, and sensor failures,

To accomplish these objectives, vibration and audio signals are acquired from multiple accelerometers and microphones located at different positions in the prototype industrial machine (Iron Butcher). The signals are efficiently represented with the Wavelet Packet Transform to capture frequency-based fault signatures and a signal-based scheme is developed to avoid the requirement of complex model for machine and associated faults. An on-line monitoring system is implemented to provide real-time machine updates every three seconds. A feature selection scheme is developed to minimize the effect of sensor uncertainty, and various classification algorithms are implemented and tested to ensure a high accuracy of fault classification.

The overall goal of the research presented in this thesis is to develop a robust fault diagnosis scheme capable of diagnosing a wide range of faults in a complex industrial machine with minimal knowledge of machine and fault characteristics. A representative industrial machine is available in the Industrial Automation Laboratory, University of British Columbia, and will be used to develop and test the schemes of the present research.

1.3. Experimental System: Iron Butcher

Fish processing is a multibillion dollar industry in North America and a major industry in the province of British Columbia, Canada. In Canada alone, the annual value of the fish processing industry is estimated to be three billion dollars. The original Iron Butcher, a machine designed at the turn of the 20th century was widely used in the industry for the head cutting operation of salmon. This machine uses a primitive design, without sensing and feedback adjustments, which leads to significant (about 10%) wastage of useful meat and also degradation of product quality. To decrease the wastage of valuable fish, improve the product quality and process efficiency, and reduce the use of labor in hazardous routine operations, the Industrial Automation Lab (IAL) at the University of British Columbia designed a machine that would replace the original Iron Butcher. The new machine, termed the "Intelligent Iron Butcher" (see Fig. 1.2) has the following important features: high cutting accuracy, improved product quality, increased productivity and efficiency and flexible automation.

Figure 1.2: Intelligent Iron Butcher.

The machine consists of the following subsystems, each with its own specific function:

1. **Electromechanical conveying system**: The conveying system is responsible for transporting the fish from the loading zone to the cutting area. It is powered by an ac (alternating-current) induction motor coupled to a speed-reducing gearbox. The ac induction motor is controlled by a Variable Frequency Drive (VFD) and a speed transmission unit with a gear ratio 106.58:1 (see Fig. 1.3(a)). The rotational motion from the output of the gearbox is translated into an intermittent linear motion via a mechanical linkage attached to a sliding mechanism. The sliding mechanism contains a row of pins that fold in one direction only (Fig. 1.3(b)), so the fish on the machine conveyor will only move in the forward direction.

(a) Gearmotor and Linkage

(b) Sliding Mechanism

Figure 1.3: Electromechanical Conveying Unit.

2. **Hydraulic positioning system**: The hydraulic system is responsible for positioning the cutter blade assembly with respect to the fish head (Fig. 1.4). There are two double acting cylinders (one for positioning each axis of the cutter assembly). Each hydraulic cylinder is actuated via a three-way solenoid valve. The overall system is powered by a hydraulic pump.

(a) Hydraulic System Schematic   (b) Cutter Table
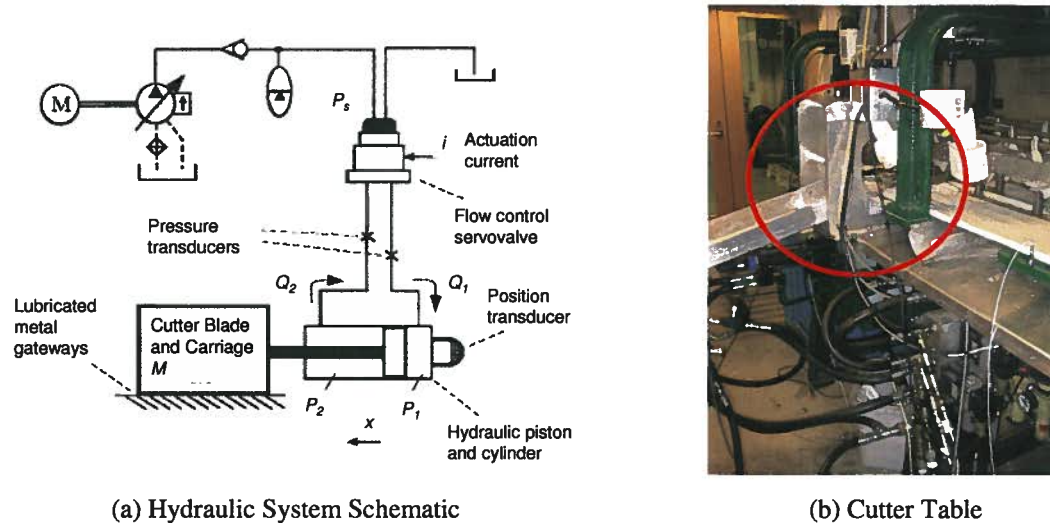
Figure 1.4: Hydraulic System.

3. **Pneumatic system**: The pneumatic system is responsible for two functions. One is to power the cutter blade, which cuts the fish head (see Fig. 1.5(a)). The other is to hold the fish stationary during the cutting operation (Fig. 1.5(b)). There are four pneumatic cylinders in total (three single acting cylinders for stabilizing the fish and one double acting cylinder for the cutting operation). The cylinders are actuated via a four-way, five-port double solenoid valve. The overall system is powered by a compressor.

(a) Cutter Blade   (b) Holding Mechanism
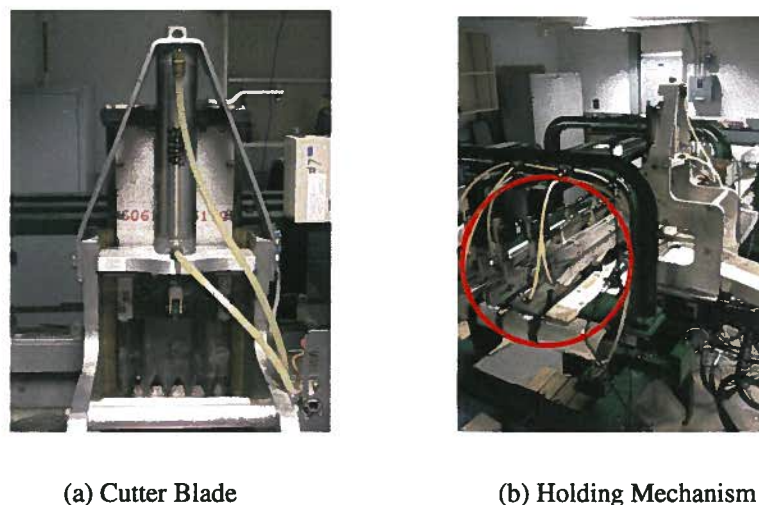
Figure 1.5: Pneumatic System.

Under normal conditions, the overall plant executes the following sequence of operations (see Fig. 1.6). Fish are manually fed by the operator into the loading zone of the Iron Butcher. The pins on the conveying table push the fish forward during the first half of the motion cycle. During the second half of the cycle, the pneumatic holder is activated and the fish is held down in place. At this time, there is one fish in the cutting zone and one fish in the standby zone. While the conveying pins move back, they do not move the fish because they are retractable in the reverse direction. The cutting operation occurs during this time. A primary CCD camera captures the image of the fish and a vision algorithm calculates the optimal cutting position to minimize fish meat wastage. The controller sets the reference x-y position and an electro-hydraulic manipulator accurately positions the cutter assembly accordingly. Once the assembly is in position, the pneumatic cutter cuts the fish head. The holder is then released and the cutter blade moves up after the fish head is cut. Then the conveyor mechanism begins the next cycle.
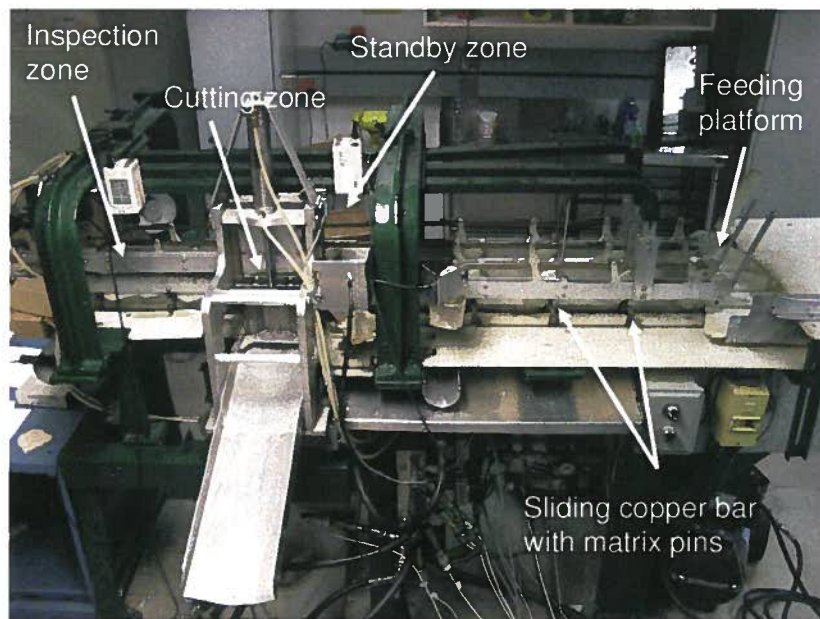
Figure 1.6: Machine Operation.

1.4. Investigated Faults

The Iron Butcher has multiple modes of failure associated with each of the subsystems. Table 1.1 summarizes some of possibilities. Since it is impractical to investigate all these faults in detail, a subset of these faults will be investigated to demonstrate the effectiveness of the proposed condition monitoring scheme. Other faults may be handled in a similar manner.

Table 1.1: Potential Faults in the Iron Butcher.

Major Sub-Systems	Potential Faults
Electromechanical Conveying System	- Induction motor failure - Gearbox failure - Linkage failure - Jammed fish
Hydraulic Cutter Assembly Positioning System	- Pump rotor/shaft failure - Motor failure - Proportional valve failure - Hydraulic actuator leakage
Pneumatic Powered Cutter and Fish Stabilizer	- Valve failure - Compressor rotor/shaft failure - Motor failure - Pneumatic actuator leakage

For the purposes of the present work, the following representative faults from three categories are investigated:

1. General on-off type faults: For testing on-off faults, the electromechanical conveying system and the hydraulic subsystem are turned off. On-off type faults can represent a catastrophic failure in the system. In the tested cases, it may be more practical to obtain other signals from the machine; e.g., current input, pressure transducer, etc. However, it should be noted that the installed sensors are not meant for this function and catastrophic failure detection may be considered an additional function of these sensors.

2. Common partial faults: For rotating machinery, faults are commonly found in three main components: shafts, bearings, and gears. For investigating these common faults, components in the gearbox (see Fig 1.7) have been modified to simulate these conditions. Detailed drawings and part numbers for the gearmotor can be found in Appendix A.

Figure 1.7: Gearmotor from SEW Eurodrive.

Bearing Damage: Bearings are one of the most important components in rotating machinery and are also the most susceptible to failure. Defects can appear in various bearing components including the outer race, inner race and the rolling elements. Bearing damage can be caused due to many reasons including excessive wear, corrosion, incorrect installation, mechanical shock and fatigue, misalignment, large electrical currents, and insufficient lubrication. To emulate bearing damage (see Fig. 1.8), the inner race in bearing #34 (in red) was ground and the rolling elements were deliberately damaged with a hammer. Also, the rolling elements and races of bearings #25, #37 and #45 (in blue) were sandblasted to simulate natural wear. See Appendix A for detailed information about the bearings.
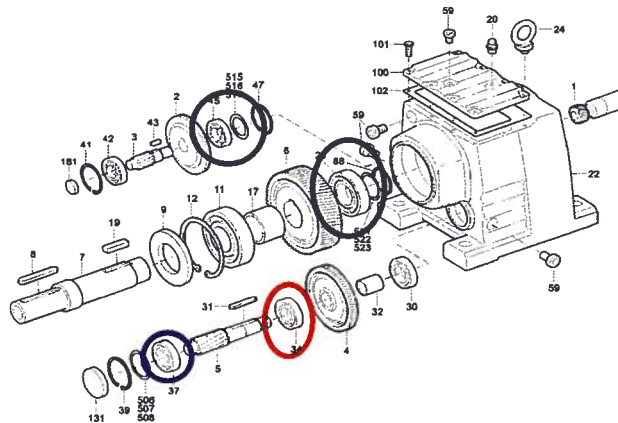
Figure 1.8: Location of Bearing Damage. © SEW Eurodrive, 2008, adapted by permission.

Shaft Misalignment: Misalignment between shafts can occur in various places in a machine. There are two kinds of misalignment: parallel misalignment and angular misalignment.

9

Parallel misalignment refers to the offset of a shaft axis from its correct position and angular misalignment refers to meeting of shaft centerlines at an angle. Misalignment can be caused due to incorrect installation and mechanical shock causing realignment. Misalignment can lead to excessive machine vibrations and high radial loads on bearings, causing premature failures in these components. To emulate parallel and angular shaft misalignment (see Fig. 1.9), shaft #17 and the outer diameter (OD) of bearing #11 was ground down 0.002" +- 0.0005" to allow a sideway shift of bearing #25. The OD of bearing #25 was ground down 0.0065" +- 0.0005" to allow shaft misalignment. Bearing #25 was shimmed on side A, resulting in the shift of shaft #7 toward the input pinion #1. Loctite was applied on the outside of bearings # 11 and # 25 to prevent the outer race from spinning. The end result of this process was the misalignment of shaft #7.
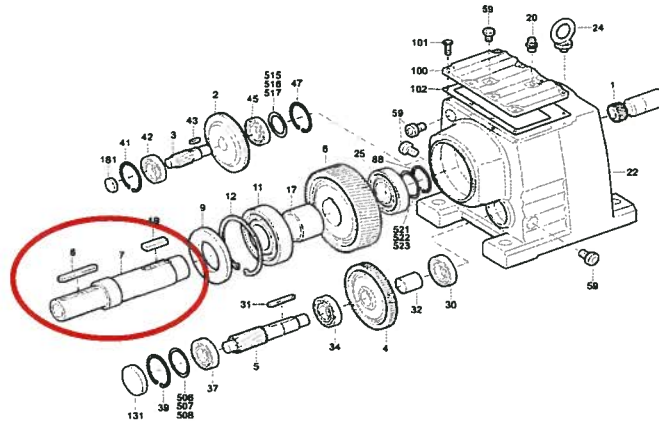
Figure 1.9: Location of Shaft Misalignment. © SEW Eurodrive, 2008, adapted by permission.

Gear Damage: Since gears transmit power from one component of the machine to another, there are significant forces on the gear teeth, making them especially susceptible to failure if there is a defective component. Gear defects can appear in various forms including non-uniform tooth wear, cracked, chipped or missing teeth, misalignment between teeth, backlash, and runout. The causes can be due to natural wear, operation outside normal range, and improper maintenance. To emulate defects in the gear teeth (see Fig. 1.10), pinion #34 (in blue) and gear #4 (in red) were hit with a hammer and the teeth were gauged with a grinder. Pinion #34 and gear #4 are helical gears with 13 teeth and 76 teeth, respectively (Appendix A).
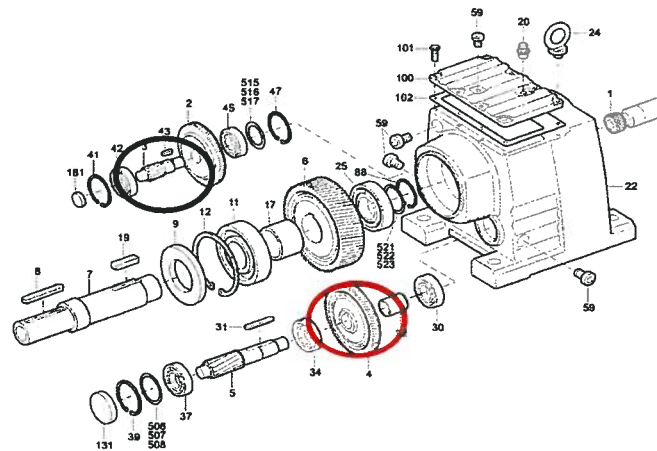
Figure 1.10: Location of Gear Damage. © SEW Eurodrive, 2008, adapted by permission.

3. Faults specific to fish cutting machine: Two possible faults specific to the fish processing machine are investigated. A dulling cutter blade and a jammed fish in the conveyor table are two likely modes of failure during machine operation. To investigate cutter blade dullness, two materials with different flexural strengths: acoustic soundboard and polystyrene, were used as the cutting material (see Fig. 1.11). The vibration profile and the cutting sound were compared to determine if the difference in cutting impact could be detected. To investigate a fish jam, a mock fish was held securely on the table so the pins would continuously hit the stationary fish, likely causing noticeable vibration in the machine.

(a) Polystyrene Insulation   (b) Acoustic Soundboard

Figure 1.11: Dull Blade Fault Simulation.

4. In addition to machine component faults, sensor faults ware also investigated. Remote locations and industrial environments may pose problematic conditions for sensors to operate

11

reliably. To determine the feasibility of the condition monitoring scheme developed in this thesis, the robustness of the scheme in the presence of sensor failures is investigated. Unlike the previous machine component faults, the sensor faults are simulated in the process of feature selection and classification by randomly setting the output of selected sensors to a zero value.

1.5. General Techniques of Fault Diagnosis

Schemes of fault diagnosis can be broadly classified into two categories: model-based schemes and model-free schemes, also known as signal-based schemes. Model-based schemes require a model of the plant being monitored to make a diagnosis whereas model-free systems only require a machine signal to make a diagnosis. The following sections outline the two approaches and discuss their relative merits.

1.5.1 Model Based Systems

Model-based systems of fault diagnosis require a "model" or approximate mathematical representation of the physical plant. The model can take different forms depending on the plant characteristics. Some forms of representation include [2]:

1. Physical equations: If the system is nonlinear and static, it can be represented in the form $\psi(y,u) = 0$. There exist similar relationships for dynamic, nonlinear and multiple-input/multiple-output systems

2. State equations of linear systems: Physical equations for dynamic, linear equations can be written as state and output equations:

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$

3. State observers: State observers allow for the approximation of the dynamic state of the system by knowing the system inputs and outputs. For the described discrete state-space model, a simple observer can be written in the form:

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + H\left[y(k) - \hat{y}(k) \right]$$

$$\hat{y}(k) = C\hat{x}(k)$$

12

4. Transfer functions: Physical equations for stationary systems can be written in the Laplace domain as a ratio of output $y(s)$ to input $u(s)$:

$$G(s) = \frac{y(s)}{u(s)}$$

5. Neural-network models: The system can be represented as a network of interconnected neurons known as a neural network. Details about neural networks are found in Chapter 4 [32].

6. Fuzzy models: The system can be represented using models derived from approximate reasoning and logic [32].

All these methods can be used to generate residual values, which measure the divergences between observed operating conditions and normal operating conditions. Once residuals are generated, different techniques can be used to make a diagnosis about the machine condition. Experienced personnel can qualitatively determine the source of the divergence, or quantitative methods can be used as well. These can range from simple limit-checks to more advanced classification techniques such as Linear Discriminant Analysis and Neural Networks.

1.5.2 Signal Based Systems

Signal Based systems require only a machine signal to make a diagnosis. Machine signals can be represented in the time-domain, frequency domain, or in the time-scaled frequency domain. The underlying physics of the machine and fault often dictate the most suitable representation, as discussed in Chapter 2. Once the signal is suitably represented, domain experts can make a diagnosis qualitatively, or more advanced pattern recognition techniques such as Neural Networks and Support Vector Machines can be utilized for the purpose [2]. Most industrial condition monitoring programs use signal-based schemes due to the time-consuming and complex nature of the system modeling process. If there are system nonlinearities, coupling between multiple subsystems, or unavailability of system information and parameter values, the development of an accurate system model can become quite challenging. The Iron-Butcher is a good example of a system that is difficult to model in its entirety. It consists of a highly nonlinear electro-hydraulic manipulator, has coupling between the three subsystems, and does not have information about all the system parameters as required to build a complete model. As a

result, the present work will focus on developing a signal-based condition monitoring scheme for complex machines like the Iron Butcher.

For the purposes of the present thesis, the task of signal-based condition monitoring can be divided into three subtasks (see Fig. 1.12). First, the signals are acquired from the machine and suitably represented. Then, the size of the feature set is reduced to facilitate the classification process. Finally, the reduced feature set is sent to a properly designed classifier, which generates a diagnosis about the machine condition.
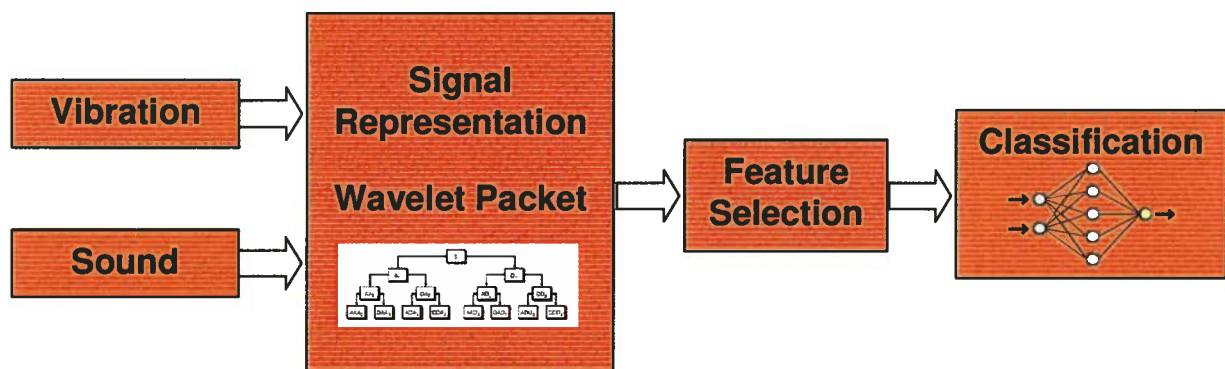
Figure 1.12: Condition Monitoring Subtasks.

1.5. Review of Previous Work

1.5.1 Multisensor Condition Monitoring

Acquiring information from multiple sensors rather than a single source can have many potential advantages including redundancy, complementarity, timeliness, and reduced cost of information. Sensor fusion can effectively reduce the overall uncertainty in a system while increasing the accuracy of signal perception. In their review of multisensor fusion applications, Luo et al. [4] describe many examples where data from multiple sources are combined using traditional and "intelligent" methods. Applications of multi-sensor technology are highly varied and include robotics, biomedical engineering, remote sensing, and equipment monitoring among many others.

