Bearing Condition Monitoring using Acoustic Emission and Vibration
- The Systems Approach -

Submitted for the Degree of Doctor of Philosophy

by

Tonphong Kaewkongka

Department of Systems Engineering
Brunel University

September 2002

Abstract

This thesis proposes a bearing condition monitoring system using acceleration and acoustic emission (AE) signals. Bearings are perhaps the most omnipresent machine elements and their condition is often critical to the success of an operation or process. Consequently, there is a great need for a timely knowledge of the health status of bearings. Generally, bearing monitoring is the prediction of the component's health or status based on signal detection, processing and classification in order to identify the causes of the problem.

As the monitoring system uses both acceleration and acoustic emission signals, it is considered a multi-sensor system. This has the advantage that not only do the two sensors provide increased reliability they also permit a larger range of rotating speeds to be monitored successfully. When more than one sensor is used, if one fails to work properly the other is still able to provide adequate monitoring. Vibration techniques are suitable for higher rotating speeds whilst acoustic emission techniques for low rotating speeds.

Vibration techniques investigated in this research concern the use of the continuous wavelet transform (CWT), a joint time- and frequency domain method. This gives a more accurate representation of the vibration phenomenon than either time-domain analysis or frequency- domain analysis. The image processing technique, called binarising, is performed to produce binary image from the CWT transformed image in order to reduce computational time for classification. The back-propagation neural network (BPNN) is used for classification.

The AE monitoring techniques investigated can be categorised, based on the features used, into: 1) the traditional AE parameters of energy, event duration and peak amplitude and 2) the statistical parameters estimated from the Weibull distribution of the inter-arrival times of AE events in what is called the STL method.

Traditional AE parameters of peak amplitude, energy and event duration are extracted from individual AE events. These events are then ordered, selected and normalised

before the selected events are displayed in a three-dimensional Cartesian feature space in terms of the three AE parameters as axes. The fuzzy C-mean clustering technique is used to establish the cluster centres as signatures for different machine conditions. A minimum distance classifier is then used to classify incoming AE events into the different machine conditions.

The novel STL method is based on the detection of inter-arrival times of successive AE events. These inter-arrival times follow a Weibull distribution. The method provides two parameters: STL and L_{63} that are derived from the estimated Weibull parameters of the distribution's shape (γ), characteristic life (θ) and guaranteed life (t_o). It is found that STL and L_{63} are related hyperbolically. In addition, the STL value is found to be sensitive to bearing wear, the load applied to the bearing and the bearing rotating speed. Of the three influencing factors, bearing wear has the strongest influence on STL and L_{63}. For the proposed bearing condition monitoring system to work, the effects of load and speed on STL need to be compensated. These issues are resolved satisfactorily in the project.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to Dr. Y. H. Joe Au for his excellent supervision during the course of this project. I must also thank him for his valuable comments, support, and motivation in the completion of the thesis; and I am also very thankful for his advice and enormous encouragement throughout the completion of my PhD.

I would also like to thank Dr. Richard T. Rakowski for his valuable contribution and helpful comments into the present research work. I am also grateful to Prof. Barry E. Jones for his valuable suggestions and for giving me the opportunity to work in the INTErSECT Faraday Partnership Flagship Project (1998-2002), "Acoustic Emission Traceable Sensing and Signature Diagnostics (AESAD)". My thanks also extend to Mr. Winfield Stewart from Corus, Middlesborough UK, who kindly provided data and equipment for some of the experimentation.

I would like to express my gratitude to the Royal Thai government and the Department of Physics, Faculty of Science, Chulalongkorn University for their support in awarding a PhD scholarship.

My special thanks go to my friends, Jirapong Lim and Pakorn Kaewtrakulpong, for their advice and comment. I would also like to thank the departmental technicians, in particular Mr. Brian Shaw for his expertise and help in setting up the test rigs used in this work.

I wish to give my wholehearted thanks to my parents, Pol. Col. Buncha and Nuleebol Kaewkongka, I am also indebted to them for their unwavering support, loving and care, and to my brothers for their persistent encouragement throughout the period of my PhD study.

Publications relating to this work

[1] **Kaewkongka T., Au Y. H. J., Rakowski R. T. and Jones B. E.** (2000) Continuous wavelet transform and neural network for condition monitoring of rotodynamic machinery. Doctoral Research Conference DRC2000, Brunel, Runnymede. 13 1-4.

[2] **Kaewkongka T., Au Y. H. J., Rakowski R. T. and Jones B. E.** (2001) Continuous wavelet transform and neural network for condition monitoring of rotodynamic machinery. Proc of IEEE Conference IMTC, Budapest. 3 1962- 1966.

[3] **Kaewkongka T., Au Y. H. J., Rakowski R. T. and Jones B. E.** (2001) Condition monitoring of rotodynamic machinery using acoustic emission and fuzzy C-mean clustering technique. Proceeding of Conference COMADEM 2001, Manchester. 49-56.

[4] **Kaewkongka T., Au Y. H. J.** (2001) Application of acoustic emission to condition monitoring of rolling element bearings. Journal of Measurement and Control. 34 245-247.

[5] **Kaewkongka T., Au Y. H. J., Rakowski R. T. and Jones B. E.** (2002) Fuzzy C-mean for bearing lubrication condition monitoring using acoustic emission, IOM Communications: Instrumentation to Innovation – Applications and Developments in Metal Production and Use, London. 3 1-3.

[6] **Kaewkongka T., Au Y. H. J., Rakowski R. T. and Jones B. E.** (2002) A comparative study of Short Time Fourier transform and Continuous Wavelet transform for bearing condition monitoring, International Journal of COMADEM, (Accepted in 2001 and to be published).

Contents

Table of Figures

Table of Tables

1 Introduction

1.1 Need for bearing condition monitoring and diagnosis

Where there are moving parts on a machine, there are bearings. It is therefore hardly surprising to say that bearings are perhaps the most common machine elements. If manufacturing plant is to function efficiently, bearings need to be kept in good repair. Over the past half century, in manufacturing industry the influence of maintenance on a company's ability to make profit has grown immensely (Au, 1990). Partly this is because manufacturing plant has become larger and more sophisticated so that maintenance work requires specialised hence expensive skilled labour. Partly it is because the ever-growing international competition has forced manufacturers to look for ways to reduce or eliminate costly equipment downtime. In addition, bearing failure often leads to catastrophic failure of the whole equipment, putting at risk the health and safety of its operators.

Maintenance is carried out in order to replace, repair, modify or service some identifiable part of manufacturing equipment and intended to ensure that it continues to operate to a specified availability for a specified time. In short, the function of maintenance is to control plant availability. Steady-state availability, at a simple level, can be understood as a measure of the time that equipment is available to do useful work (Kelly, 1978) and may be defined in terms of the mean time between failure (MTBF) and the mean time to repair (MTTR) as,

$$\text{Steady-state availability} = \text{MTBF}/(\text{MTBF}+\text{MTTR})$$

As is evident in this formula, availability can be increased by increasing MTBF and reducing MTTR. Whilst the maintenance manager can do very little to affect MTBF, which is related to the problem of the equipment's own inherent reliability and maintainability decided at the equipment design stage, the manager can reduce MTTR through corrective and preventive maintenance, and hence improve the equipment's availability.

Corrective maintenance is the maintenance that is carried out when the equipment fails or falls below a required level of acceptability, while in operation. This approach of allowing failures to happen can be costly particularly for equipment with high capital and running cost, as is often the case nowadays (Collacott, 1977; Barron,

1996). As mentioned earlier, such incidence of failure can also have other consequences relating to the health and safety of operating personnel.

Preventive maintenance is that which is carried out at predetermined intervals with the intention to reduce the likelihood of the equipment's condition falling below an acceptable condition. Preventive maintenance can be either time-based or condition-based (Mellor, 1988; Williams, 1994). Time-based preventive maintenance uses a fixed interval derived from historical life-history data of the part being maintained. Consequently there is always a risk that the equipment may be 'over-maintained', introducing faults that arise from the maintenance action itself, or 'under-maintained', allowing failure to occur before maintenance work is done. Determination of such a preventive maintenance programme is a difficult management problem.

Condition-based maintenance requires the identification and implementation of an effective condition monitoring technique that can detect incipient failure (Kelly, 1978; Michael, 1979). The more sensitive the technique, the more is the lead time available for planning the maintenance work that requires the organisation and direction of resources (men, spares and equipment) and information (technical description of equipment and maintenance methods).

There are two different kinds of condition monitoring methods, direct and indirect. In the context of bearing monitoring, a direct method typically involves some visual inspection for signs of wear on the surfaces of various elements of a given bearing. As can be readily understood, this approach is very time-consuming and impractical in a manufacturing plant because bearings tend to be rather inaccessible in machines. Even if such an approach is possible, it may not be desirable because the mere act of dismantling and re-assembling a bearing often induces a problem that requires future maintenance effort.

An indirect method relies on the use of observable signals arising from a bearing while in operation. Sensors are used to produce signals containing information on a bearing's condition and the quality of the signals is expressed in terms of the signal-to-noise ratio. A high signal-to-noise ratio is desired, which is often achieved by placing a sensor as close to the given bearing as possible. Successful condition

monitoring suggests that there must be changes in the signals being monitored for signals that are static bear no information. The basic idea of condition monitoring is to identify the changes in the signal between normal and abnormal operating conditions and reveal as much information contained in the signal as various signal processing techniques allow.

There are undoubtedly many advantages of proper condition monitoring. Chief amongst them are:

- Preventing unexpected breakdown often with serious operational, health and safety and environmental consequences.
- Reducing unnecessary maintenance work, often the source of maintenance-induced faults, and hence lowering the cost of maintenance.
- Minimising spares holding by being able to predict their requirements in good time.
- Maximising productivity by increasing equipment availability because incipient failure can be detected so that there is better work control matching resources (men, spares, equipment) to maintenance work load, both predictive and corrective.

It should be mentioned that a bearing's life could be estimated using statistical means (Donald, 1957). Statistical bearing life estimation predicts the fatigue life of a bearing while operating in normal condition. However, its application is problematic if unusual operating conditions such as bearing overload are encountered since overloading shortens a bearing's life dramatically. Condition monitoring can overcome such problem because it bases its decision for maintenance action on the actual condition of the given bearing. Two commonly used techniques for bearing condition monitoring are the vibration and acoustic emission (AE) techniques, both being the subject of research reported in this thesis.

The systems approach proposed in this research comprises the multi-sensor data fusion methods, which make use of vibration and acoustic emission (AE) sensing methods. By employing both methods based on different underlying physical principles, a greater dynamic range of operating speed can be covered. The systems approach is adopted not only for fault detection, involving early damage

identification, but also for fault diagnosis, providing information about the causes of machine failures.

1.2 Aims and objectives

The main aim of the research is to design, test and evaluate a condition monitoring system that uses both vibration (acceleration) and acoustic emission signals for monitoring the health of rotating machines operating over a wide speed range and subjected to different loading conditions.

To achieve the aim the following objectives are defined:

- To carry out a complete literature review of research work on bearing condition monitoring and diagnosis and the relevant knowledge and theory to enable a deeper understanding of different types of bearing (Chapter 2) and vibration and AE techniques (Chapter 3).
- To adopt a systems approach to design a multi-sensor monitoring system based on the concept of multisensor data fusion using the two complementary sensing methods: vibration and AE methods and to provide theoretical background and strategies of integrating or 'fusing' data obtained from sensors (Chapter 4).
- To research into the theory of and develop a novel method, called the Shape-to-Life (STL) method, using the Weibull distribution of the inter-arrival times of AE events for bearing condition monitoring (Chapter 4).
- To establish the measurement requirements and then to identify and use the corresponding equipment and techniques for sensing, pre-processing and analysing signals generated from the bearing; this includes selection of transducers and filters for signal acquisition and conditioning (Chapter 5).
- To implement suitable data fusion techniques on the acoustic emission and acceleration signals obtained from two test rigs (high speed light-duty and low speed heavy-duty) and to validate the performance effectiveness of the monitoring system for a wide dynamic range of rotating speed and load conditions (Chapter 6).
- To use simulated vibration and AE signals to evaluate the relative performance between the two joint time- and frequency- domain signal processing

techniques, CWT and STFT; and to use simulated AE events to establish the reliability and robustness of the STL method to a variable AE detection threshold (Chapter 7).

- To propose a bearing condition monitoring scheme using the STL method, formulating guideline for speed and load compensation and for the setting of an alarm level for bearing condition monitoring (Chapter 8).

The objective of taking measurement of AE and vibration signals is not only to achieve an effective condition monitoring system for a wide range of speeds but also to obtain a quantitative measurement that is highly sensitive to machine faults and hence to provide a route to its causes as machine diagnosis. Therefore, experimental work is performed for both simulated bearing faults and actual bearing life test. This is to permit validation of the proposed system as an effective machine diagnostic method. In particular, the bearing life test is to provide realistic bearing faults as encountered when a bearing is operated from new to its final failure. Conclusions and the validity of the systems approach are presented in Chapter 9, which also includes a recommendation for future work, if a fuller understanding and improvement to the proposed system are to be achieved.

2 Background to bearings and condition monitoring

In order to fully explore the application of bearing condition monitoring using multisensor and data fusion techniques, it is necessary to review the current state of the art of bearings. This chapter is divided into six sections. The first section introduces the subject of bearings and their types. The second section discusses lubrication. The third section gives a definition of bearing life and its estimation. The fourth section gives a review of bearing failures and their causes. The fifth section introduces maintenance strategies and benefits of its use. The sixth and last section provides a classification of condition monitoring techniques.

2.1 Introduction to bearing and its types

Whenever there are machines, there are bearings. By definition, a bearing is a device that supports a rotating shaft or spindle, which slides relative to a sustaining component (Donald, 1957). This relative movement is dictated by the requirements of the mechanism of which the bearing is a part. This is where friction occurs, resulting in energy loss whenever the machine elements are driven.

Generally, bearings can be classified into two categories: plain bearings and rolling element bearings. The classification is based on the nature of friction mechanism that occurs at the contacting interface.

2.1.1 Plain bearings

A plain bearing relies on the relative motion between plain surfaces that are either flat or cylindrical. Figure 2.1 gives an example of the plain bearing. In order that motion is possible with minimum resistance, friction between the two surfaces must be reduced as much as possible. This can be achieved by smoothing the two surfaces and by introducing a film of lubricant between them.

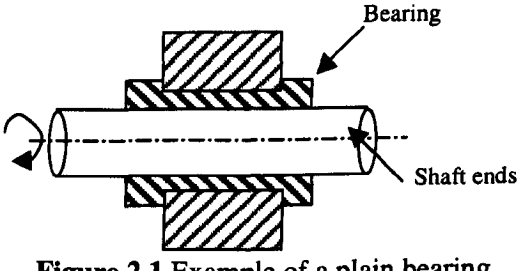


Figure 2.1 Example of a plain bearing

The material of a plain bearing is always softer than that of the shaft or slider it supports. This allows the bearing to wear out rather than the shaft or slider. The design of plain bearing assemblies is usually such that the bearing material itself can be easily replaced.

In order to achieve satisfactory operation of a plain bearing, maintenance of a correct lubricant film is required. Therefore, the bearing surface is grooved to act as reservoirs so that constant distribution of lubricant across the bearing surface is maintained.

2.1.2 Rolling element bearings

By definition, rolling element bearings, also called rolling bearings, include all forms of bearings that utilise the rolling action of balls or rollers to provide minimum friction in the constrained motion of one body relative to another. Both ball and roller bearings have been used extensively as part of rotating machine elements. Rolling bearings are used to permit rotation of a shaft relative to some fixed structure. Some rolling bearings permit translation of a fixture in the direction governed by a stationary shaft; and a few rolling bearing designs permit a combination of relative linear and rotary motion between bodies.

Rolling elements, either balls or rollers, are held between two raceways as shown in Figure 2.2. A soft metal cage or retainer separates the rolling elements and ensures that they are evenly spaced. Sliding friction is replaced by rolling friction; the latter always produces far lower frictional resistance.

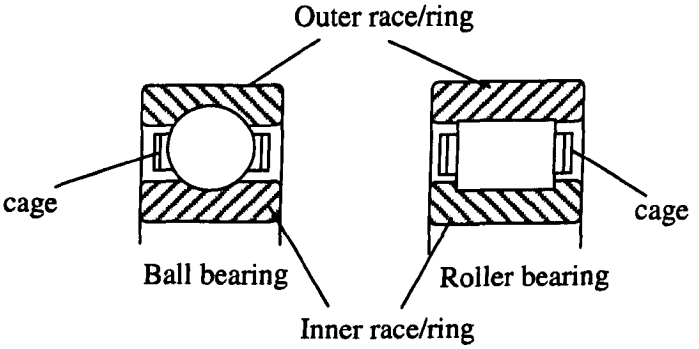


Figure 2.2 Examples of rolling element bearings

The choice of balls or rollers as the rolling elements depends on the operating condition and loads. Ball bearings can operate at higher speeds without overheating. They are less expensive for lighter loads, have lower frictional resistance at light loads and are available in a wider range of sizes. On the other hand, roller bearings can carry heavier loads because of the larger size of rollers, are less expensive for heavier loads and more readily tolerate shock pulses and impact loading.

The variation in the area of contact between balls and rollers gives rise to differences in performance. The ball has a small area of contact that approximates to a point depending on the loading and hence deformation of the ball and raceway. In contrast, the roller has a greater area of contact, which approximates to a line. The difference is illustrated in Figure 2.3. The larger area of contact of the roller enables it to carry heavier loads and withstand impact better, but causes an increase in the frictional resistance at low loads, as contrasted with ball bearings giving lower resistance at low loads. However, ball bearings suffer greater deformation at high loads, resulting in a much larger frictional resistance than do roller bearings.

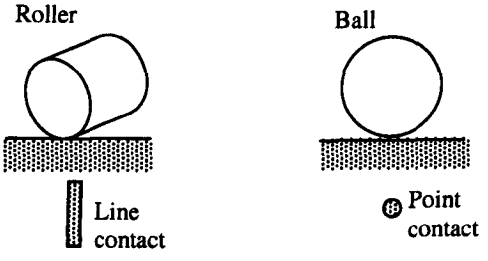


Figure 2.3 Examples of contact points

2.2 Lubrication

One of the important considerations for smooth bearing operation is the lubrication. Lubrication is used to reduce the friction between two contacting surfaces. Sliding friction between the rolling elements on the one hand and the cage and raceways on the other are accommodated by the lubricant. In bearing operation, contact loadings are high, but the geometry of the bearing is such that hydrodynamic lubrication can be achieved during rotation and a thin film of lubricant can be maintained continuously (Jeffrey, 1991).

Moreover, the lubricant can provide the protection of raceways and rolling elements against corrosive pitting and rust. Not only rusting can produce pitting, the oxide itself is highly abrasive (Collacott, 1977). The lubricant also protects against contamination from water and dirt. The lubricant must additionally be chemically neutral and non-corrosive. The lubricant also acts as a heat transfer medium to maintain an even temperature throughout the bearing. Typically the choice of lubricant lies between the use of grease and oil, which will be discussed in the next sections.

2.2.1 Grease lubrication

Most lubricating greases comprise oil compounded with one or more metallic soaps. In general, the oil used depends on the viscosity required and the type of soap depends on the anticipated temperature of the bearing at operating conditions. During bearing operation, oil is gradually released from the grease to lubricate the bearing. The metallic soaps are generally made up of lithium, sodium or calcium greases, which prevent water from penetrating through mechanical seals without losing their lubricating properties. An anti-rust additive can also be included in its formulation in order to help inhibit the corrosive activity. Consequently, grease types are classified according to the type of soap base used to thicken the oil (Arvid, 1945; Barron, 1996).

Calcium-base (or lime)

Calcium-based greases are low in cost and are used for general purposes when bearing operating conditions such as running speeds and temperatures are not too high. These greases have properties that make them non-soluble in water and prevent water from penetrating through seals. However, their uses are limited to applications below 70 degrees Celsius.

Sodium-base (soda)

Sodium-based greases are not water repellent, but can absorb water without losing their lubricating properties. Normally, these greases have a high dropping point (see below) of 150 degrees Celsius. They can be used for machine elements near a heat source.

Lithium base

Lithium greases are resistant to water and also have particular advantages of not easily washed out of bearings even by large amount of water. They can tolerate higher speeds and temperatures than do lime greases and may operate at temperatures up to 150 degrees Celsius or even higher, given speeds are not too great.

2.2.2 Properties of greases

The properties of greases, being semi-solid rather than pure liquid, can be distinguished with the following set of parameters (Jeffrey, 1991):

Hardness: Since greases are semi-solid then they can be categorised as ranging from hard to soft. The level of hardness is obtained from the results of a penetration test.

Dropping point: This is the temperature where the grease transforms its physical property from semi-solid into liquid as may happen to a solid when it reaches its melting point.

Pumpability: The pumpability property is a measure of how well the grease can be passed through a system.

Stability: This property indicates the ability of grease to maintain its characteristics, such as its thickness or hardness, with time and operating conditions.

Water resistance: The water resistance is an important property that determines whether or not the grease will dissolve in water. This can help prevent bearings being penetrated by water through mechanical seals.

2.2.3 Oil lubrication

Even though grease has many advantages, oil lubrication is in many cases more convenient and more effective. For example, oil is preferred when the bearing is enclosed in a common housing with other components for which oil is essential. An example is the enclosed transmission system where operating temperatures or speeds are too high for grease and where minimal friction at the point of contact is required (Arvid, 1945).

To make use of oil in rotating machines, the simplest way is to lubricate a rolling element bearing by running it in an oil bath. This method is effective when operating at slow to medium speeds. However, the level of oil should not be higher than the centre of the rolling element at its lowest point in order to avoid churning of oil. At high speeds, it is essential to reduce as much as possible the amount of oil in the bearing housing, at the same time maintaining effective lubrication. Drip feed may be used for this purpose. When operating at high speeds and heavy loads, pump-fed circulation from a central oil reservoir can be used in order to assure that an adequate supply of lubricant can be maintained.

2.2.4 Properties of oils

The properties of lubricating oils (liquids) can be measured by the following parameters (Jeffrey, 1991):

Viscosity: The viscosity is the most important characteristic, which governs the thickness of a fluid and its flowability. This can also be described as resistance to flow. Generally when temperature increases, viscosity decreases and vice versa.

Viscosity index: This is the rate of change of viscosity with temperature. Ideally, it is desirable to have constant viscosity over a wide range of temperatures. A high viscosity index gives an indication that the oil is more likely to have such a property. In contrast, a low viscosity index indicates that the oil tends to thin out when temperature increases.

Oxidation resistance: Since oil comprises hydrocarbon, when it is exposed to atmosphere especially at increased temperatures it tends to absorb oxygen. This causes chemical changes in the oil, which can worsen its lubricating property.

Emulsification: This property of being an emulsion, a mixture of water and oil, is undesirable because it can degrade the lubricating property. Emulsification is a measure of the tendency of oil to mix immediately with water. On the other hand, demulsification is a measure of the readiness of oil to be separated from water in an emulsion.

2.3 Bearing life

The term 'life' can be given as that period of service, terminated by fatigue phenomena (Arvid, 1945). The fatigue phenomenon is caused by the cyclic stresses produced by rolling surfaces after long periods of running and its onset depends on the magnitude of loading under rotation. Fatigue appears initially as a crack and develops into a flaked or spalled area on one or the other of the load carrying surfaces.

2.3.1 Life expectancy

Life is measured in the number of revolutions of the bearing or in the number of hours of operation at a certain speed of rotation. The actual life attained by the individual bearings can vary significantly. In a large batch of bearings, the life of the last bearing in operation may be 20 to 40 times longer than that of the first failed bearing.

Since the extent of the influence of the individual factors on the life expectancy is unknown, the anticipated life of an individual bearing cannot be predicted. Life expectancy (Eschmann, 1958) is therefore a statistical value, which can only be gained from a large number of bearings, identical from a statistical point of view. A typical life expectancy distribution curve, obtained from extensive time consuming running tests with large series of identical bearings, is shown in Figure 2.4.

It is shown that the shortest actual life attained by all bearings is approximately one-tenth of the average life, while the longest actual life is three to four times the average life.

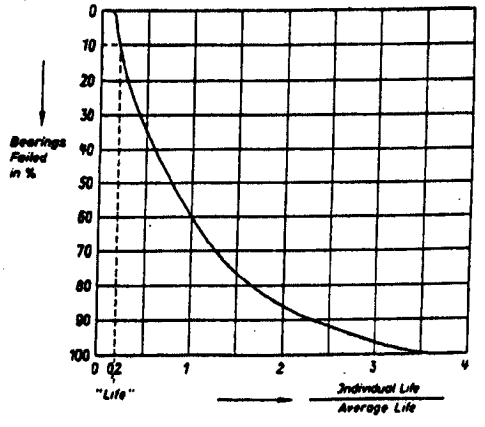


Figure 2.4 Cumulative distribution curve for the actual life (Eschmann, 1958)

2.3.2 The definition of rating life

Since there is a need for a common fixed value for the calculation of bearing life for the permissible percentage of bearing failure, the rating life is then established by bearing manufacturers in order to enable consumers to design their installations reliably and yet economically.

The rating life is the standard method of reporting the results of many tests of bearings. It represents the life that 90% of the bearings would attain at a rated load. The rating life is referred to as the L_{10} life at the rated load. The rating life, L_{10}, of a group of apparently identical ball or roller bearings is defined as the number of revolutions (or hours at some given constant speed) that 90% of a group of bearings will complete or exceed before the first evidence of fatigue develops. This rating life also corresponds to approximately 1/5 of the average life as shown on the cumulative distribution curve of Figure 2.4 (Eschmann, 1958).

2.3.3 Life equation

To select a bearing from a manufacturer's catalogue, the bearing load carrying capacity and bearing geometry need to be taken into account. For a given bearing the load carrying capacity is given in terms of *the basic dynamic load rating* and *the basic static load rating*.

The basic static load rating, C_0, is the load the bearing can withstand without any permanent deformation of any component. If this load is exceeded it is likely the bearing races will be indented by the rolling elements. Consequently, the operation of the bearing would be rather noisy and impact loads on the indented area would generate rapid wear and progressive failure of the bearing would occur.

The basic dynamic load rating, C, is the constant radial load which a bearing can endure for 1×10^6 revolutions. The life of a bearing with basic dynamic load rating C with a load P can be computed by (Peter, 1998)

$$L = \left(\frac{C}{P}\right)^k \textbf{ million revolutions} \qquad \textbf{Equation 2.1}$$

where k = 3 for ball bearings and k = 3.33 for roller bearings.

2.4 Bearing Failures

A rolling element bearing cannot operate forever. It has a definite life, which is determined by the number of rolling contact cycles, and the load applied on the raceways by the rolling elements. Bearing failure may occur from a variety of causes (Nisbet, 1978). The various types of failure according to causes will be discussed in the following section.

2.4.1 Fatigue

Unless operating conditions are ideal and the fatigue load limit is not reached, the bearing life is limited. The initiation of fatigue appears as a function of the number of revolutions performed by the bearing and the magnitude of the load. Fatigue is the result of cyclical shear stresses and it appears immediately below the load-carrying surface (Eschmann, 1958; Barwell, 1979). This usually occurs in rolling element bearings due to the cyclical effect of the rolling elements, and in plain bearings subjected to fluctuating load patterns. After a time these stresses cause cracks which gradually extend up to the surface. As the rolling elements pass over the crack fragments of material break away and this is known as *flaking* or *spalling*. The flaking progressively increases in extent and eventually makes the bearing unserviceable.

Flaking

The phenomenon called flaking initiates by the development of small fatigue cracks in the surface of the rings or rolling elements. These cracks cause material fragments to break away from the raceways. The first flakes are typically small. However, when the bearing is used continuously, the flaked area spreads by the breaking off of more material on the surfaces. In the end it may stretch over the entire surface of the loaded zone (Arvid, 1945).

In the first stages flaking looks like a small bruise on the surface as shown in Figure 2.5. When the rolling elements pass over the flaked area, flaking then develops and spreads out over the entire raceway as shown in Figure 2.6.

Figure 2.5 Inner ring with initial flaking (Arvid, 1945)

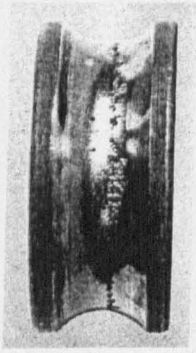


Figure 2.6 Inner ring with spreader flaking (Arvid, 1945)

2.4.2 Faulty installation

The most common fault in installation is excessive internal preloading caused by tight fits or thermal expansion. With excessive preload, the contact pressure increases resulting in noisier operation and an increase in temperature (Charles, 1998).

An unbalanced preload can occur when one of the rings is deformed by an improperly fitted shaft or housing. This results in stress increase in the two opposite regions and shortening the life of the bearing. A bearing not designed to tolerate any tilting of the rings may fail prematurely if it is run with an excessive preload or misalignment. In this condition, the rolling elements moving round the stationary ring will be constantly

constrained to change their speed and direction. This creates friction with the cage and raceways, hence overheating.

Cracks

Crack formations can be of several kinds. A heavy load of short duration can develop a crack. The most common cause is rough treatment when the bearings are being mounted or dismounted (Arvid, 1945; Donald, 1957). Excessive forces applied directly against the ring may cause fine cracks to form. The tensile stresses arising in the rings as a result of excessive drive-up generate cracks when the bearing is put into operation. Cracks can also occur when bearings are heated and then mounted on shafts manufactured to the wrong tolerance. Such cracks are as shown in Figure 2.7.

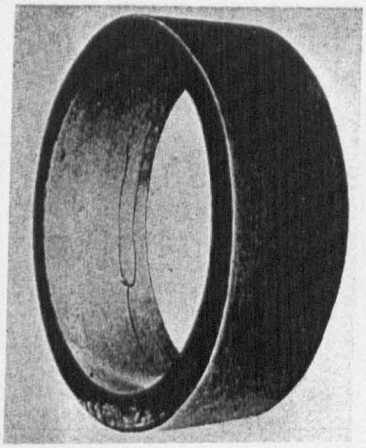


Figure 2.7 Outer ring with cracks (Arvid, 1945)

2.4.3 Contamination

Another frequent cause of bearing failure is the ingress of dirt or other foreign matters, which often can be abrasive, into the bearing housing or between rolling element and raceways (Barwell, 1979; Jeffrey, 1991). Where bearings are lubricated and connected to other machine elements such as a gearbox, it is difficult to avoid the admission of fine metallic matters from the gearbox. The abrasive particles can generate wear on the bearing, whilst metallic particles have a generally similar effect but cause more indentation and pitting.

Wear

In general, wear does not occur in ball and roller bearings as a result of their normal use. Wear occurs as a result of the ingress of abrasive particles into the bearing, which can cause wear of the raceway and flanges as well as of the cage (Nisbet, 1978; SKF, 1994). Also poor lubrication can cause wear. Vibration in bearings that are not running also gives rise to wear.

Abrasive particles such as grit that has entered the bearing can cause wear of raceway, cage and rolling elements. The surfaces become dull according to the coarseness and nature of the abrasive particles. These particles gradually increase as material is worn away from the running surfaces and cage. When a bearing is rotating, the process of wear is accelerated. Finally, the surfaces become worn rendering the bearing unserviceable. The wear caused by abrasive particles is as illustrated in Figure 2.8.

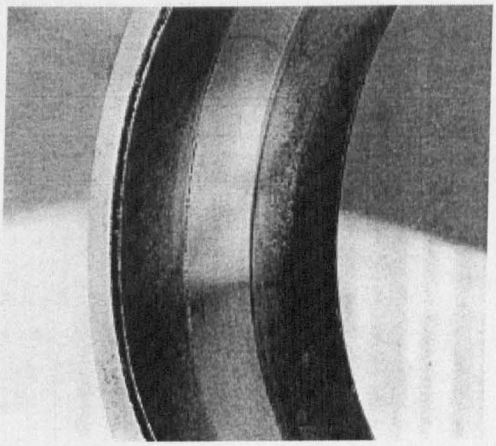


Figure 2.8 Outer ring of a roller bearing worn by abrasive particles (SKF, 1994)

Indentations

Indentations occur when foreign particles get into the bearing and are pressed between rolling elements and rings (Michael, 1979; SKF, 1994). When the particles are rolled into the raceways by the rolling elements, dents will be formed; the particles in question are not necessarily hard. Indentation can also occur when the bearing is subjected to abnormal loading while stationary. Figure 2.9 illustrates such indentations.

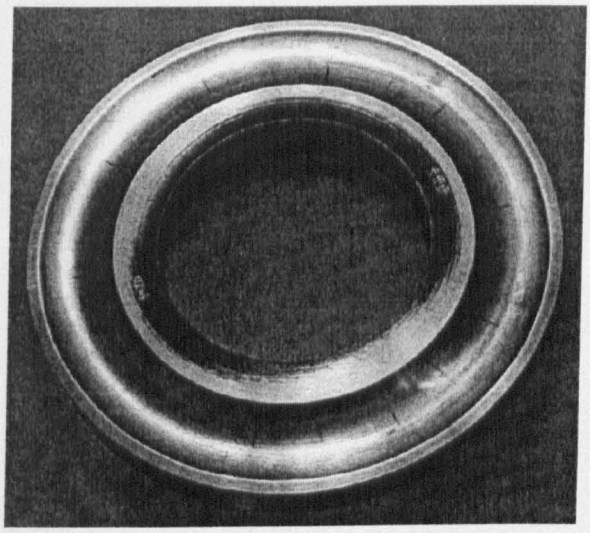


Figure 2.9 A washer subjected to overloading while stationary (SKF, 1994)

2.4.4 Defective lubrication

Lubrication plays a major role in the proper functioning of rolling element bearings. Inferior lubricants containing abrasive particles can cause wear and lead to ultimate failure (Nisbet, 1978). In this case, lubricants fail to protect the working surface because of its inferior lubricating and water resistant characteristics. Applying too much grease or oil can also lead to churning and rapid eventual breakdown. By contrast, bearings operating with too little grease or oil result in excessive friction, hence smearing and seizing.

Smearing

Smearing is a special kind of seizing in its early stage. Smearing can develop between rolling elements and raceways as a result of relative sliding as opposed to rolling, under load with inadequate lubrication (Nisbet, 1978; SKF, 1994). When smearing takes place, the material is generally heated to such a high temperature that re-hardening takes place. This generates stress concentrations that may lead to cracking or flaking. Smearing may also arise when the rolling elements are subjected to severe acceleration on their entry into the loaded zone. An example of smearing is shown in Figure 2.10.

Figure 2.10 Smearing on a roller bearing surface from a spherical roller bearing – 100 x magnification (SKF, 1994)

Surface distress

If an oil film between rolling elements and raceways becomes too thin, the peaks of the surface roughness will come into contact with each other, causing small cracks in the surfaces (SKF, 1994). The phenomenon is known as *surface distress*. However, this kind of crack is different from fatigue cracks that are first initiated beneath the surface leading to flaking. By contrast, the surface distress cracks are microscopically small and increase very slowly to such a size that they interfere with the smooth running of the bearing. Surface distress is shown in Figure 2.11.

Figure 2.11 Surface distress in the form of a band encircling on a rolling element (SKF, 1994)

Insufficient lubricant can also cause wear. Metal to metal contact occurs because it is not possible for an oil film with sufficient load carrying capacity to form (Nisbet, 1978; SKF 1994). This gives the surfaces a mirror-like finish. If the lubricant is totally used up, the temperature will rise rapidly and cause the bearing to seize. An example of wear caused by inadequate lubrication is shown in Figure 2.12.

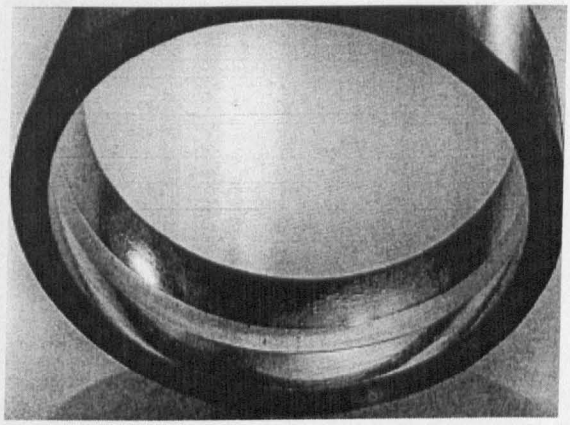


Figure 2.12 Outer ring of a roller bearing worn by inadequate lubrication (SKF, 1994)

2.4.5 Corrosion

Corrosion, especially rust formation, is the enemy of all rolling element bearings. Rust formation in bearings is mostly the result of the entrance of water or humidity. When water or corrosive agents reach the inside of the bearing in some certain quantities such that the lubricant cannot provide protection for the steel surface, this process will soon lead to deep seated rust (Nisbet, 1978; SKF, 1994). In addition, moisture will also condense from the air inside the bearing due to temperature variations. An example of corrosion is shown in Figure 2.13.

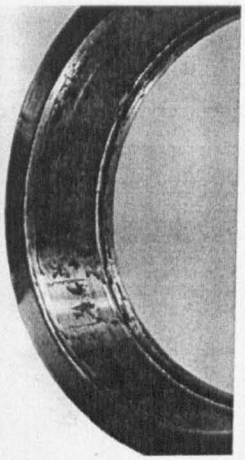


Figure 2.13 Outer ring with deep seated rust (SKF, 1994)

2.4.6 Fretting corrosion (False brinelling)

For a stationary bearing, there is no lubricant between rolling elements and raceways. Fretting corrosion also referred to as false brinelling, takes place initially at the edges of the contact area between the raceways and rolling elements (Nisbet, 1978; Jeffrey, 1991; SKF 1994). The absence of a lubricant film gives rise to metal-to-metal contact and vibration can lead to relative movements between the rolling elements and the rings. As a result of these movements, small particles break away from the surfaces leading to the formation of depressions in the raceways as shown in Figure 2.14.

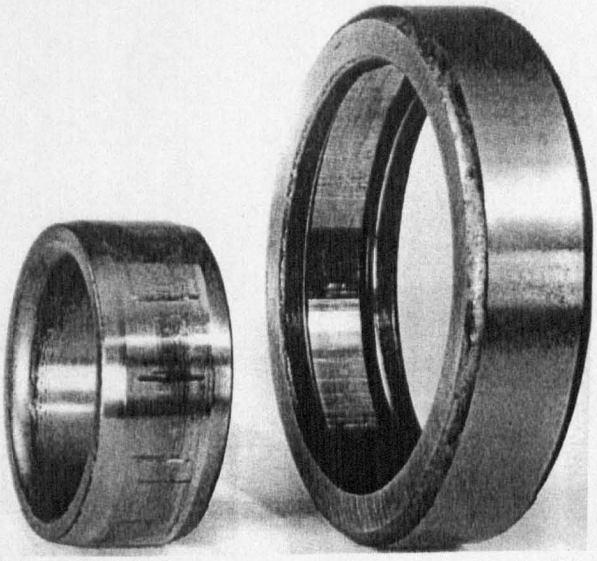


Figure 2.14 Inner and outer ring of a cylindrical roller bearing exposed to vibration (SKF, 1994)

2.4.7 Electrical discharge damage

Electric discharge damage is formed when an electric current passes through a bearing such as in electric motors, either by current leakage or by induction effects. If the potential drop is more than about 0.5 volt, a spark will occur through the thin lubricant film and ultimately damage the bearing components (SKF, 1994). The surface of the damage is similar to electric arc welding. The passage of electric current usually leads to the formation of fluting in bearing raceways. Rollers are also subjected to fluting, while there is only dark discoloration of balls. An example of fluting caused by the passage of electric current is shown in Figure 2.15.

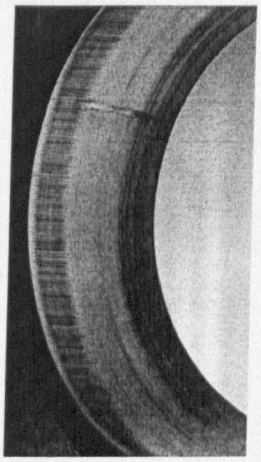


Figure 2.15 Fluting on outer ring caused by electric current (SKF, 1994)

2.5 Maintenance strategies

Condition monitoring and diagnosis (Au, 1990) has played a major role in industry as an evaluation of the condition and its cause of machines. This can be performed by extracting information from machines to indicate operating condition of the machines or instruments. Successful condition monitoring will enable them to be operated and maintained with safety and economy.

Traditionally condition monitoring can be referred to as predictive maintenance. In industry, the maintenance strategies can be classified in terms of the different approaches: breakdown maintenance, regular preventive maintenance and predictive maintenance (Collacott, 1977).

2.5.1 Breakdown maintenance

Breakdown maintenance is one that allows machines to be run until they fail, and then repaired. This is the simplest approach to maintenance and it is very effective where the consequence of failure and cost of machine elements are low (Barron, 1996). It can also be considered as a default maintenance action because the possibility of unpredicted breakdown always exists.

The drawback is severe when the consequence of failure and the price of machine parts are high. Unpredicted failure can alter the production plans and can be costly in terms of lost output. However, it may be the only way if the system fails without giving any warning of impending failure.

2.5.2 Regular preventive maintenance

Regular preventive maintenance is a better strategy compared to breakdown maintenance. In this strategy, a machine is stopped for its condition to be checked after a predetermined time has been reached. This is an improvement of the breakdown maintenance since the regular maintenance decreases the possibility of unplanned stoppages (Williams, 1994; Barron, 1996).

The major disadvantage of regular preventive maintenance is that it still cannot prevent unexpected failures that occur in a random fashion. These unplanned failures

can interrupt production and maintenance activities. If spare parts are not ready, restoration cannot be carried out.

2.5.3 Predictive maintenance

Predictive maintenance is the condition-based maintenance. In this strategy preventive maintenance is carried out at what may be irregular intervals but they are determined by the actual condition of the machine component or system. The main purpose of predictive condition monitoring is to give an early warning as a result of knowing that a machine component or system has become abnormal (Williams, 1994).

Successful condition monitoring can provide benefits such as saving on the capital costs due to fewer standby systems and service spare parts, operation cost due to lower machine downtime or unproductive time and maintenance cost due to less unplanned maintenance and more effective plant operation. It also results in the increased system availability, which relates to the time that a system is available for production, by minimising the time for repair or lost to shutdowns. Hence, it also leads to more consistent quality and improvement of the product because of the more efficient machine operation that can be achieved.

The implementation of an effective bearing condition monitoring requires accurate and reliable fault detection and diagnosis. As mentioned earlier, only using statistical life estimation as preventive maintenance is not sufficient since in the real situation variable conditions often occur such as load and speed variations. This can be misleading if the actual condition of the bearing is unknown. Therefore, research into, and the continuing use of condition monitoring as a maintenance tool has led to the development of many fault detection and diagnosis techniques which will be discussed in following section. Amongst the various techniques, the most popular techniques are vibration and acoustic emission; both will be presented in Chapter 3.

2.6 Various techniques for condition monitoring

2.6.1 Manual inspections

In manual inspection the maintenance operators use their senses of sight, touch, and hearing to make an assessment and interpretation of the operating conditions of machine (Michael, 1979). This method requires experience on the part of the operator and his constant attendance on the machine. This can simply be carried out by dismantling the machine components and looking to check for the obvious defect, hearing the unusual sound generated from the machine, touching to check for overheating and smelling to investigate burning caused by overheating. Its advantage is the simplicity and low cost.

2.6.2 Temperature monitoring

Temperature monitoring is available and important for components that generate, transfer or store energy such as heat. It involves the use of thermocouples or resistance thermometers to perform measurement on bearing temperatures. Thermocouples use wires of dissimilar metals such as copper or constantan, which can produce a voltage proportional to temperature in the linear range. Resistive devices measure resistance changes that occur with changes in temperature (Jeffrey, 1991).

In the case of bearings, temperatures tend to fluctuate according to load so any preset warning levels must take account of normal maximum temperature. The warning level should be based on a rise above normal rather than an arbitrary maximum. This method can indicate warning of several hours before breakdown occurs and can be used to raise an alarm and shut down the machine.

2.6.3 Spectrographic oil monitoring

Spectrographic oil monitoring (Barron, 1996) is a very effective method for assessing the condition of the oil based on its sample and the components that come into contact with it; the sample can be used as a representative of the total contents of the system. The applications of this method are for machines where remote transducer mounting on specimen is difficult to carry out, or perhaps where many moving components are too close together such as in compressors, gas turbines, gearboxes and low speed equipment.

This method requires regular sampling so that trends can be established. Wear will produce a slight and steady increase of the metal levels; the sudden increase occurs when there is malfunction of components. Typical wear indicators are copper, zinc, aluminium, iron, carbon and chromium; contamination indicators include silicon, sodium and boron. One of the limitations of this method is that large particles present in the oil may lead to misleading results.

2.6.4 Particle retrieval

Particle retrieval is used to analyse the wear debris present in the lubricating oil. However, instead of measuring the metal level it is more concerned with the physical size of the particles. Wear debris can be sampled using magnetic plugs in the system downstream of bearings to capture ferrous particles and by using oil filter to collect non-ferrous particles. Then a microscope is used to inspect the physical size and shape of the debris (Mellor, 1988).

During the run-in period of machine elements, large quantities of very small particles can be expected. In normal operating condition, a smaller quantity of wear would be expected which is normally below 25 nanometre in size. A steady increase of particles larger than 0.25 mm over three or more samples gives indication that wear occurs at a rapid rate and failure may be imminent (Jeffrey, 1991).

2.6.5 Corrosion monitoring

Corrosion monitoring can be used in the plant where environment and process fluid can lead to corrosion of certain parts, leading to catastrophic breakdown (Williams, 1994). The corrosion may lead to deterioration in performance of machines and is usually a greater problem for structural components than for rotating machines.

Since corrosion proceeds as an electro-chemical process, the technique for monitoring is directed at the identification of the electrical currents set up by the corrosive activity. A probe that consists of a thin wire is inserted into the process fluid. As the wire corrodes, its cross sectional area reduces and the electrical resistance increases. This can be done with a suitable bridge circuit to give a continuous signal related to the rate of corrosion (Collacott, 1977; Michael, 1979).

3 Review of vibration and AE for bearing monitoring

Research has previously been carried out into bearing condition monitoring and diagnosis using acoustic emission and vibration techniques. In order to obtain a robust and reliable system for bearing condition monitoring, an understanding is important of how a signal is generated from a source, how the signal is propagated through the transducer, what material is used for sensing and its sensitivity and frequency ranges. Knowledge of signal transduction for vibration and acoustic emission needs to be established as vibration and acoustic emission transducers are the primary devices of the monitoring systems. Signal processing techniques that have been used for bearing condition monitoring will then be presented in the latter part of this chapter.

3.1 Introduction to vibration

With increased global competition in production and the greater pressure on making better use of machines, the importance of effective methods for condition monitoring and faults detection in machinery became apparent in the 1970s. Vibration monitoring has been applied successfully for monitoring rotating machinery. Vibration based maintenance is widely accepted as an appropriate maintenance strategy in many sectors of industry involving high cost machines or machines that have long replacement lead time. Machine vibrations produced by industrial mechanical components are key indicators of their health and operating conditions. Many continuous process industries have also found vibration monitoring as effective solutions to avoid unscheduled operation interruption and minimise maintenance cost.

Classical machine components monitored included bearings, shafts, gears and drive belts for which fault mechanisms such as unbalance, misalignment shaft and defects and looseness in bearings can be identified (Wallace, 1970). Given the fact that all machines that contain moving components may vibrate, the cyclical excitation forces transmitted to adjacent components within machines can give rise to motion (vibration). Therefore, the nature and magnitude of the vibrations on a component of the machine can give valuable information for its mechanical condition. The waveforms of vibration signals from rotating machinery are often recorded and analysed by processing the data using various techniques. Each different technique gives some information about the condition of the machine.

To measure machine vibration, it is important to understand the source of machine vibration and its signal characteristics so that meaningful information of the signal can be obtained. Therefore, a review of detection of machine vibration and its signal processing techniques is given in the next section.

3.1.1 Source of machine vibration

All moving components in a machine can produce vibration whose nature and magnitude can provide a meaningful signature of its mechanical condition. Vibration arises from cyclical forces within the machine. These forces can be caused by the configuration of the machine, or by the presence of some defects. Due to the fact that no machine structure is infinitely rigid, machine elements move cyclically in proportion to these excitation forces.

Mechanical noise is produced by the action of some vibrating or cyclically impacting machine elements causing compression waves to transmit through the surrounding materials, which can be solid, liquid, or gas. These compression waves can be detected by hearing, when they result in a phenomenon called sound or noise. Both can be generated from the same source, and contain valuable information about the condition of the machine. The frequencies can vary from a fraction of a Hz up to several thousand Hz.

In rotating machinery, consider a machine incorporating among its many components a ball bearing that has a small pit in one of its races. When a rolling element comes into contact with the pit, a small force impulse is then generated. For every revolution of the moving race, a predictable number of balls will hit the pit, and thus impulses will be produced at some known multiple of the shaft running frequency. Over a period of time, the race defect will deteriorate and the magnitudes of the impulses will increase, thus producing higher levels of vibration.