Condition based monitoring schemes using multiple sensors have been implemented in various applications and capacities in industry and academia. Of particular interest, Lang and de Silva [5] developed a condition monitoring scheme for the Iron Butcher in 2008. They used an accelerometer, microphone, and a CCD camera to obtain vibration, sound and vision signals. They used the Fourier transform approach to process sound and vibration signals and a SIFT algorithm for object tracking. A neuro-fuzzy classifier was designed to classify three types of on-off machine conditions. Although the scheme was able to detect these conditions with high accuracy, partial faults were not investigated and the robustness of the scheme was not verified under faulty sensor conditions. The present research will address both these issues and attempt to improve upon the existing condition monitoring scheme.

1.5.2 Signal Processing and Wavelet Analysis

Conventionally, the Fast Fourier transform (FFT) has been used to represent machine signals in the frequency domain. However, the wavelet transform has recently gained popularity as a powerful and computationally efficient representation for condition monitoring applications. Wavelet transform has the advantage of simultaneously providing time and frequency information about the signal, which is very useful for analyzing non-stationary signals common in many condition monitoring applications. In their review of current wavelet applications in condition monitoring, Peng and Chu [6] describe many possible uses of the wavelet transform. It can be used as a tool to compress signals, identify system parameters, detect singularities, denoise signals, and generate fault features. Wavelet analysis has been used to successfully diagnose electrical and mechanical faults in a variety of machine components including bearings, gears, motors, and pumps.

Bearings are some of the most important and common components in rotating machinery. In the area of health monitoring in roller bearings, numerous researchers have applied the discrete wavelet transform and wavelet packet transform to decompose vibration signals. Lin et al. [7] applied a Morlet wavelet and threshold denoising to detect impulses caused by faulty gears and bearings. Purushotham et al. [8] used the discrete wavelet transform and Hidden Markov Models for detecting single-point and multiple-point defects in roller bearings with up to 99% accuracy. Rubini and Meneghetti [9] compared the envelope spectrum and the discrete wavelet transform for detecting faults in roller bearing elements.

Similarly, the condition monitoring of gears with wavelet analysis has been researched as well. Wang and McFadden [10] applied the continuous wavelet transform to asses tooth damage in a helicopter gearbox. Sung et al. [11] used the Discrete Wavelet Transform to locate tooth defects in gear systems at high accuracies. The results showed that the DWT was able to perform better than the Short Time Fourier Transform, especially when the faulty gear ran at comparable speeds to other gears. Lin and Zuo [12] used an adaptive wavelet filter to decompose acceleration signals and detect fatigue cracks in gear teeth.

In addition to applying the wavelet transform to acceleration signals for fault detection, the wavelet transform has been used to decompose sound signals. Shihbata et al. [13] used the Discrete Wavelet Transform to create Symmetric Dot Patterns, which is a visualization technique for sound signals. Although not as effective as vibration-based monitoring, the transformed signals were able to capture fault signatures of fan bearings. Also, Lin [14] applied the Morlet wavelet to denoise sound signals and detect abrasion in engine bearings and pushrods. Wu and Chan [15] decomposed sound signals with the wavelet packet transform to diagnose gearbox faults at a high accuracy as well.

In conclusion, a review of existing literature suggests that applying the wavelet transform to acceleration and sound signals is an effective approach for diagnosing a wide variety of faults in rotating machinery.

1.5.3 Feature Reduction

Since the wavelet analysis generates a large number of coefficients, a feature reduction scheme is necessary to reduce the size of the feature set. Feature reduction will lower the computational burden and improve the classification accuracy [16]. The two main approaches to feature reduction are (somewhat ambiguously) termed feature extraction and feature selection. In feature extraction, statistical/numerical "transformation" methods such as Principal Components Analysis (PCA), Independent Components Analysis (ICA), and Linear Discriminant Analysis (LDA) are applied to the initial feature set to reduce its dimensionality. However the transformation can be computationally expensive and lead to numerical instabilities, particularly

when multiple sensors are used, and the resulting feature size is very large [17]. Feature selection is a more appropriate choice for feature reduction in fault diagnosis applications.

Yen and Lin [17] proposed two statistical methods of feature selection in vibration monitoring of a helicopter gearbox. After decomposing the signal with WPT and FFT, they proposed two feature selection algorithms, PWM and KNK, to reduce the size of the feature set for input into a neural network. They found that by reducing the size of the feature set from 256 to 2 features for each sensor, they were able to classify 8 types of faults at a high accuracy.

Saito and Coifman [18] developed an extension to the "Best Basis" algorithm termed Local Discriminant Bases. This method selects a subset of bases from a collection of orthonormal bases which is best able to separate signals from different classes. The orthonormal bases can be constructed by using wavelet packets or other time-frequency decompositions. The method for selecting these bases employed relative entropy as a cost function for maximizing class seperability. Along the same lines, Liu and Ling [19] developed an extension to the "matching pursuit" algorithm for selecting wavelet coefficients in fault diagnosis applications. Termed "Informative Wavelets," the algorithm uses mutual information as a criterion to search for the best wavelet coefficients.

Tafreshi and Sassani [20] developed a fault diagnosis scheme for detecting knock conditions in a single cylinder diesel engine. After applying the wavelet packet decomposition to acceleration signals, the feature selection methods: Local Discriminant Bases, Mutual Information, and Dictionary Projection Pursuit were compared. LDB and DPP were both shown to have better classification performance and lower computational speeds than MI. Also a novel method for constructing the energy entropy map was proposed to increase the performance of the LDB and DPP algorithms [21].

In addition to LDB and DPP, several search methods have been proposed for variable feature selection. Genetic algorithms, in particular, have been proposed as an effective tool for searching a large feature set and selecting the best features [22]. Jack and Nandi [23] proposed a wrapper-based Genetic Algorithm for reducing the feature set size in condition monitoring applications. The fitness function simply uses the classifier % accuracy to select the best feature set. The algorithm suffers the same setbacks as other wrapper feature selectors: it can be computationally

very expensive when selecting from large feature spaces, and there is a risk of overtraining the pattern recognition process to the particular feature set that is analyzed. In view of these issues, a filter selection algorithm is more likely to be effective for large feature sets and does not risk overfitting the features to the analyzed data set.

Of particular interest, Leong and Yen [24] developed a filter feature selection scheme using a genetic algorithm and LDB for diagnosing faults in a helicopter gearbox. After decomposing vibration signals with the WPT, a filter selection scheme with size and discriminant ability as criteria for the GA objective function was implemented. The method was able to achieve lower feature numbers while providing similar classification accuracies as PCA, ICA, PWM, KNK, and LDB. Although showing promising results, there are two drawbacks with the proposed algorithm. Firstly, the scheme does not utilize sensor redundancy. Given that the feature set size and the discriminant ability are the only two criteria for feature selection, it is entirely possible that all features may be chosen from one sensor if the features in that sensor have a high relative entropy value. Also, there are no rules for selecting the objective function weights. As the performance of the feature selector is highly dependant on these weights, it is difficult to generalize its performance and benchmark against other methods. In the research of the present thesis, to accommodate sensor failure, a modified genetic algorithm and a tuning procedure are developed and tested.

1.5.4 Classification

Several methods are available to classify the machine condition using acquired signals. These range from simple statistical algorithms such as the k-Nearest Neighbor classifier to more complex techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). In recent years, radial-basis function networks (RBFNs) and SVMs have found many applications in condition monitoring due to their capabilities as universal approximators and capacity to perform nonlinear classification.

Nandi and Jack [25] compared an ANN and SVM for detecting faults in rolling-element bearings. For fault features generated from time series and spectral analysis, they found that the average value SVM had a slightly higher training classification accuracy but a lower test classification accuracy than the ANN. Samanta [26] compared an ANN and SVM for detecting

faults in gears. The two methods showed similar classification performances, with the SVM performing slightly better for a larger feature set, but the training time for the SVM was substantially lower. Lv et al. [27] compared ANN and SVM for diagnosing power transformer faults based on features extracted from Dissolved Gas Analysis. They also found that the SVM had higher classification accuracies and a lower training time.

In comparing the two approaches, studies indicate that the SVM performs better than the ANN and requires less training time. However, the possibility of corrupted data arising from sensor failure has not been investigated. The present research compares the performance of the two classifiers under conditions that simulate both machine failure and sensor failure.

1.6. Organization of Thesis

The present chapter discussed the rationale and motivation for needing an on-line condition monitoring scheme and outlines the goals and challenges of the research presented in this thesis. The experimental system that is considered in the present context (an industrial Iron Butcher) and the faults that are introduced and studied were described. General techniques for condition monitoring were described and the choice of a signal-based technique was justified. Finally, a detailed literature review highlighting relevant past work in condition monitoring, signal processing, feature reduction and classification was presented.

Chapter 2 describes the signals acquired from the investigated machine and the instrumentation installed on the machine for condition monitoring. The foundation of the wavelet transform and wavelet packet decomposition is presented and its advantages over conventional Fourier transform methods are also discussed. Finally, the generation of a shift-invariant feature vector is presented and a method for evaluating the discriminant ability is chosen.

Chapter 3 describes and compares various methods of feature reduction. Local Discriminant Bases, a common algorithm for frequency-based feature selection, is described and the drawbacks are argued. A novel feature selection method using Genetic Algorithms is proposed and a tuning procedure is developed. Finally, both algorithms are implemented and experimental results are discussed and compared.

Chapter 4 introduces the concept of pattern recognition. The theoretical foundations of Radial Basis Function Networks and Support Vector Machines are presented. The two methods are implemented and experimental results are discussed and compared.

Chapter 5 concludes this work by proving a synopsis of the presented research and outlining the major contributions made in the thesis. Finally, suggestions are made for further research in this field.

Chapter 2

Signal Processing and Feature Representation

2.1. Instrumentation

This section describes the instrumentation used for condition monitoring in the Iron Butcher. Figure 2.1 shows a schematic representation of the monitoring equipment and the corresponding information flow.
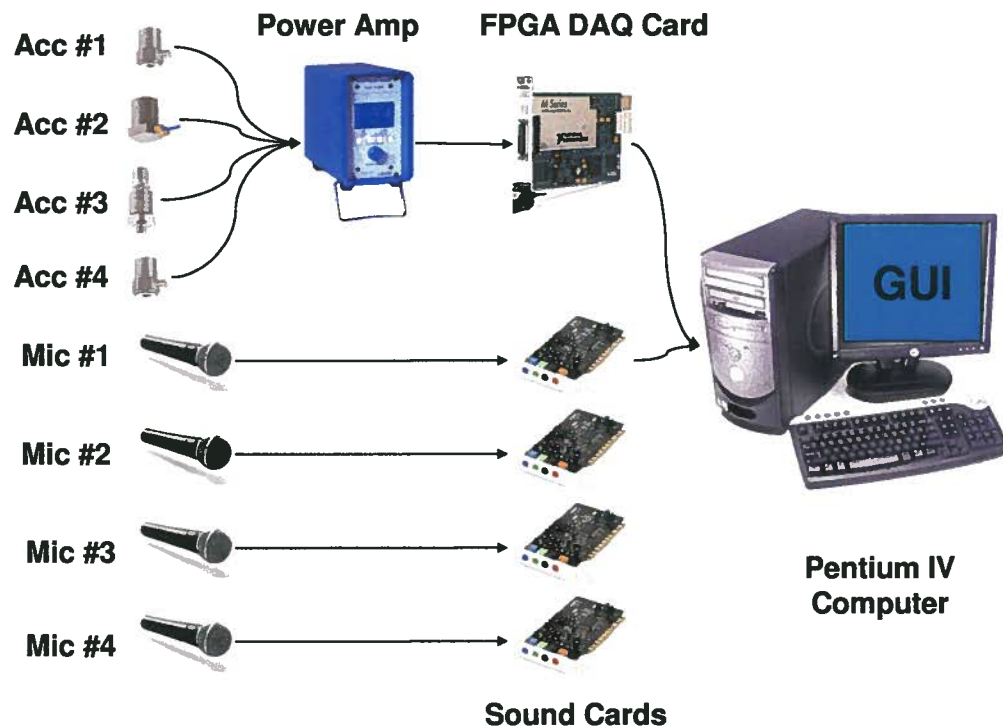
Figure 2.1: Signal Acquisition Schematic.

To obtain vibration signals from the machine, four single-axis piezoelectric accelerometers from Kistler Instruments are used. A charge amplifier from Kistler Instruments is used to amplify the signal into a mV voltage reading. The conversion factor from acceleration to voltage is 100 mV/g. Since the motor speed will have the highest frequency in the machine (45 Hz) at its normal operation, the sampling rate is set at 1 kHz, so multiple harmonics can be captured in accordance with Nyquist's sampling theorem.

Figures 4.1 show the location of the accelerometers. Two accelerometers are mounted on the gearbox to primarily capture fault characteristics of the bearings, gears and the shaft within the gearbox. There have been many studies conducted on the optimal location of the accelerometer mounting for gearbox condition monitoring. While there have been no established guidelines on the exact placement, a suggested approach is to mount the accelerometer radially with respect to the axis of rotation [1]. One accelerometer is mounted on the frame of the machine to capture catastrophic failures and conveyor vibrations. One accelerometer is placed near the cutter table to obtain information about the dulling of the blade.

(a) Accelerometers #1 and #2   (b) Accelerometer #3   (c) Accelerometer #4
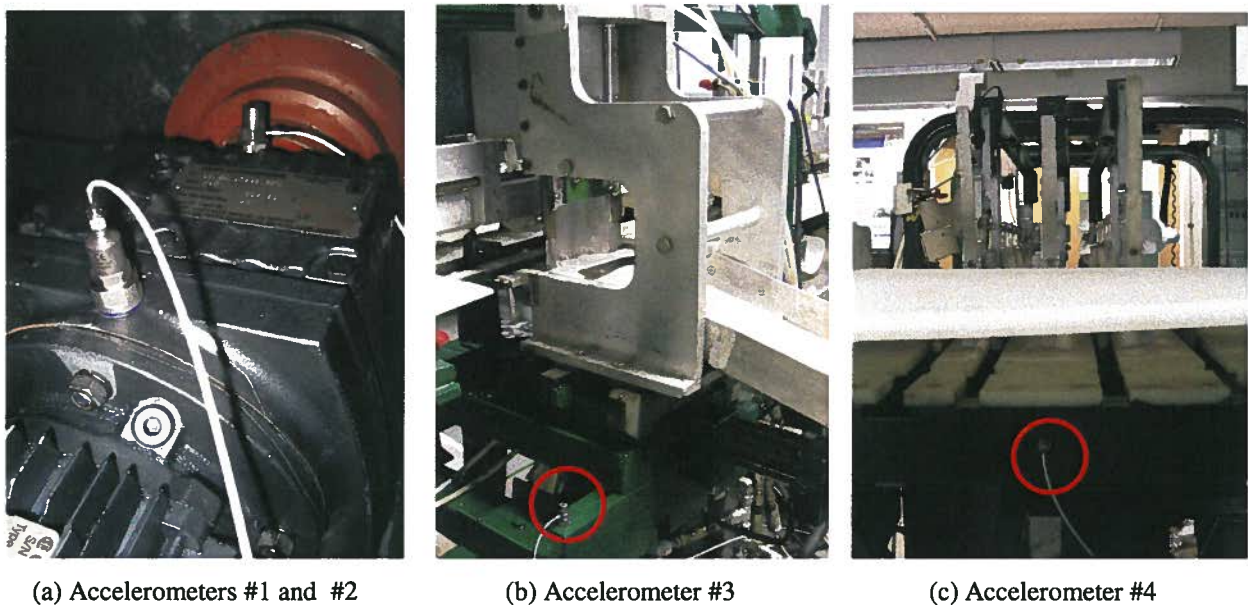
Figure 2.2: Accelerometer Positions on the Iron Butcher.

The acceleration signals are logged by a field programmable gate array (FPGA) data acquisition board from National Instruments. The FPGA DAQ board differs from the regular DAQ board in that a FPGA is used to control the device functionality rather than a ASICS board. The FPGA board offers many advantages over traditional Data acquisition boards; for example, complete control over synchronization and timing of signals, on-board decision making abilities, and true multi-rate sampling. With respect to on-line condition monitoring, a major advantage of using an FPGA board is that all data acquisition functions are hard wired in the FPGA, reducing the processor load for complex signal processing and classification computations.

Four wideband, capacitive type microphones are used to capture machine sound. The microphones capture acoustic pressure waves in the air and give a corresponding voltage reading, which is read by a computer sound card. Since most computer sound cards only accept one microphone input at a time, three additional sound cards were added to the existing computer. According to maintenance technicians at SEW Eurodrive, most gearbox conditions can be heard and diagnosed by an experienced operator. Therefore, the sampling rate was chosen to capture most of the human audible spectrum (20 Hz to 40,000 Hz); however, the sampling rate was limited to 30 kHz due to real-time processing limitations of the current computer. Figure 2.3 shows the location of the microphones. Two microphones are placed near the gearmotor, one microphone is placed above the fish-cutting machine, and one microphone is placed near the cutter blade.

(a) Microphones #1 and #2

(b) Microphone #3

(c) Microphone #4

Figure 2.3: Microphone Positions.

The signal processing functions and Graphical User Interface (GUI) are programmed in LabVIEW. For the classification computations, the ANN and SVM algorithm are expressed in MATLAB code that is embedded in the LabVIEW Virtual Instrument. The GUI has three sections: a controls section for adjusting the signal sampling rates and triggering the acquisition, a plot section where raw data and spectral data are plotted, and a diagnostics section where LED displays indicate the status of the machine. Appendix B has further information about the VI and software. The signal processing functions and diagnostic computations are performed every three seconds, thereby determining the effective refresh rate of the machine status.
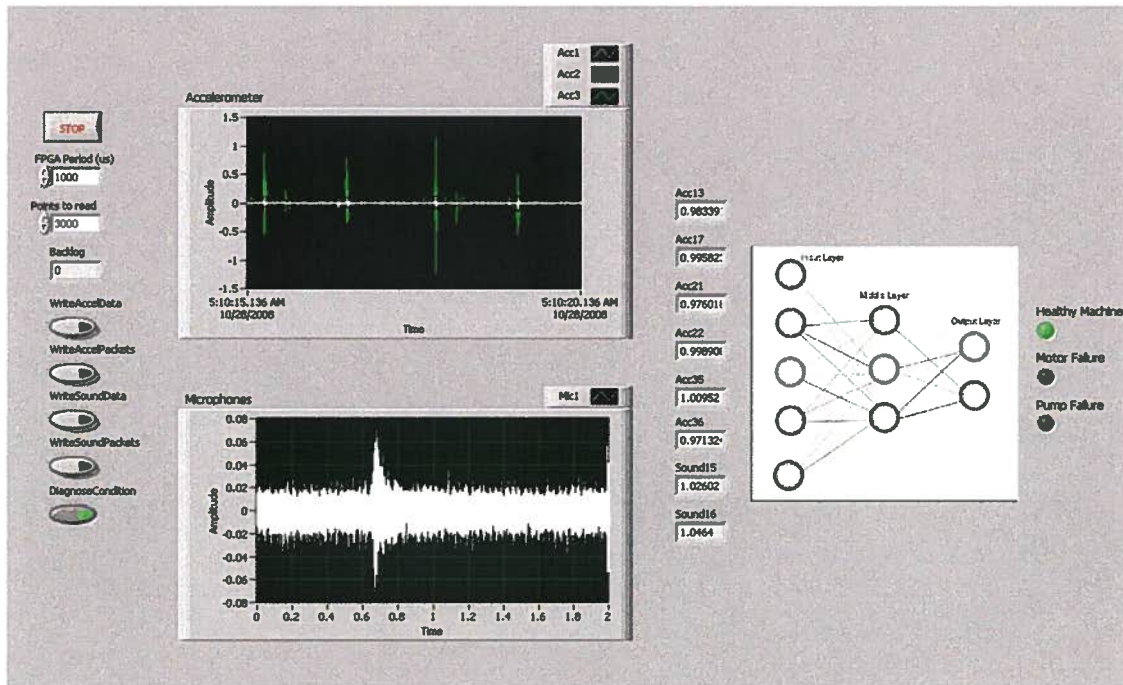
Figure 2.4: LabVIEW Graphical User Interface.

2.2. Signal Representation

The common method used to diagnose faults in reciprocating machinery is to represent the signal in the frequency domain using the Fourier Transform. This method decomposes a signal into constituent sinusoidal signals at different frequencies. Specifically,

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} \, dt \qquad (2.1)$$

where $F(\omega)$ is the Fourier transform of signal $f(t)$.

24

To avoid redundancy and reduce computational expense, the discrete version of Fourier transform is implemented in the form of the Fast Fourier Transform (FFT):

$$F_n \equiv \sum_{n=0}^{N-1} f_k e^{-\frac{2\pi i k n}{N}} \qquad (2.2)$$

where F_n is the discrete Fourier transform of the signal f_k.

Although the Fourier transform is adequate in many applications of signal processing, it has a major disadvantage. In the transformation from the time domain to the frequency domain, all information about time is lost (hidden). If the signal is non-stationary and has characteristics that change over time due to drifts, trends, abrupt events, transients or other occurrences, the Fourier analysis can become less effective. As an improvement to the Fourier transform approach, Gabor in 1946, proposed a windowing technique known as the Short Time Fourier Transform (STFT). By considering small sections of the signal in sequence and performing the Fourier transform on them, the STFT maps the signal into a two-dimensional function of time and frequency.
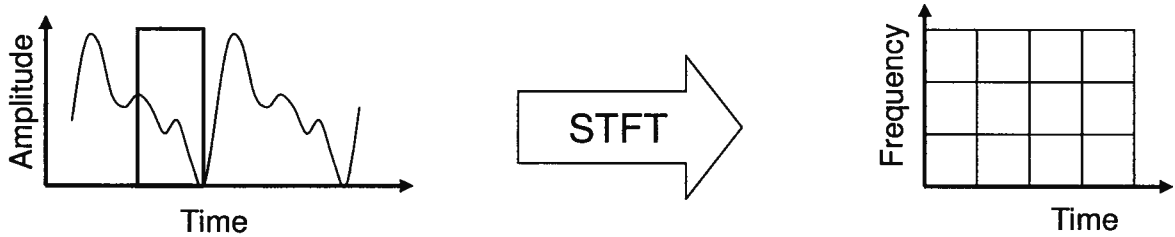
Figure 2.5: Short Time Fourier Transform.

The formal definition for STFT is:

$$STFT(\tau, \omega) = \int_{-\infty}^{\infty} s(t) g(t - \tau) e^{-j\omega t} dt \qquad (2.3)$$

where $STFT(\tau, \omega)$ is the Fourier transform of the signal $s(t)$, that was previously windowed by the function $g(t)$ with respect to the time shift variable τ.

Although the STFT gives both time evolution and the frequency spectrum of the signal, it has two major drawbacks: it has a fixed resolution with respect to the time window size at all frequencies and there are no orthogonal bases for computing the STFT. These drawbacks result

in limited precision achieved due to the window size and reduced efficiency of the algorithms used for computing the STFT. The second drawback is especially important since fault diagnosis schemes are implemented in real-time and computational cost/speed is a high priority in the present application.

Originally introduced by Grossman and Morlet in 1984, wavelets are a class of irregular, asymmetric functions that have no analytical expression to describe them. Unlike sinusoidal functions, wavelet functions have a finite duration and the average value is always zero. Due to their unique properties, wavelets have found success in a number of areas including data compression, image processing, and time-frequency spectral estimation.

The formal definition for the wavelet transform is:

$$W_x(a,b;\psi) = a^{-1/2} \int x(t)\psi^*\left(\frac{t-b}{a}\right) dt \qquad (2.4)$$

where a is the scale parameter, b is the time parameter, $\psi(t)$ is an analyzing wavelet and $\psi^*(t)$ is its complex conjugate.

As the above formulation shows, the wavelet analysis provides a time-scale view of the signal (see Figure 2.6). In providing a time-scale view, it allows the use of variable window sizes when analyzing a signal. Conveniently, the high-frequency information can be analyzed with a short time interval and the low-frequency content can be analyzed with a long time interval.
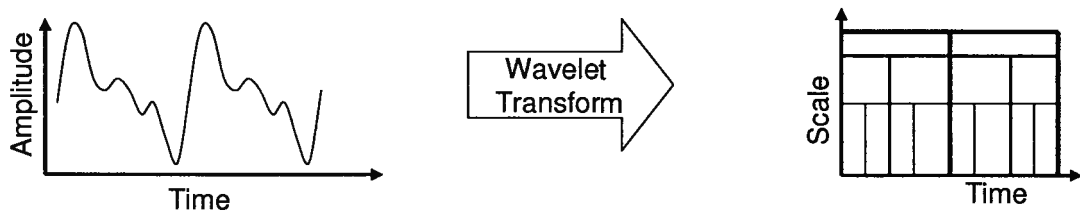
Figure 2.6: Wavelet Transform.

The choice of the analyzing wavelet depends on the particular signal processing application and the associated requirements. There exist several families of wavelets as developed by various researchers, each family with unique properties and associated advantages and disadvantages. For example, the Biorthogonal wavelet has a linear phase, which is useful for signal reconstruction, and Symlets are symmetrical, which is a useful property for image dephasing.

26

Among the most commonly used wavelet families in condition monitoring are the Daubechies, Biorthogonal, Symlet and Coiflet wavelets. There are no clear guidelines for selecting the analyzing wavelet in condition monitoring applications, so different wavelets are tested in the present application to determine if the particular choice of wavelet has an effect on classification performance.

(a) Harr Wavelet   (b) DB04 Wavelet   (c) Bior4.4 Wavelet   (d) Coif4 Wavelet
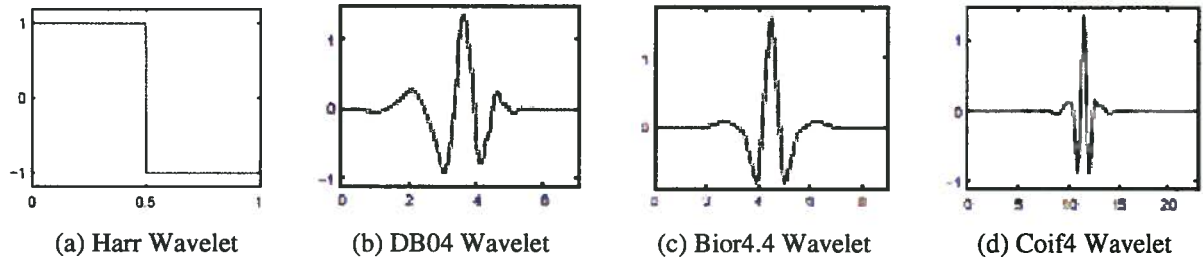
Figure 2.7: Different Analyzing Wavelets.

Figure 3.1 shows a scalogram of the machine signal from accelerometer #1 during regular machine operation. The vertical axis represents the scale of the analyzing wavelet and the horizontal axis gives the time of the signal. The color of the map corresponds to the magnitude of the coefficients at each scale. Since the period of the conveyor motion is large, there is a much higher correlation between the coefficients at the highest scales.
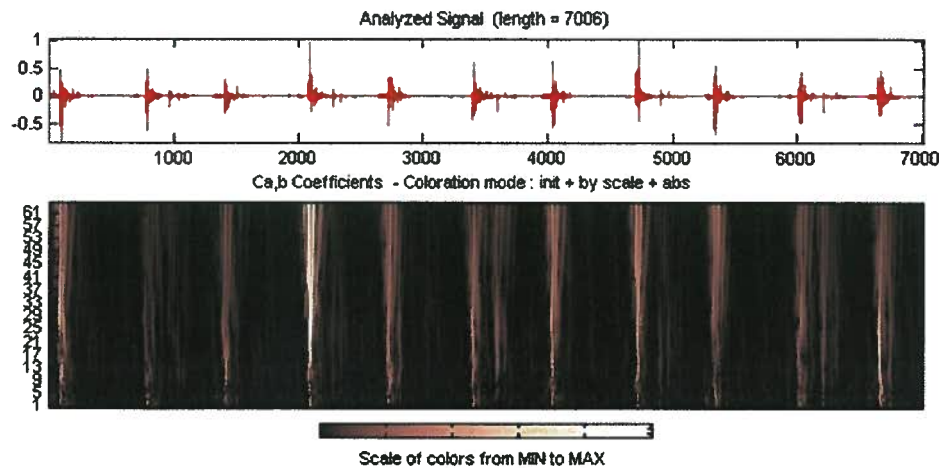
Figure 2.8: Scalogram of Signal from Accelerometer #3.

Since the CWT (Continuous Wavelet Transform) is computationally expensive and contains redundant information, a subset of scale parameters a and b is chosen to efficiently represent the signal with no loss of information. The parameters a and b are discretized as $a = a_0^m$ and

27

$b = na_0^m b_0$ where m and n are integers. The discretization of the scale and time parameters results in the Discrete Wavelet Transform (DWT), defined as:

$$W_x(m,n;\psi) = a_0^{-m/2} \int x(t)\psi^* \left(a_0^{-m}t - nb_0 \right) dt \qquad (2.5)$$

An efficient scheme to compute the DWT using cascaded filters was developed by Mallat and is known as the Fast Wavelet Transform (FWT). This involves the introduction of a scaling function $\phi(t)$ and the subsequent calculation of wavelet $\psi(t)$ from $\phi(t)$ [17]. We have

$$\phi(t/2) = \sqrt{2} \sum_k h_k \phi(2t - k) \qquad (2.6)$$

$$\psi(t/2) = \sqrt{2} \sum_k g_k \phi(2t - k) \qquad (2.7)$$

where g_k and h_k are elements of the coefficient vectors of quadrature mirror high pass and low pass filters, respectively, and k is a time localization parameter. As a result of this relationship, the wavelet transform can be applied to a signal by using filters only without the need for wavelets or scaling functions. In the DWT, only the coefficients from the low pass filter (approximation) are passed through subsequent filters. The high frequency coefficients (details) are not considered to contain as much information and are left as is.

a) Fast Wavelet Transform   b) Filter Bank
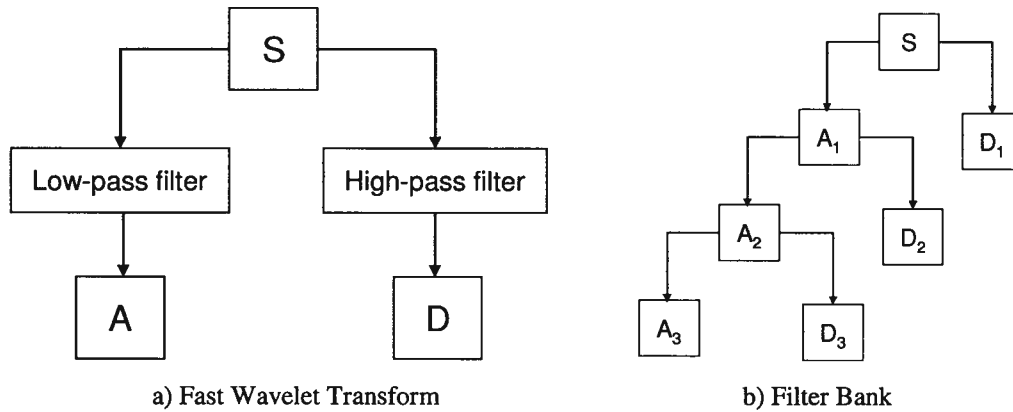
Figure 2.9: Discrete Wavelet Transform.

In developing a condition monitoring scheme that can be generalized for various machines and fault characteristics, it is important to recognize that frequency bandwidth of interest may not be known beforehand. If the fault characteristics are present in a narrow, high-frequency bandwidth,

28

the Discrete Wavelet Transform may not analyze the characteristics with sufficient resolution. The Wavelet Packet Transform (WPT) is a generalization of the DWT, where both the approximation and the details are split into further nodes (see Figure 2.10). This allows the signal to be represented as any combination of the approximation and details nodes.
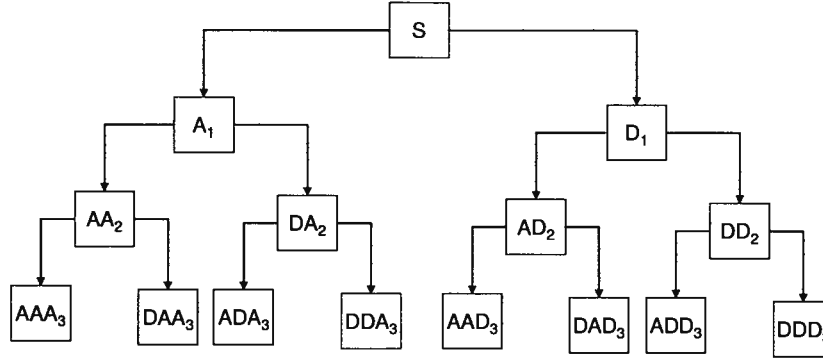
Figure 2.10: Wavelet Packet Transform.

To represent this transformation, a wavelet packet function is defined as

$$W_{j,k}^{n}(t) = 2^{j/2} W^{n}(2^{j} t - k) \tag{2.8}$$

where n is the oscillation parameter, j is the scale parameter, and k is the translation parameter. The first two wavelet packet functions are the scaling function and the basic wavelet function, respectively, as given by

$$W_{0,0}^{0}(t) = \phi(t) \tag{2.9}$$

$$W_{0,0}^{1}(t) = \psi(t) \tag{2.10}$$

All subsequent wavelet packet functions can be described by the following set of recursive relationships:

$$W_{0,0}^{2n}(t) = \sqrt{2} \sum_{k} h_{k} W_{1,k}^{n}(2t - k) \tag{2.11}$$

$$W_{0,0}^{2n+1}(t) = \sqrt{2} \sum_{k} g_{k} W_{1,k}^{n}(2t - k) \tag{2.12}$$

Using these definitions, the wavelet packet coefficients of a function f can be determined by

$$w_{j,k,n} = \int f(t) W_{j,k}^{k}(t) dt \tag{2.13}$$

By decomposing both the high frequency and low frequency components, we obtain a rich library of orthonormal bases that contain time and frequency information about the stationary and nonstationary characteristics of a signal.

2.3. Experimental Results

Figures 2.11 to 2.22 show a selection of the acquired acceleration and sound signals and the corresponding discrete wavelet decomposition of these signals. The accelerometer and microphone positions correspond to those shown in Figures 2.2 and 2.4.Even though the WPT is used for feature generation, the DWT is better suited in familiar problems and therefore used for illustrative purposes. The signals are decomposed into 4 levels using the DB04 wavelet. Here S is the raw signal in millivolts, a_4 is the approximation (corresponding to low frequencies) at the fourth level and d_x is the detail (corresponding to high frequencies) at the x level. The y-axis gives the raw millivolt reading and the x-axis represents the time count in 1 ms for acceleration and 0.033 ms for sound signals.
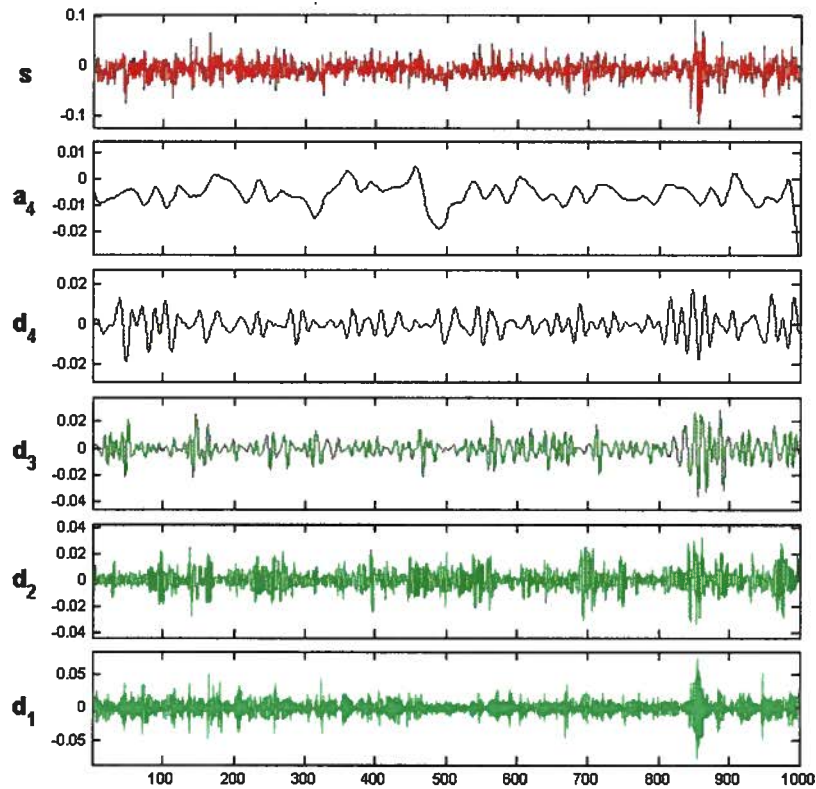
Figure 2.11: Accelerometer #2 Signal for Baseline Condition.
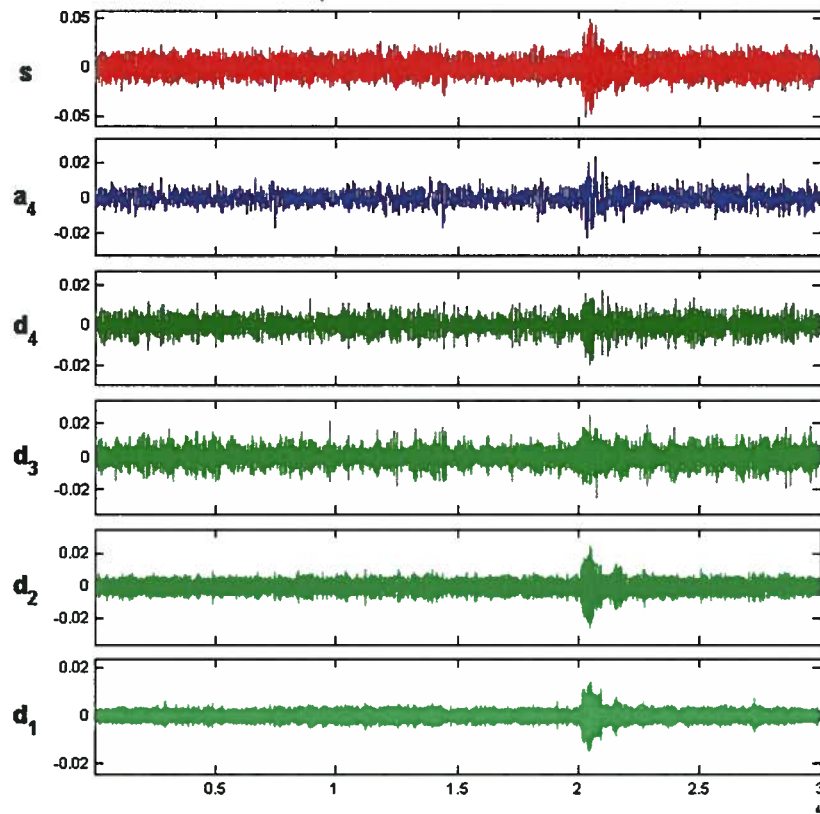
Figure 2.12: Microphone #1 Signal for Baseline Condition.
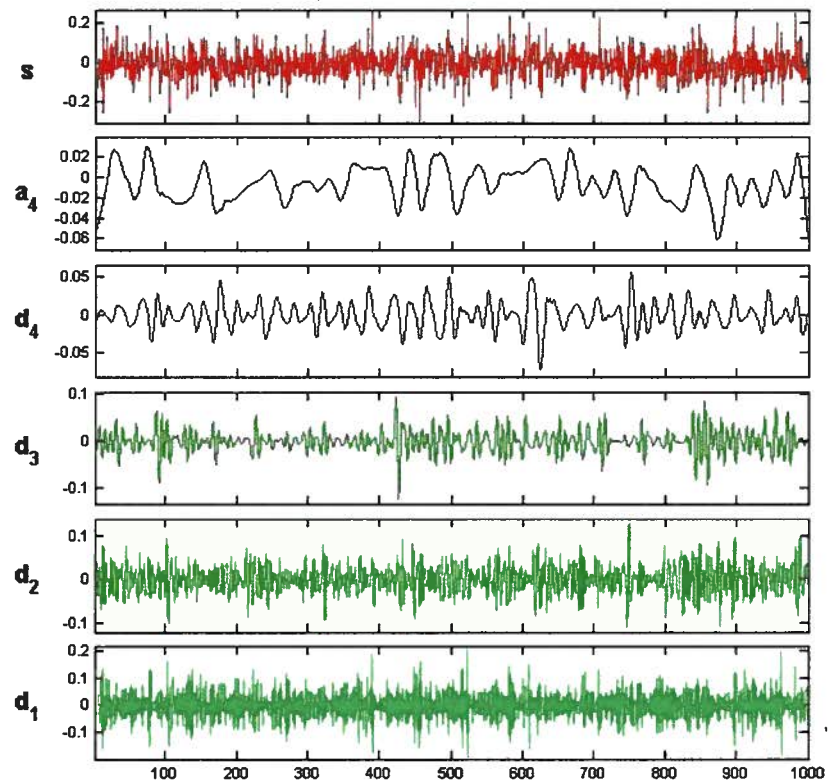
Figure 2.13: Accelerometer #2 Signal for Faulty Gear Condition.