3.1.2 Vibration signal transduction

To detect vibration, two types of transducers are used: *seismic*, and *proximity* (Michael, 1979). Seismic transducers produce a signal proportional to the absolute motion in space and are capable of detecting the majority of machine defects such as unbalance, misalignment, bent shafts, loose components and rubs. In contrast, proximity transducers generate a signal proportional to the relative motion between the tip of transducer and the point of interest on a vibrating surface.

A proximity (relative motion) transducer is used for sensing displacement between the mounting points, usually the bearing housing, and the point of interest, usually the rotating shaft. The transducers work on the principle of eddy current loss related to the displacement between the tip of the transducer and the target surface as illustrated in Figure 3.1. The coil embedded in the tip of the probe is fed with an A.C. current, whose frequency is of the order of 2 MHz, and the eddy current losses are measured by a bridge circuit, which can produce a D.C. level proportional to the displacement or gap (Broch, 1980, Smith 1989). A significant advantage of proximity transducer is that being mounted in a machine casing it is less exposed to hostile and dirty environments and consequently less likely to be damaged by knocks. The frequency of operation ranges from a fraction of a Hz up to about 1000 Hz.

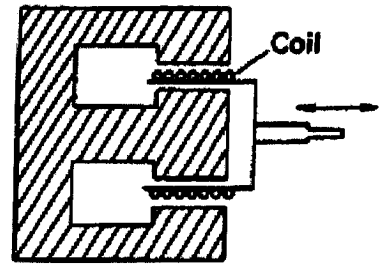


Figure 3.1 Example of proximity transducer (Smith, 1989)

Seismic transducers generate an output signal proportional to the absolute velocity or acceleration relative to a fixed point as illustrated in Figure 3.2. The original type of velocity transducer in the form of a spring-mounted coil produces a signal proportional to velocity. This type of transducer uses the property that a wire when cutting lines of flux in a magnetic field will generate a voltage proportional to the field intensity and the velocity of the wire. They are useful for monitoring the unbalance on rotating machinery. The frequency response is approximately from 10 Hz to 1000 Hz. However, velocity transducers are less sensitive to high frequency vibration compared to acceleration transducers also referred to as accelerometers.

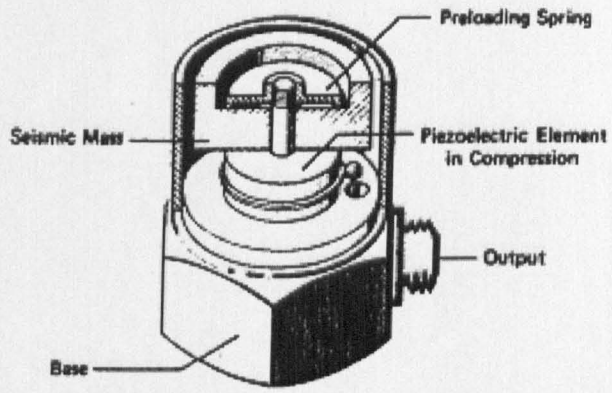


Figure 3.2 Example of seismic transducer (Broch, 1980)

The most widely used seismic transducers are of the acceleration type because of the wider frequency range of operation that they afford and their availabilities in a wide variety of general-purpose applications. Accelerometers have sensing elements that are typically piezoelectric crystals loaded with a small inertia mass and rigidly mounted in a casing. The most important part of the accelerometer is its piezoelectric element that is generally made from a polarised ferroelectric ceramic. It is capable of generating an electric charge directly proportional to strain whenever a load is applied to put the piezoelectric element in tension, compression or shear. The voltage output is proportional to the acceleration over a wide range of frequency, up to the point where the output starts to rise due to the effect of resonance of the inertia mass supported on the crystal. The frequency response can vary from 10 Hz to about 30000 Hz. In recent years, accelerometers have become available with built-in matching charge amplifiers

within accelerometer casings and their sizes are not very much larger than those without amplifiers (Charles, 1998).

Generally, machines of the same type and condition tend to vibrate at constant vibration velocity, independent of speed. Similarly acceleration levels tend to increase and displacement levels tend to decrease with increasing speed. It is advantageous to choose the parameter that gives the flattest frequency spectrum in order to draw out the details over most of the dynamic range of the measuring instrumentation. Therefore, the choice of the type of sensors depends on the frequencies that are to be measured. In summary, displacement, velocity and acceleration measurements are best suited to low frequency, mid-range frequency and high frequency applications respectively.

Preamplifiers and filters are also used for signal conditioning. A preamplifier is used to increase the voltage or charge level of the signal in order to drive alarms or meters, or to match other connecting electronic circuits. A filter used in vibration monitoring is an electronic circuit that allows only certain frequency components of the signal to pass. High pass and low pass filters reject those frequency components that lie below and above some predetermined frequencies respectively. A band-pass filter allows only those frequency components within a predefined band to propagate.

3.1.3 Vibration signal processing

Overall level monitoring

Overall level monitoring technique is sometimes called 'all pass monitoring'. It is used to check the trend of the magnitude of vibration for the general condition of the complete machine. Such monitoring needs to be performed from the initial condition or the base line of the machine so that its irregular condition can be indicated as a changing trend. One drawback is that if a machine fault is not severe or the signal is insensitive to the fault, the overall level may not show any significant change.

Spectral analysis

Spectral analysis can be used to diagnose defects that occur in the bearing by analysing a vibration signal in terms of its component frequencies. The Fourier transform of a function x(t), which represents an analogue signal with finite power, is defined as (Randall, 1977),

$$F[x(t)] = X(f) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}dt \qquad \textbf{Equation 3.1}$$

Causes of defects can usually be identified with a narrow-band spectrum analysis in which the spectral peaks are compared at the calculated frequencies. Since defects in bearing components produce known characteristic vibration frequencies, some of which can be expressed as a multiple of the running speed of the machine, deterioration of a machine component generates an increase in vibration level at its characteristic defect frequencies. These frequencies are dependent on the dimensions of the bearing, the number of rolling elements and the speed of rotation (Michael, 1979); Appendix A lists the formulae for calculating the various bearing frequencies.

Shock pulse method (SPM)

The SPM technique (SPM Instrument AB, 1970) relies on the fact that damage, contamination and other defects in bearings will cause mechanical impacts. The technique works by detecting the development of a mechanical shock wave generated by the impact between two masses. At the instant of impact, mechanical contact occurs and a compression (shock) wave develops in each mass. The SPM method is based on the events occurring in the mass during the extremely short time period after the first particles of the colliding bodies come in contact. To observe the shock pulses, the vibration signal is passed through a band-pass filter in order to isolate one of the resonances. Then the filtered signal is converted into a train of impulses by a pulse circuit converter. By observing the increase in the level and rhythm of impulses, it is possible to establish whether the bearing element is damaged.

Kurtosis

The Kurtosis method (Rogers, 1979; Pachaud, 1997; Shiroishi, 1997) is used to analyse the machine condition by monitoring the statistical parameter known as the kurtosis, which can be derived from the statistical moments of the probability density

function of the recorded signal. Kurtosis is a statistical measure relating to the 'peakiness' of the distribution of vibration amplitude. Specifically, kurtosis is the fourth moment, μ_4, about the mean *E(X)* divided by the fourth power of the standard deviation σ.

$$Kurtosis = \frac{\mu_4}{\sigma^4} = \frac{E[(X - E[X])^4]}{E[(X - E[X])^2]^2} \qquad \textbf{Equation 3.2}$$

where the symbol E is the expectation operator and X is the vibration amplitude. Kurtosis is used as an indicator of defects inducing shocks. A normal bearing has a normal distribution function of vibration amplitudes and the corresponding kurtosis value is equal to three. As wear develops, the kurtosis value of the vibration amplitude distribution grows above 3.

Cepstrum analysis

Cepstrum analysis can be used to provide information about frequencies that repeat themselves in a frequency spectrum. The cepstrum (Tang, 1991) is calculated by finding the inverse Fourier transform of the complex natural logarithm of the power spectrum of a signal to increase the significance of lower amplitudes:

$$x(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left[X(e^{j\omega})\right] e^{j\omega n} d\omega \qquad \textbf{Equation 3.3}$$

A cepstrum is used to highlight periodicities in the spectrum, in the same way that the spectrum is used to highlight periodicities in the time waveform. The harmonics in the spectrum are summed into one single peak in the cepstrum, thus making faults identification easier to identify. For example, gear systems tend to generate sidebands since gear rotational frequency can modulate the tooth meshing component, producing families of sidebands. These appear as spikes in the frequency spectrum, equally spaced on each side of the tooth meshing frequency by an amount equal to the gear rotation frequency and its multiples. Cepstrum analysis can determine the sideband spacing in such situations.

Orbit analysis

To perform an orbital analysis, two proximity transducers are placed around the shaft 90 degrees apart and the output displays the motion of the centre of a rotating shaft in the two mutually perpendicular directions, as shown in Figure 3.3. Orbits are useful in the analysis of the shaft vibration in order to reveal machine faults (McCormick, 1997). The shaft orbit can provide basic amplitude and phase lag information indicating wear in a bearing, such as shaft imbalance, misalignment and oil whirl. However, since the orbits are constructed in the time domain, they are affected by noise and surface roughness, often leading to incorrect monitoring.

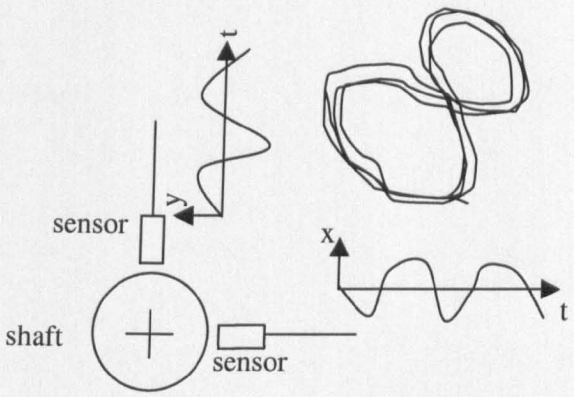


Figure 3.3 Arrangement of a typical shaft orbital motion

3.2 Introduction to acoustic emission

The technology of acoustic emission (AE) has its beginning in 1950 marked by the work of Joseph Kaiser. During the decades of the 1950s and 1960s, research into the fundamentals of acoustic emission created a huge growth of interest resulting in the rapid development of instrumentation for characterisation of AE behaviour in many materials (Swindlehurst, 1973). In the 1970s, acoustic emission was more widely recognised for its unique capabilities as a non-destructive test (NDT) method for monitoring dynamic processes; its use has been steadily growing in various industrial sectors. In the 1980s computers provided a fundamental and powerful component for both instrumentation and data analysis (Thomas, 1996).

The term *acoustic emission* can be used to refer to both a technique and a phenomenon upon which the technique is based. Acoustic emission transducers allow the more sensitive detection of sounds of higher frequencies and lower intensities such as produced from crack propagation, dislocation motion, fibre debonding in composites and some phase changes (Scruby, 1987). Acoustic emission is a natural phenomenon of sound generation applied to the spontaneously generated elastic wave produced within a material under stress. In other words, if there is a sudden release of strain energy within a solid, caused, for example, by the growth of a crack or plastic deformation, then some of the energy is dissipated in the form of elastic waves. Typically, the proportion of energy depends on the nature of the source such as how localised it is and how rapidly the release takes place.

To be able to use AE as an effective condition monitoring technique, an understanding of the AE source is necessary. The different sources of acoustic emission will be described in the following section.

3.2.1 Sources of acoustic emission

Acoustic emission occurs in the form of transient bursts that correspond to a localised micro failure within the body of the material. They are inherently broadband and contain frequencies from DC to the MHz region. In addition, source mechanism relating to the crack extension and plastic deformation, active in defect growth such as fatigues, is of high-frequency stress waves source (Holroyd, 1993). Typically the frequencies lie within the range of 20 kHz to 1 MHz.

In the case of a growing crack, arguably the most important AE source, the formation of new crack faces must be accompanied by sudden changes in stress and displacement of material in the vicinity of the crack (Scruby, 1987). Varying stresses and strains must, by definition, act as sources of stress (elastic) waves as shown in Figure 3.4.

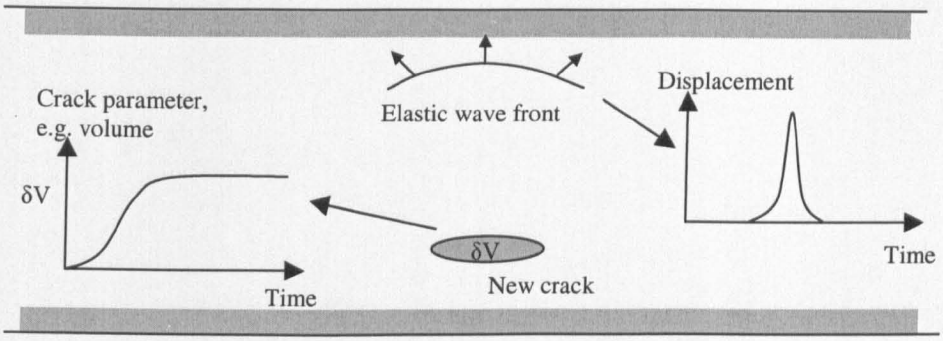


Figure 3.4 Example of an AE source as crack growth

Generally, at a microscopic level stress wave activity originates as individual randomly occurring transients. These bursts of AE events may occur individually in clusters or quasi continuously depending on the type of machine, its running speed and severity of machine degradation. The stress wave corresponding to each transient burst propagates readily through the metallic structure of most machines creating reflections at the boundaries defined by the geometry of the machine structure.

Ideally a rotating machine operating perfectly would produce no mechanically generated high frequency stress waves. Since no machine is perfect and there are potentially other sources of activity, this does mean that machinery in very good condition can still generate a continuous type of AE signal as in the form of overlapping multiple low level bursts. This has the appearance of a signal having a low mean level with little or no distinct transient superimposed on it.

In a bearing, when a rolling element in a bearing encounters a race defect, normally a small pit or crack, there will be an impulsive change of load on the system locally and a transient burst will propagate, appearing as discrete random transients superimposed on the signal. Acoustic emission signals propagate in longitudinal and shear modes inside the bearing, and in the Rayleigh mode on the surface of the raceways (Bashir, 1999). After its propagation from the point of the growing crack, many reflections will ensue at the surfaces of the bearing. The signal will probably be highly complex by the time it reaches the mounting point of the transducer. In the foregoing, it is

assumed that the wave initiated from the crack growth is a bulk wave and then converts to a Rayleigh wave, which travels onwards to the measurement point.

3.2.2 Acoustic emission signal transduction

In order to capture AE events, a transducer or sensor is required to convert very small surface displacement (on the order of picometres) into electrical signals that can be amplified and recorded. In principle, an acoustic emission transducer is used to transform the mechanical energy in an elastic wave into an electrical voltage signal. The operations of most AE transducers depend on the piezoelectric effect. Piezoelectric effect is preferred because of its high sensitivity and relative immunity to mechanical noise; and the robustness of the sensing element allows its use in many industrial applications.

Most AE transducers (Swindlehurst, 1973) have a sensing element in the form of a thin disc of piezoelectric material, normally lead zirconate titanate (PZT) that can convert mechanical deformation into corresponding electrical voltage as shown in Figure 3.5. This disc is metalised on both sides for electrical contact, and mounted in a metal cylinder to provide shielding from electromagnetic interference. When the disc is unbacked and undamped or lightly damped it behaves as a resonator for incident elastic waves so that the output signal is usually a decaying sinusoid. The chosen frequency is dictated by the thickness of the element. Backing the element with an attenuating medium such as epoxy, which is normally loaded with tungsten to match its impedance, creates a heavily damped transducer with a more broadband response (McFadden, 1984).

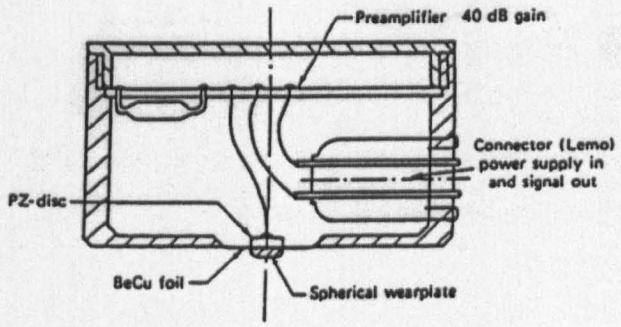


Figure 3.5 Example of a piezoelectric transducer (Swindlehurst, 1973)

Typically the useful frequency responses of AE transducers lie in the range of 100 kHz to 1 MHz. The frequencies are well matched to specimens with thickness lying between a few millimetres and a fraction of a metre, since the thickness controls the specimen resonance and the dominant Lamb wave frequencies. Therefore, in order to select a suitable transducer for an application, it is obviously advantageous to optimise sensitivity by matching frequencies. Broadband AE transducers normally comprise a piezoelectric element whose natural frequency is above the pass band of interest. Hence, to monitor in a range from 100 kHz to 1 MHz, a natural frequency between 1 and 2 MHz might be selected. Damping such a resonator can give a reasonably flat response below the natural frequency (Scruby, 1987).

To attain high sensitivity, the piezoelectric transducer must be mounted to the specimen under observation in such a manner that the acoustic emission energy can pass into the transducer with minimum attenuation at the contact surfaces between the transducer and specimen. The required intimate mechanical contact can be obtained on the flat surface by means of a couplant at the interface. The couplant is any material, which helps the propagation of acoustic waves across the interface; in contrast, a bond is a couplant which physically holds the transducer to the surface of the specimen. Examples of a bond include a thin film of grease, oil, epoxy adhesive and non-attenuating fluid.

One drawback with the use of piezoelectric disc element is that it produces a broadband response only for waves incident perpendicular to its surfaces. When destructive interference occurs, the output is distorted. The bandwidth is also reduced at other orientations. All these can lead to a degradation of performance of the transducer. However, these effects are insignificant at lower frequencies unless the transducer is very large.

In practical measurement, a resonant transducer is used so that the sensitivity can be improved. There is only a significant gain if the bandwidth of the signal is much less than the bandwidth of the noise. When the bandwidth of the AE signal is wide, then the gain is insignificant because both the background noise and the signal are broadband. Nevertheless, broadband transducers are chosen when waveform analysis is desired and prior knowledge of the frequency of interest is unknown.

A preamplifier is the first stage in the amplification of the acoustic signals. The preamplifier should be low noise, and matched to the transducer response as nearly as possible. By far the most common type of filter used in practice is band-pass filtering because reducing bandwidth can improve signal-to-noise ratio. Typically, high frequency rejection reduces electronic noise. Likewise low frequency rejection assures that mechanical vibration, which can be an order of magnitude higher, does not interfere with the results.

3.2.3 Acoustic emission signal processing

There is a need to interpret the AE signals so that it can be related to the actual condition of machine components. Conventional AE parameters (Pollock, 1973), as shown in Figure 3.6, can be extracted from the acoustic signal and they are the ringdown count, rise time, time of arrival, event duration, energy and peak amplitude.

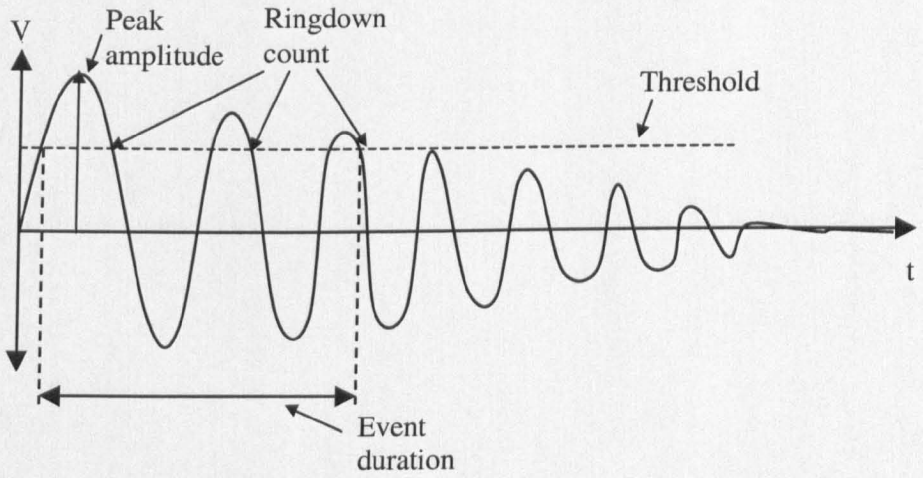


Figure 3.6 Examples of traditional AE parameters

To characterise these parameters and also to eliminate 'noise', a threshold is selected so that events that rise above the threshold are counted. Evidently, the threshold level influences the values of some of these parameters.

Ringdown count

Ringdown count is one of the earliest parameters employed using a discriminator. This can be obtained by passing an acoustic emission signal through the discriminator that triggers a count whenever the voltage of the signal rises above the predetermined threshold value. If the intensity of the signal is high, then the counts will be large.

Time of arrival

The time of arrival of an acoustic emission event is the time instant at which the event first crosses the predetermined threshold.

Peak amplitude

Peak amplitude of a signal may be taken as a measure of the signal's intensity. The peak amplitude of an AE event is the maximum excursion of the corresponding voltage signal from the zero level. The peak amplitude is independent of the threshold setting. This parameter is often used to assess the severity of the AE source.

Event duration

Event duration is dependent on the amplitude threshold. It is the time interval between the first and last threshold crossing for an individual event. The signal duration is equal to the period of each cycle times the number of counts.

Energy

The energy of an AE event is the energy contained in the corresponding voltage signal and, strictly speaking, is not the true energy of the event itself. Energy analysis procedures involve squaring and integrating the time signal, the signal energy being proportional to the area under the ensuing curve. However, the energy, which is converted from the sensor, can be related to a sudden release of strain energy from the emission source such as crack growth, impact and friction.

3.3 Previous work on bearing condition monitoring

Research has previously been carried out into the use of vibration and acoustic emission for bearing condition monitoring. And there are a number of monitoring methods reported. It is thought beneficial to provide a review of these methods by way of comparison in terms of performance and application. Based on the representation

of a signal, its analysis can be carried out in one of three domains: time domain, frequency domain and joint time- and frequency- domain.

3.3.1 Time domain analysis

Early work in this area by Butler (1973) investigated the detection of damaged rolling bearings using the shock pulse method to vibration signals. He assumed that when a rolling element comes into contact with a damaged area of the raceway or with debris in bearing, mechanical shocks are then generated and transmitted through the bearing structure. He concluded from the experiments that a shock pulse meter provides quantitative and reliable results. This method is efficient but has the disadvantage that when there is more than one fault in a machine, the impulse repetition will not correspond to one single fault, leading to incorrect diagnosis.

Yoshioka (1982) described an acoustic emission system for source location. The experiment was conducted using an acoustic emission sensor to detect AE signals and a magnetic detector as a source locator. The magnetic detector counts the number of passing gear teeth in the outer race of the retainer to determine the position of the ball in the raceway. The results showed the coincidence between the positions of detected AE counts and positions of defects in the raceway that were measured by a tool-maker's microscope.

Comparative studies of various vibration and acoustic emission measurement methods (Rogers, 1979; Tandon et al., 1992; Holroyd, 1993; Tandon et al., 1999; Choudhury et al, 2000) were given. Measurements were conducted on good new bearings and bearings with simulated defects in their elements. Vibration measurements included the overall vibration acceleration, envelope-detected acceleration, kurtosis, crest factor and shock pulse; acoustic emission measurements included ringdown counts, event duration and peak amplitude. The results indicated that, in general, the detection of defects at lower speeds is highest with acoustic emission and lowest with the vibration shock pulse method.

Mellor (1988) and Heng et al. (1998) studied rolling element bearing condition monitoring using vibration and sound. The crest factor and the distribution of moments including kurtosis and skewness were used. In the absence of bearing

damage, the high frequency vibration generated from a rolling element bearing is random in nature with a corresponding Gaussian or normal amplitude distribution. In the presence of damage, the high amplitude impulses generated by the impacts distort the shape of the amplitude distribution from being Gaussian.

Bashir et al. (1999) and Zhang (2000) investigated the release of acoustic emission energy during fatigue crack growth. They presented mathematical models to describe the acoustic energy associated with a fine crack growth in a bearing in an attempt to provide a better description of the failure mechanism. The models predicted that the amplitude (in volt) of impulse function is proportional to the corresponding height (in metres) of the bearing surface displacement.