31
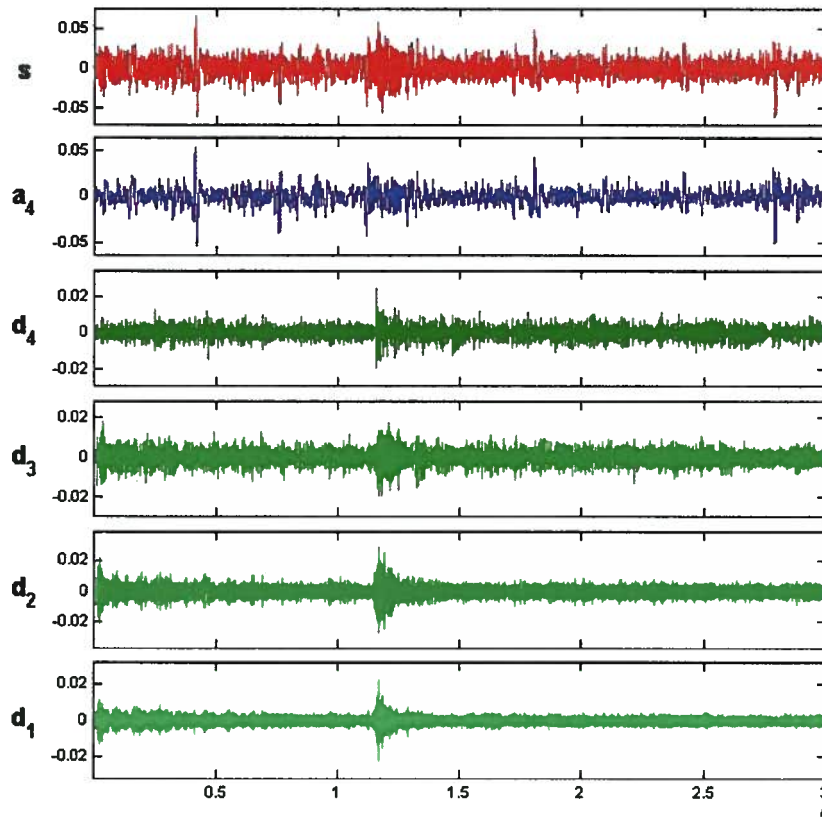
Figure 2.14: Microphone #1 Signal for Faulty Gear Condition.
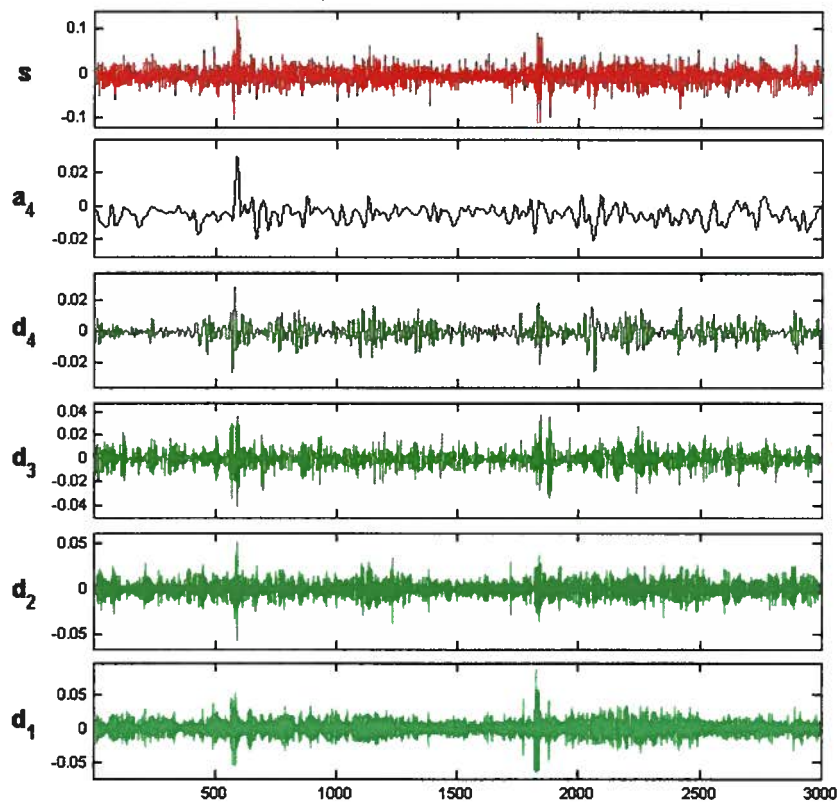
Figure 2.15: Accelerometer #2 Signal for Faulty Bearing Condition.
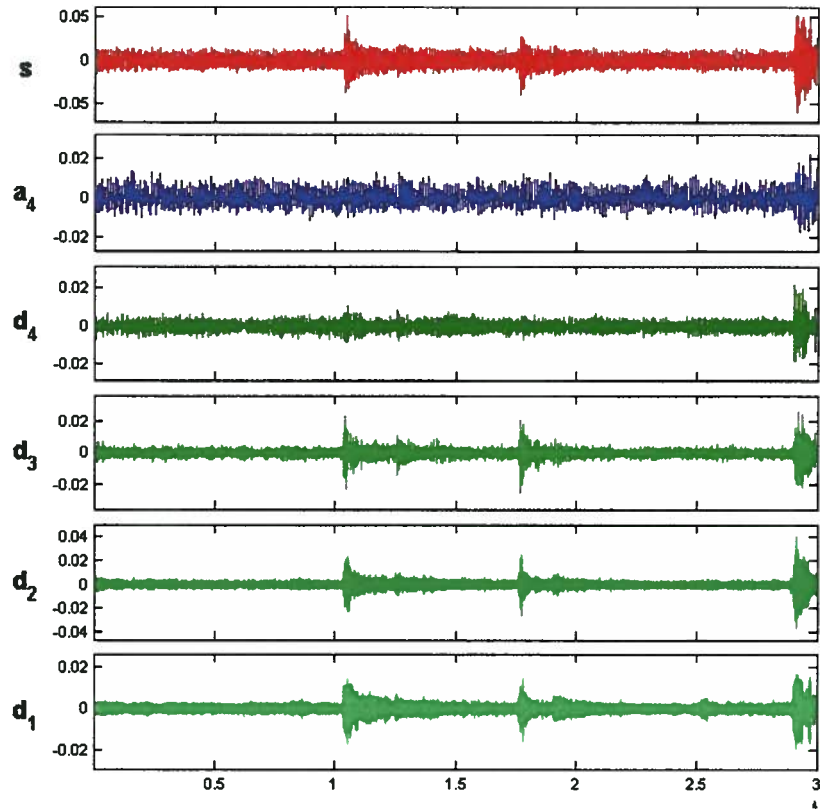
Figure 2.16: Microphone #1 Signal for Faulty Bearing Condition.
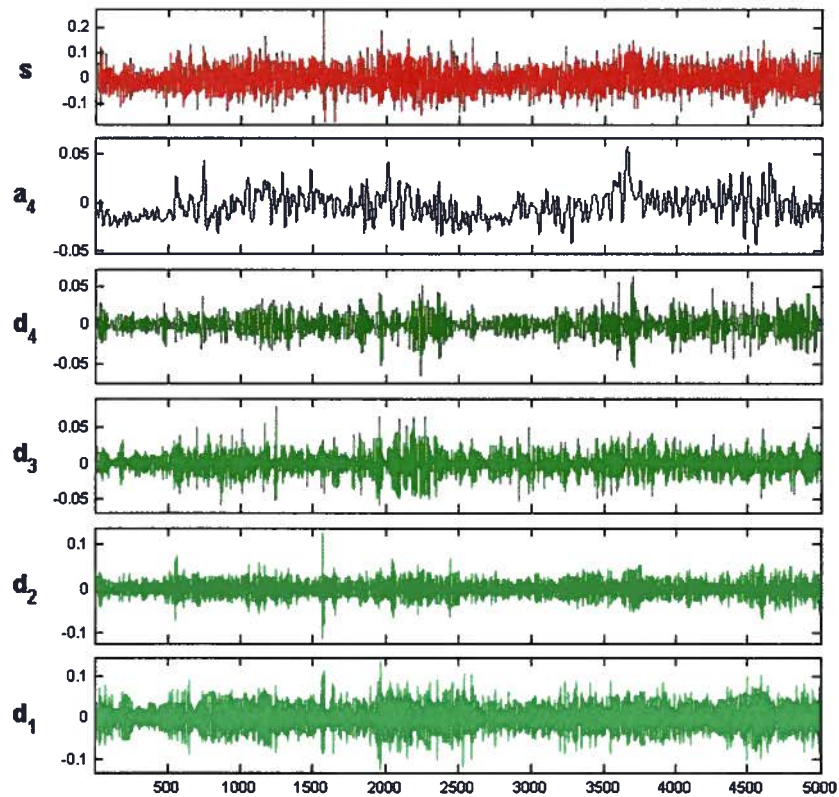
Figure 2.17: Accelerometer #2 Signal for Shaft Misalignment Condition.
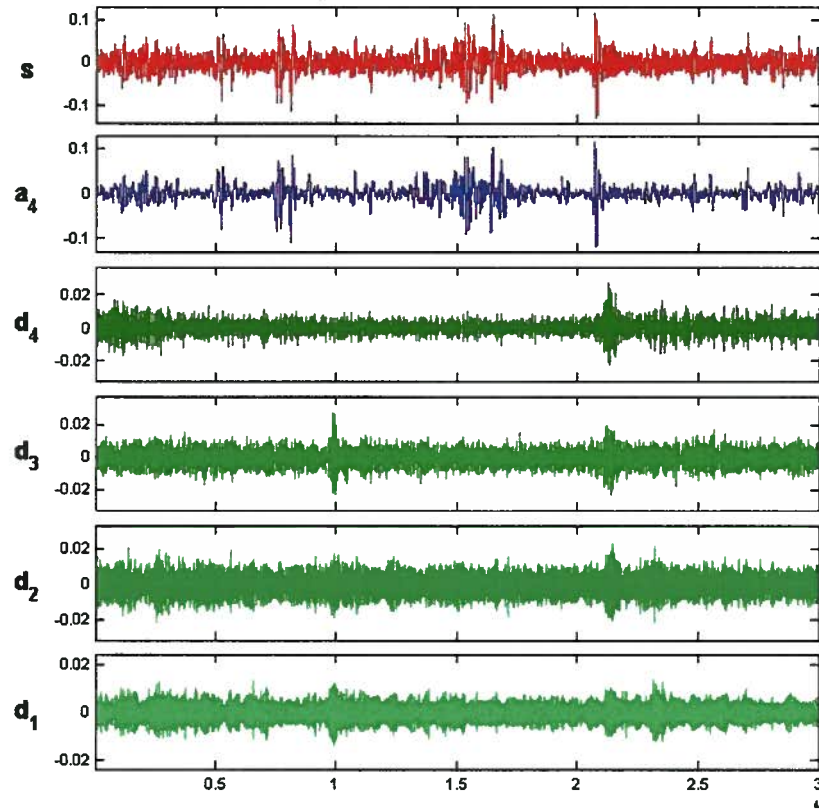
Figure 2.18: Microphone #1 Signal for Shaft Misalignment Condition.
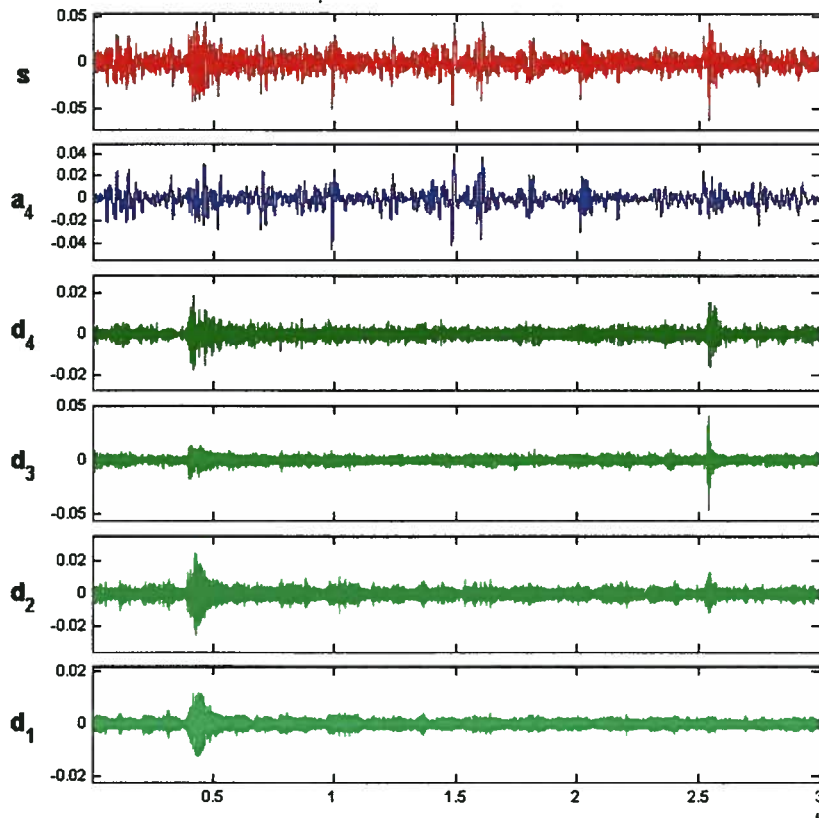
Figure 2.19: Microphone #1 Signal for Hydraulic System Fault Condition.
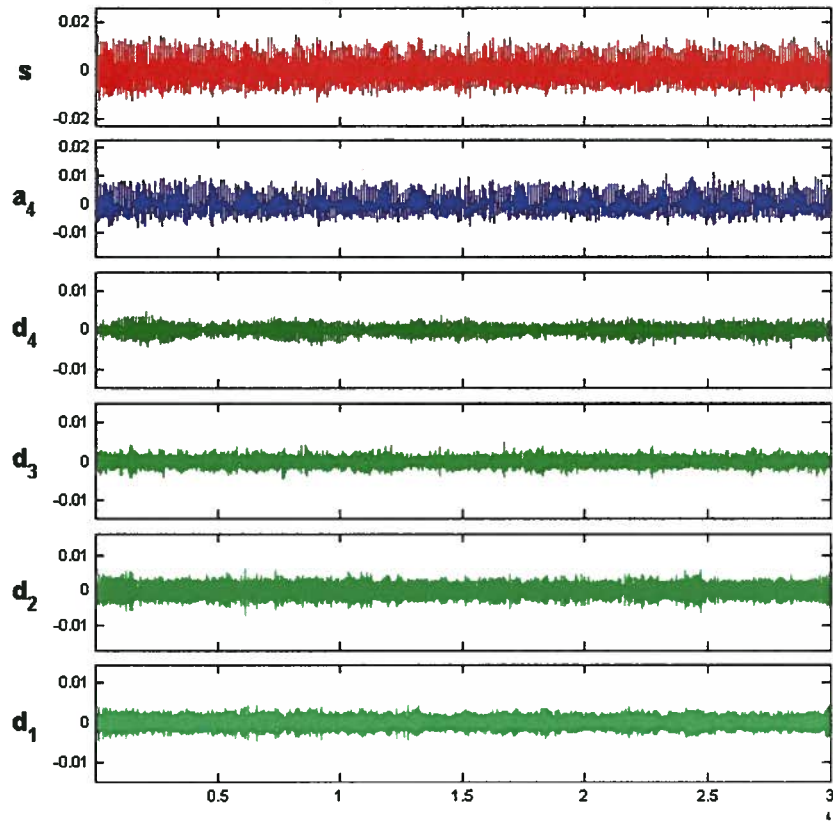
Figure 2.20: Microphone #1 Signal for Motor Fault Condition.
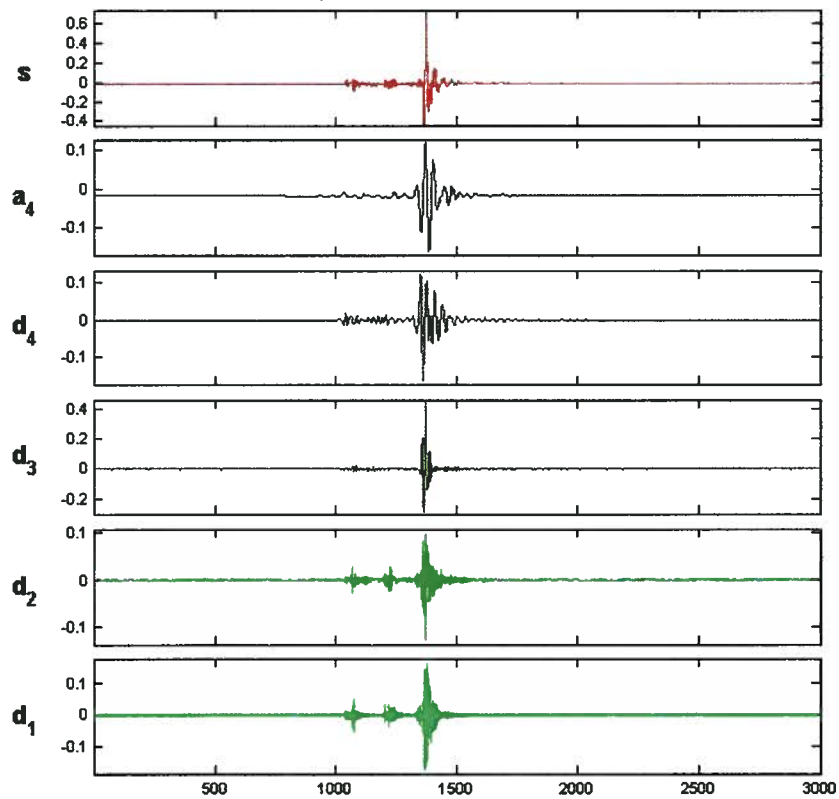
Figure 2.21: Accelerometer #3 Signal for Sharp Cutter Blade.

35
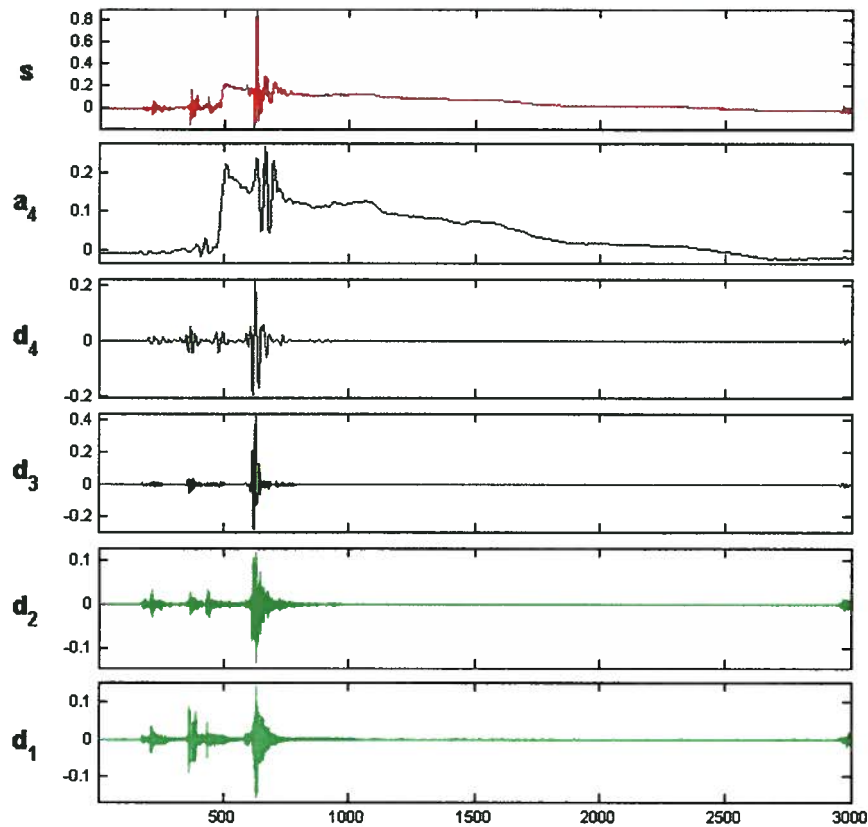
Figure 2.22: Accelerometer #3 Signal for Dull Cutter Blade.

As the gearbox contains 6 gears and 5 bearings in total, it is difficult to exactly correlate the signal readings with the emulated faults in view of the rotating components running at different frequencies and associated harmonics. Also, the precise nature of the faults is unknown because the gear units came preassembled with damaged components. However, the signal decompositions show that the wavelet transform is able to effectively differentiate between most of the different conditions. Comparing Figures 2.11 and 2.13, the peak amplitude in the approximation for the faulty gear is double the baseline amplitude and the frequency of the peaks corresponds approximately to the gear meshing frequency, indicating a problem with the gear teeth. Comparing Figures 2.11 and 2.15, there is a large periodic impulse in the detail bandwidths for the faulty bearing, possibly arising from rolling elements passing over inner race defects. In Figure 2.17, the approximation itself is oscillatory at a very low frequency. Since the misaligned output shaft is rotating at the lowest frequency in the gearbox, one can correlate the oscillation with the imbalanced force arising from the misaligned shaft. In Figure 2.19, the signal energies at all scales of the decomposition are lower, resulting from the missing sound of the hydraulic pump. In Figure 2.20, the periodic conveyor sound is missing and the only sound

present is from the hydraulic pump. Figures 2.21 and 2.22 show the impulses from the cutting blade breaking materials with different strengths. As expected, the sharp blade generates a clean material breakage and the dull blade has to exert the force over a longer time to achieve the same effect.

2.4. Feature Representation

One major disadvantage of the wavelet packet transform is the lack of transform invariance in the wavelet bases. Two signals that are shifted slightly in time can have significant differences in coefficient representations. As a result, wavelet packet coefficients cannot be used directly as reliable feature representation means for on-line condition monitoring systems. Yen [24] and Tafreshi [20] also describe difficulties of using the coefficients directly for feature representation. One way to solve the feature representation problem is to define node energy as the sum of all coefficients in the node. By choosing a large window, the effect of any signal shifts will be minimized. We have

$$E_{j,k} = \sum_n w^2_{j,k,n'}$$
(2.14)

where w are the coefficients in node j,k of the wavelet packet tree.

By computing the energy of each node, we define a unique feature for each frequency band of the wavelet decomposition. Once the energy is computed, the features are preprocessed to eliminate disproportionate differences between classes and to improve the classification performance. A simple unit range scaling is used to find the normalized feature $\tilde{x}$; thus,

$$\tilde{x} = \frac{x - l}{u - l}$$
(2.15)

where x is the mean, l is the lower bound, and u is the upper bound of the raw features across all classes.

2.5. Discriminant Measure

The discriminant ability of a feature can be defined as a measure of how differently two sequences p and q are distributed. In the application of pattern recognition, it can be described as

the ability to differentiate between two classes. There are many statistical measures of discriminant ability [28]. Some of the most common measures include:

- Generalized f-divergence-based distance measures
- Mean distance-based distance measures
- Contrast type distance measures
- Model validation distance measures
- Entropy-based distance measures

According to Saito and Coifman [18], a natural choice for the discriminant measure in wavelet-based pattern recognition applications is relative entropy. Before discussing relative entropy, the concept of entropy is introduced. Entropy can be viewed as an energy concentration of a coordinate vector or a measure of how much information a signal contains.

Shannon entropy is defined as:

$$H(p) \equiv \sum_{i=1}^{n} p_i \log p_i \qquad (2.16)$$

where p is a nonnegative sequence with $\sum p_i = 1$

Because applications of pattern recognition are concerned with the ability to differentiate between signals, a discriminant version of entropy, the relative entropy, is often used:

$$I(p,q) \equiv \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \qquad (2.17)$$

where p and q are two nonnegative sequences with $\sum p_i = \sum q_i = 1$.

Because the discriminant measure in Eq. 2.17 is not symmetric and does not satisfy the triangle inequality, the discriminant measure will depend on which class is defined as p and which class is defined as q. The symmetric version of relative entropy, also known as J-divergence, is used:

$$J(p,q) \equiv \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i} \qquad (2.18)$$

For measuring the discriminant abilities between multiple C classes (as in the present work), the sum of pairwise combinations of relative entropy will be used:

$$D\left(\left\{p^{(c)}\right\}_{c=1}^{C}\right) \triangleq \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} D(p^{(i)}, p^{(j)}) \qquad (2.19)$$

where D is the discriminant measure of the signals p

2.6. Analysis of Feature Variation

Before the feature selection process was implemented, a simple statistical test was implemented to confirm that the differences in features were statistically significant. To test whether the feature distributions overlapped, the 95% confidence interval of the mean was calculated using

$$C.I. = \tilde{x} \pm 1.96 \frac{s}{\sqrt{n}} \qquad (2.20)$$

where s is the standard deviation for the features in a class and n is the sample size.

The statistical analysis revealed that differences between cutter dullness conditions and fish jamming conditions were not statistically significant; i.e., none of the sensors could detect any faults for these conditions when the machine was in operation.

Note: As a result, these faults were not considered in the following chapters for the development of pattern recognition algorithms.

Interestingly, these faults could be detected in isolation (when the other subsystems were turned off) and when there were dedicated sensors for detecting these faults.