Holroyd (1993), Yoon et al. (1995), James (1995) and Miettinen et al. (2000) carried out experiments using acoustic emission for bearing monitoring. They found that AE is an effective condition monitoring method for early fault detection in bearings. The results showed that the appearance of distinct random transients indicated the degradation of bearing condition. As degradation intensified the number of bursts and the mean signal level also rose. The dramatic reduction of AE events was achieved when pump bearings were re-greased.

3.3.2 Frequency domain analysis

McFadden (1984) explored the use of acoustic emission on monitoring bearings at low speeds. Experiments were carried out on a test bearing operating at low speed (about 10 rev/min). He found that AE is sensitive to loading condition: a change in load at constant speed results in an increase output of AE signals. He also concluded that the non-uniform frequency response characteristics of a typical AE transducer makes it less suitable for general purpose monitoring of bearing at high speeds unless frequencies are matched.

Experimental studies for bearing monitoring based on a frequency spectrum (Jun, 1995; Mechefske, 1995; Liu, 1996; Pineyro et al, 2000) were also conducted. The results showed that employing fast Fourier transform (FFT) has the potential of revealing information on the sources of failures. The peaks in frequencies define the type of fault and the amplitude of the peaks indicates the severity of the fault.

Although the frequency spectrum appears to be the best indicators of machine conditions, the defect frequency may be close to frequencies excited by other machine components leading to incorrect interpretation.

Qu et al. (1989) proposed the holospectrum technique for rotor surveillance and diagnosis. The holospectrum method employed FFT spectra of rotor vibration both in horizontal and vertical directions. It can provide information not only about peak frequency and magnitude, but also the phase relationships. Figure 3.7 illustrates some examples of holospectrum. It is composed of circles, ellipses and lines. A pure circle is obtained if the amplitudes between the vertical and horizontal components are equal and their phases are 90 or 270 degrees. An ellipse is obtained if there is imbalance in

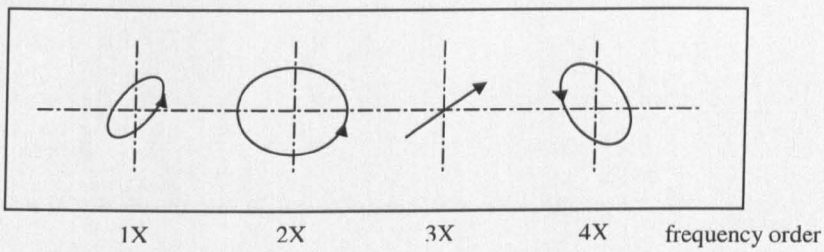


Figure 3.7 Examples of the two-dimensional holospectrum

the shaft. A line is obtained if the phase lag between two elements of machine is 0 or 180 degrees and the slope of the line depends on their amplitude proportion. Results showed that this method is very sensitive to rotor imbalance or temporary bending of shafts.

Dalpiaz et al. (2000) used cepstrum for fault detection in gear. Since gear vibration spectra commonly show sidebands of the meshing frequency, which typically arise from the modulation of the tooth meshing frequency by gear rotating frequency, it is difficult to quantify the degree of modulation by means of spectral analysis. They used cepstrum analysis to identify the periodic structures in spectrum, thus giving a more accurate interpretation of the sideband level over the whole spectrum.

Envelope detection of acoustic emission and vibration signals is also an important technique for condition monitoring of rotating machinery. Mellor (1988), Sato et al.

(1991) and Geropp (1997) used the signal envelope detection waveforms and frequency spectra to identify bearing faults caused by periodic shock impulses that can be considered as amplitude modulated signals. Figure 3.8 illustrates the envelope detection procedure. The input signals were band-pass filtered and rectified to reduce mechanical noise. To extract the carried frequency, the signal was demodulated and then passed through a low-pass filter. What remains is the frequency component of the signal caused by the repetitive defect frequency.

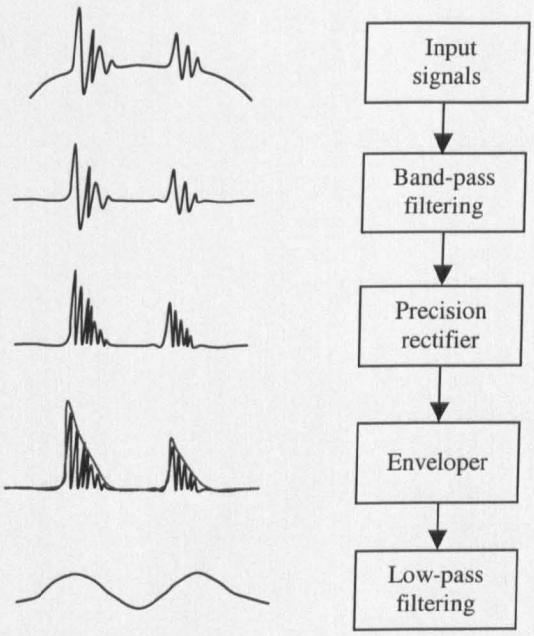


Figure 3.8 Diagram of signal enveloping process

Shiroishi et al. (1997) and Li et al. (1999) investigated a defect detection method for rolling element bearings. They used a signal processing scheme which was a combination of high frequency resonance technique (HFRT) and adaptive line enhancer (ALE) techniques as shown in Figure 3.9. The HFRT, similar to envelope detection, comprised band-pass filtering to allow only the resonance frequencies to pass through and demodulation to provide a demodulated signal in the absence of low frequency mechanical noise. The ALE has the effect of increasing the detectability of a periodic defect signal by reducing wideband noise of the obtained demodulated signal and enhancing the envelope spectrum of the defect signal with clear peaks at the harmonics of the characteristic defect frequency.

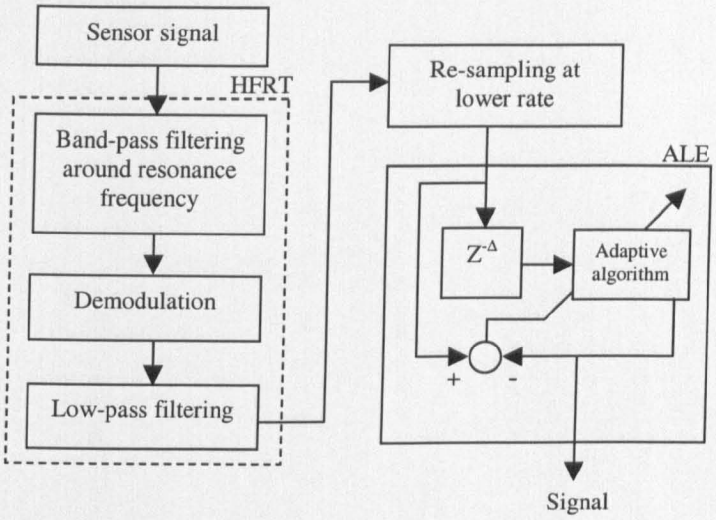


Figure 3.9 HFRT and ALE block diagram

3.3.3 Joint time- and frequency- domain analysis

Dalianis et al. (1997), Lee et al. (1997), Stazewski et al. (1997b), and Han (1999) used the Wigner-Wille distribution for bearing condition monitoring. The Wigner-Wille distribution approach is one of the time-frequency methods. They claimed that this method could reveal the type of defect more accurately than the original FFT spectra in a gearbox such as tooth-cracking or pitting. Wang (1993) and Oehlmann et al. (1997) demonstrated the use of Wigner-Wille distribution that produces a complicated time-frequency representation of signals. They also proved that application of the Spectrogram (STFT) for the early detection of damage in gears has some advantages over the Wigner distribution because of its simplicity.

The short time Fourier transform (STFT), the simplest method of time-frequency analysis, has also been used for rotating machines faults detection (Staszewski et al., 1997a; Conforto et al., 1999; Safizadeh et al., 2000). STFT can give a clear representation of a signal in the time-frequency plane and a simple interpretation of the energy variation due to damage. However, it has the drawback that high resolution of frequency and time cannot be obtained simultaneously.

The discrete wavelet transform (DWT) was used to predict occurrences of defects in ball bearings (Mori et al., 1996; Li et al., 1996; Paya 1997; McFadden et al., 1999; Shibata et al., 2000). The DWT, a time- and frequency- domain method, is particularly potent for detecting subtle time localised changes. Vibration signals from a dented ball bearing were experimentally investigated. Impulsive responses occur when a ball rolls over the pre-spalling part of the raceway. The results showed that the DWT coefficients are a good indication of bearing failure because an increase in the impulse causes a corresponding increase in the wavelet coefficients.

Liu (1999) and Lin (2001) suggested a de-noising method for machine engine fault diagnosis. The de-noising technique uses Morlet wavelet transform to remove the background noise. The FFT method was then applied to the de-noised signal. The FFT spectra of different machine conditions can be distinguished. The results showed that this method could help the operator to diagnose the machine condition correctly, even if the signal was masked by noise.

Condition monitoring methods using a continuous wavelet transform (CWT) on rotating machines were proposed (Wang, 1993; Dalpiaz et al.,1997; Ypma et al., 1997; Boulahbal et al., 1999; Wang 2000). In the case of a reciprocating machine, the wavelet analysis can reveal how the vibration frequency content varies within the machine cycle, thus enabling the transient dynamic phenomena to be clearly detected and precisely located. In the gearbox application, the wavelet maps were found to display distinctive features in the presence of a cracked tooth.

4 Theoretical background of the systems approach

In this chapter, a case will be made for the adoption of the systems approach for bearing condition monitoring. Such an approach is based on the concept of multisensor data fusion (Joshi, 1999). The main considerations are that with this approach, the ensuing monitoring system must perform better than other monitoring systems in terms of its ability to give an early indication of bearing incipient failure and of its reliability and robustness. To achieve this, an understanding of the functions of sensory data processing, data fusion and interpretation are of importance. This is because these functions yield the intelligence in the systems by 'fusing' or integrating multiple sources of information in the condition monitoring process (Esteban et al., 1999).

4.1 Introduction to multisensor integration

Multisensor integration usually refers to the use of multiple sensors in the system. Successful integration requires knowledge of the task to be accomplished and the effective representation of that knowledge. The key challenges in multisensor integration are multisensor data fusion, multisensor planning and multisensor systems architecture (Joshi, 1999).

Generally speaking, multisensor data fusion is the combination of data from multiple transducers from measurements into a coherent internal representation. The reasons that multisensor data fusion is employed in the proposed bearing condition monitoring are as follows (Hu et al., 1999a; Esteban et al., 1999).

Accuracy: The results obtained based on a set of sensors are more likely to be accurate than those based on a single sensor. Moreover, the use of multisensor can reduce the uncertainty of measurements.

Reliability: Employing multiple sensors increases fault tolerance and improves systems reliability. Even if a single sensor fails to operate, another sensor can still provide useful information. In contrast, monitoring using a single sensor cannot provide any valid data when a fault occurs, if that sensor has failed.

Robustness: Multiple sensors based on different physical principles can be utilised to improve the effective range of situations in which the system operates, hence enabling the system to handle a larger dynamic range of machine operating conditions. In addition, using complementary sensors to observe different aspects or characteristics can give better inferences of the operating status, which are difficult to make based on an individual sensor.

Timeliness: Multisensor signals can be processed in parallel thus giving faster results than monitoring via processing signals in sequence from a set of separate sensors.

Multisensor planning usually addresses the acquisition of transducer data by deciding what data to acquire and how to acquire it. This also involves the choice of transducers that are suitable for the measurement, their mounting positions at which a high signal-to-noise ratio can be obtained, their preamplifiers and filters that match the instruments and operating frequency ranges.

A sensible multisensor architecture will enable complementary sensors with non-overlapping ranges to operate together as a single sensor with greater capability and robustness. This is because transducers may fail or operate not at its best if the operating range is violated. A flexible multisensor architecture should be able to accommodate the changes in the dynamic range of machine operating conditions (Hu et al., 1999b).

In the context of multisensor fusion, this project used both vibration and acoustic emission techniques for bearing condition monitoring and diagnosis. The vibration technique has been established to have the capability of detecting the different causes of bearing failure based on monitoring the bearing's characteristic defect frequencies. However, its range of usefulness is limited to about 20 Hz up to 20 kHz. If a bearing runs at speeds below the frequency range, then the vibration technique becomes powerless. On the other hand, acoustic emission may provide a satisfactory monitoring solution for low speed rotation. This is because AE is sensitive to defect initiation and growth and is minimally affected by the dynamics of the machine structure.

Clearly using AE and vibration transducers can cover a much larger speed range of rotating machines. Such fusion can increase the reliability and hence robustness of the monitoring system as it is more tolerant to faults in sensors. In addition, the results obtained are likely to be more accurate as measurement uncertainty is reduced.

In the proposed systems approach, the vibration technique refers to the joint time- and frequency- domain analysis that will reveal the characteristic signatures of different bearing operating condition. The acoustic emission technique employs the detected AE parameters in the traditional Euclidean distance classification, and the raw AE time signals will be characterised using a novel technique producing a parameter known as the shape-to-life (STL) value. The schematic diagram of the proposed approach is shown in Figure 4.1. Their theoretical backgrounds and details of the techniques are given in the following sections.

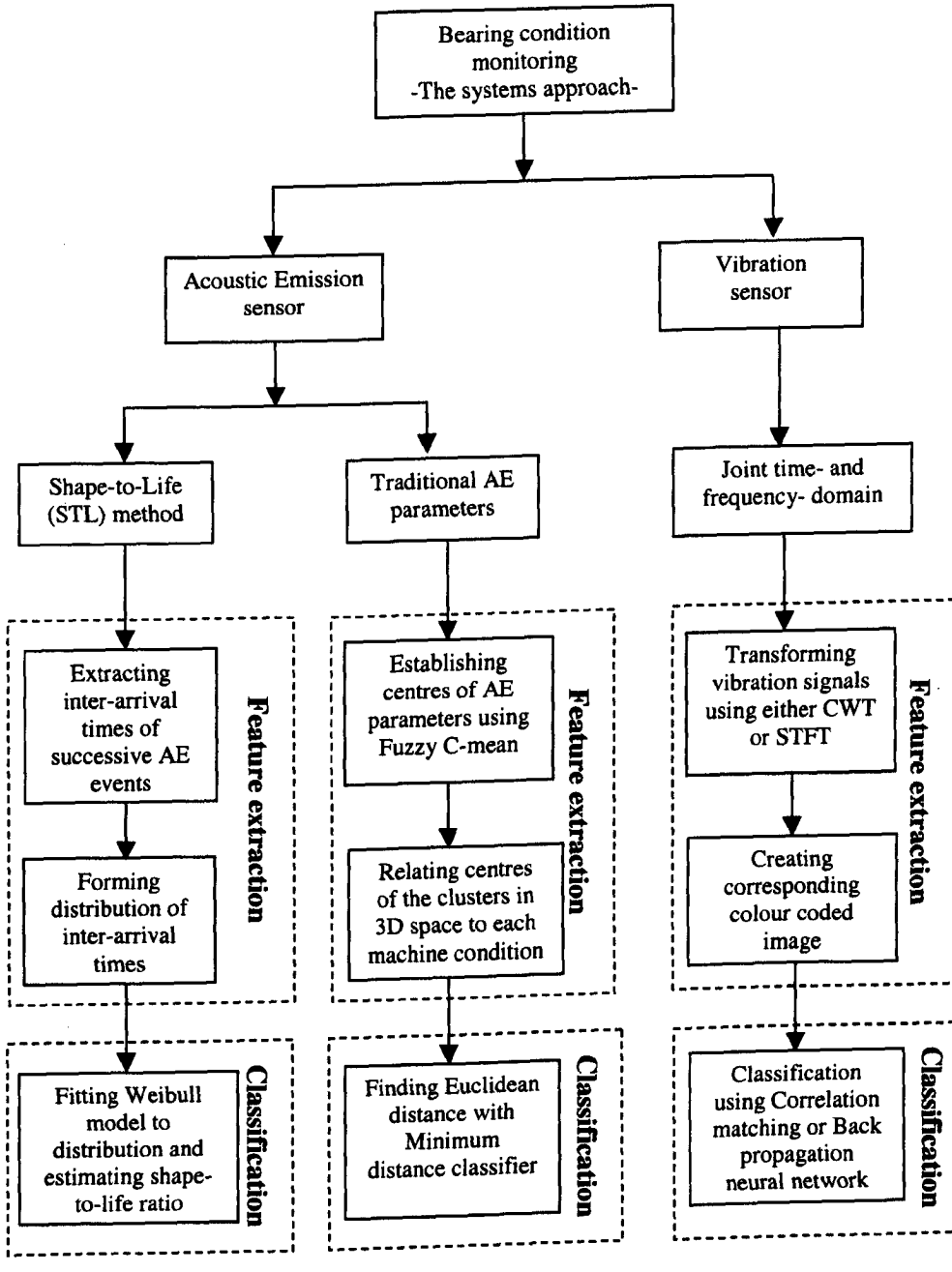


Figure 4.1 The schematic diagram of the proposed systems approach for bearing condition monitoring and diagnosis

4.2 Vibration analysis

4.2.1 Basic concept

The original idea of the proposed method was developed from the assumption that the conventional methods using either time or frequency domain technique are incapable of revealing the transient events which occur in the real environment. In bearing application, if there is a defect, it will produce such transient events. Typical features of the time signal are the RMS value, peak value, crest factor, kurtosis and the shape, size and orientation of a bearing locus derived from orbital analysis. These features, once established to be related to the defect being monitored, often yield satisfactory results. However, if the signal generation mechanism is complex, time-domain methods are often not refined enough.

Frequency domain methods work best when the signals being monitored are periodic in nature, as is the case for vibration on a rotating shaft supported in bearings. Since a bearing defect produces a signal that is periodic at the characteristic defect frequency, frequency-domain analysis appears to be a suitable tool for condition monitoring. Examples of the frequency-domain techniques include spectrum analysis, cepstrum analysis, high frequency resonance technique (HFRT) and holospectrum. Among them, spectrum analysis seems to be the most common and dominates the fault diagnosis scene.

The main limitation of spectrum analysis is that although a local transient will contribute towards the overall frequency spectrum, its location on the time axis is lost. There is no way of knowing whether a particular frequency component exists throughout the life of the time signal or during just one or a few selected periods. Unfortunately, many monitoring situations demand knowledge of not just the frequency composition of a transient but also its time of occurrence. For instance, when a rolling element passes a localized defect in a bearing, it generates a transient in the measured signal, as does the contact of a damaged tooth with other teeth in a gearbox. A machine with rapidly varying speed is another example of transient events. For these reasons, techniques that retain both time and frequency perspectives are preferable.

When a rolling element runs over a localised defect, it produces a transient shock pulse, repeating itself periodically in the vibration signal because of the continuous rotation. To extract those transient events from the signal, the Short Time Fourier Transform (STFT) and Continuous Wavelet Transform (CWT) techniques are considered. These two techniques are used to produce a three dimensional surface plot with the time, frequency (or scale for CWT) and magnitude as the x, y and z axes, if the magnitude on the z axis is converted into a corresponding colour, the 3-D plot then becomes a 2-D colour-coded image, and its pattern represents a signature for a particular bearing operating condition. Correlation matching is then used as a tool to classify the bearing condition. Figure 4.2 shows the processing sequence in the proposed bearing condition monitoring scheme using vibration.

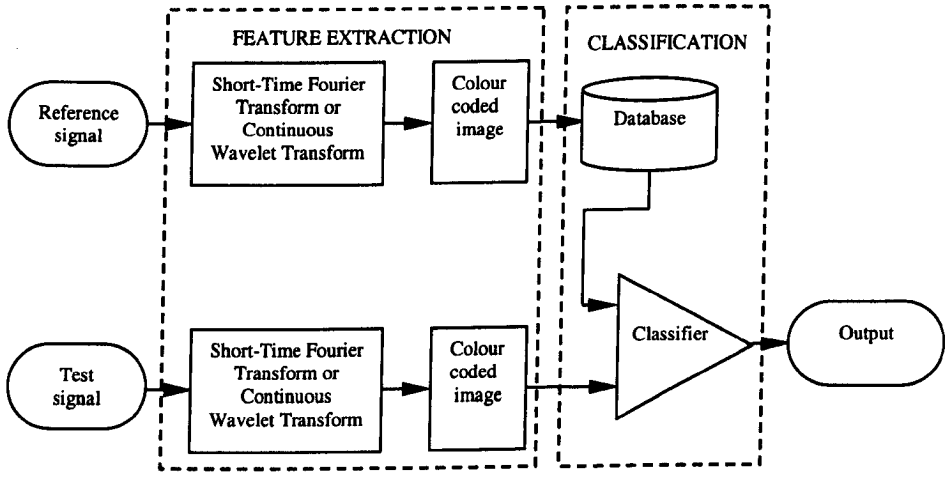


Figure 4.2 The block diagram using vibration in the proposed bearing condition monitoring

4.2.2 Short-Time Fourier transform

The Fourier transform of a signal x(t) is defined as (Randall, 1977)

$$F[x(t)] = X(f) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}\,dt \qquad \textbf{Equation 4.1}$$

For Equation 4.1 to be valid, the signal x(t) being transformed must be stationary, which means that its amplitude distribution does not depend on absolute time. In other words, the moments of the distribution –mean, variance, and so on - are stationary. A signal with localized events is clearly not stationary.

To overcome this difficulty, the signal is divided into segments such that within each segment the stationary property is approximately true and so Equation 4.1 can be applied. The Short-Time Fourier Transform (STFT), also known as the Windowed Fourier transform (WFT), does exactly that. STFT uses a constant-width time window, and the segment of the signal exposed is transformed into the frequency domain (Safizadeh et al, 2000). As the window slides to a new position along the time axis, the Fourier transform is again computed. This is repeated until the whole duration of the signal is covered. Mathematically, STFT can be expressed as a function of the frequency ω and the position b along the time axis; thus,

$$F(\omega,b)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} f(\tau)g(\tau-b)e^{-j\omega\tau}d\tau \qquad \textbf{Equation 4.2}$$

This is the Fourier transform of the function f(t) windowed by the function g(t) for all b. A disadvantage of this method is that the time and frequency resolutions are constant as determined by the constant size of the window used and the fixed number of points of the signal exposed by window. Consequently, when the signal has a wide frequency bandwidth, STFT cannot give the high resolutions for frequency and time simultaneously.

The function $F(\omega, b)$ in Equation 4.2 is depicted as a two-dimensional colour-coded map where the x-axis denotes b, which, for a time signal, is time itself, and the y-axis denotes ω, the frequency. The colour at a point (ω, b) represents $F(\omega, b)$.

4.2.3 Continuous wavelet transform

The Continuous wavelet transform (CWT) converts a time signal $f(t)$ into the joint time- and scale- domain (Rioul, 1991). Unlike the discrete wavelet transform (DWT), the CWT can operate at every scale, from that of the original signal up to some chosen maximum scale. In addition, the CWT is also continuous in terms of translation of the analysing wavelet in that it is shifted smoothly over the whole time axis of the analysed signal. CWT, therefore, converts a one-dimensional time signal into a two-dimensional pictorial image with time and scale as its orthogonal axes. In this respect, the end product of CWT is comparable to that of STFT where the axes involved are

related to time and frequency. Frequency and scale are the two parameters that are inversely related: the greater the frequency, the smaller is the scale.

The basis function used is a mother wavelet, $\psi(t)$, analogous to the sine or cosine function used in Fourier Transform. CWT is defined by (Gaul et al, 1997; Suzuki et al, 1996)

$$F_{\Psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\Psi\left(\frac{t-b}{a}\right)dt \qquad \textbf{Equation 4.3}$$

The quantity in Equation 4.3

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \qquad \textbf{Equation 4.4}$$

is the wavelet function. In the present study, a commonly used orthogonal set of wavelets created by Ingrid Daubechies, db4 (Daubechies, 1990) was chosen and a typical db4 wavelet is shown in Figure 4.3. While the choice is somewhat arbitrary, it is noted that the db4 wavelet bears some resemblance to acceleration signals observed from vibrating mechanical structures.

In the wavelet function of Equation 4.4, the position variable b is the translation parameter, same as b in the STFT definition, Equation 4.2. Whereas STFT uses a frequency variable ω, CWT involves a scale variable a. Scaling is a primary characteristic of wavelet analysis. The mother wavelet function $\psi(t)$ in Equation 4.3, besides being a basis function, also plays the role of a window function, equivalent to $g(t)$ in Equation 4.2. The scale variable a is loosely the reciprocal of the frequency variable ω; changing a controls the frequency band. Variable b controls the size of the window. It is therefore possible to adjust the resolutions of time and scale (frequency) independently by changing b and a.