Since there are no clear guidelines for picking the optimal analyzing wavelet, different wavelets are tested to check if the choice of wavelet function significantly affects the discriminant measure. The top 16 bases with the highest discriminant power were calculated using each wavelet decomposition. The results are as shown below:

DB04: 26 25 31 27 28 18 23 30 32 19 29 21 24 13 05 06

Haar: 25 31 26 27 28 18 19 32 30 23 21 29 11 16 13 22

Bior4.4: 27 25 26 32 28 31 19 21 18 23 24 06 30 29 22 05

Coif4: 28 25 26 27 32 31 30 18 23 06 29 05 21 19 13 24

Sym4: 28 26 27 25 32 23 31 18 30 19 21 29 13 05 06 20

Note: The discriminant calculations are different from those presented in Chapter 3 because these ranking were generated by using only one sample of acceleration and microphone signals because of data processing limitations.

As the discriminant rankings show, there is very little difference (qualitatively) in the discriminant measures of the different wavelet decompositions. As a result, the feature selection algorithm and the subsequent classification procedure are presumably unaffected by the choice of the analyzing wavelet. DB04 is chosen as the analyzing wavelet due to its popular use in other applications of condition monitoring and the possibility of more standardized comparisons with existing research.

Chapter 3

Feature Selection

3.1. Feature Reduction

Feature Reduction is one of the most important components of the pattern recognition process. There are two main reasons to reduce the dimensionality of the feature space: to decrease the computational expense and increase the classification accuracy [16]. Feature reduction techniques can be divided into two broad categories: Feature extraction and feature selection. Feature extraction reduces a feature space of dimensionality m to a subspace of dimensionality $d < m$ by applying a linear or non-linear transformation. Principal Components Analysis (PCA), also known as the K-L expansion, is a popular technique for feature extraction. Other variations of feature extractors have been proposed for dealing with non-Gaussian distributed data and applying nonlinear transformations. These techniques include Independent Components Analysis (ICA), Nonlinear PCA, and Neural Networks.

Although feature extraction methods provide better classification accuracies than feature selection methods, there are some drawbacks with applying feature extraction in condition monitoring applications. If the dimensionality of the feature is exceedingly large, the computational cost of the transformation will be correspondingly large. This may not be suited for on-line condition monitoring applications where the frequency of machine diagnosis needs to be sufficiently high. Also, once the transformation is applied, the physical meaning of the transformed features is lost. When sensor failure is considered, it may be beneficial to know which sensors are providing useful information and how the system will perform if those sensor signals are not accurate.

In light of these issues, feature selection is better suited for condition monitoring applications with multiple signal sources. Feature selection reduces a feature space of dimensionality m to a subspace $d < m$ by selecting a subset of the features that minimizes the classification error. Feature selection methods can be categorized as filters or wrappers. Wrapper selection uses the

classifier to score the performance of the selected features. Different search strategies can be used to guide the selection process including exhaustive searches, forward and backward selection, and stochastic methods such as genetic algorithms. Although simple to implement and often produce good results, wrappers methods run the risk of overfitting the model and features to a particular data set and classifier [29]. In the present work, rather than using wrappers to select features, they are used as a tool for evaluating the performance of classification techniques, as described in Chapter 4.

Filter selection methods select a subset of features independently of the classifier according to some predetermined criteria. Feature selection uses a metric to predict the performance of a feature subset. For classification problems, discriminant measures of class seperability are often used as filter selection metrics. In the next two sections, Local Discriminant Bases, a complementary feature selection approach to WPT, is discussed and a Genetic Algorithm-based feature selection approach is proposed.

3.2. Local Discriminant Bases

Coifman and Wickerhouser [30] originally developed the popular and widely used "Best Basis" algorithm for optimal wavelet packet tree decomposition. In the method, using entropy as the cost function, the optimal structure of the wavelet packet tree is determined by exploiting the additive property of entropy and utilizing a "divide and conquer" approach. Since the original algorithm was developed for signal compression applications and focused on optimal signal representation, Coifman and Saito [18] proposed "Local Discriminant Bases," an algorithm specifically designed for signal classification applications. In their method, rather than using entropy, an additive discriminant measure is used as the cost function to maximize class seperability while minimizing the representation size.

To apply the local discriminant bases algorithm, the signal needs to be represented by a collection of orthonormal bases $\Omega_{j,k}$ which can be obtained from wavelet packet decomposition or local trigonometric transforms.
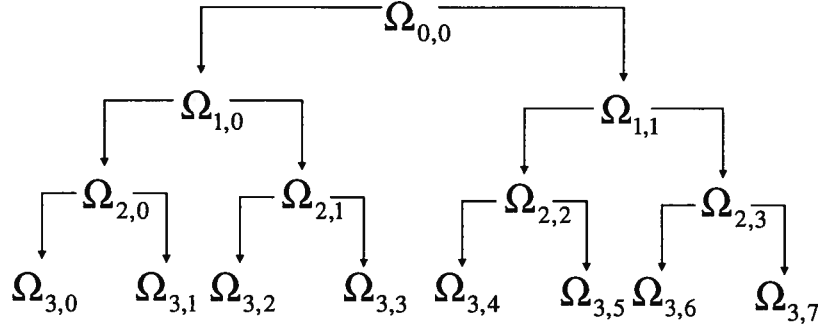
Figure 3.1: Orthonormal Bases from WPT.

The algorithm for the LDB is described as follows [18]:

Given a training dataset $\mathcal{T}$ consisting of C classes of signals $\left\{\left\{x_i^{(c)}\right\}_{i=1}^{N_c}\right\}_{c=1}^{C}$ where

$B_{j,k}$ are the basis vectors of subspace $\Omega_{j,k}$ and

$\Delta_{j,k}$ is an array containing the discriminant measures of subspace $\Omega_{j,k}$;

Step 1: Once the signal is decomposed into a dictionary of orthonormal bases, specify the maximum depth of decomposition J and discriminant measure D

Step 2: Set $A_{J,k} = B_{J,k}$ and $\Delta_{J,k} = D\left(\left\{\Gamma_c(J,k,\cdot)\right\}_{c=1}^{C}\right)$ for $k = 0,\ldots\ldots,2^{j-1}$

Step 3: Determine the best subspace $A_{j,k}$ by

$$Set\ \Delta_{j,k} = D\left(\left\{\Gamma_c(j,k,\cdot)\right\}_{c=1}^{C}\right)$$

If $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$,

then $A_{j,k} = B_{j,k}$

else $A_{j,k} = A_{j+1,2k} \oplus A_{j+1,2k+1}$ and $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$ (where $\oplus$ is a direct sum)

Step 4: Rank the Basis functions in order of discriminant ability

Step 5. Use the bases with the highest discriminant function for constructing the classifier.

To summarize the process, the algorithm starts by evaluating the discriminant measures of the terminal nodes as specified by the maximum level of decomposition. The sum of the two "children" node discriminant measures is compared with the discriminant measure of the parent node, which is one level higher on the Wavelet Packet tree. If the summed discriminant values of

43

the children are higher than that of the parent, the children nodes are kept as the "best bases." If the discriminant value of the parent node is higher, the children nodes are discarded and the comparison process repeats itself for the parent node which will now be the children node. This sequence of actions will repeat until the highest level of the tree is reached, resulting in a decomposition that has the maximum value of relative entropy.

Although LDB has been used extensively in different pattern recognition problems, the algorithm is not designed to be used for feature selection from multiple signal sources (i.e., multiple time-frequency energy maps). Once the "best" features for each signal source are determined, there are no guidelines on how to select the features across different signal sources. Relative entropy cannot be used on its own for feature selection because if one signal source has a high relative entropy measure for all of its bases, all the features will be selected from that signal source. This can be problematic for two reasons:

1. Multiclass relative entropy (Eq. 2.19) is not a perfect measure of the overall discriminant ability because it can easily be biased by large individual relative entropies.
2. If a sensor with large relative entropies fails, the classification scheme will fail, defeating the original purpose of multisensor fusion.

In the next section, a genetic algorithm is proposed to search the feature space for the best feature set according to three important criteria.

3.3. Genetic Algorithm for Feature Selection

Introduced by Holland in 1975, genetic algorithms are a class of derivative-free optimization algorithms that imitate the process of natural selection in genetics. Compared to traditional calculus-based schemes, genetic algorithms have two major advantages: they are applicable to discrete problems and they are less likely to get trapped in local optimums [32]. Also, compared to other enumerative techniques such as dynamic programming, genetic algorithms are better capable of handling large complex problems without breaking down or suffering from the *"curse of dimensionality."*

A simple genetic algorithm optimization procedure consists of the following steps:

1. **Initialization:** An initial population of chromosomes is randomly generated

2. **Selection:** The fitness values all chromosomes in the current population are evaluated. The chromosomes with the highest fitness values are selected for reproduction as they will have a higher probability of mating and creating the next generation of "better" chromosomes.

3. **Reproduction:** The chromosomes mate using the crossover operation, to produce the next generation. Genetic operators such as mutation can be applied to increase the diversity of the population

4. **Termination:** Steps 2 and 3 are repeated until a certain condition is reached, upon which the algorithm is terminated.

Further details about Genetic Algorithms and Evolutionary Computing are found in [32].

A genetic algorithm for robust condition monitoring systems is proposed and detailed as follows:

1. **Initialization:** The features are represented as a binary sequence of genes. If the gene is a 0, that means that the feature is not selected. If the gene is a 1, that means that the gene is selected as a feature.

| 01101010011 | 01101010011 | 01101010011 | 01101010011 |

| Accelerometer #1 | Accelerometer #2 | Accelerometer #3 | Microphone #3 |

The genes are randomly initialized and a population of 100 chromosomes is created.

2. **Selection:** For the fitness function, three intuitive criteria specific to feature selection in condition monitoring applications are proposed:

a. Size: Choosing a small feature set will reduce the complexity of the classifier, improve the speed of computation during the signal processing (because only a subset of the signal is processed) and improve the speed of classification. For the fitness function, a normalized metric for the size of the feature is defined as:

$$S = \left(\frac{N_{genes=0}}{N_{Total}} \right)$$

(3.1)

where $N_{genes=0}$ is the number of unselected features and N_{Total} is the total number of features.

b. Discriminant Ability: The chosen feature set must have high discriminant ability. The discriminant ability of a feature can be defined as a measure of how differently two sequences p and q are distributed. In the application of pattern recognition, it can be described as the ability to differentiate between two classes.

For the fitness function, a normalized metric for the discriminant ability of the feature set is defined as:

$$Dis = \left(\frac{RE_{genes=1}}{RE_{Total}} \right) \tag{3.2}$$

where RE is a measure of relative entropy as given by Eq. 2.19.

Practically, the relative entropy of the feature set can be found by multiplying a diagonal matrix of individual relative entropies by the transpose of the chromosome vector:

$$\begin{bmatrix} RE_1 & 0 & 0 & 0 \\ 0 & RE_2 & 0 & 0 \\ 0 & 0 & RE_3 & 0 \\ 0 & 0 & 0 & RE_4 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = RE_{Total} \tag{3.3}$$

c. Diversity: The chosen features should be spread out between different sensors, so that if one sensor fails or is corrupted, the detection scheme will still be able to function with a minimal loss in classification accuracy. The standard deviation of sensor feature size is proposed as a measure of feature set diversity:

$$\sigma = \sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (F_i - \mu)^2} \tag{3.4}$$

where N_s is the total number of sensors, F_i is the feature size of sensor i and μ is the mean feature size of the sensors.

One problem with this measure of diversity is that a feature size of zero is a solution with the highest diversity. To avoid this triviality, a modified normalized metric for the feature diversity is defined as:

$$Div = \left(\frac{\sigma}{N_{genes=1}} \right) \tag{3.5}$$

Using these criteria, a multi-objective fitness function is defined as:

$$F = \alpha_1(S) + \alpha_2(Dis) + \alpha_3(Div) \tag{3.6}$$

46

where α_1, α_2 and α_3 are the weights for the three objective functions and $\sum_k \alpha_k = 100$.

Since the performance of the classifier will be dependant on the weights of the fitness function criteria, an intuitive tuning procedure is proposed to optimize the weight selection:

i. *Start the procedure with a high α_1 (weight for size criteria) and a high $\alpha_2 : \alpha_3$ (discriminant weight to diversity weight) ratio.*

ii. *Classify the uncorrupted data. If the classification accuracy is high, proceed to step 3. If the classification accuracy is low, repeat step 1 with a lower α_1*

iii. *Classify the corrupted data. If the classification accuracy is high, the chosen feature set is good and the algorithm can be terminated. If the corrupted classification accuracy is low, proceed to step 4.*

iv. *Decrease the $\alpha_2 : \alpha_3$ ratio and reclassify the corrupted data. If the classification accuracy is high, the chosen feature set is good and the algorithm can be terminated. If the classification accuracy is low, repeat step 1 with a lower α_1.*

This tuning procedure attempts to choose a feature set with the smallest number of features and the highest classification accuracy under both machine failure and sensor failure conditions.

3. **Reproduction.** A rank-based selection procedure is implemented to determine the best parents and an elitism strategy is used to preserve the best individuals in a population. Scattered crossover ($p = 0.8$) and Gaussian mutations ($p = 0.2$) are utilized to increase the population diversity [32].

4. **Termination.** The algorithm terminates when the chromosomes have evolved for 100 generations or the change in the fitness function value between generations is less than 10^-3

3.4. Experimental Results

As described in Chapter 2, the signals from four accelerometers and four microphones are decomposed to four levels using the DB04 wavelet. This results in a candidate space of 128 features. For the raw sensor data, 300 samples (50 samples from each class) are used as training inputs and 300 samples (50 samples from each class) are used as testing inputs. For the corrupted sensor data, 300 samples are used for training inputs and 300 samples are used as training inputs as well. However, the corrupted sensor data are modified to simulate faulty sensor conditions. In the 300 training samples and 300 testing samples, each sample has one sensor turned off (zero value), simulating a catastrophic sensor failure.

To test the feature selection algorithm, a radial basis function network (Chapter 4) is used as the classifier. The code for the Genetic Algorithm and Neural Network classifier is written using MATLAB toolboxes. Table 3.1 summarizes the steps of the tuning procedure. Here Z refers to the percentage of raw sensor data that is classified correctly and Z_c refers to the percentage of corrupted data that is classified correctly.

Table 3.1: GA-Feature Selection Tuning Procedure.

$\alpha_1, \alpha_2, \alpha_3$	Feature No	Z (%)	Z_c (%)	Action
90, 8, 2	117	68	61.5	$\downarrow \alpha_1$
80, 16, 4	110, 120	61.5	56.5	$\downarrow \alpha_1$
70, 24, 6	79, 104, 120	84	79.5	$\downarrow \alpha_1$
60, 32, 8	40, 79, 96, 104, 120	95	91.5	$\downarrow \alpha_1$
55, 35, 10	10, 32, 48, 51, 79, 88, 104, 113	96	93.5	$\downarrow \alpha_1$
52, 38, 10	40, 69, 71, 79, 84, 88, 96:98, 102:104, 106, 110, 113:115, 117:120, 122	99.5	95.5	$\downarrow \alpha_2 : \alpha_3$
52, 30, 18	16, 27, 36, 50, 79, 92, 104, 113	100	99	.....

Figures 3.2 to 3.6 show the optimization procedure for selected steps of the tuning procedure. The upper graph is a plot of the best and mean fitness values of the population as a function of the number of generations the population has evolved. The lower graph shows the chromosome with the highest fitness function when the algorithm terminates. As discussed earlier, if gene is 1, the feature is selected and if the gene is 0, the feature is not selected.

Figure 3.2: Feature Selection at $\alpha_1 = 80$, $\alpha_2 = 16$ and $\alpha_3 = 4$.

Figure 3.3: Feature Selection at $\alpha_1 = 60$, $\alpha_2 = 32$ and $\alpha_3 = 8$.

Figure 3.4: Feature Selection at $\alpha_1 = 52$, $\alpha_2 = 35$ and $\alpha_3 = 10$.
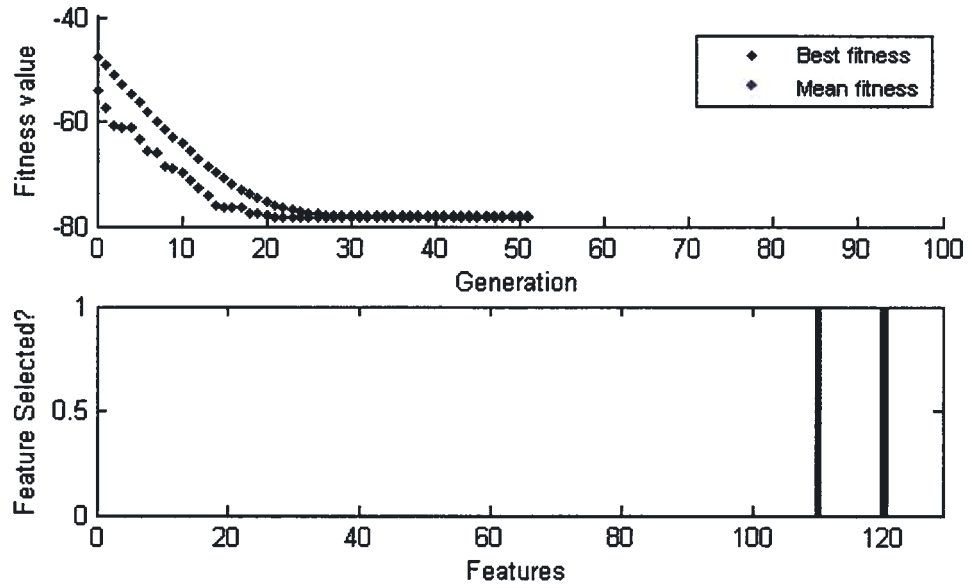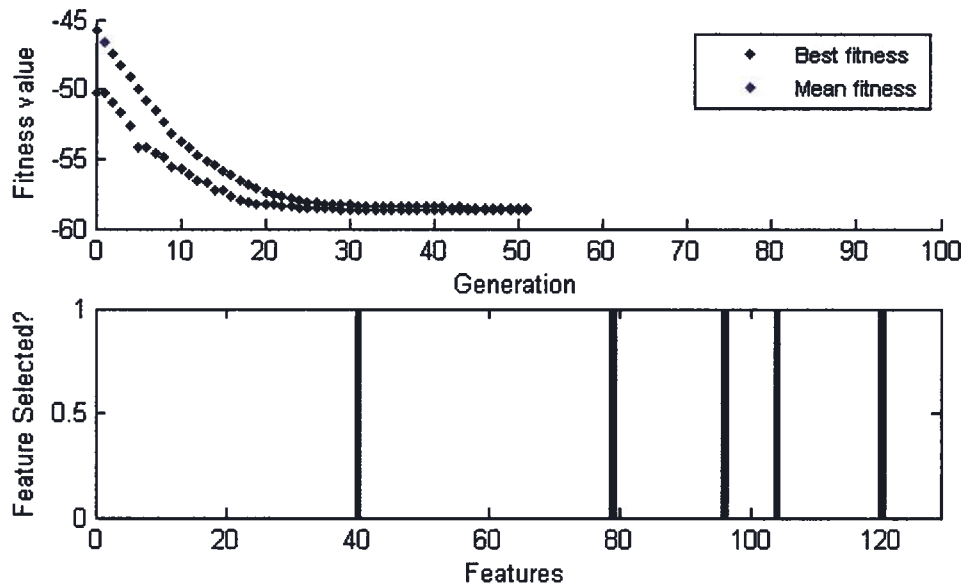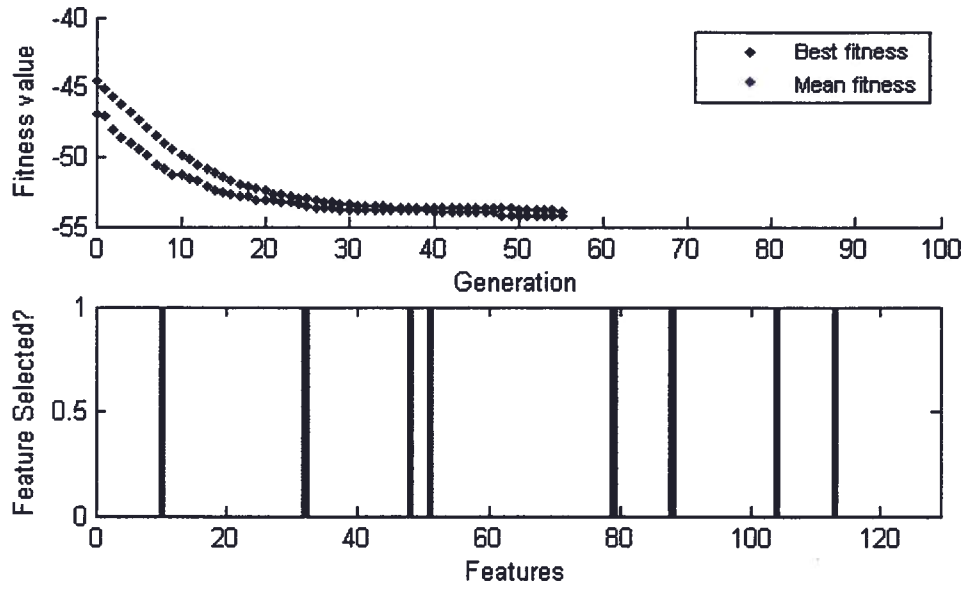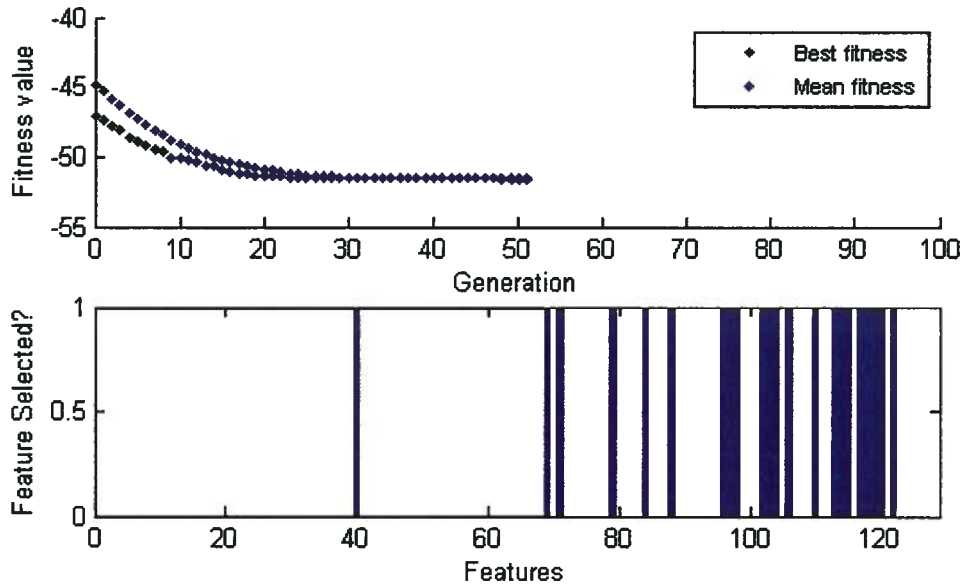

Figure 3.5: Feature Selection at $\alpha_1 = 52$, $\alpha_2 = 38$ and $\alpha_3 = 10$.
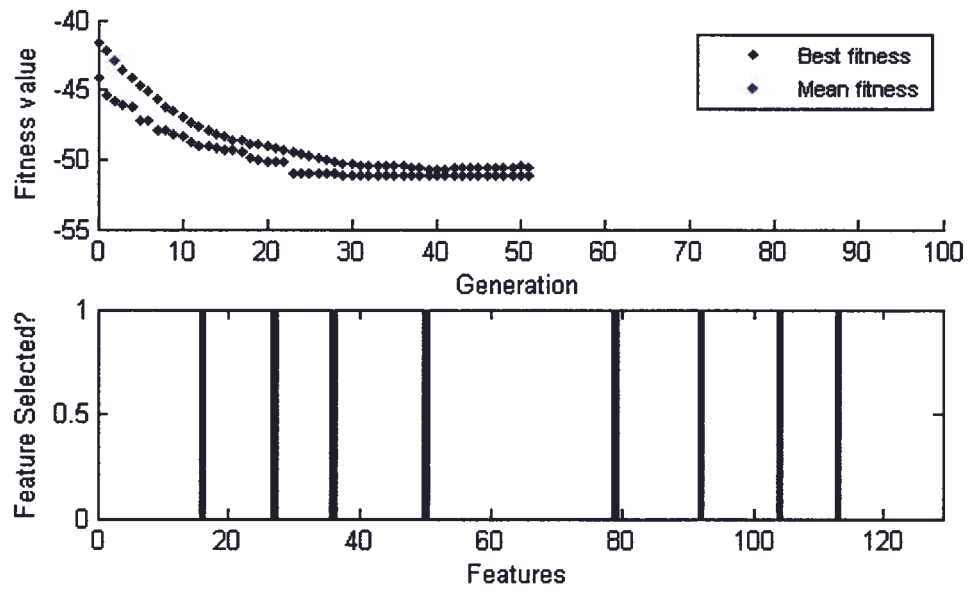
Figure 3.6: Feature Selection at $\alpha_1 = 52$, $\alpha_2 = 30$ and $\alpha_3 = 18$.