The function $F_{\psi}(a, b)$ in Equation 4.3 is depicted as a two-dimensional colour-coded map where the x-axis denotes b, which, for a time signal, is time itself, and the y-axis denotes a, the scale. The colour at a point (a, b) represents $F_{\psi}(a, b)$.

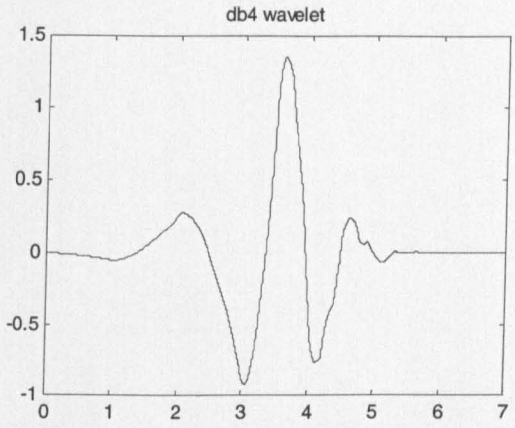


Figure 4.3 The db4 wavelet function

4.2.4 Image processing

Image processing is performed using the method of thresholding or binarising (Gonzales, 1993). It is applied to the grey scale CWT image to convert the colour of each pixel into either black or white. A binary image has the obvious advantage that when classification using either correlation matching or neural network is done, the computational time will be much shortened, as multiplication involving a 0 or 1 is much easier to perform.

4.2.5 Correlation matching

Correlation matching is used for pattern recognition in the proposed method. The two-dimensional transformed images obtained from either STFT or CWT are used to compute the similarity coefficient between the reference and test images. The method works as follows.

Given an image f(x,y) of size M by N pixels and a reference image w(x,y) of size JxK pixels, one starts by placing the geometric centre of the reference image w(x,y) on the pixel on the top-left corner of the image f(x,y) and then compute the correlation coefficient based on the pixels that overlap between the two images. One then shifts the geometric centre of w(x,y) one pixel to the right along the first row of f(x,y) at which another correlation coefficient is calculated. Each pixel in the first row is visited in this fashion; and when the first row is completed, other rows are visited in turn until the whole image f(x,y) is covered. The result is an MxN correlation

coefficient matrix. The correlation coefficient in row s and column t of the matrix is given by (Gonzales, 1993).

$$\gamma(s,t) = \frac{\sum_x \sum_y \left[f(x,y) - \bar{f}(x,y)\right]\left[w(x-s,y-t) - \bar{w}\right]}{\left\{\sum_x \sum_y \left[f(x,y) - \bar{f}(x,y)\right]^2 \sum_x \sum_y \left[w(x-s,y-t) - \bar{w}\right]^2\right\}^{1/2}}$$

Equation 4.5

where s = 0,1,2,...,M-1, t = 0,1,2,..,N-1, $\bar{w}$ = the average value of pixels in w(x,y) (computed only once), and $\bar{f}(x,y)$ = the average value of f(x,y) in the region coincident with the w(x,y). Furthermore, it should be noted that the summation is taken over the image coordinates common to both f and w. Figure 4.4 illustrates the procedure. The correlation coefficient γ(s,t) is in the range of –1 to 1.

For perfect match correlation,

$$f(x,y) = w(x,y), \Rightarrow \gamma(s,t) = 1$$

, perfect mismatch correlation

$$f(x,y) \neq w(x,y), \Rightarrow \gamma(s,t) = -1$$

and for less than perfect correlation,

$$1 > \gamma(s,t) > -1$$

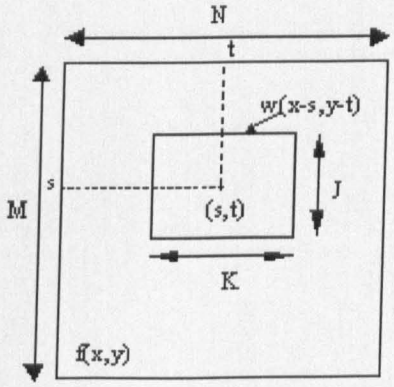


Figure 4.4 Correlation matching of f(x,y) and w(x,y) at point (s,t)

4.2.6 Back-propagation neural network

The back-propagation neural network (BPNN) (Zhang et al, 1996; Paya et al, 1997; Subrahmanyam, 1997; Vyas, 2001) is used as a tool for classification of machine operating conditions. The BPNN architecture is shown in Figure 4.5. The network has an input layer, a hidden layer and an output layer. The extracted feature values of the signals provide the inputs to the neural network; and the output nodes represent the machine operating conditions.

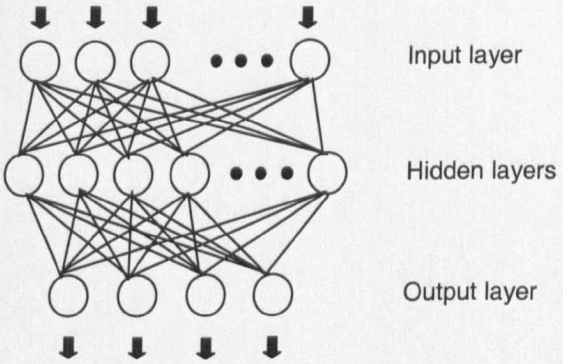


Figure 4.5 Back-propagation neural network architecture

Through a process of trial and error based on minimizing the mean square error (MSE), the choice of number of nodes and hidden layers are arbitrary and depend on the user. The value *NET* appearing at a node in a layer is computed by summing the products of all inputs leading to that node with their corresponding weights plus a bias. This value then forms the argument of a transfer function f that produces an output for the node. In vector notation, given the input vector X_i, the weight vector W_i, and the bias vector θ_i, the output is given by

$$OUT = f(NET) = f(\sum_{i}^{n} X_i W_i + \theta_i) \qquad \textbf{Equation 4.6}$$

The transfer function f is the commonly used sigmoid function defined as

$$f = \frac{1}{1+e^{(-NET)}} \qquad \textbf{Equation 4.7}$$

The sigmoid function acts as an output gate that can be open (0) or closed (1). The computing process as described is schematically shown in Figure 4.6.

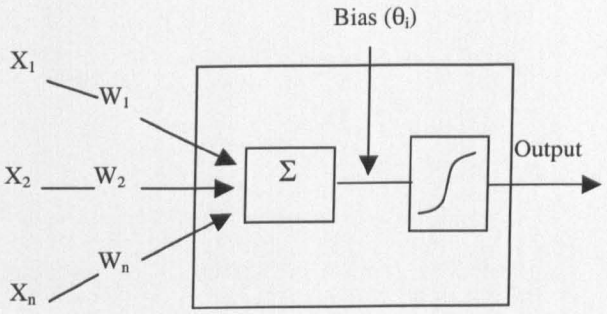


Figure 4.6 Schematic process of BPNN

The back-propagation algorithm is used to obtain the correct weights and biases in a training process. A set of training data with known outputs is fed into the network. The weights are initially set to random values; the biases are fixed at unity. The input data are presented to the network; outputs are calculated and compared with the desired outputs. The normalized mean square error (MSE) is then calculated and propagated back to adjust the weights on the neural connection. This process is repeated for a large number of epochs until the error is relatively low and acceptable, when the network is considered to have classified the test set correctly.

4.3 Acoustic emission analysis

4.3.1 Basic concept

Acoustic emission (AE) is a non-destructive testing technique, which is used for detecting plastic deformation, defect growth in metals and composites (Holroyd, 1993). It is based upon the detection of high frequency stress wave (structure borne sounds) that are naturally generated by the failure of materials at microscopic level. In addition to AE sources of crack extension and plastic deformation that are active in defect growth, AE can also be generated from other source mechanisms such as impact, friction, rubbing and cavitation. For bearings, sources of acoustic emission might stem from mechanical looseness (impacts), inadequate lubrication (friction and impacts), overloading (friction and impacts), fatigue failure (crack growth and friction) and wear of surfaces (friction, rubbing and impacts).

Previous research (McFadden et al., 1984; Tandon et al., 1992; Yoon et al., 1995; James et al., 1995) has demonstrated that acoustic emission is an effective tool for early detection of damage due to metallic contact. Consequently AE monitoring is superior to vibration monitoring in that the former can detect subsurface crack growth whereas the latter can at best detect only after a defect has emerged on the surface of a structure.

In this research, acoustic emission is used in two different ways: one traditional and established whilst the other novel. In the first way, traditional AE parameters such as peak amplitude, event duration and AE energy are classified using the technique called fuzzy c-mean clustering. This technique extracts the Euclidean centre from each cluster (corresponding to a particular machine operating condition) and uses the minimum distance classifier to decide to which cluster, and hence the operating condition, the AE event should belong. The novel way, developed in this study, involves the creation of a distribution of inter-arrival times of AE events and calculation of the shape-to-life (STL) value from the distribution. This STL value is used as the signature of the machine operating condition.

4.4 Bearing condition monitoring with traditional AE parameters

An AE event is characterised using parameters such as ring-down count, rise time, event duration, energy and peak amplitude. A threshold is used in order to eliminate 'noise' and only events that rise above the threshold are counted. Evidently, the threshold level affects the value of some of these parameters. A typical example is the event duration. By definition, it is the time that the envelope of an AE event is above the threshold. When the threshold level is high, the event duration will be shorter.

As was discussed in section 3.2.3, the peak amplitude of an AE event is the maximum excursion of the corresponding voltage signal from the zero level. The energy of an AE event is the energy contained in the corresponding voltage signal and, strictly speaking, is not the true energy of the event itself. Energy corresponds to the integration of voltage squared over time.

The objectives of the reported work are:

1. To represent AE events from different machine operating conditions in terms of their event duration, peak amplitude and energy as a point in a three-dimensional space;
2. To establish the centres of the clusters for the four conditions in this three-dimensional space using the fuzzy c-mean clustering technique; and
3. To classify an AE event from an unknown condition by computing the minimum Euclidean distance of this event from the respective centres.

The methodology for machine condition monitoring and recognition is as shown schematically in Figure 4.7.

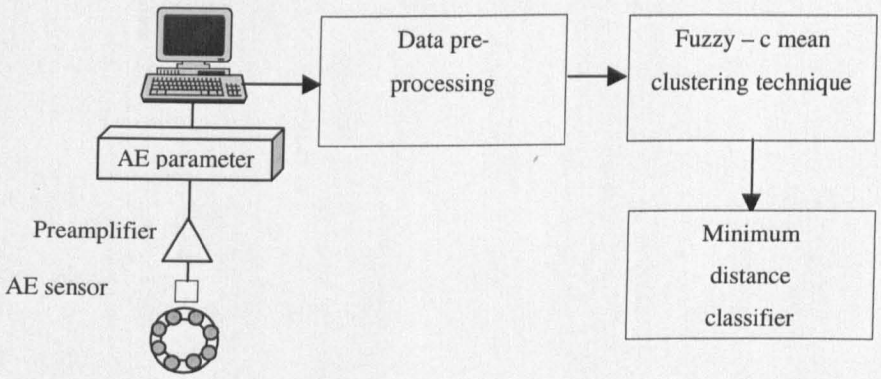


Figure 4.7 Fuzzy c-mean clustering on traditional AE parameters for machine condition monitoring

4.4.1 Pre-processing of data

For each condition, recordings of the captured AE parameters of event duration, peak amplitude and energy were made. These recordings were then divided into two sets. The first set served as the reference generated from the training exercise whereas the second set provided the test samples for validating the classifier obtained from the training process.

Both the training sets and the testing sets were processed as follows:

1. Sort the events in each set in the descending order of event duration.
2. Discard the first ten AE events in the sorted list as they may contain outliers, which if included, would distort the characteristics the AE events in a sub-set.
3. Select the next five AE events from the remaining list for subsequent clustering analysis.
4. Normalise each feature in a unit vector.

The new data sets are now much shorter, and the corresponding parameters of event duration, peak amplitude, and energy will be further analysed using the fuzzy c-mean clustering technique.

4.4.2 Fuzzy c-mean clustering technique

Fuzzy c-mean is an iterative technique for data clustering. The user decides on the number of clusters that a data set is to be separated into, initialises a proximity matrix and defines an error threshold for the stop condition of the iteration (Liu, 1996; Ross, 1995; Mechefske, 1998). In this study, the implementation of fuzzy C-mean was achieved using a MATLAB program. The source code is given in Appendix C1.

The proximity matrix contains membership values of an individual AE event that it does or does not belong to a particular machine condition. Initial values assigned to this matrix are arbitrary and binary logic values are generally used. Thus full membership is represented by 1 whereas non-membership by 0.

The first iteration generates estimated locations of the cluster centres and a refined proximity matrix in which the membership values become fuzzified, which mean that they now have values between 0 and 1. With successive iterations, the estimate for the cluster centre's locations will be more and more accurate and the proximity matrix will be updated. The iteration will stop when the change in the norm of the proximity matrix from its previous iteration becomes less than the designated error threshold. The cluster centres returned from the last iteration are taken to be the 'best' estimates.

Specifically, the fuzzy – c mean algorithm consists of the following steps (Ross, 1995):

1. Fix the number of c-cluster centres and a threshold ε for the stop condition in step 4. Initialise the proximity matrix $U^{(0)}$.

2. Update the c-cluster centres $\{v_i^{(r)}\}$ according to the current proximity matrix, using

$$v_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} \cdot x_{kj}}{\sum_{k=1}^{n} \mu_{ik}^{m'}} \qquad \textbf{Equation 4.8}$$

where μ_{ik} is the membership of the k^{th} data point in the ith class, and m' is the weighting parameter (the arbitrary value of $m' = 2$ was used).

3. Update the proximity matrix for the r^{th} iteration, $U^{(r)}$ according to previous cluster centres, using

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{2/(m'-1)} \right]^{-1} \qquad \textbf{Equation 4.9}$$

where $d_{ik} = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2 \right]^{1/2}$ is the distance measured.

4. If the objective function, as defined below, is less than the threshold ε, then stop; else, go to step 2.

$$\left\| U^{(r+1)} - U^{(r)} \right\| \leq \varepsilon \quad \text{otherwise set r = r+1} \qquad \textbf{Equation 4.10}$$

4.4.3 Minimum distance classification

A minimum distance classifier (Tou, 1974) was used to determine the class of machine condition to which a particular AE event belongs. It works by computing the Euclidean distances of the characteristic features from the centres of the clusters for the different machine operating conditions. The AE event is considered to belong to the cluster whose centre is closest. In other words, the distance is defined as an index of similarity so that minimum distance is equivalent to maximum similarity. For example, if the four centres have co-ordinates (X_{i1}, Y_{i2}, Z_{i3}) where i = 1, 2, 3, 4, and the AE event (X_1,Y_2,Z_3) as shown in Figure 4.8, the minimum Euclidean distance is then

$$D = \min\left(\sqrt{(X_{i1} - X_1)^2 + (Y_{i2} - Y_2)^2 + (Z_{i3} - Z_3)^2}\right) \qquad \textbf{Equation 4.11}$$

The minimum-distance classifier was used, because it readily allows an AE event of an operating condition to be added into the system. New clusters can be readily added to existing ones without affecting the current clusters. The source code is given in Appendix C2.

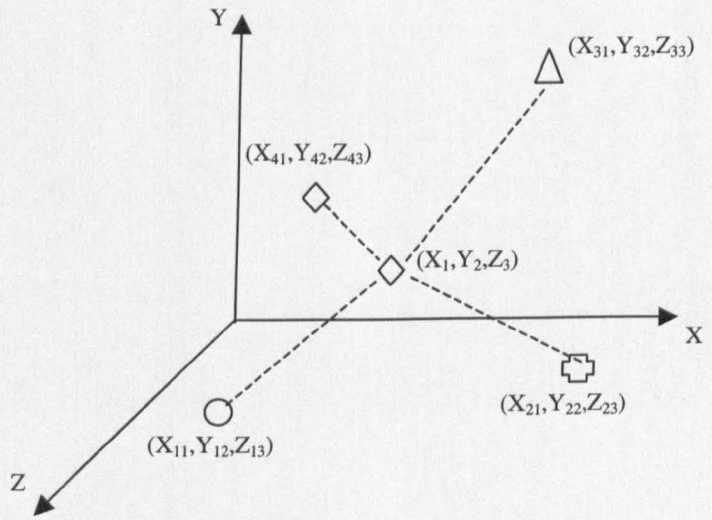


Figure 4.8 Minimum distance classification

4.5 Shape-to-life (STL) method

4.5.1 Inter-arrival times

Figure 4.9 show an AE signal consisting of seven AE events. Dedicated AE measuring instruments - for example, the AET5500 – capture each event that goes above a predefined threshold and extract various AE parameters together with a time stamp of the event. The time difference between two consecutive AE events is called the inter-arrival time between them. A high-speed data acquisition system – for example, the LABVIEW on PC – captures the whole time signal, Figure 4.9, and the inter-arrival time, as indicated in the diagram, can be easily computed from the discrete-time signal file using, for example, a MATLAB program.

The collection of inter-arrival times forms a distribution, which, as suggested in this thesis, is a form of Weibull distribution. As can be seen in the next section, the inter-arrival times of AE events should follow a Weibull distribution because AE events can be regarded as 'failures' on a microscopic scale.

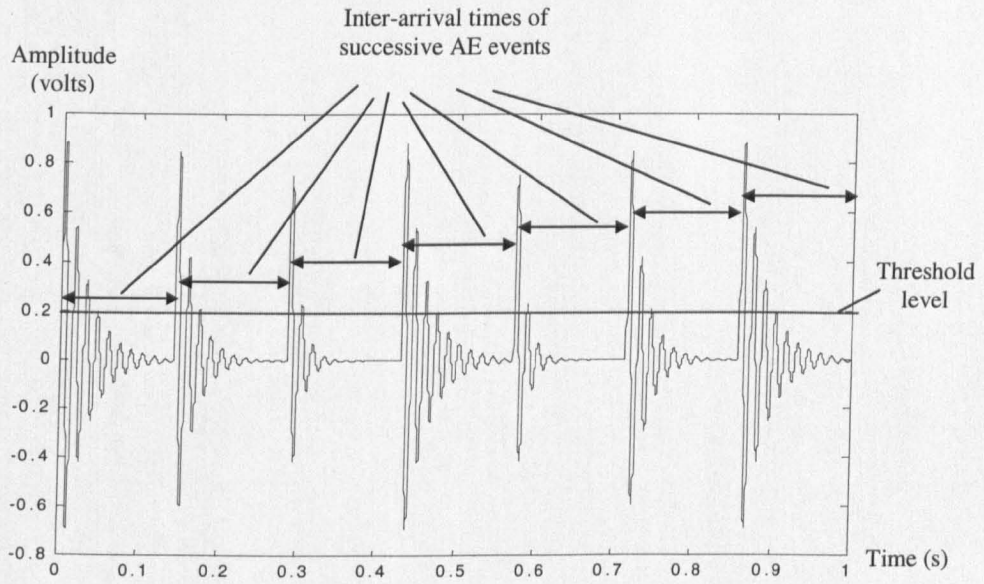


Figure 4.9 Inter-arrival times of successive AE events

4.5.2 Model of inter-arrival times of AE events

The Weibull distribution has been used in reliability engineering to model times to failure (O'Connor, 1991; Gertsbakh, 1989; Elsayed, 1996; John et al., 2000). Its usefulness lies in the simplicity that a single probability density function can be made to represent the time of failure arising from different modes of failure (running-in, random and wear-out). Since an instance of AE can be considered a kind of failure on a microscopic scale, the Weibull distribution is most likely to be suitable also for representing the probability of inter-arrival of AE events at a sensor. In this sense, the inter-arrival time is equivalent to time to failure. The justification for using the Weibull distribution is set out in this section.

The cumulative probability, F(t), of inter-arrival times of AE events must, by virtue of its definition, increase monotonically with the time interval t, starting with a probability value of zero at time t = 0 and approaching unity as time t tends to infinity. The form of this curve, known as the cumulative probability curve, for mathematical convenience, is represented by an exponential function as

$$F(t) = 1 - e^{-\phi(t)} \qquad \textbf{Equation 4.12}$$

Equation 4.12 is graphed in Figure 4.10, where it can be seen that $\phi(t)$, a function of time t, determines the precise form of the curve.

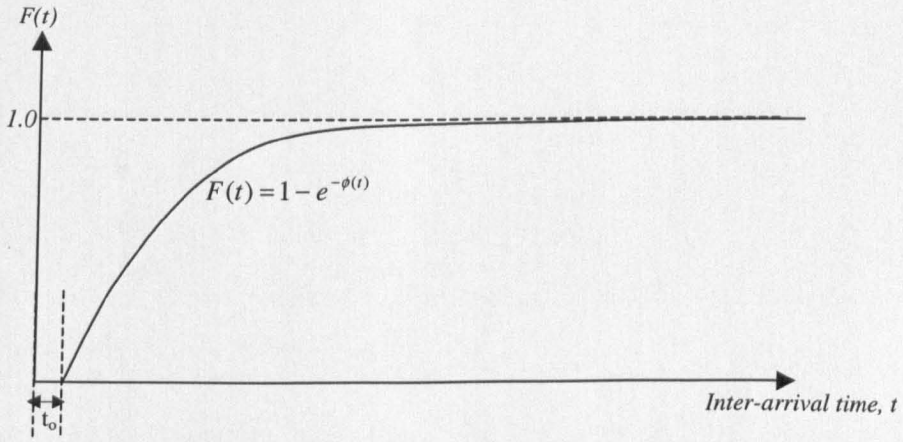


Figure 4.10 Cumulative probability of inter-arrival time F(t)

When AE event occurs, it occupies a finite time interval during which the occurrence of yet another event cannot be distinguished with an AE measuring instrument. Therefore, the inter-arrival time that can be measured of the subsequent AE cannot be shorter than the event duration of the current one. In effect, this means that a subsequent detectable event cannot occur in a time less than the event duration of the current one. In the context of cumulative probability F(t), its value has to be zero below some threshold time t_o, known as the guaranteed life.

The function $\phi(t)$, which defines the precise form of the cumulative probability F(t), should be non-dimensional because it is an exponent of the constant *e* in Equation 4.12.

Taking account all of the foregoing considerations, it is reasonable to suggest that $\phi(t)$ take the form

$$\phi(t) = \left(\frac{t - t_o}{\theta}\right)^{\gamma}$$

where t_o = the guaranteed life

θ = the characteristic life, and

γ = the shape parameter.

Thus, the inter-arrival times has the cumulative distribution function (cdf) given by

$$F(t) = 1 - e^{\left[-\left(\frac{t-t_o}{\theta}\right)^{\gamma}\right]} \quad \text{for } t>t_o \qquad \textbf{Equation 4.13}$$

and the corresponding probability density function (pdf) of

$$f(t) = \frac{\gamma \cdot (t - t_o)^{\gamma-1}}{\theta^{\gamma}} \cdot e^{\left[-\left(\frac{t-t_o}{\theta}\right)^{\gamma}\right]} \quad \text{for } t>t_o \qquad \textbf{Equation 4.14}$$

It is noted that Equation 4.13 and 4.14 are, indeed, the respective of cdf and pdf of the Weibull distribution. In these equations, $(t - t_o)$ denotes the 'quiet' zone, which marks

the period after one AE event has dropped below the threshold and before the occurrence of the next.

The parameter θ is referred to as the characteristic life in the Weibull distribution when used to describe the probability of time to failure in reliability work. This term is appropriate if, as mentioned before, an AE occurrence is regarded as a microscopic failure, which has been accepted to be the case. Characteristic life is therefore the characteristic AE inter-arrival time in this context. If the quiet zone ($t - t_o$) is as long as the characteristic life θ, then at the inter-arrival time of $t = t_o + \theta$, the cumulative probability has the value of

$$F(t_o + \theta) = 1 - e^{(-1)} = 0.63$$

In other words, if a hundred inter-arrival times were collected, 63 would have a value less than $t_o + \theta$.

The shape factor, γ, in Equation 4.13 and 4.14 is used to express the various patterns of the inter-arrival time distribution, some of which are shown in Figure 4.11. For $\gamma = 1$, $f(t)$ becomes an exponential probability density function. When $\gamma = 2$, the density function is known as a Rayleigh distribution. When $\gamma = 3.43$, the density function becomes a normal distribution.

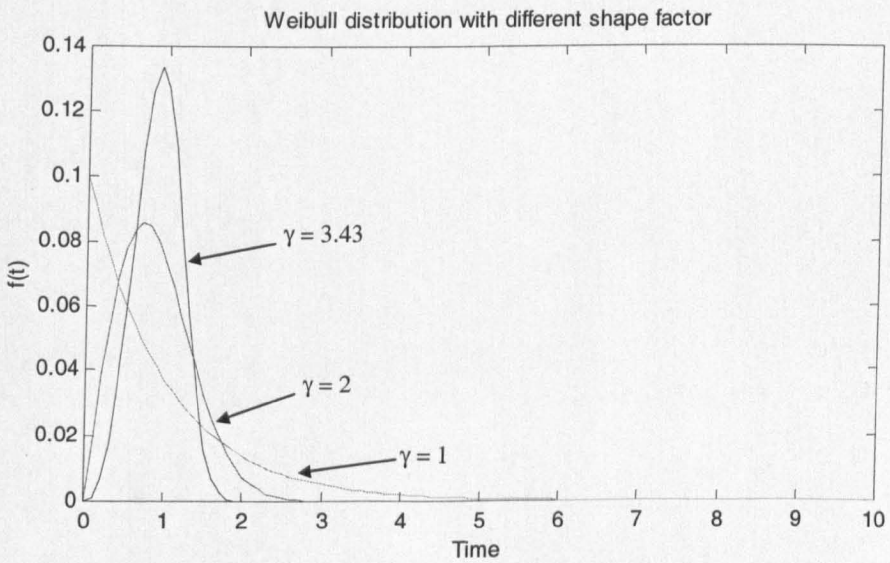


Figure 4.11 Weibull distribution with different shape factors

4.5.3 Definition of STL and L_{63}

Once a collection of inter-arrival times of an AE signal is formed, a Weibull distribution is then fitted to it and the Weibull parameters, shape (γ) and characteristic life (θ) and guaranteed life (t_o), can be estimated. Note that γ, being the exponent in Equation 4.13, is non-dimensional, θ and t_o have the dimension of time and hence the unit seconds.

The STL is defined to be the ratio of shape to characteristic life. In symbols, it is

$$STL=\frac{\gamma}{\theta} \qquad \textbf{Equation 4.15}$$

It is obvious that STL has the unit of s^{-1}.

L_{63} denotes the time duration within which 63% of the inter-arrival times of the distribution lies. As explained at the end of Section 4.5.2, the duration is

$$L_{63} = t_0 + \theta \qquad \textbf{Equation 4.16}$$

4.5.4 Weibull parameters estimation

To estimate the Weibull parameters; the shortest duration (t_o), the shape (γ) and characteristic life (θ) parameters of inter-arrival times that follow Weibull distribution, the method of least squares which operates on the cumulative distribution function (cdf) is used. This is because of the problem with using a pdf for parameter estimation is that the pdf is seldom smooth. Only when the sample size is very large then the pdf will give the appearance of a smooth curve. The cdf overcomes this problem. The Weibull cdf, as given in Equation 4.13 has a far simpler mathematical form than its pdf, Equation 4.14. Furthermore, the complement of the Weibull cdf, known as the cumulative reliability probability, in the field of Reliability Engineering (Karl, 1975; Elsayed, 1996), has an even simpler form, as

$$R(t) = e^{\left\{-\left(\frac{t-t_0}{\theta}\right)^{\gamma}\right\}} \quad \text{for } t > t_o \qquad \textbf{Equation 4.17}$$

The method of least squares works on the principle of minimising an objective function defined as the sum of squares of the deviations between the cdf and R(t) of

Equation 4.18 where guesses of the three parameters are initial made. Through a procedure of varying the three parameters' values, the objective function is reduced to a preset small value (ε). The objective function in this case is

$$\sum_{j=0}^{n-1}\left[R_j - \exp\left(-\left(\frac{t_{input}-t_o}{\theta}\right)^{\gamma}\right)\right]^2 \le \varepsilon \qquad \textbf{Equation 4.18}$$

The program for implementing this algorithm is written in MATHCAD; its listing is given in Appendix C5.

5 Experimental design and measurement of the systems approach

This chapter describes the measurement system and test rigs used in the research into bearing condition monitoring and diagnosis on rotating machines operating under a variety of conditions including variations of speeds and loads, different types of simulated bearing defects and ultimately the natural wear and damage of bearings arising from fatigue failure. Discussion will be presented in three stages. Firstly, there will be the consideration of the experimental design that gives consistent and reliable results. Secondly, the issue of choice of suitable measurement instruments will be explored in the context of extracting meaningful information from the measurement process. Finally, the design of two test rigs, representative of rotating machines, will be presented; data obtained from running these test rigs will be used to validate the proposed condition monitoring and diagnostic approach.

5.1 Experimental design

The purpose of the design of the experiment is mainly to investigate the capability and limitation of the proposed systems approach. Therefore, the criterion for the experimental design is based on how effective the multisensor data fusion approach will establish the relationship between of the machine bearing operating conditions and the relevant information extracted from the sensors – acoustic emission and vibration.

The diagram showing the multisensor measurement system is given in Figure 5.1. In this system, signals from both transducers are pre-amplified and filtered in order to increase the gain signal-to-noise ratio before being passed through to the final data processing units.

The preamplifier and filter need to be chosen according to the range of frequency of interest and to fulfil the impedance matching requirement of measurement electronic circuitry so that a high signal-to-noise ratio can be achieved.

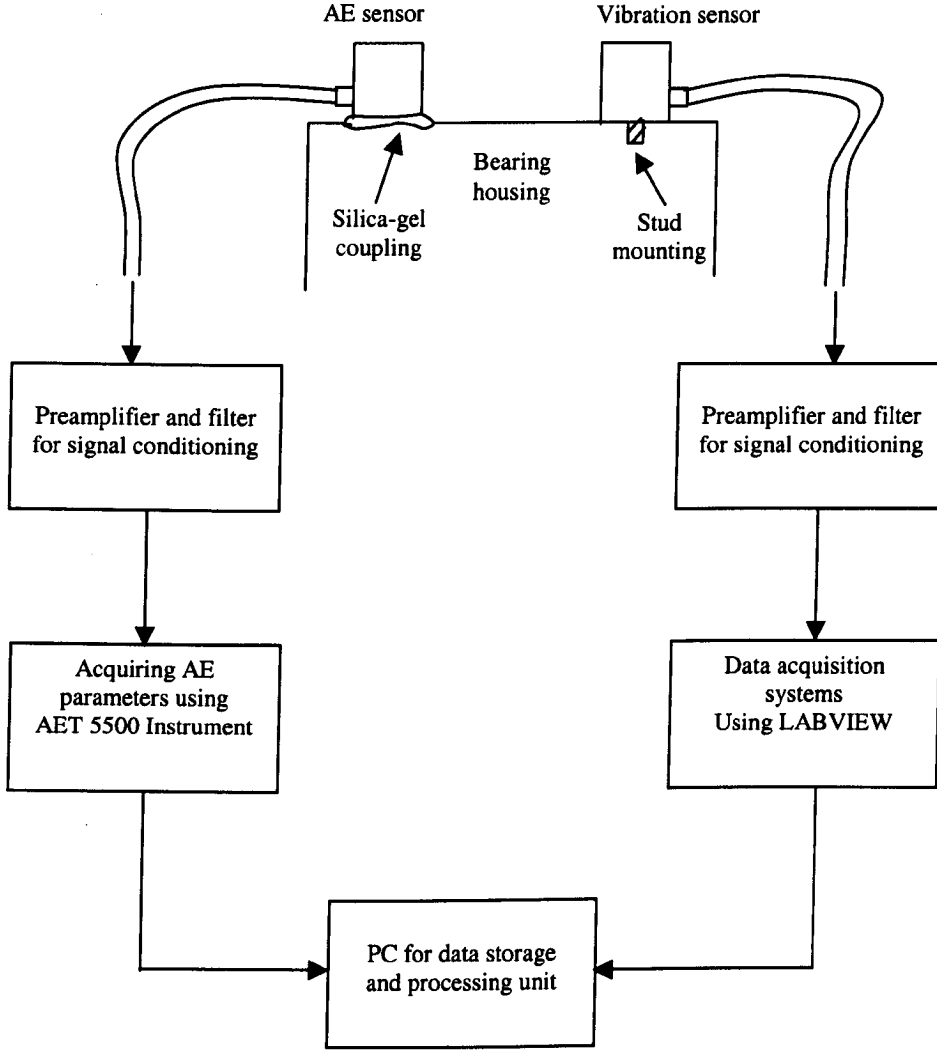


Figure 5.1 Schematic diagram of the experimental design and measurement of systems approach

5.2 Measurement equipment

Performing monitoring requires: 1) sensing devices or transducers that are capable of acquiring useful information, preamplifiers and filters; 2) test rigs that are capable of delivering various controlled machine operating conditions and 3) data acquisition systems that are capable of recording all relevant information in order for it to be post-processed in subsequent analysis.

5.2.1 Accelerometer

To measure vibration, an accelerometer, PCB 333A12 shear mode, was used to measure the radial vibration at the bearing. PCB 333A12 has the sensitivity of 952 mV/g considered relatively high when compared to industrial accelerometers whose values are typically between 10 and 100 mV/g. The reason for choosing the high sensitivity accelerometer is that it can provide a higher signal-to-noise ratio thus giving a better discrimination of background noise. The frequency range of PCB 333A12 is between 1 to 3000 Hz; the operating temperature ranges from –50 to 82 degrees Celsius.

The accelerometer was mechanically fastened to the bearing housing by means of a threaded stud. Such coupling with the accelerometer mounted to a very smooth surface, generally yields the highest mechanical resonant frequency and, therefore, the broadest usable frequency range (Michael, 1979). The position of the accelerometer mounting was chosen to be at the top of the bearing housing since the housing has the lowest dynamic stiffness in this direction.

5.2.2 Accelerometer preamplifier

A filter (KEMO model) with a built-in amplifier is used to process the signals, which are then sampled into a PC. The preamplifier module is used to increase the gain of the voltage signal. The filter module is an electronic circuit, which allows only certain frequency components of the signal to propagate. High pass and low pass filters reject those frequency components, which lie, respectively, below and above some predetermined frequencies.

5.2.3 AE transducer

A wide band transducer (model WD, PAC) was chosen for AE measurement because of the large dynamic frequency range of the measured signal. The operating frequency range of the transducer is from 100 kHz up to about 1 MHz. From the manufacturer's calibration curve, the sensitivity against frequency is relatively smooth and flat over this range so that the transducer can provide a constant output signal throughout the operating frequency. The sensor is enclosed in a stainless steel case and the ceramic sensing disc is coated to minimise EMI interference. Its operating temperature ranges from –45 to 80 degrees Celsius. Because of the wide frequency range, wide band transducers produce signals that permit meaningful frequency analysis.

Silicon-gel is used as a couplant, which ideally forms a thin layer connecting the transducer to the coupling surface. The couplant fills any gaps caused by surface roughness and eliminates air voids to ensure good acoustic transmission. Before mounting the AE transducer, the surface of specimen needs to be clean and flat so that it facilitates maximum couplant adhesion as uniform coupling results in a greater sensitivity to AE signals. The sensor is positioned on a flat and smooth surface machined on the top of the bearing housing.

5.2.4 AE preamplifier

The AE signal detected by the piezoelectric element in the sensor needs to be amplified before transmission to the measurement circuitry. A preamplifier is connected, via a short cable close to the transducer and it provides a suitable gain and high cable drive capability. A preamplifier (model 2100/PA 60) is used with a gain of 60 dB, and its built-in band pass filter set at 100 kHz – 1 MHz. The band-pass filter operates as a frequency window that eliminates the frequency components lying outside the window. The 2100/PA preamplifier's output is limited to 10 V (peak to peak); the connection to a signal cable is via a lemo type connector.

5.2.5 AET 5500 data acquisition system

AET 5500 is a general purpose instrument that can monitor AE events generated from any AE source. AET 5500 can be configured from one to eight active channels for monitoring and processing acoustic emission parameters such as ringdown counts, events, event duration, peak amplitude, rise time, voltage level (RMS) and AE energy.

These AE parameters can be stored in a PC to permit further data analysis at a later time. Signal processing in AET 5500 is performed by a 16-bit processor mainframe.

When the amplified and buffered signals reach the mainframe, it may again amplify the signal (post amplification) before processing, depending on the signal characteristics that need to be monitored and the signal strength. The AET system allows a user to adjust the threshold at a fixed or automated (floating) level. The automatic (floating) threshold allows the threshold level to increase and decrease in response to changing levels of background noise such that the recording of only significant AE events can be assured.

The system also provides time duration information of the AE events produced when the threshold is crossed. The clock that determines the time duration is a unit that counts the pulses of a time base signal. After the signal is processed, it is sent to the intelligent graphics terminal (IGT) where a user can view a graphical representation or statistical table of the desired acoustic emission data.

5.2.6 Data acquisition system

The data acquisition system is based upon a Pentium PC with a 1 GHz - CPU speed which is fitted with a high performance National Instruments NI 6110 data acquisition card. The NI 6110 card comprises four 12-bit input resolution channels, simultaneously sampled analogue-to-digital (A/D) input channels with scalable input limits and adjustable sampling rates up to 5 M samples/s, both controllable by software.

Typically the detected signals from both AE and vibration sensors were recorded on the data acquisition card simultaneously with different sampling frequencies, for example, sampling at 2000 Hz from the accelerometer and at 3 MHz from the acoustic emission sensor. Then the signal is converted into a digital format that can be stored in the PC so that it can be further processed and analysed.

The software used was LABVIEW (version 5.1). This is a graphical programming software package that enables the creation of a measurement program for data acquisition system. The virtual measurement program provides a graphical interface

for a user to define and control the data acquisition process to operate at various sampling rates and durations of recording.

5.2.7 Hall effect sensor

A Hall effect sensor is used on the first test rig (description in Section 5.3) to determine the speed of the rotating shaft. The Hall effect sensor can produce a convenient external trigger for delimiting the acceleration or AE signals to correspond to a cycle of rotation. The sensor produces a pulse every time a piece of ferrite material, or the like, passes in front of the sensor. The sensor is mounted next to the drive end bearing housing. The duration between two successive pulses gives the cycle time of rotation of the shaft and its reciprocal is the shaft rotating speed. Increase the speed of rotation and the period between pulses decreases. Likewise, decrease the speed of rotation and the period increases.

5.2.8 Photo detector sensor

A Photo sensor is used to give a rotation reference signal for the second test rig (description in section 5.4). The sensor produces a pulse every time the light ray emitted from a light source (LED) that is reflected back to photo detector. A reflective tape is attached to the rotating shaft to provide a shiny surface for the light reflection. The photo sensor is mounted to point at the side of the rotating mechanical coupling. The signal from the sensor is also used as a trigger to start the simultaneous collection of data when there is more than one channel involved.

5.2.9 RMS circuitry

A raw AE signal, being of high frequency, demands a high sampling rate for its faithful capture. In many instances, there is no need to study the raw signal. One common form of a converted AE signals is its rms, which has the characteristic that the broad outline of the original signal is still retained. More importantly, the rms equivalent, being of a much lower frequency, can be captured into a data acquisition system with a much lower sampling rate, thus taking up much less memory for data storage. A RMS circuitry was built for this purpose. Its role can be considered as providing a level based signal processing which involves the generation of a DC coupled dynamic signal related to the magnitude of the amplified and filtered AC

signal from the transducer. In this study, a time constant of 100 milliseconds was chosen as the averaging time to determine the information content of the signal.

5.3 Test rig for multi-faults at high rotating speed

A first test rig was designed and built in order to produce multi-fault operating conditions from rotodynamic machinery. From now on this test rig will be referred to as a high-speed light duty test rig. Examples of multi-faults are: 1) the out-of balance forces condition can be generated using various masses at different eccentricities attached on rotating disc, 2) the misalignment condition can be simulated by adjusting the spindle from one moving bearing literally from the other, and 3) the defective bearing can be artificially generated using an electric discharge pen.

The design of the test rig incorporates a rotating shaft supported at two points: a deep groove ball bearing at the drive end and a tapered bore bearing at the non-drive end. A sketch of the test rig is illustrated in Figure 5.2. The shaft is manufactured from steel with a diameter and length of 20 mm and 300 mm respectively. The two bearings are a FAG 20205K.T.C3 which is a self-aligning single-row taper-bore roller bearing and a FAG 6304 2ZR.C3 which is a deep grove ball bearing. They are mounted in bearing housings which in turn were attached to a base plate.

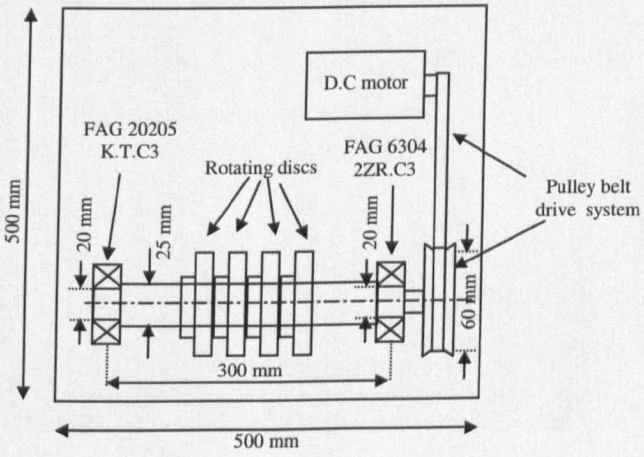


Figure 5.2 Sketch of the high-speed light duty test rig

The non-drive end incorporates a tapered bored bearing fitted using an adapter sleeve. The bearing used at the non-drive end is a barrel roller bearing (FAG 20205K.T.C3). The reasons for using barrel roller bearing are because it is suitable for applications where high radial load carrying capacity and the compensation of misalignment are required. Barrel roller bearings of the basic design are available either with cylindrical or with tapered bore. The bearings with cylindrical bore have normal radial clearance, those with tapered bore an increased radial clearance (C3).

The plummer block is used at the non-drive end, resulting in a quick release mechanism for the tapered bore bearing in order to facilitate a comparison between a new bearing and a defect-simulated bearing. An adapter sleeve is also used for fastening the tapered bore bearing on the shaft. This facilitates the shafts to be machined to larger tolerances than would be the case if the bearing is seated directly on the shaft. The bearing rests against a shaft shoulder; the tapered sleeve is pushed into the bearing bore until the radial clearance is reduced to the required amount.

The drive end housing is manufactured from steel. This is to accommodate a single row deep grove ball bearing (FAG 6304 2ZR.C3). The single row deep grove ball bearing is chosen because it can accommodate radial and thrust loads for rotating machines at high speed. Sealed deep grove ball bearing is maintenance free. However, its self-aligning capability of deep grove ball bearing is somewhat limited, thus a well-aligned bearing seat is required. Misalignment impairs the smooth running of the balls, inducing additional stress in the bearing and consequently reducing the bearing service life.

The shaft is driven by a D.C motor via a pulley belt drive system at the drive end. The driven pulley is made from aluminium bar to incorporate with a 5 mm pin hole V-belt. The spindle is driven by the D.C motor running at 20 rev/sec. The motor delivers a constant torque at all speeds. The shaft is designed to withstand dynamic loading effects from centrifugal forces arising from rotating unbalance masses.

Four identical rotating discs are manufactured from aluminium to generate out-of-balance conditions. To secure a disc onto the shaft, a grub screw is used which presses

a brass insert onto the shaft that reduces the risk of imprinting the shaft, thus being able to slide and rotate a disc along the length of the shaft.

The base plate is manufactured from aluminium of dimension 500 mm x 500 mm x 10 mm. Four circular struts are fitted with rubber at the bottom of each corner of the bed to isolate the mechanical vibration induced by the test rig from the workbench.

An accelerometer (PCB 333A12) and an AE sensor (WD1000) are mounted on a non-drive end at the top of bearing housing. A Hall effect sensor is used to determine the speed of the rotating shaft and also give a reference rotation signal.

The test rigs provides facilities to produce the four machine operating conditions characterised by:

1. The rotating shaft dynamically balanced (referred to as 'balance shaft'),
2. The rotating shaft dynamically unbalanced in one plane to the extent of 65 x 10^{-5} kg.m at mid-span of the shaft (referred to as 'unbalanced shaft'),
3. The shaft with misalignment achieved from moving one bearing laterally by 1 mm relative to the other (referred to as 'misaligned shaft') and
4. The roller bearing seeded with a defect on the outer raceway of 1 mm diameter produced with an electric discharge pen (referred to as 'defective bearing').

5.4 Test rig for a bearing life test and loading variation at low speed

A second test rig was designed intended for different loading conditions and built in order to evaluate the proposed monitoring scheme. From now on this test rig will be referred to as a low-speed heavy duty test rig. The test rig was originally borrowed from Corus; it was used in this research to study the effect of variation of loading conditions on AE and vibration at slow rotating speed. The original rig was then modified such that a radial load could be applied to the rotating shaft using hydraulics and the speed of the shaft can be lowered still further using an inverter and motor controller. A sketch of the low-speed heavy duty test rig is shown in Figure 5.3.

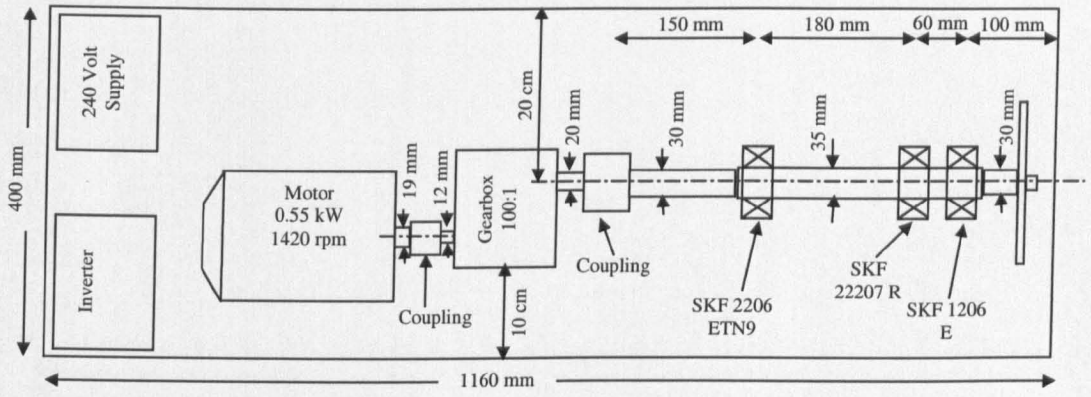


Figure 5.3 Sketch of the low-speed heavy duty test rig

The design of the test rig comprises a rotating shaft supported at three points: a double-row self-aligned ball bearing at the drive end, a spherical roller bearing at the applied load position near the non-drive end and a single row self-aligned ball bearing at the non-drive end. The shaft is manufactured from steel of 35 mm diameter. The three bearings are a SKF 2206 ETN9 which is the double-row self-aligned ball bearing, a SKF 22207 E which is a spherical roller bearing and a SKF 1206 E which is a single row self-aligned ball bearing. They are mounted in bearing housings that in turn were attached to a base plate.

A three-phase A.C motor (model ABB AA082001ASA) is chosen to replace the original motor so that adjustable speed and high torque can be achieved. The motor is foot-mounted onto the base plate. The connection of the three-stator phase windings

of the motor is delta (Δ) so that this would accept a three-phase input voltage range of 220-240 Volts for a 50 Hz supply frequency. The power of the motor is rated at 0.55 kW (or 0.75 hp).

An inverter (or A.C motor speed controllers) is used to enable the 3-phase induction motor to have its speed varied. The inverter (model Allen-Bradley Range: B161 Series) is used to provide adjustable speeds by means of voltage deviation. Its analogue speed control is of range 0-10 Volts or 4-20 mA. The selected inverter has its application power range at 0.75 kW. It can be set up and operated using the integral keypad. The shaft is driven by a controllable inverter running at 0.23 rev/sec.

A worm wheel reduction gear (model Ondrives: PP60-100) is used to replace the original gearbox in order to decrease speed of the shaft by an A.C motor. The gear can produce a reduction ratio of 100:1 from the motor shaft rotation speed. This also gives sufficient driving torque at low speed to overcome the high radial load. The flexible mechanical couplings are used on both input and output shafts of the reduction gearbox to incorporate with the drive shaft from electric motor and the rotating shaft. They are drilled to fix with keyways on both shafts.

The base plate is manufactured from steel of 400 mm x 1160 mm x 2 mm. At the bottom, the two U-shape steels are mounted to the bed in order to prevent the plate from deformation by an excessive load.

Figure 5.4 illustrates the modified loading system. The applied radial load is intended to cause accelerated fatigue failure on the bearing. The hydraulics system comprises a hydraulic cylinder, hydraulic hand pump, high-pressure hydraulic hose (1 metre), gauge mounting block and hydraulic pressure gauge. The reason for using hydraulics system are that it can provide enough radial force to accelerate a bearing life and its simplicity in designing the corporate components as a frame shape over shaft to enable the applied force radially downwards onto the non drive end of the shaft. The technical drawings of the modified loading system are given in Appendix B.