3.5. Discussion

As Table 3.1 shows, the raw dataset and corrupted dataset classification accuracies increase as α_1 decreases. When a sufficiently high raw dataset classification accuracy is achieved, the $\alpha_2 : \alpha_3$ ratio is decreased, so the features are better spread out over the sensors and the corrupted dataset classification accuracy increases. The final feature set has 8 features spread out over each of the 8 sensors. If an algorithm (e.g., LDB) is used to rank discriminant measures and select a feature set based on the ranking alone, the results given in Table 3.2 are obtained.

Table 3.2: Feature Selection Based on Relative Entropy.

Features	Sensors used	Z (%)	Z_c (%)
120	Mic #4	25.5	25
120, 113	Mic #4	46.5	44.5
120, 113, 104	Mic #4, Mic #3	62.5	59.5
120, 113, 104, 110	Mic #4, Mic #3	74.5	75
120, 113, 104, 110, 117	Mic #4, Mic #3	77.5	77.5
120, 113, 104, 110, 117, 102	Mic #4, Mic #3	80	79
120, 113, 104, 110, 117, 102, 119	Mic #4, Mic #3	81	78.5
120, 113, 104, 110, 117, 102, 119, 115	Mic #4, Mic #3	79.5	78.5
120, 113, 104, 110, 117, 102, 119, 115, 103	Mic #4, Mic #3	79.5	77.5
120, 113, 104, 110, 117, 102, 119, 115, 103, 97	Mic #4, Mic #3	81.5	79
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98	Mic #4, Mic #3	84	81.5
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98, 118	Mic #4, Mic #3	83	79
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98, 118, 106	Mic #4, Mic #3	83	79.5
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98, 118, 106, 79	Mic #4, Mic #3, Mic#2	93.5	90.5
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98, 118, 106, 79, 122	Mic #4, Mic #3, Mic#2	93.5	86.5
120, 113, 104, 110, 117, 102, 119, 115, 103, 97, 98, 118, 106, 79, 122, 121	Mic #4, Mic #3, Mic#2	93	89

As Table 3.2 indicates, Microphones #4 and #3 have the largest discriminant measures. If the features were chosen based on the relative entropy ranking alone, the raw data set and corrupted dataset classification accuracies are worse than that from the filter selection algorithm proposed earlier. This can be attributed to the two drawbacks mentioned in section 3.1: The multiclass

relative entropy rankings can be easily biased by a large difference between individual classes, and sensor redundancy is not utilized. The features for microphones #3 and #4 indicate a large difference between the baseline conditions, hydraulic system fault conditions, and the electrical motor fault conditions. As a result, the discriminant rankings for these features are high and the aforementioned faults are classified at a high accuracy. However, the microphones are not able to differentiate the other faults in a reliable manner and cannot be used as the sole sources of features for the classification process. Some other criteria (such as sensor diversity) is required to diagnose different kinds of faults under conditions of ideal sensor and unreliable sensor.

Chapter 4

Classification

4.1. Classification

In the context of pattern recognition, classification refers to the process of categorizing data on the basis of one or more traits. Mathematically, classification can be considered a mapping from a feature space x to a label y. Traditionally, there have been three theoretical approaches to classifier design [16]. The most basic approach is to classify data based on similarity. Once there is a good measure of similarity, patterns can be classified on their degree of similarity to existing patterns. This is the underlying concept of template matching and distance-based classifiers. Another approach is to use posterior probabilities to determine the likelihood of a pattern belonging to a certain category. This is the approach used in Baye's rule and logistic classifiers. The third approach is to construct decision boundaries that directly minimize classification error criteria. This approach is considered the most powerful and is well suited for dealing with noisy data and high dimensionality feature spaces [16]. In particular, Radial Basis Function Networks and Support Vector Machines have emerged as popular techniques of nonlinear classification due to their excellent classification accuracies and generalization abilities. The following sections introduce the theoretical foundations of these two classifiers.

4.2. Support Vector Machines

Originally introduced by Boser, Guyon and Vapnik in 1992, Support Vector Machines (SVM) are a powerful set of supervised learning methods for solving problems of classification and regression. A geometrical explanation of the SVM algorithm can be given. Specifically, it constructs a hyperplane that maximizes the margin between two classes of data inputs. The following explanations provide the derivation of the nonlinear, least squares classifier used in the present work [33-35].

Linear Binary Classification

To demonstrate the classification problem, consider the case of classifying the following data into two separate classes:

$$(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)'$$

where $x \in R^N$ is the input data vector and $y \in (-1, +1)$ is the target or known class for x.

Assuming the data is linearly separable, a line (for the case $D = 2$) or a hyperplane (for the case $D > 2$) can be drawn such that it separates the data into two different classes (see Fig. 4.1). The constructed hyperplane has the form

$$w \cdot x + b = 0 \tag{4.1}$$

where,

w is a normal vector to the hyperplane

$\dfrac{b}{\|w\|}$ is the perpendicular distance from the origin to the hyperplane.
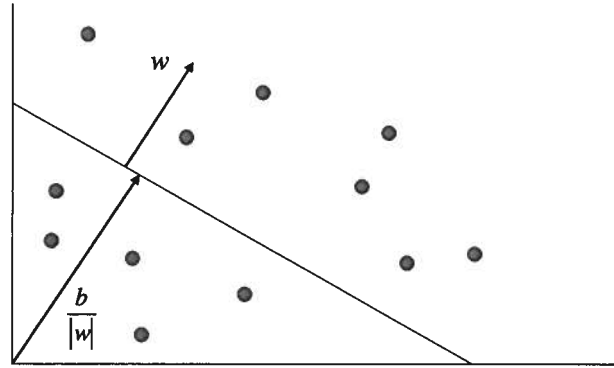
Figure 4.1: Separating Hyperplane.

Given a set of training inputs, the separation into two classes can be described by the following set of conditions:

$$x_i \cdot w + b = \geq 1 \text{ for } y_i = +1 \tag{4.2}$$

$$x_i \cdot w + b = \leq -1 \text{ for } y_i = -1 \tag{4.3}$$

Combining these two inequalities into a single condition, we obtain:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \tag{4.4}$$

We define support vectors on the points that lie closest to the separating hyperplane. Then two planes H_1 and H_2 are defined (see Fig. 4.2) such that they lie on these support vectors and satisfy the conditions:

$$x_i \cdot w + b = +1 \text{ for } H_1$$

$$x_i \cdot w + b = -1 \text{ for } H_2$$
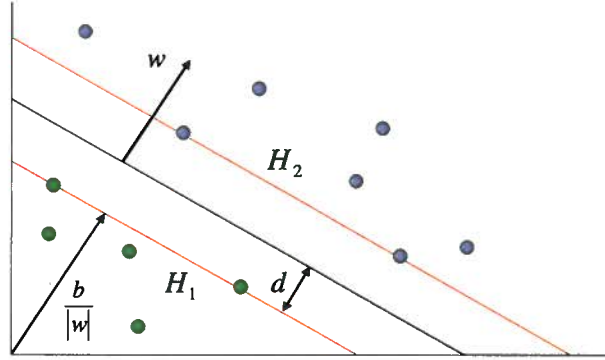
Figure 4.2 Support Vector Machine Principle.

The distances between the separating hyperplanes to H_1 and H_2 are represented as two equivalent distances d_1 and d_2, known as the SVM margin. In order to maximize the distance from the hyperplane to the closest points, it is clear that this margin will have to be maximized. Relating it to Eq 4.4, the margin is equal to $\dfrac{1}{w}$ and the problem is reformulated as the following optimization problem with constraints:

$$\text{Minimize } \|w\| \text{ subject to } y_i(x_i \cdot w + b) - 1 \geq 0 \qquad (4.5)$$

Rewriting the problem such that it can be solved by dynamic programming, we obtain the following problem:

$$\text{Minimize } \frac{1}{2}\|w\|^2 \text{ subject to } y_i(x_i \cdot w + b) - 1 \geq 0 \qquad (4.6)$$

To accommodate the constraints and to ensure that the training inputs are represented as dot products between vectors (see nonlinear classification), a Lagrangian switch is made and the problem is reformulated using Lagrange multipliers α, where $\alpha_i \geq 0 \forall_i$:

$$L_P \equiv \frac{1}{2}\|w\|^2 - \alpha[y_i(x_i \cdot w + b) - 1 \forall i]$$

$$\equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{L} \alpha_i \qquad (4.7)$$

To find the solution to the Lagrangian problem, L_P is differentiated with respect to w and b, and then the derivatives are set to zero:

$$\frac{\partial L}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{i=1}^{L} \alpha_i y_i x_i \qquad (4.8)$$

$$\frac{\partial L}{\partial b} = 0 \quad \rightarrow \quad w = \sum_{i=1}^{L} \alpha_i y_i \qquad (4.9)$$

By replacing the equation for w in the primal form of the Lagrangian L_P, the dual Lagrangian form L_D is obtained and is maximized; thus,

$$L_D \equiv \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \text{ subject to } \sum_{i=1}^{L} \alpha_i y_i = 0 \text{ and } \alpha_i \geq 0 \qquad (4.10)$$

The above formulation can be solved using quadratic programming to find α, which can then be substituted in Eq 4.10 to find w and b. The resulting classifier has the form:

$$y(x) = \text{sgn}[\sum_{i=1}^{\#SV} \alpha_j y_i y_j x_i \cdot x_j + b] \qquad (4.11)$$

where index i counts from one to the number of support vectors.

Linear Non-seperable Classification

Often, there exist data that are not fully separable for various reasons including incompleteness, unreliability, and noise. To accommodate misclassifications, the margin constraints in Eq. 4.4 can be relaxed by introducing a slack variable ξ_i:

$$y_i (x_i \cdot w + b) - 1 + \xi_i \geq 0 \text{ where } \xi_i \geq \forall i \qquad (4.12)$$

The objective function can be redefined to include this relaxation of constraints:

$$\text{Min } \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{L} \xi_i \text{ subject to } y_i (x_i \cdot w + b) - 1 + \xi_i \geq 0 \ \forall_i \qquad (4.13)$$

where C is a positive real constant that represents the trade-off between slack variable penalty and margin size.

The Lagrangian is reformulated as

$$L_P \equiv \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{L} \xi_i - \sum_{i=1}^{L} \alpha_i [y_i (x_i \cdot w + b) - 1 + \xi_i] - \sum_{i=1}^{L} u_i \xi_i \qquad (4.14)$$

57

Setting $\dfrac{\partial L}{\partial w} = 0$, $\dfrac{\partial L}{\partial b} = 0$ and $\dfrac{\partial L}{\partial \xi_i} = 0$, the dual form of the Lagrangian which can be solved by

quadratic programming, is obtained:

$$L_D \equiv \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \text{ subject to } \sum_{i=1}^{L} \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C \quad (4.15)$$

Nonlinear Classification:

In 1995, Vapnik was able to extend the linear classification technique to perform nonlinear classification by mapping the input data to a high dimensional space in which it might be separable. To do so, a kernel function is defined such that $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ where $\varphi(x)$ is a nonlinear mapping from the feature space to the Hilber space H (see Fig. 4.3). The replacement of (x) with $\varphi(x)$ and expression of a kernel as the inner product of $\varphi(x)$ is commonly referred to as the "Kernel Trick." This enables us to work in high dimensional spaces with explicitly performing computations in that space.
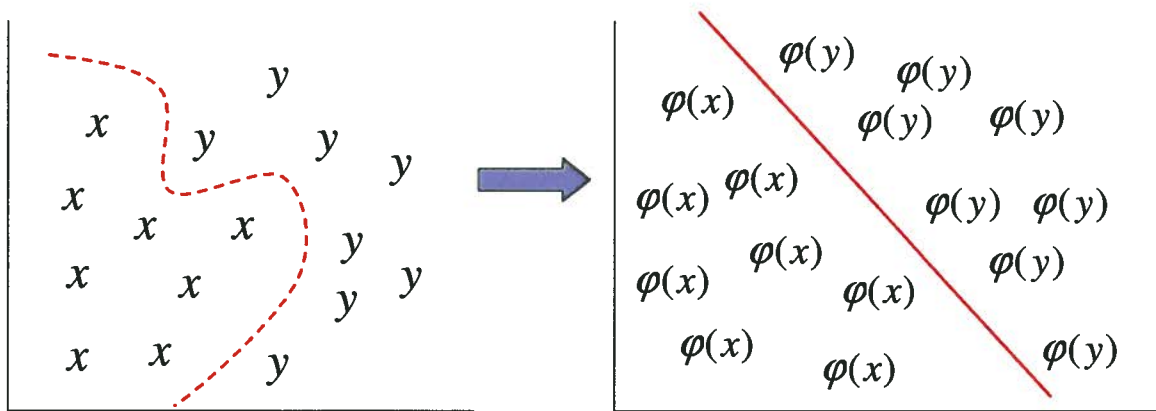

Figure 4.3: Nonlinear Mapping φ.

Many types of Kernels can be used to map the input data into a higher space. Among the choices include:

Linear Kernel: $K(x_i, x_j) = x_i \cdot x_j$

Radial Basis Function: $K(x_i, x_j) = \exp\left(\dfrac{-\| x_i - x_j \|^2}{\sigma^2}\right)$

Polynomial Kernel: $K(x_i, x_j) = (x_i \cdot x_j + 1)^d$

Multilayer Perceptron Kernel: $K(x_i, x_j) = \tanh(k_1 x_i \cdot x_j + k_2)$

Since the RBF Kernel is considered a good first choice for many classification applications, it will be used in the present thesis for comparison against the RBFN [25-27]. However, in the present work, the linear and polynomial kernels will also be tested for wrapper feature selection.

A similar procedure to linear classification is followed to obtain the nonlinear classifier. The primal form of the Lagrangian is now written as:

$$L_P \equiv \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{L}\xi_i - \sum_{i=1}^{L}\alpha_i [y_i(\varphi(x)\cdot w + b) - 1 + \xi_i] - \sum_{i=1}^{L}u_i\xi_i \qquad (4.16)$$

By setting the Primal Lagrangian derivatives to zero, the subsequent dual form is obtained:

$$L_D \equiv \sum_{i=1}^{L}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \text{ s.t. } \sum_{i=1}^{L}\alpha_i y_i = 0 \text{ and } 0 \le \alpha_i \le C \qquad (4.17)$$

The resulting classifier has the form:

$$y(x) = \text{sgn}[\sum_{i=1}^{\#SV}\alpha_j y_i K(x_i, x_j) + b] \qquad (4.18)$$

Least Squares Classification

Suykens [33] proposed a modification to the original SVM algorithm so that the solution is obtained by solving a linear set of equations rather than using Quadratic Programming. The optimization problem was redefined as:

$$\text{Min } \frac{1}{2}\|w\|^2 + \gamma\frac{1}{2}\sum_{i=1}^{L}e_i^2 \text{ subject to } y_i(\varphi(x)\cdot w + b) = 1 + e_i \qquad (4.19)$$

The original formulation by Vapnik was modified by changing the inequality constraint in Eq. 4.4 to an equality constraint and using a square loss function for error variable e_i. The Primal Lagrangian is written as

$$L_P \equiv \frac{1}{2}\|w\|^2 + \gamma\frac{1}{2}\sum_{i=1}^{L}e_i^2 - \sum_{i=1}^{L}\alpha_i [y_i(w\cdot x + b) - 1 + \xi_i] - \sum_{i=1}^{L}u_i\xi_i \qquad (4.20)$$

Setting the derivatives to zero: $\frac{\partial L}{\partial w} = 0$, $\frac{\partial L}{\partial b} = 0$, $\frac{\partial L}{\partial e_i} = 0$ and $\frac{\partial L}{\partial \alpha_i} = 0$ we obtain a linear Karush-Kuhn-Tucker (KKT) system [33]:

$$\begin{bmatrix} 0 & y^T \\ \hline y & \Omega + I/\gamma \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \qquad (4.21)$$

where $\Omega = Z^T Z$ and the kernel trick is applied in the Ω matrix, $\Omega_{ij} = y_i y_j K(x_i x_j)$

As seen above, the linear KKT has a square system with a unique solution (for full rank matrix) which is much easier to solve than the convex optimization problem required by Eq. 4.17.

Finally the classifier takes the form:

$$y(x) = \text{sgn}[\sum_{i=1}^{\#SV} \alpha_j y_i K(x_i, x_j) + b] \tag{4.22}$$

As this formulation shows, there are two parameters that need to be tuned for designing a soft margin, least squares SVM with radial basis kernel function: the regularization parameter γ and the kernel parameter σ. These parameters are obtained by performing a "leave-one-out" cross validation procedure on the sample data set [34]. In leave-one-out cross validation (see Fig. 4.4), the classifier is trained multiple times with all but one of the training samples. The missing sample is used to test the obtained classifier. The error value will guide the choice of subsequent hyperparameters.

Note: In the wrapper feature selection test, it is very time consuming to cross-validate for each feature subset. Therefore, the parameters are tuned only once for the neural network feature inputs at the same conditions.
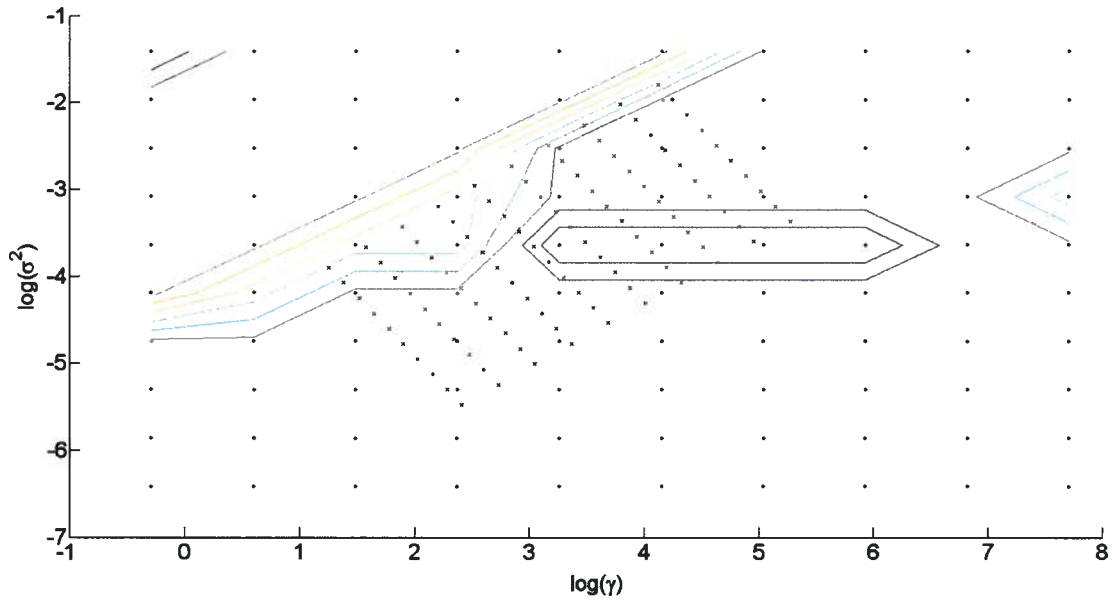
Figure 4.4: Leave-one-out Cross Validation to Find Hyperparameters γ and σ.

4.3. Radial Basis Function Network

Radial basis function networks are a special type of feedforward neural networks which use radial basis functions as the activation function. The basic structure of an RBFN (see Fig. 4.5) consists of an input layer, one hidden layer with a radial basis activation function and an output layer [32]. The connection weights between the units of the input layer and the units of the hidden layer are all equal to 1. Researchers have shown that an RBFN with a sufficient number of hidden nodes can be used as a universal approximator, and this is a useful feature for fault diagnosis due to the stochastic nature of the feature selection process. The downside of using an RBFN is that some classification problems may require a large number of hidden layer neurons to achieve satisfactory results.
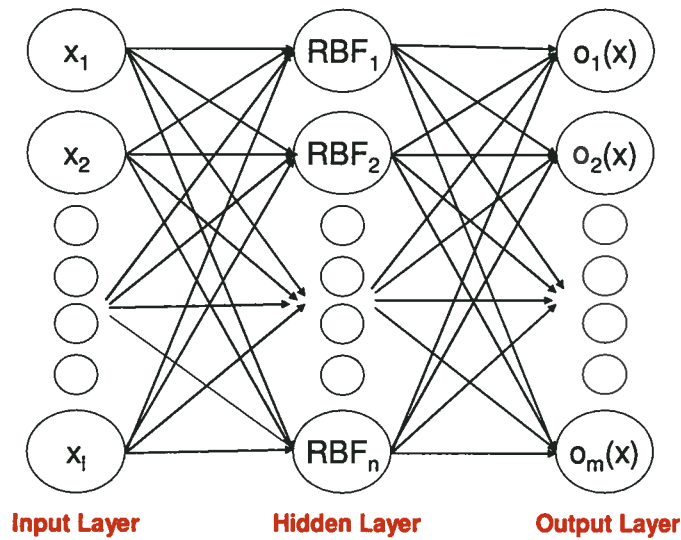
Figure 4.5: Structure of the Radial Basis Function Network.

In this structure, there is a nonlinear transformation between the input layer and the hidden layer, and a linear transformation between the hidden layer and the output layer. This allows the input space to be cast nonlinearly into a higher dimension space in which it might be linearly separable. The nonlinear transformation has the following characteristics: it is symmetrical, has a maximum at the center of the activation function, and has positive values that decrease from the center. The resulting output of the function will be bounded and localized as a result. The general form the RBF function can be written as:

$$g_i(x) = r\left(\frac{\|x - v_i\|}{\sigma_i}\right) \qquad (4.23)$$

61

where x is the input vector, v_i is the vector describing the center of function g_i, and σ_i is the unit width parameter of g_i.

A commonly used function for g_i is the Gaussian kernel function described by

$$g_i(x) = \exp\left(\frac{-\| x - v_i \|^2}{2\sigma_i^2}\right) \qquad (4.24)$$

The output of the RBFN with n neurons in the hidden layer and r output units can be described by:

$$o_i(x) = \sum_{i=1}^{n} w_{ij} g_i(x) \qquad j = 1,\ldots,m \qquad (4.25)$$

where w is connection weight between the i-th perceptron and the i-th output and g_i is the activation function.

There are three parameters that one can use to train the network once the structure of the network is selected: the center and width (normalization parameter) of the radial function, and connection weights between the hidden layer and the output layer. The traditional method to finding these parameters is to employ a two stage approach. First the center and width of the radial function are determined using an unsupervised clustering algorithm; k-means clustering in this case. Then the connection weights are found using a supervised learning algorithm; backpropogation in this case.

4.4. Experimental Results

Support Vector Machines and Radial Basis Function Networks are tested with two feature selection schemes. The feature space is identical to what is used in section 3.3. There is a candidate feature space of 128 features resulting from 8 sensor signals decomposed into 4 levels each. For the raw sensor data, 300 samples (50 samples from each class) are used as training inputs and 300 samples (50 samples from each class) are used as testing inputs. For the corrupted sensor data, 300 samples are used for training inputs and 300 samples are used as testing inputs, with every sample of training and testing having one sensor turned off.

4.4.1 Filter Selection

The features selected in section 3.3 for the RBFN are tested for the SVM to compare the classification accuracies for uncorrupted and corrupted data sets (see Table 4.1).

Table 4.1: Filter Feature Selection.

$\alpha_1, \alpha_2, \alpha_3$	Feature No	NN Z (%)	NN Z_c (%)	SVM Z (%)	SVM Z_c (%)
90, 8, 2	117	68	61.5	80	79
80, 16, 4	110, 120	61.5	56.5	69	57
70, 24, 6	79, 104, 120	84	79.5	77.5	79.5
60, 32, 8	40, 79, 96, 104, 120	95	91.5	96	94.5
55, 35, 10	10, 32, 48, 51, 79, 88, 104, 113	96	93.5	96.5	97
52, 38, 10	40, 69, 71, 79, 84, 88, 96:98, 102:104, 106, 110, 113:115, 117:120, 122	99.5	95.5	98	94.5
52, 32, 18	16, 27, 36, 50, 79, 92, 104, 113	100	99	99	98.5

As Table 4.1 shows, the SVM performs better than the neural network approach for the lower feature set sizes, but the neural network approach is able to achieve slightly higher classification accuracies at larger feature set sizes for both corrupted and uncorrupted data sets.

4.4.2 Wrapper Selection

In addition to testing the classification accuracies of the features selected in section 3.3, a wrapper selection process is used to determine the optimal feature set. Wrapper feature selection aims to maximize the prediction accuracy of the classifier. To do so, a search strategy can be implemented to search the feature space for candidate features, and the selection of features is determined by the classification accuracy of the classifier. For the search strategy, a genetic algorithm is executed in the same manner as described in section 3.2. However the fitness function is evaluated in the following way: First, the classifier is trained with the training data for the selected features. Then, the classifier is tested with validation data using the selected features, and the % of data that is accurately classified is calculated. The fitness function is defined simply as

$$F = \% \text{ accuracy} + \text{small size penalty} \qquad (4.26)$$

There is a small size penalty in the fitness function to ensure that if multiple feature sets give the same highest classification accuracy, the smallest feature set will be chosen.

Different kernel functions are tested to determine if the choice has an impact on the classification accuracy and the size of the obtained feature set. A linear and 2nd order polynomial kernels are tested against the radial basis function kernel under normal and faulty sensor conditions. The results are summarized in Table 4.2. The kernel hyperparameters are tuned once with the features selected for the artificial neural networks (ANN) and are left constant during the feature selection algorithm. In the first column of Table 4.2, the classifier type and the data set are indicated, and the second and third columns indicate the highest classification accuracy achieved and lowest number of features required. Figures 4.6 to 4.9 show the optimization procedure for the RBFN and RBF-kernel SVM.

Table 4.2: Wrapper Feature Selection.

Classification	Features	Accuracy
RBFN	4	100 %
RBFN Corrupted	6	100 %
RBF-SVM	3	100 %
RBF-SVM Corrupted	11	100 %
Lin-SVM	20	89.5 %
Lin-SVM Corrupted	39	91.5%
2nd Poly-SVM	22	83.5 %
2nd Poly-SVM Corrupted	53	79.5 %

As Table 4.2 shows, the RBFN and the RBF-SVM provide very similar performance. They are both able to achieve high classification accuracies while reducing the feature set size substantially for normal sensor data and faulty sensor data. However, the linear kernel and polynomial kernel yield much worse performances than that from the RBF kernel. They are only able to achieve accuracies of 89.5 % and 91.5% for normal and corrupted sensor data, respectively. This could be due to two reasons: one possibility is that these kernels are more sensitive to hyperparameters than the RBF kernel. Another possibility is that they do not perform as well as the RBF-kernel SVMS under conditions of noisy data and missing inputs. Also the training time in the order of increasing length was as follows: ANN, RBF-SVM, Linear Kernel SVM, and Polynomial Kernel SVM. The SVM training times were significantly longer than the ANN training time. This observation is not consistent with other research [27]. This could be attributed to the algorithms programming inconsistencies in different software environments.
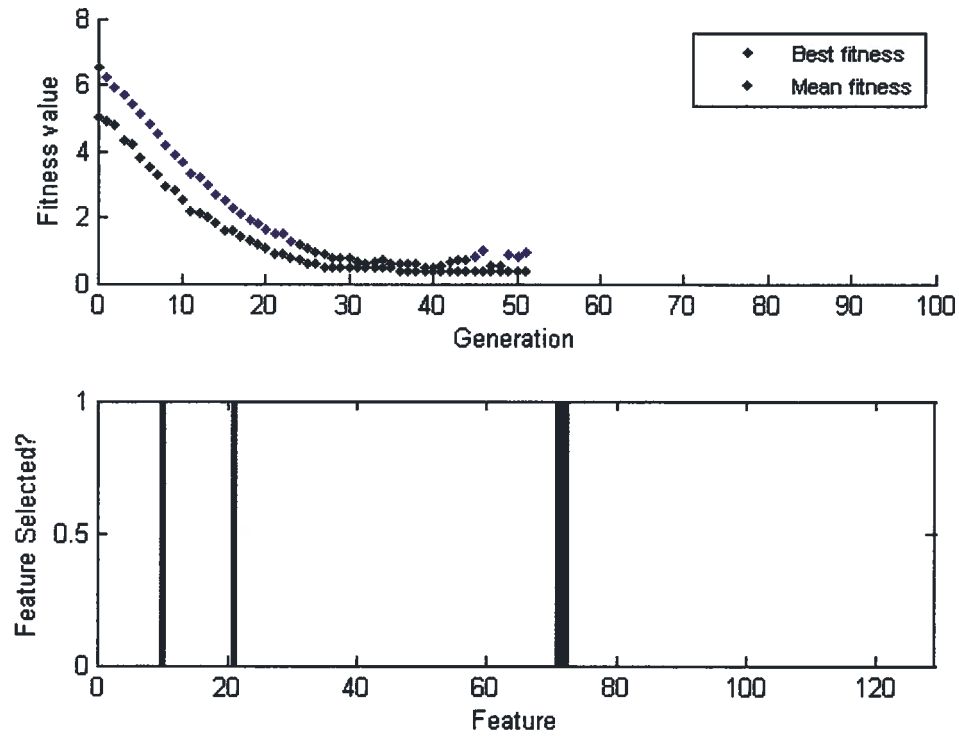
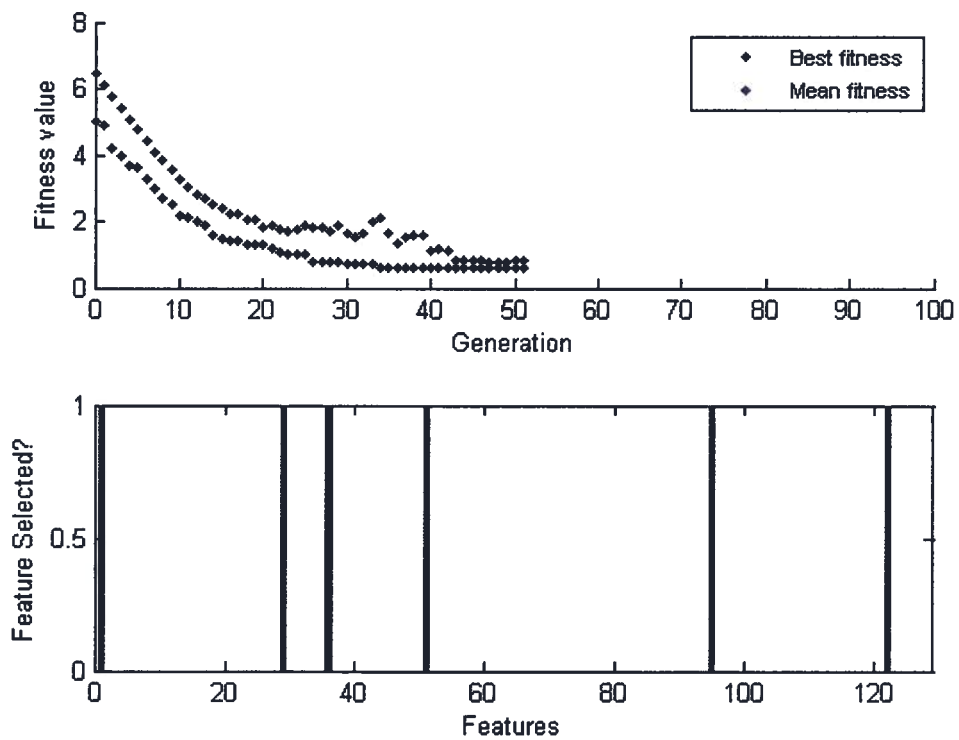Figure 4.6: Wrapper Feature Selection for RBFN.

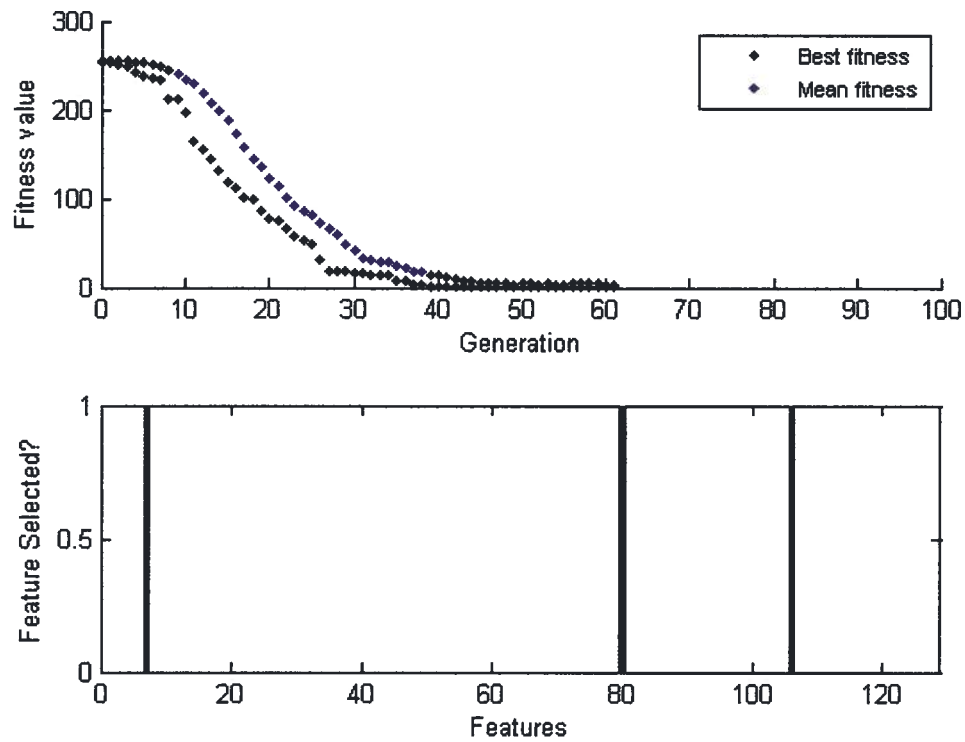Figure 4.7: Wrapper Feature Selection for RBFN w/ Corrupted Sensor Data.

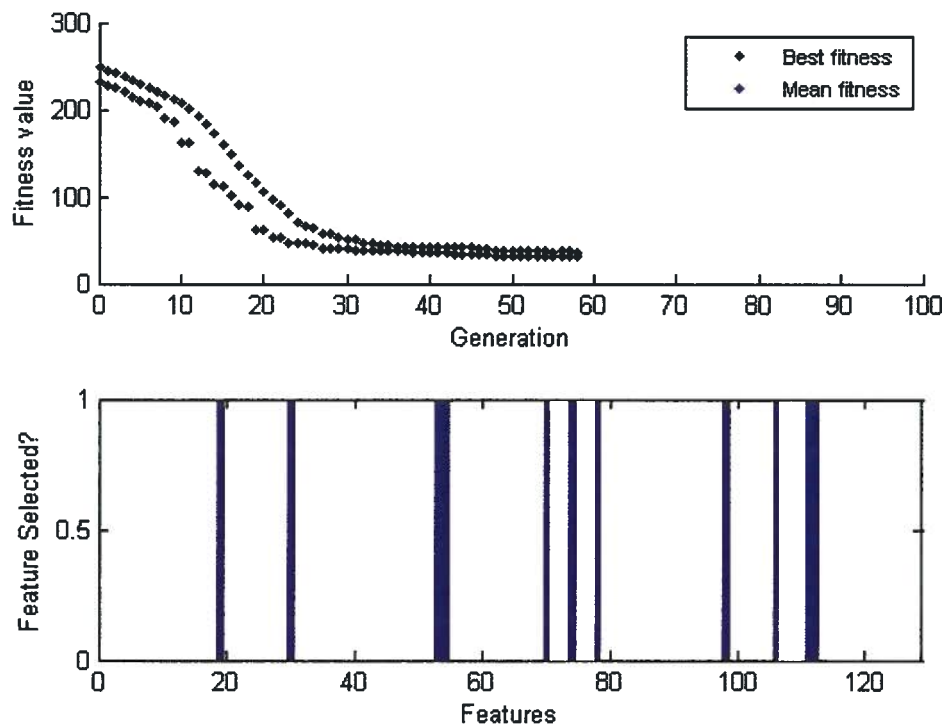Figure 4.8: Wrapper Feature Selection for RBF-SVM.

Figure 4.9: Wrapper Feature Selection for RBF-SVM w/ Corrupted Sensor Data.

66

Chapter 5

Conclusions

5.1. Synopsis and Contributions

In this thesis, a multisensor-based condition monitoring scheme was developed and tested on an industrial fish processing machine. Two on-off catastrophic faults, three gearbox faults, and two other partial faults were physically implemented and sensor faults were simulated for evaluating the developed methodology. The machine was instrumented with four accelerometers and four microphones to continuously acquire vibration and sound signals. The signals were represented with the wavelet packet decomposition and node energies were used to generate a feature vector. Different analyzing wavelets were tested and it was determined that the choice did not significantly impact the pattern recognition process. A simple statistical analysis indicated that the differences between the feature vectors for blade dullness and fish jam were not significant; therefore, not included in the data set for pattern recognition.

To improve the classification accuracy and reduce the computational cost, a multi-objective genetic algorithm and tuning procedure was developed, which reduced the dimensionality of the feature space. Experimental tests demonstrated the effectiveness of the scheme and also showed the drawbacks of using the discriminant measure alone as a means for reducing the feature set size. Two classifiers, Radial Basis Function Networks and Support Vector Machines, were introduced and tested using input features from the proposed filter selection scheme and a wrapper feature selection scheme.

The classifiers were tested under conditions of ideal sensor data and corrupted sensor data. The RBF-kernel SVM and the RBFN performed well under both conditions but the linear-kernel SVM and the polynomial-kernel SVM were neither able to achieve high classification accuracies nor small feature subsets. This could be due the nature of the data (noisy and incomplete) or the classifier sensitivity to the tuning parameters. Also, the SVM training procedure took significantly longer time than the neural network training procedure.

The main contributions of this work can be summarized as follows:

- Development of condition monitoring instrumentation and software development to implement an on-line condition monitoring scheme capable of updating the machine status every three seconds with respect to 6 potential machine defects

- Successful application of wavelet packet decomposition to capture fault signatures of 6 machine conditions

- Development of a feature selection method using genetic algorithms and associated tuning procedure, with demonstrated advantages over conventional methods

- Comparison of two classification schemes under conditions of healthy sensors and corrupted sensors

5.2. Future Directions

The implemented fault testing could be expanded to include further types and severities of faults including electrical faults and hydraulic system faults, thus creating a comprehensive fault diagnosis system for the Iron Butcher. In particular, the gear and bearing faults should be investigated with different severities of defects to fully verify the capabilities of the diagnosis system. In the present thesis, the sensor faults were treated as on-off type, which is not always the mode of sensor failure in practice. Different sensor conditions such as saturation and measurement noise may be simulated in the future. Also, the current inability to differentiate between blade sharpnesses and fish jam conditions needs to be investigated further, possibly leading to research in better sensing methods.

More analysis can be performed on the correlation between the physical faults and the signal processing to optimize the feature generation process. In the present work, it is assumed that the wavelet analysis technique is sufficient for capturing the fault signatures. Comparisons can be made with other techniques such as the FFT, STFT and Hilbert Transform. Also, further analysis

may be performed on the physical interpretation of the fault signatures, rather than relying on pattern recognition methods alone.

The genetic algorithm-based feature selection method may be tested against different data sets other the one used, and the validity of the tuning procedure could be further investigated. The performance may be benchmarked against other feature extraction methods such as Principal Components Analysis and Linear Discriminant Analysis, to quantify the assumptions made in Chapter 3 about classification accuracy and computational expense.

In the area of classifier design, an effort may be made to further understand the performance benefits of one classifier over another. Currently, all conclusions about classifiers are made empirically, but a better understanding of the classification mechanism will enable one to design classifiers more confidently for condition monitoring applications.

References

[1] C. W. de Silva, *Vibration and Shock Handbook*. Boca Raton, FL: CRC Press, 2005.

[2] J. Korbicz, J. M. Koscielny, W. C. Cholewa and Z. Kowalczuk, *Fault Diagnosis: Models, Artificial Intelligence, Applications*. New York: Springer, 2004.

[3] Consortium of the project Offshore M&R (NNE5/2001/710), Advanced Maintenance and Repair for Offshore Wind Farms using Fault Prediction and Condition Monitoring Techniques. Kassel, Germany : ISET, 2005.

[4] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," *IEEE Sensors J.*, vol. 2, pp. 107-119, Apr. 2002.

[5] H. Lang, Y. Wang, and C. W. de Silva, "An automated industrial fish cutting machine: Control, fault diagnosis and remote monitoring," *Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on*, pp. 775-780, 2008.

[6] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, pp. 199-221, 3. 2004.

[7] J. Lin, M. J. Zuo, and K. R. Fyfe, " Mechanical fault detection based on the wavelet de-noising technique," *Journal of Vibration and Acoustics*, vol. 126, pp.9-16, 2004.

[8] V. Purushotham, S. Narayanan, and S. A. N. Prasad, "Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition," *NDT E Int.*, vol. 38, pp. 654-664, 12. 2005.

[9] R. Rubini and U. Meneghetti, "Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings," *Mechanical Systems and Signal Processing*, vol. 15, pp. 287-302, 3. 2001.

[10] W. J. Wang and P. D. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *J. Sound Vibrat.*, vol. 192, pp. 927-939, 5/23. 1996.

[11] C. K. Sung, H. M. Tai, and C. W. Chen, "Locating defects of a gear system by the technique of wavelet transform," *Mechanism and Machine Theory*, vol. 35, pp. 1169-1182, 8/1. 2000.

[12] J. Lin and M. J. Zuo, "Gearbox fault diagnosis using adaptive wavelet filter," *Mechanical Systems and Signal Processing*, vol. 17, pp. 1259-1269, 11. 2003.

[13] K. Shibata, A. Takahashi, and T. Shirai, "Fault diagnosis of rotating machinery through visualization of sound signals," *Mechanical Systems and Signal Processing,* vol. 14, pp. 229-241, 3. 2000.