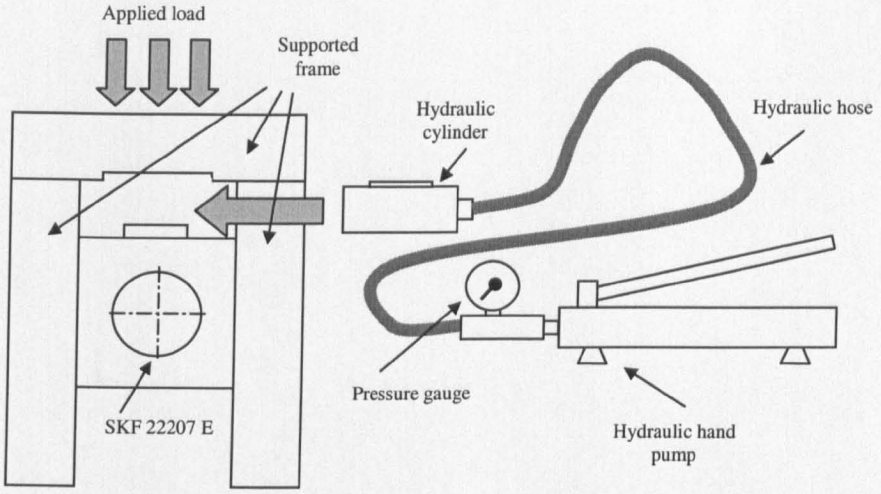


Figure 5.4 Sketch of the modified hydraulic loading system

In order to design the hydraulic loading system, the following criteria are considered. The applied radially force provided by a low height hydraulic cylinder, Figure 5.5. The used low height hydraulic cylinder (model RS200-8775) combines a minimum closed height with maximum stroke of 10 mm to provide a short lifting force in very confined work area.

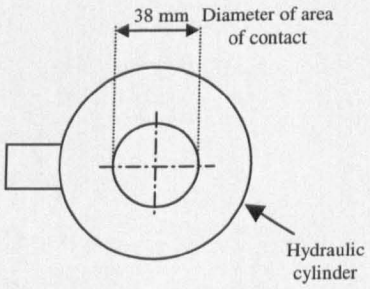


Figure 5.5 Sketch of a hydraulic cylinder

The area of contact of hydraulic cylinder is calculated from its radius

$$A = \pi r^2 \qquad \textbf{Equation 5.1}$$

$$A = 1134.11(mm^2) \qquad \textbf{Equation 5.2}$$

The maximum working pressure from the hydraulic hand pump is 700 bars. Therefore, the maximum applied force can be calculated from

$$F = P.A \qquad \textbf{Equation 5.3}$$

$$F = 700x100x\frac{10^3}{10^6}(N/mm^2)xA \qquad \textbf{Equation 5.4}$$

$$F = 70(N/mm^2)x1134.11(mm^2) \qquad \textbf{Equation 5.5}$$

$$F = 79387.7N \approx 8100kg \qquad \textbf{Equation 5.6}$$

The vibration and AE are measured in the vertical direction on the bearing housing with the respective PCB 333A12 accelerometer and a wideband acoustic transducer. A photo detector sensor is used to determine the speed of the rotating shaft and also give a reference rotation signal. The test rig is run under the following conditions:

1. the rotating shaft dynamically balanced with no external load,
2. the rotating shaft dynamically balanced with stepwise increasing loads up to 300 bars applied radially downwards to the non drive end of the shaft.

When a bearing is subjected to radial load over a long period, fatigue failure will develop due to the cyclical stress; such failures are common in industry.

5.5 Conclusion

The measurement system and test rigs used in this research have been discussed. To design a multisensor measurement system, both AE and accelerometer were used to perform such the task in the sense that utilising the two transducers will enable complementary for their non-overlapping ranges so that they can operate together as a single sensor with greater fault detection capability. The selections of measurement equipments are of importance in order to obtain meaningful signals or information from the two transducers such as a high signal-to-noise ratio and impedance matching, leading to a successful condition monitoring. Finally, the details of the design of two test rigs, so called a high-speed light duty and a low-speed heavy duty test rigs, are given. The high-speed rig, which can provide four different machine operating conditions: artificial fault simulations, was run at 20 rev/sec. By contrast, the low-speed rig was operates at 0.23 rev/sec. This rig was used to provide a validation of a bearing service life test when it was subjected to radially overload forces.

6 Experimental results of the systems approach

This chapter describes experiments on condition monitoring using vibration and acoustic emission signals conducted on two test rigs: high-speed light duty and low-speed heavy duty. The purpose is to establish the suitability of the monitoring methods for a range of machine conditions. The presentation of this chapter is in four parts. Firstly, vibration condition monitoring on the high-speed light duty test rig using joint time- and frequency- domain analysis is discussed. Secondly, AE condition monitoring using the fuzzy c-mean clustering technique on the high-speed light duty test rig is given. Thirdly, AE data provided by Corus, Middlesborough, for clean and contaminated lubricant bearing conditions are classified using the fuzzy c-mean clustering technique. Finally, the inter-arrival times between AE events obtained from experiments conducted on both test rigs are studied in terms of the shape of distribution modelled as the Weibull distribution with the corresponding shape, guaranteed and characteristic lives estimates; this leads to a novel method of condition monitoring using the ratio of shape to life parameters (STL) and L_{63} as indicators of bearing condition.

6.1 Vibration condition monitoring using CWT and BPNN

The aim of this experiment is to investigate the applicability of continuous wavelet transform (CWT) on vibration signals for condition monitoring of high speed rotating machine. As demonstrated in Section 4.2.1, CWT, a joint time and frequency domain technique, is capable of revealing hidden transients that occur when rolling elements come into contact with a defect. The approach adopted was to transform the vibration signals into colour coded images with CWT for different machine conditions and to perform classification using a back-propagation neural network (BPNN).

6.1.1 Vibration experimental set up

The high-speed light duty test rig was used to validate the proposed approach. It was set to run at 20 rev/sec. The test bearing was a FAG 20205K.T.C3 self- aligning single row taper-bore bearing. The vertical radial vibration at the bearing was measured using a PCB 333A12 accelerometer, located on top of the non-drive end housing. The acceleration signals were first low-pass filtered at 1 kHz for anti-aliasing and then sampled into a PC with the Labview data acquisition software. The sampling rate used in the experiments was 2,000 samples/second.

6.1.2 Vibration experimental results

Vibration measurements were obtained from four different machine conditions as presented in Section 5.3. For each condition, twenty signals were collected and then divided into two equal sets of ten each. The first set was used for training the neural network while the second set for testing. The four different machine conditions are: (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing. Samples of these raw time domain signals are as shown in Figure 6.1.

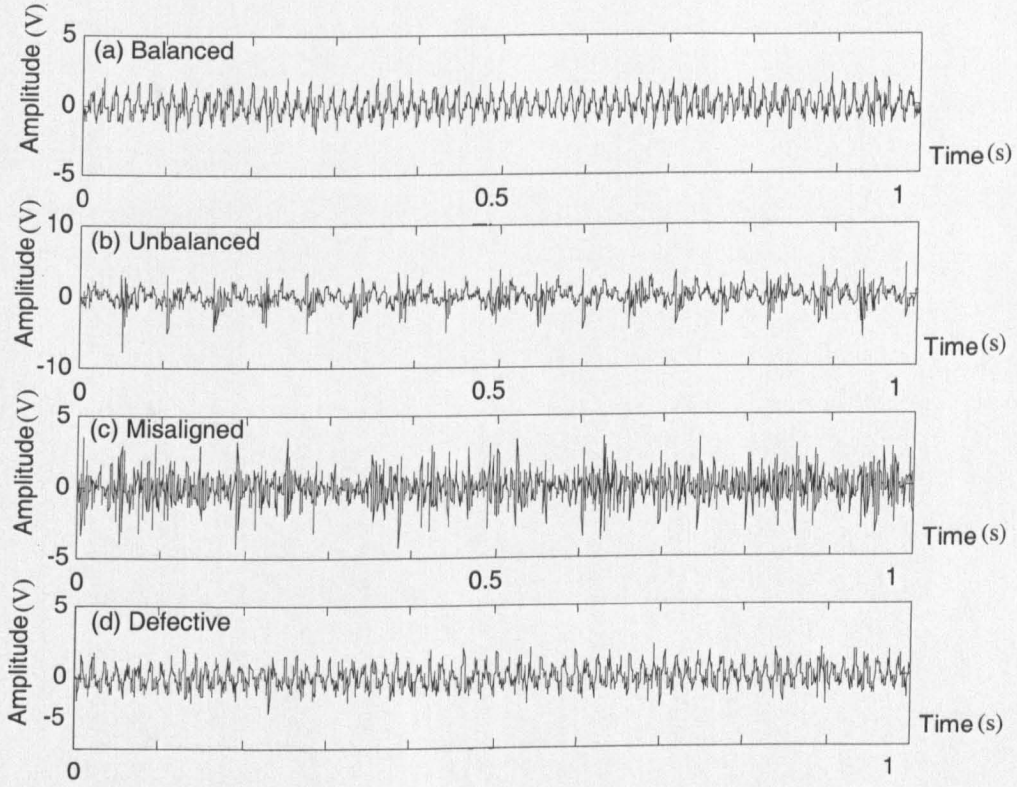


Figure 6.1 Acceleration signals from four machine conditions: (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing

The coefficients $F_{\Psi}(a,b)$ in the CWT method were calculated in the manner as described in Section 4.2.3. They were then displayed as a grey-scale map with the vertical and horizontal axes denoting respectively the scale a and position b, the latter being equivalent to time t, as shown in Figure 6.2 for the four different conditions. Although there is general similarity between them, subtle differences are clearly visible.

The CWT grey-scale maps were then 'binarised' using thresholding. The selected threshold value in this experiment was 128, which is half of the range of grey scale values (0-255). With this threshold, the intensity of each pixel was converted into either black or white, thus producing a binary two-toned image. Such an image has the obvious advantage that when classification using neural networks is done, the computation time will be much shorter, as multiplication involving a 0 or 1 is much easier to perform.

Figure 6.3 (a) and (b) show a sample image before and after binarising. The binarised CWT image of a signal would be used for the training and testing sessions of the neural network.

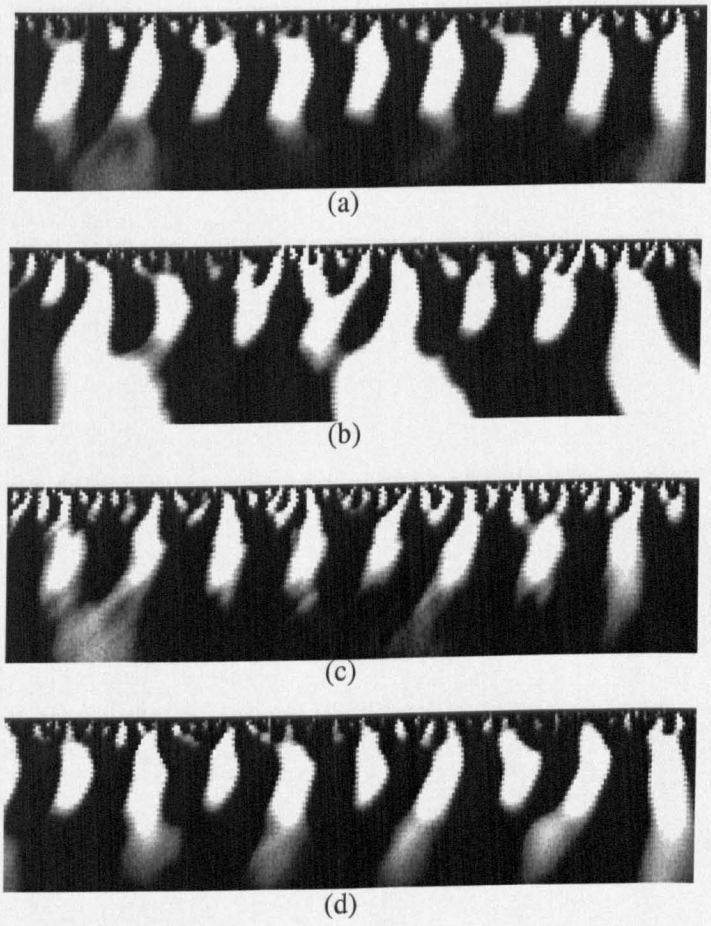
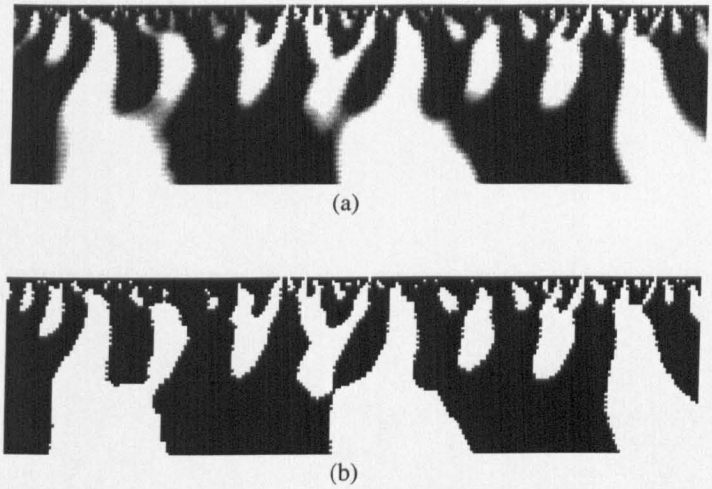


Figure 6.2 CWT transformed images of (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing



(a)

(b)

Figure 6.3 Examples of (a) an original image and (b) its binary image

A back-propagation neural network (BPNN), with the architecture as shown in Figure 6.4 was used to classify machine conditions. The network has an input layer, two hidden layers and an output layer. The values of the pixels composing the CWT binary image was used as inputs to the neural network in both training and testing. Since the image consisted of 64 scales and 250 time intervals, thus giving a resolution of 64x250 = 16000 pixels, the number of input nodes is also 16000. The output layer comprised 4 nodes, representing the four machine conditions of balanced, unbalanced and misaligned shafts, and defective bearing. Through a process of trial and error based on minimizing the mean square error (MSE), the choice of 12 nodes in the first hidden layer and 18 nodes in the second hidden layer were considered optimal.

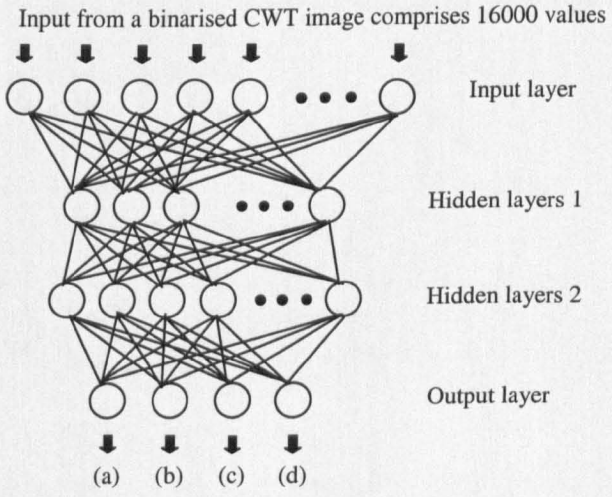


Figure 6.4 Back-propagation neural network architecture for (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing

Both the training set and testing sets comprised 40 signals made up of 10 signals from each of the four machine conditions. After the BPNN had been trained using the training set to recognise the four machine conditions correctly, it was then tested against the testing set. The test results are tabulated in Table 6.1 where each row shows the values produced at the respective four output nodes of the neural network, Figure 6.4. These values, ranging from about 0 to 1, represent the likelihood that the test condition belongs to one of machine conditions. For example, the first test signal obtained from a 'balanced' condition, producing the first row of numerical values in Table 6.1, was classified as 'defective bearing'. This is evidently incorrect and so labelled as 'W' (for wrong) in the last column. Had the classification been correct, the label would be 'R' (for right).

It can be seen from the last column of Table 6.1 that the test signals have been classified correctly 36 out of 40 times, with a recognition rate of 90%. 'Binarising' or thresholding gives the benefit of noise removal and reduced computational effort as mentioned earlier. However, it is obvious that the threshold level influences greatly the recognition rate: if the level is pushed to the extremes, 0 or 255, the corresponding binarised CWT image will turn out to be completely white or black.

Table 6.1 Classification results from BPNN

	Values from output nodes of BPNN				
Test set	**(a) Balanced shaft**	**(b) Unbalanced shaft**	**(c) Misaligned shaft**	**(d) Defective bearing**	**Classification result**
Balanced shaft	0.22	-0.01	0	0.79	W
	0.94	0	0	0.06	R
	0.36	0.19	0	0.34	R
	0.92	-0.08	0	0.23	R
	0.98	0	0	0.03	R
	0.96	0.03	0.03	-0.03	R
	0.86	-0.02	0	0.18	R
	0.73	0	0	0.27	R
	0.2	-0.01	0	0.81	W
	0.98	0.05	0.01	-0.02	R
Unbalanced shaft	0.04	0.35	0.22	0.26	R
	0.04	1	0	-0.04	R
	-0.12	0.84	0	0.45	R
	0.01	0.89	0	0.08	R
	-0.01	0.97	0	0.04	R
	-0.01	0.19	0.29	0.47	W
	0	1.01	0	-0.01	R
	-0.08	0.92	0.01	0.18	R
	0.1	0.7	0.04	0.09	R
	-0.01	1.05	0	-0.05	R
Misaligned shaft	0	0	0.99	0	R
	0	0	1	0	R
	0.01	0.07	0.95	0.12	R
	0	0.02	1	-0.02	R
	0	0	1	0.01	R
	0.01	-0.03	0.95	0.05	R
	0	0	1	0	R
	0	0	0.99	0	R
	0	0	0.99	0	R
	0	0.04	1	-0.04	R
Defective bearing	-0.01	0.08	0.33	0.51	R
	0.02	0.01	0	0.94	R
	-0.08	-0.1	0.1	1.1	R
	0.3	0.34	-0.02	0.3	W
	-0.02	0.13	0.26	0.6	R
	-0.05	0	0.42	0.67	R
	0.24	-0.02	0	0.78	R
	-0.09	0.32	0.29	0.52	R
	0	-0.05	0.44	0.64	R
	0.38	-0.1	-0.02	0.88	R

Note: R = classified correctly , W = misclassified

6.2 AE condition monitoring using fuzzy c-mean clustering

This section describes the application of the fuzzy c-mean clustering technique on AE for machine condition monitoring. AE signals were measured from the high-speed light duty test rig and AE parameters were then extracted and used as characteristic features of the machine condition. In this experiment, the same four machine conditions were investigated. The fuzzy c-mean clustering technique works by first identifying the centres of clusters in the AE feature space for the four machine conditions based on the data from a training set. Then during testing, the point corresponding to the test condition in the feature space will have its distances calculated from the established cluster centres; the point is then assigned to the nearest cluster.

6.2.1 AE experimental set up

Experiments were conducted on the high-speed light duty test rig running at 20 rev/sec. The test bearing was a self-aligned single-row roller bearing (FAG 20205K.T.C3). The test rig provided four machine conditions corresponding to (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing. The details of the set up were given in Section 5.3.

An acoustic emission transducer of the wideband type from Physical Acoustic Corporation (PAC) was mounted on the non-drive end bearing housing with a silicone gel couplant. The AE signal from the transducer was amplified 60 dB and band-pass filtered - 100kHz to 1 MHz - with a PAC preamplifier before entering the AET5500 (a micro-processor based system from Acoustic Emission Technology Corp.) for converting the signal into AE parameters.

As discussed in Section 3.2.3, the correct selection of a threshold is important in order to obtain a high signal-to-noise ratio. In this experiment, a floating threshold was chosen because such a threshold, floating just above the background noise level, eliminates much of the background noise.

6.2.2 AE experimental results

For each condition, twelve recordings each of about 30-second duration were made and they captured the AE parameters of event duration, peak amplitude and energy.

These recordings were then divided into two sets of 3 and 9 each. The first set served as the reference generated for the training exercise whereas the second set provided the test samples for validating the classifier obtained from the training process. Then the AE parameters were pre-processed using the four steps as described in Section 4.4.1 and summarised here. They are: (1) sorting AE events in descending order of the AE event duration, (2) discarding the first ten events as outliers, (3) selecting the next five AE events from the remaining list and (4) normalising the feature vector of each event into a unit vector. As is evident from the procedure described, the number of feature vectors used is only 5 and so the whole computation process is made very fast.

Since the AE parameters of event duration, peak amplitude and AE energy were used, each feature vector had only three elements. The clusters could therefore be displayed in three-dimensional graphs, as Figures 6.5 and 6.6. Figure 6.5 shows the AE events, 15 of each condition or 60 in total, for the four different machine conditions from the training sets. Figure 6.6 shows the same AE events but this time with clusters identified and their centres computed.

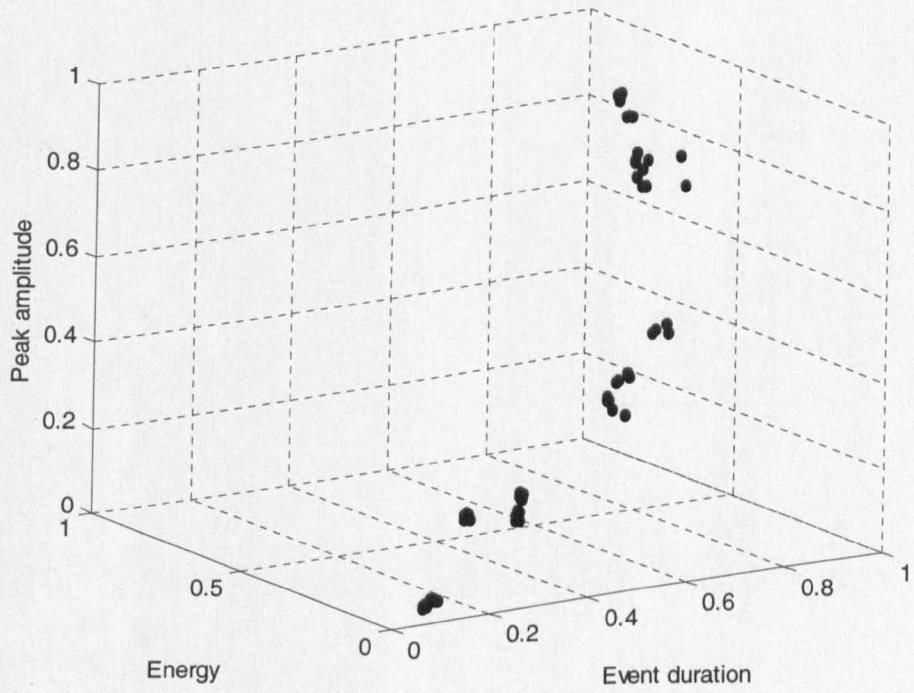


Figure 6.5 AE events for the four different machine conditions from the training sets

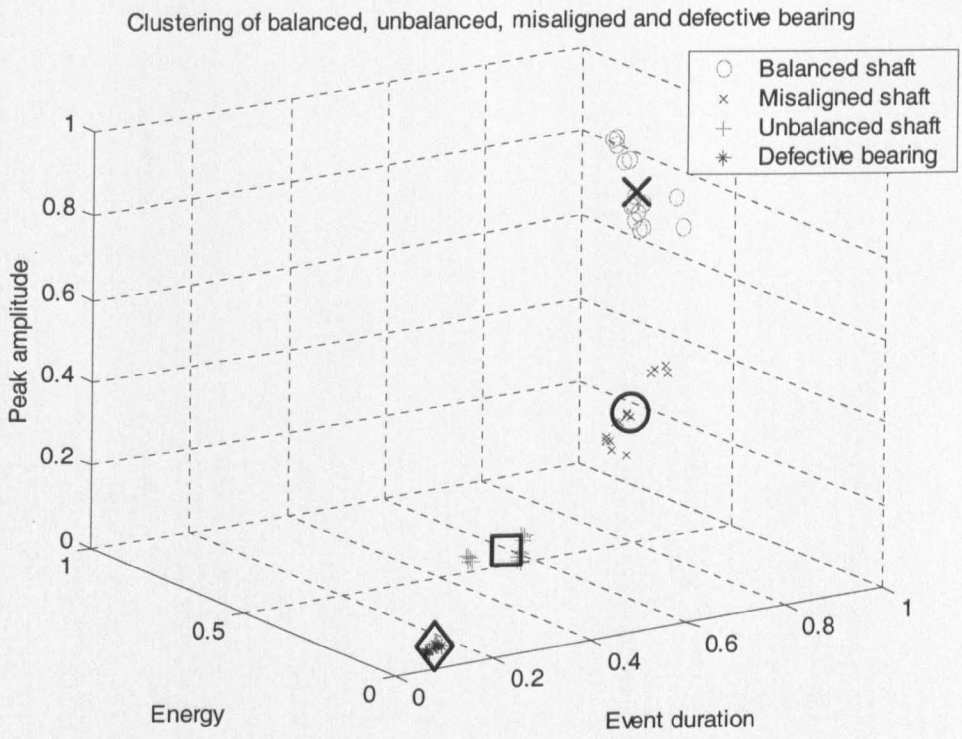


Figure 6.6 Established cluster centres using training data set from four different operating conditions

Figure 6.7 shows the results from the test set. They comprised the pre-processed 45 AE events in total for each four different conditions. A minimum distance classifier was then used for classification. It works by calculating the Euclidean distances of the AE event (expressed as a point in the three-dimensional space of event duration, peak amplitude and energy) from the centres of the established clusters for the four machine conditions, Figure 6.6. The AE event under test is considered to belong to the cluster whose centre is closest. In other words, maximum similarity corresponds to minimum distance.