[14] J. Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis," *NDT E Int.,* vol. 34, pp. 25-30, 1. 2001.

[15] J. Wu and J. Chan, "Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network," *Expert Syst. Appl.,* vol. 36, pp. 8862-8875, 7. 2009.

[16] A. K. Jain, R. P. W. Duin, and Jianchang Mao, "Statistical pattern recognition: a review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 22, pp. 4-37, 2000.

[17] G. G. Yen and K. -. Lin, "Wavelet packet feature extraction for vibration monitoring," *Industrial Electronics, IEEE Transactions on,* vol. 47, pp. 650-667, 2000.

[18] N. Saito and R. R. Coifman, Local discriminant bases and their applications, *Journal of Mathematical Imaging and Vision,* v.5 n.4, p.337-358, Dec. 1995.

[19] B. Liu and S. -. Ling, "On the selection of informative wavelets for machinery diagnosis," *Mechanical Systems and Signal Processing,* vol. 13, pp. 145-162, 1. 1999.

[20] R. Tafreshi, " Feature extraction using wavelet analysis with application to machine fault diagnosis," Ph.D. dissertation, The University of British Columbia, Vancouver, BC, Canada, 2005.

[21] R. Tafreshi, F. Sassani, H. Ahmadi, and G. Dumont, " An Approach for the Construction of Entropy Measure and Energy Map in Machine Fault Diagnosis," *Journal of Vibration and Acoustics,* vol. 131, 024501, 2009.

[22] Yang and V. Honavar, Feature Subset Selection Using A Genetic Algorithm, *Feature Extraction, Construction and Selection: A Data Mining Perspective,* pp. 117-136, 1998, second printing, 2001.

[23] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," *Vision, Image and Signal Processing, IEE Proceedings,* vol. 147, pp. 205-212, 2000.

[24] G. G. Yen and W. F. Leong, "Fault classification on vibration data with wavelet based feature selection scheme," *ISA Trans.,* vol. 45, pp. 141-151, 4. 2006.

[25] L. Jack and A. Nandi, "Support vector machines for detection and characterization of rolling element bearing faults," *Proc. Inst. Mech. Eng. Part C,* vol. 215, pp. 1065-1074, 01/01. 2001.

[26] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing,* vol. 18, pp. 625-644, 5. 2004.

[27] G. Lv, H. Cheng, H. Zhai, and L. Dong, "Fault diagnosis of power transformer based on multi-layer SVM classifier," *Electr. Power Syst. Res.,* vol. 75, pp. 9-15, 7. 2005.

[28] M. Basseville, Distance Measures for Signal Processing and Pattern Recognition, *Signal Processing,* vol. 18, pp. 349-369, 1989.

[29] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Machine Learning Res. (Special Issue on Variable and Feature Selection),* vol. 3, pp. 1157-1182, 2003.

[30] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *Information Theory, IEEE Transactions on,* vol. 38, pp. 713-718, 1992.

[31] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, pp. 153-158, 1997.

[32] F. O. Kharray and C. W. de Silva, *Soft Computing and Intelligent Systems Design.* New York, NY: Addison_Wesley, 2004.

[33] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines,* Singapore: World Scientific, 2002.

[34] B. Schölkopf and A. J. Smola, *Learning with Kernels.* Cambridge, MA: MIT Press, 2001.

[35] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, no. 2, pp. 121-167, 1998.

[36] P. M. Bentley and J. T. E. McDonnell, "Wavelet transforms: an introduction," *Electronics & Communication Engineering Journal,* vol. 6, pp. 175-186, 1994.

[37] R. M. Rao and A. S. Bopardikar *Wavelet Transforms: Introduction to Theory and Applications* Reading, MA: Addison-Wesley, 1998.

[38] A. Jensen, A. la Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform.* Berlin: Springer-Verlag, 2001.

Appendix A: Gearbox Data

This section gives detailed information about the parts in the gearmotor.
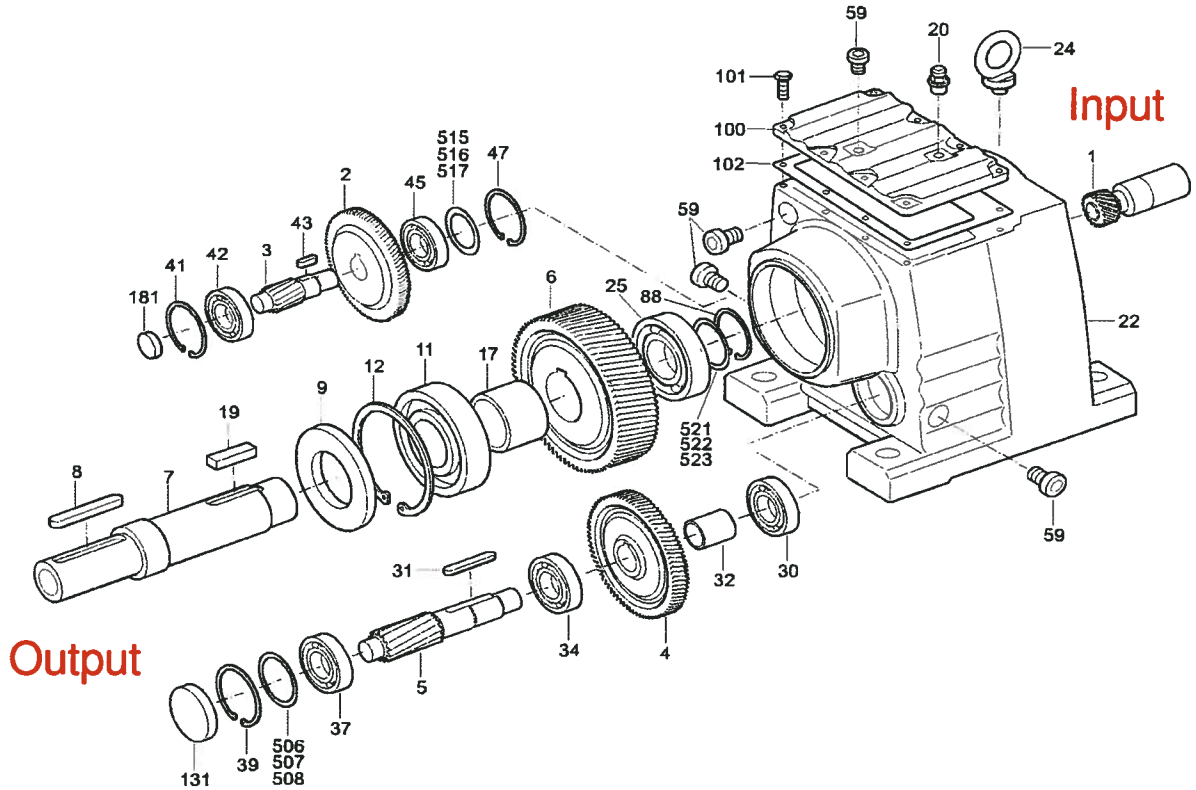
Figure A.1: Gearbox Exploded View. © SEW Eurodrive, 2008, adapted by permission.

Table A.1: Gear Specifications.

Part #	Gear Type	# Teeth
1	Helical pinion	22
2	Helical	69
3	Helical pinion	13
4	Helical	76
5	Helical pinion	16
6	Helical	93

Table A.2: Bearing Specifications.

Part #	Bearing Type	Manufacturer Series
34	Roller Bearing	SKF NCF3004V
45	Ball Bearing	SKF 6202
25	Ball Bearing	NACHI 6206
37	Ball Bearing	NTN 6303

Parallel Helical Reducer
R57A

① R57A with 1.375 in. Dia. Shaft   ③ R57A with 35mm Dia. Shaft
② R57A with 1.250 in. Dia. Shaft   ④ R57A with 30mm Dia. Shaft
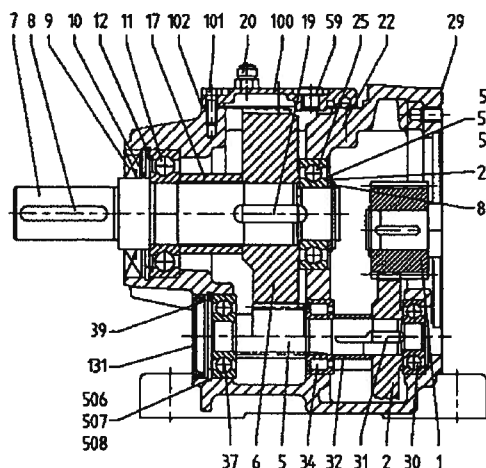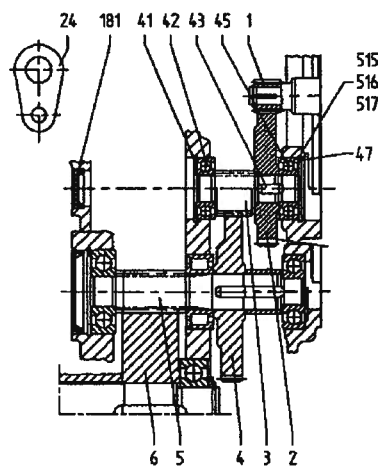
2-gear stages   3-gear stages

When ordering parts, please supply nameplate data with serial number or S.O. number, model number, description of part and part number. All gears are stamped with a part number.

X - As required
Shaded items denote recommended spare parts.
⑤ - Sealant for all flanged sealing surfaces.

Item	Part Name		Description	Part No	Qty	Item	Part Name		Description	Part No	Qty
1	Pinion				1	31	Key	2 Stage	B6x6x16mm	011 601 7	1
2	Gear				1	31	Key - Hardened	3 Stage	B6x6x14mm	011 614 9	1
3	Pinion Shaft				1	32	Distance Piece			643 342 1	1
4	Gear				1	34	Roller Bearing		NCF3004V	013 690 5	1
5	Pinion Shaft				1	37	Ball Bearing		6303	010 507 4	1
6	Gear				1	39	Retaining Ring		47x1.75mm Internal	010 318 7	1
7	Output Shaft	①	1.375 in. Dia.	641 345 5	1	41	Retaining Ring		35x1.5mm Internal	010 314 4	1
7	Output Shaft	②	1.250 in. Dia.	043 759 X	1	42	Ball Bearing		6202	010 485 X	1
7	Output Shaft	③	35mm Dia.	641 344 7	1	43	Key-Hardened		B5x5x10mm	011 483 9	1
7	Output Shaft	④	30mm Dia.	043 644 5	1	45	Ball Bearing		6202	010 485 X	1
8	Key	①	5/16x5/16x1-13/16 in.	806 923 9	1	47	Retaining Ring		35x1.5mm Internal	010 314 4	1
8	Key	②	1/4x1/4x1-11/16 in.	806 922 0	1	59	Drain Plug		M10x1mm	011 426 X	5
8	Key	③	A10x8x56mm	010 029 3	1	88	Retaining Ring		30x1.5mm External	010 277 6	1
8	Key	④	A8x7x50mm	010 023 4	1	100	Inspection Cover			643 209 3	1
9	Oil Seal		BA-SF40x72x10/6mm	017 758 X	1	101	Hex Head Screw		M6x16mm	010 105 2	6
10	Oil Seal		B1-SF40x66x8/10mm	017 434 3	1	102	Gasket			643 504 1	1
11	Ball Bearing		6207-Z-J	011 287 9	1	131	Closing Cap		47x7mm	012 497 4	1
12	Retaining Ring		72x2.5mm Internal	010 322 5	1	181	Closing Cap		37x10mm	010 689 5	1
17	Spacer			641 346 3	1	506	Shim		37x47x0.1mm	010 373 X	X
19	Key-Hardened		B10x8x32mm	010 057 9	1	507	Shim		37x47x0.3mm	010 397 7	X
20	Breather Plug		M10x1mm	013 030 3	1	508	Shim		37x47x0.5mm	012 340 4	X
22	Gearcase			641 341 2	1	515	Shim		25x35x0.1mm	010 369 1	X
23	Thrust Washer		S30x42x2.5mm	010 348 9	1	516	Shim		25x35x0.3mm	010 393 4	X
24	Lifting Lug			164 412 2	1	517	Shim		25x35x0.5mm	010 416 7	X
25	Ball Bearing		6206	010 489 2	1	521	Shim		30x42x0.1mm	010 385 3	X
29	Anaerobic Sealant	⑤	Tube	910 255 8	X	522	Shim		30x42x0.3mm	010 409 4	X
30	Ball Bearing		6302	010 506 6	1	523	Shim		30x42x0.5mm	012 345 5	X

SEW-EURODRIVE, INC.

3/2001

Figure A.2: Gearbox Manufacturer Catalogue. © SEW Eurodrive, 2008, adapted by permission.

Appendix B: Instrumentation

B.1. Signal Acquisition Hardware

This section gives the specifications of the accelerometers, amplifier and DAQ Board used for acquiring vibrations signals from the machine.

Table B.1: DAQ Board Specifications.

Analog Input Specification	NI PCI-7833R
Number of Channels	8 SE/8 DI
Sampling Rate	200 kS/s/ch
Resolution	16 bits
Maximum Voltage Range	-10..10 V
Range Accuracy	7.78 mV
Range Sensitivity	0.305 mV
Minimum Voltage Range	-10..10 V
Range Accuracy	7.78 mV
Range Sensitivity	0.305 mV
On-Board Memory	196 kB
I/O Connector	68-pin VHDCI female

Table B.2: Amplifier Specifications.

Specification	Kistler 5134B
Sensor Excitation Current (mA)	15 max
Sensor Signal Voltage (V)	24
Frequency Range (Hz)	0.1...68000
Output Signal (V)	10 max
Operating Temperature Range (ºF)	32....140
Width (in)	2.791
Height (in)	5.071
Depth (in)	7.331
Mass (kg)	1.75

Table B.3: Accelerometer Specifications.

Specifcation	8702B25	8704B100	8728A500	8730A500
Range (g)	± 25	± 100	± 500	± 500
Sensitivity (mV/g)	± 200	± 50	10	10
Frequency Range (Hz)	1...8000	0.5...10000	2...10000	2...10000
Resolution (mgrms)	2	6	20	10
Shock (g)	2000	2000	5000	5000
Transverse Sensitivity (%)	1.5	1.5	1.5	1.5
Operating Temperature Range ºF	-67...212	-67...212	-67...248	-67...2248
Non linearity (% FSO)	± 1	± 1	± 1	± 1

Temp. coef. Of sensitivity (%/ºF)	-0.0306	-0.0306	-0.034	-0.034
Mass (g)	8.7	7.5	1.6	1.9
Supply (mA)	2…20	2…20	2…20	2…20
Height (in)	0.799	0.98	0.41	0.74

B.2. Interface Software

This section shows the front panel and back panel for the LabVIEW software interface developed in this work.
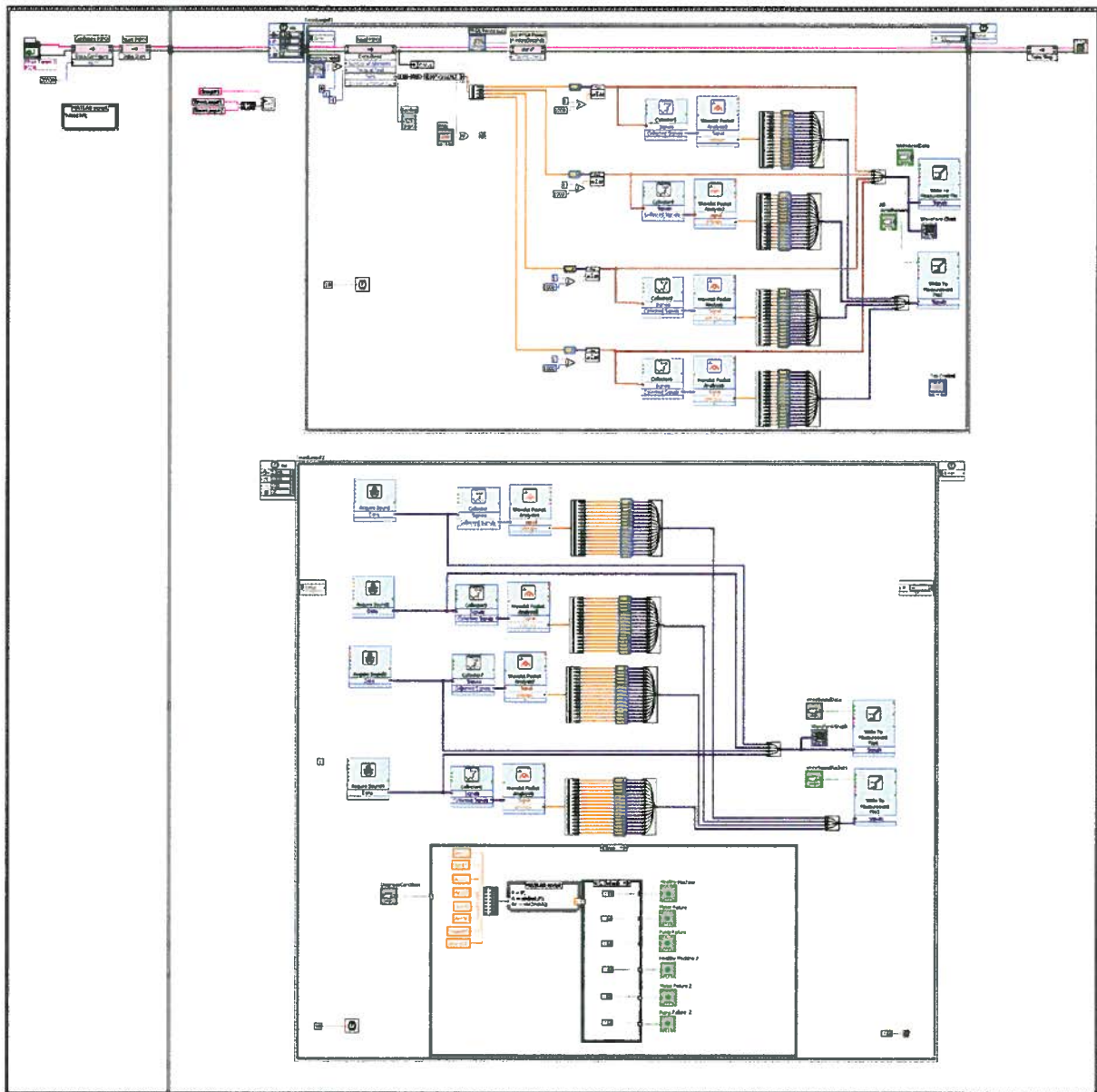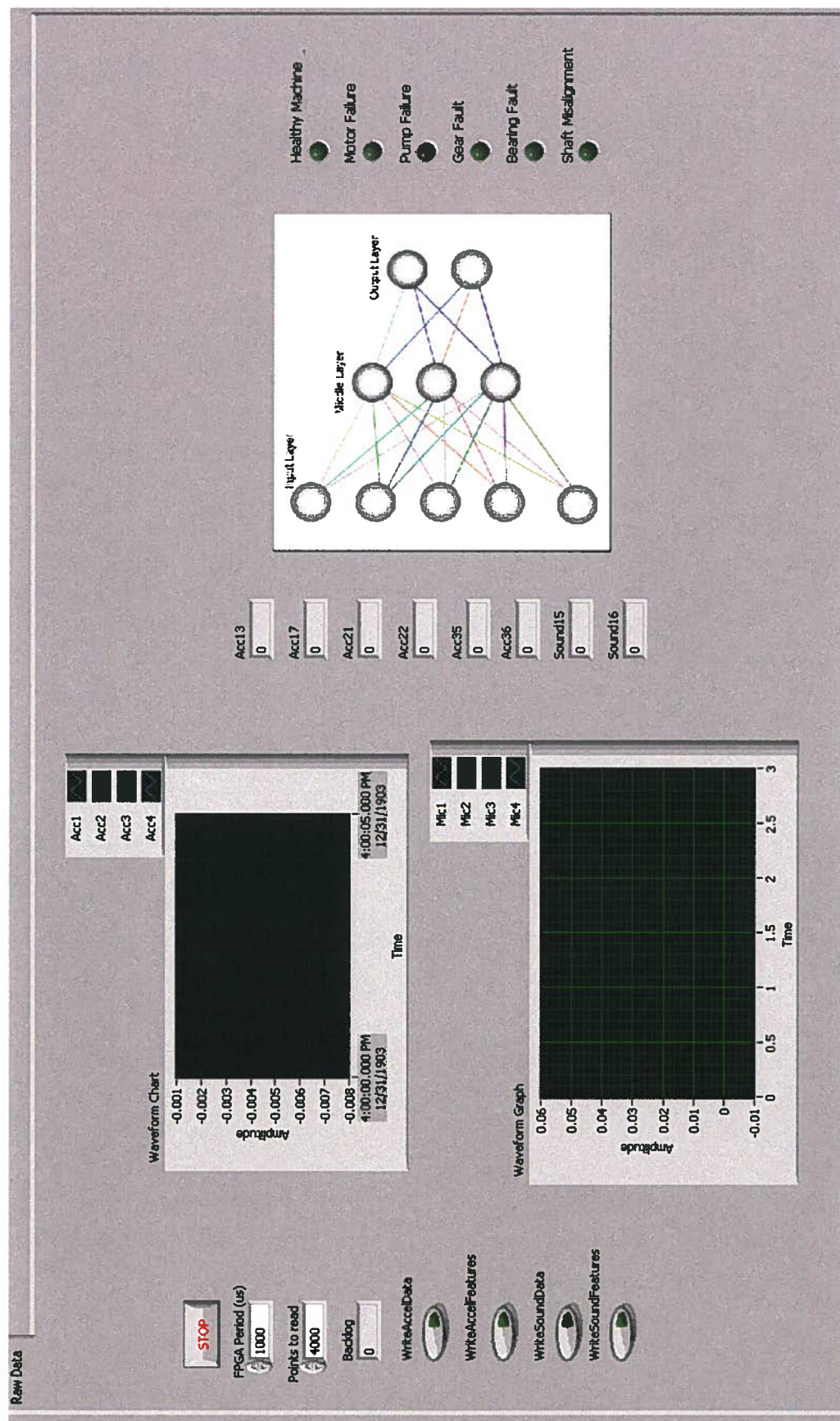

Figure B.1: Block Diagram for LabVIEW Interface.

Figure B.2: Front Panel for LabVIEW Interface.