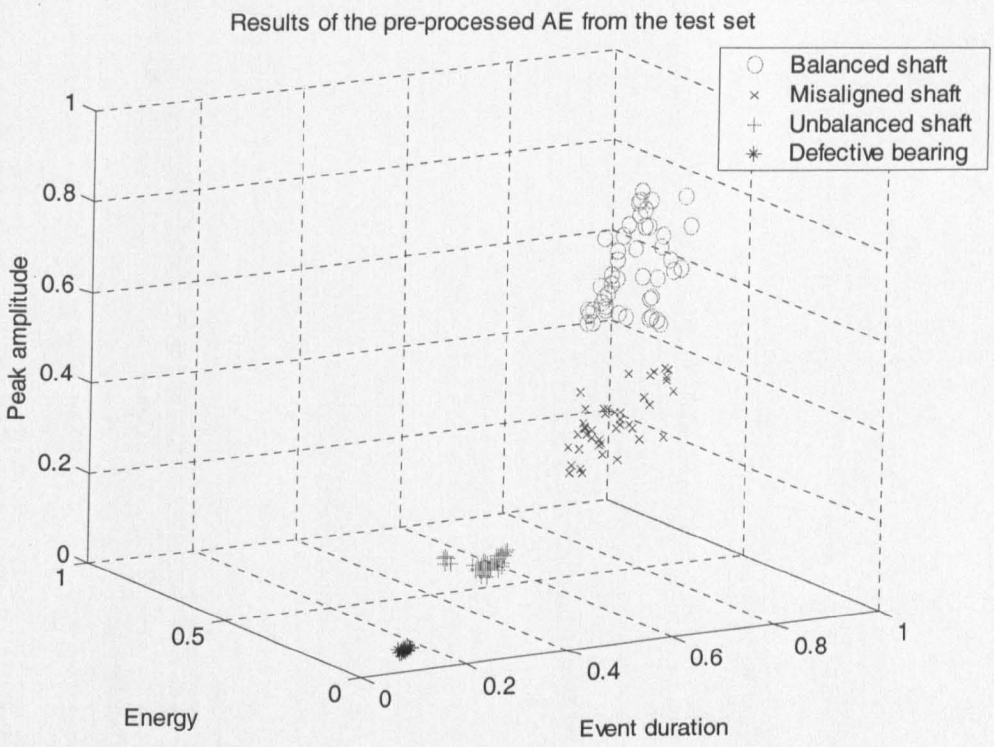


Figure 6.7 Three-dimensional graph of the pre-processed AE events from a test set

Classification results of 36 recordings (9 recordings x 4 conditions) from the test samples of all four types of machine conditions are given in Table 6.2. It shows the output values from the minimum distance classifier. The events were classified correctly 35 times out of 36, a recognition rate of 97 percent. The only error occurs when the unbalanced shaft condition was misclassified as that due to a defective bearing.

It can also be noted that the normal machine condition of 'balanced shaft' is very distinctive from the other three abnormal conditions, as their distance values (highlighted in the 'balanced shaft' column, Table 6.2) are all very small in comparison with others. In other words, this approach has very little risk of raising a false alarm.

Table 6.2 Minimum distance classification results

Bearing conditions	**Balanced shaft**	**Unbalanced shaft**	**Misaligned shaft**	**Defective bearing**
Testing set #1				
Balanced shaft	**1.47**	807.90	507.80	1162.30
Unbalanced shaft	670.60	**136.28**	164.10	490.60
Misaligned shaft	478.60	328.40	**28.07**	682.74
Defective bearing	861.81	**55.13**	355.41	299.32
Testing set #2				
Balanced shaft	**4.34**	810.90	510.81	1165.30
Unbalanced shaft	821.35	**15.16**	314.90	339.80
Misaligned shaft	516.47	290.31	**10.52**	644.71
Defective bearing	1003.8	197.11	497.41	**157.34**
Testing set #3				
Balanced shaft	**28.28**	778.76	478.58	1133.20
Unbalanced shaft	839.19	**32.51**	332.83	321.93
Misaligned shaft	468.16	338.71	**38.41**	693.09
Defective bearing	1002.30	195.59	495.80	**158.90**
Testing set #4				
Balanced shaft	**5.70**	802.85	502.82	1157.30
Unbalanced shaft	841.6	**35.47**	335.11	319.61
Misaligned shaft	400.91	405.84	**105.69**	760.26
Defective bearing	1038.90	232.27	532.50	**122.20**
Testing set #5				
Balanced shaft	**70.65**	736.04	435.92	1090.50
Unbalanced shaft	773.04	**33.84**	266.61	388.11
Misaligned shaft	421.84	384.97	**84.73**	739.38
Defective bearing	1031.20	224.54	524.80	**129.90**
Testing set #6				
Balanced shaft	**15.98**	790.71	490.61	1145.10
Unbalanced shaft	761.76	**44.95**	255.50	399.37
Misaligned shaft	372.86	433.85	**133.77**	788.28
Defective bearing	1028.20	221.56	521.71	**133.04**
Testing set #7				
Balanced shaft	**40.24**	766.44	466.34	1120.90
Unbalanced shaft	868.07	**61.40**	361.74	293.05
Misaligned shaft	391.57	415.12	**115.40**	769.55
Defective bearing	1222.80	416.14	716.40	**61.70**
Testing set #8				
Balanced shaft	**4.28**	804.99	504.82	1159.40
Unbalanced shaft	1046.90	**240.23**	540.50	114.21
Misaligned shaft	461.38	345.31	**45.76**	699.74
Defective bearing	1108.10	301.40	601.72	**53.21**
Testing set #9				
Balanced shaft	**2.77**	809.05	508.92	1163.50
Unbalanced shaft	905.94	**99.29**	399.50	255.20
Misaligned shaft	477.31	329.39	**29.99**	683.82
Defective bearing	1081.20	274.57	574.80	**79.90**

6.3 AE condition monitoring of bearing lubrication

One of the most important aspects of smooth bearing operation is the assurance that a film of lubricant such as grease is always present between bearing surfaces, so that friction and surface wear can be minimised. Contaminated grease or the lack of lubricant may lead to ineffective bearing operating or eventual breakdown of a machine. Therefore, it is also important to be able to monitor reliably the condition of lubrication. The proposed method of applying the fuzzy c-mean clustering technique to the pre-processed AE (as discussed in Section 4.4.2) was validated for lubricant condition monitoring at low operating speed.

The main purpose of this study was to categorise two bearing lubrication conditions: clean grease and contaminated grease. The source of contamination is the slag from the steel making process. Experiments were performed on a main bearing at Corus, Middlesborough. Attached to the bearing, the AE sensors captured elastic waves generated in the material which was subjected to stress or friction as is the case when a rolling element comes into contact with wear debris.

6.3.1 Experimental set up for lubrication condition monitoring

The equipment used for processing the AE signals was a LOCAN AT acoustic emission system. The AE transducers were R15I with an integral preamplifier and the transducers had a resonant frequency of 150 kHz and a measurement range of 60 – 600 kHz. The transducers were attached to the bearing using magnetic clamps with silicone grease as couplant between the transducer and the bearing housing. The bearing was FAG 515518K and the lubricating grease was of the Molub Alloy type. The bearing was run at the rotating speed of 0.5 rev/min.

6.3.2 Experimental results for lubrication condition monitoring

At each lubrication condition, twelve recordings, each of 30-second duration, were taken. These were then split into two groups of 4 and 8 recordings. The first group would be used for training; the second group for testing. Following the procedure in Section 4.4.1, only five AE events in each recording were selected, each event specified in terms of its event duration, energy and peak amplitude.

The fuzzy c-mean clustering technique was used for classification implemented by means of a MATLAB program. Its source code is given in Appendix C1.

The classification results are shown as a three-dimensional graph in Figure 6.8. It can be seen that the test results of 40 AE events (or 8 recordings with 5 AE events in each) were all classified correctly, a 100% recognition rate.

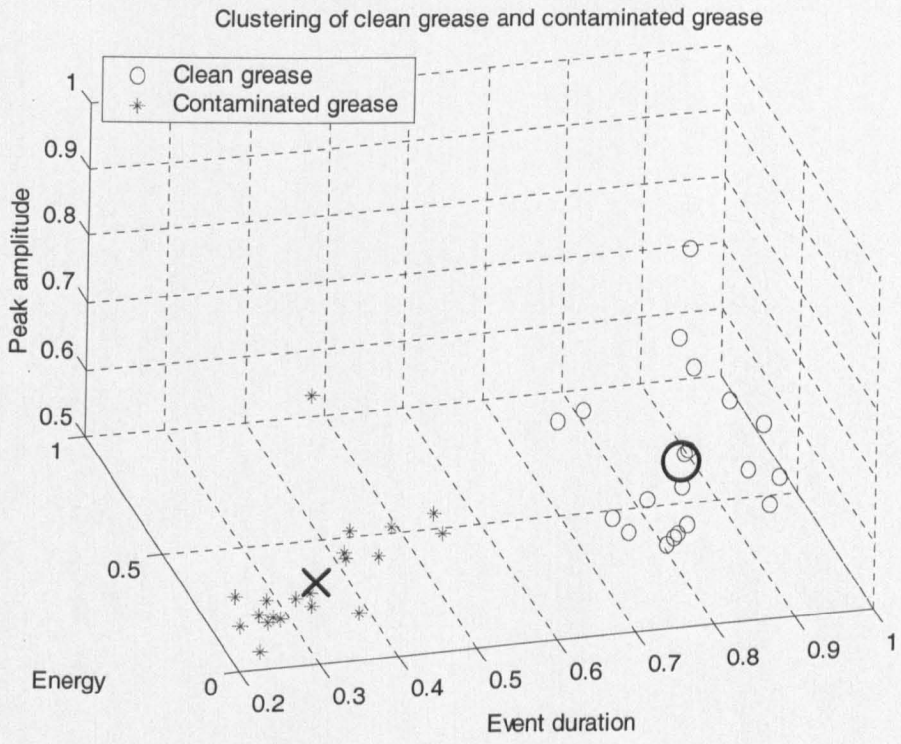


Figure 6.8 Classification results after applying the fuzzy c-mean clustering technique with AE events displayed in terms of the three AE parameters as axes of the three-dimensional graph.

6.4 AE condition monitoring using the STL method

An investigation of a novel method, shape-to-life (STL) ratio, was carried out on the two test rigs. The aim was to explore the use of STL over a range of rotation speeds from 20 rev/sec down to 0.23 rev/sec. Employing the two test rigs, the high-speed light duty and the low-speed heavy duty, can provide a larger range of operating conditions for loading and speed variations. The sensitivity, reliability and robustness of the STL method can be established from this investigation. Presentation in this section was divided into four parts. Firstly, the STL method was applied to four different machine conditions created on the high-speed light-duty test rig with a view to evaluate its discriminatory power. Secondly, the effect of bearing loading on STL and L_{63} was investigated using the low-speed heavy-duty test rig. Thirdly, progressive bearing wear was studied on the low-speed heavy-duty test rig to determine its effect on STL and L_{63}. Finally, the effect of speed variation on STL and L_{63} was determined on the low-speed heavy-duty test rig and an empirical equation was provided to facilitate comparison of STL's across different speeds.

6.4.1 Four faulty conditions on the high-speed light duty rig

The main purpose of the experiments was to study the performance of the STL method when applied to different conditions of machines operating at high speed. Four machine conditions were examined: a balanced shaft, a misaligned shaft, an unbalanced shaft and a defective bearing. For each condition, inter-arrival times of successive AE events were extracted and a distribution was formed. As proved in Section 4.5.2, this is a Weibull distribution characterised by three parameters, the guaranteed life t_o, characteristic life θ and shape γ. t_o is the threshold time corresponding to the shortest AE event duration plus the 'dead' zone specified to separate the end of one event from the next. In this work, L_{63}, being the sum of t_o and θ, is used because L_{63} denotes the time at which 63% of the inter-arrival times of AE events will have occurred and hence is more meaningful than θ. The STL is a ratio of the shape parameter to characteristic life, that is, γ/θ, as was explained in Section 4.5.3. The values of the three Weibull parameters of a distribution were then estimated using the method of parameters estimation as explained in Section 4.5.4.

6.4.1.1 Experimental set up for the STL method

Experiments were conducted on the high-speed light duty test rig, running at 20 rev/sec. Details of the set up are as described in Section 6.2.1. For each machine condition, ten recordings each of about 30-second duration were made at one-hour intervals. Each recording captured the times of arrival of successive AE events. The inter-arrival times of successive AE events were then determined from the times of arrival of successive AE events as noted in Section 4.5.2.

6.4.1.2 Experimental results for the STL method

Given a Weibull distribution from each recording, the Weibull parameters of life (θ), shape (γ) and guaranteed life t_o were estimated. Using $L_{63} = \theta + t_o$, L_{63} was then calculated. The estimated parameters for the four different machine conditions are shown in Table 6.3.

The shape parameter (γ) from Weibull distribution describes the pattern of inter-arrival times of AE events produced from microscopic failure. All four different machine conditions return a value of around 1 for the shape parameter, suggestive of a random failure pattern.

As L_{63} indicates the time below which 63% of the inter-arrival times will fall, the shorter the L_{63} interval, the more active the AE emission, and hence the greater the number of microscopic failure within a given time period. As can be seen from Figure 6.10, the balanced condition gives L_{63} = 0.25 second approximately. In contrast, the defective bearing condition gives L_{63} = 0.02 second approximately, being the lowest of all conditions. This means that when the test rig was operating under the defective bearing condition, the AE rate was the highest compared to other conditions, having 63% of the AE inter-arrival times within 0.02 second.

Table 6.3 STL and L_{63} values from high-speed light duty test rig at different machine conditions

Condition	Time (hr)	Life (s)	Shape	t_0 (s)	STL(s^{-1})	L_{63}(s)
Balanced shaft	1	0.2090	1.1650	0.0009	5.5742	0.2099
	2	0.2490	0.8050	0.0065	3.2329	0.2555
	3	0.2380	1.1930	0.0085	5.0126	0.2465
	4	0.2220	1.1620	0.0098	5.2342	0.2318
	5	0.2550	1.2000	-0.0083	4.7059	0.2467
	6	0.1980	1.1190	-0.0047	5.6515	0.1933
	7	0.2300	1.2200	-0.0099	5.3043	0.2201
	8	0.2380	1.4600	-0.0130	6.1345	0.2250
	9	0.2100	1.1360	-0.0130	5.4095	0.1970
	10	0.2020	1.1780	-0.0057	5.8317	0.1963
Unbalanced shaft	1	0.0630	1.7870	0.0015	28.3651	0.0645
	2	0.0690	1.3750	0.0059	19.9275	0.0749
	3	0.0660	2.2690	-0.0068	34.3788	0.0592
	4	0.0590	1.5500	0.0081	26.2712	0.0671
	5	0.0640	1.5050	0.0048	23.5156	0.0688
	6	0.0550	1.7090	0.0099	31.0727	0.0649
	7	0.0560	1.8320	0.0067	32.7143	0.0627
	8	0.0610	1.4860	0.0067	24.3607	0.0677
	9	0.0610	1.4960	0.0099	24.5246	0.0709
	10	0.0580	1.6470	0.0100	28.3966	0.0680
Misaligned shaft	1	0.1080	1.2370	0.0099	11.4537	0.1179
	2	0.0900	1.2580	0.0100	13.9778	0.1000
	3	0.1060	1.0020	0.0100	9.4528	0.1160
	4	0.1180	1.0480	0.0099	8.8814	0.1279
	5	0.1180	1.0310	0.0100	8.7373	0.1280
	6	0.1360	0.9900	0.0100	7.2794	0.1460
	7	0.1150	1.0610	0.0087	9.2261	0.1237
	8	0.1270	1.0260	0.0099	8.0787	0.1369
	9	0.1110	0.9880	0.0098	8.9009	0.1208
	10	0.0980	1.1440	0.0098	11.6735	0.1078
Defective bearing	1	0.0140	0.9780	0.0058	69.8571	0.0198
	2	0.0120	0.7720	0.0076	64.3333	0.0196
	3	0.0140	1.0330	0.0055	73.7857	0.0195
	4	0.0120	0.7220	0.0075	60.1667	0.0195
	5	0.0098	0.7160	0.0078	73.2481	0.0175
	6	0.0097	0.7670	0.0074	79.3257	0.0170
	7	0.0130	0.9450	0.0064	72.6923	0.0194
	8	0.0100	0.8140	0.0074	81.4000	0.0174
	9	0.0092	0.6380	0.0079	69.1899	0.0171
	10	0.0130	0.8390	0.0068	64.5385	0.0198

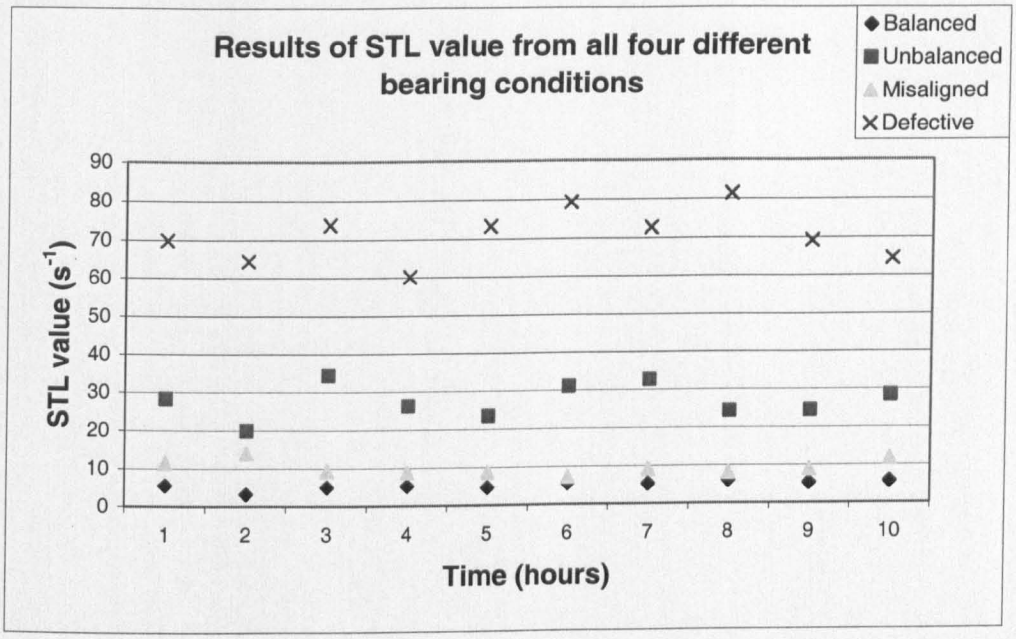


Figure 6.9 Progression of STL with time from high-speed light duty test rig at four different machine conditions

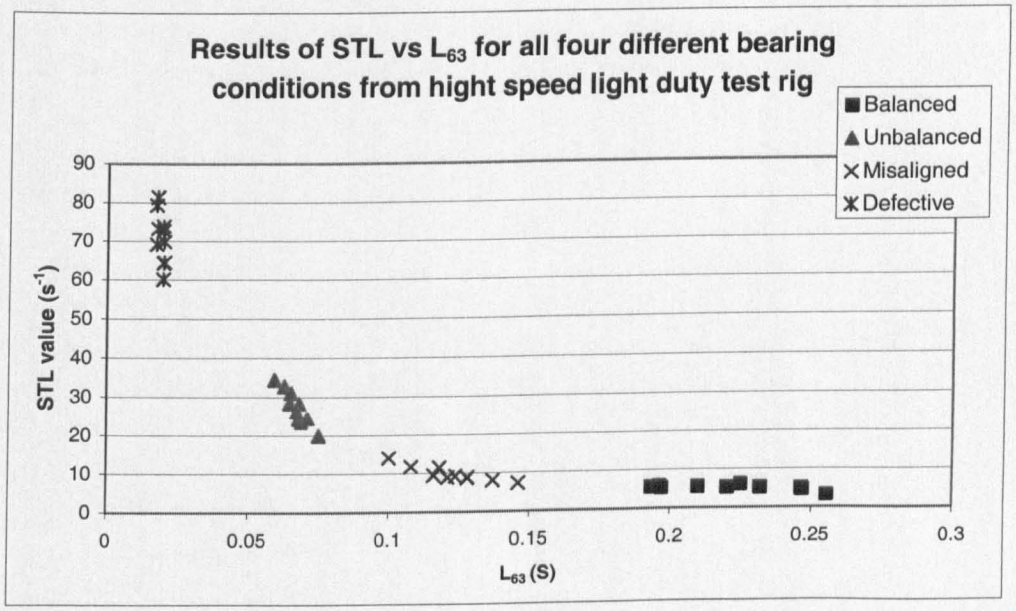


Figure 6.10 STL versus L_{63} values from high-speed light duty test rig at four different machine conditions

The STL values from all four machine conditions are plotted against the time of operation in Figure 6.9. As can be seen from the graph, the values of STL for each machine condition remained more or less constant throughout the whole period of the ten recordings. This may be due to the fact that the machine conditions did not change in that period of time or that the STL values were not sensitive to changes in machine conditions.

The means and standard deviations of STL values for the four machine conditions are tabulated in Table 6.4. It can be seen that their means are different with the defective bearing having the highest STL of 70.8 s^{-1} compared with 5.2 s^{-1} for the balanced shaft, a ratio of 13.6 to 1. Considering the fact that the defective bearing had a mere 1 mm indentation on the outer race, the difference is all the more remarkable. It can also be seen that the unbalanced shaft gives an STL of 27.4 s^{-1}, which is more than 5 times as large as that for the balanced shaft. The misaligned shaft gave a mean STL of 9.8 s^{-1}, which is just under 1.9 times the value of the balanced shaft.

Table 6.4 STL results from four different machine conditions

Condition	STL value (s^{-1})	
	Mean	**S.D**
(a) Balanced	5.2091	0.8041
(b) Unbalanced	27.3527	4.4974
(c) Misaligned	9.7662	2.0065
(d) Defective	70.8537	6.6757

6.4.2 Loading effects on the STL on the low-speed heavy duty rig

Experiments were carried out on the low-speed heavy duty test rig in order to determine the sensitivity of the STL value to the levels of load acting on the bearing. In many applications, bearings experience load variation, leading to failure of the bearing if a safe load is exceeded. The load was increased from an figure that a bearing can sustain until over the specified safe load as recommended by the manufacturer. The STL values were calculated from the Weibull distribution of the inter-arrival times of AE events and then correlated to each loading condition.

6.4.2.1 Experimental set up for loading effects on the STL

The configuration of the test rig was given in Section 5.4. The shaft was driven at 4.7 rev/sec. SKF 2206 ETN9 was chosen for the test because it can sustain static load up to 6700 N (59.23 bar). Loads were varied in steps from 0, 40, 60 to 80 bars. A high capacity hydraulic ram was used for exerting a variable radial load on the bearing. For each loading condition, ten recordings, each of 30-second duration, were made at hourly intervals.

The AET 5500 was used to provide the time of arrival of each AE event. A wideband AE transducer (UT1000) was fixed to the non-drive end housing, Figure 6.11. An AECL/2100 preamplifier with a 60 dB gain and band-pass filter at 100k – 1MHz was chosen. An automatic threshold was set at 0.1 Volt with a 2.51 post-amplified gain.

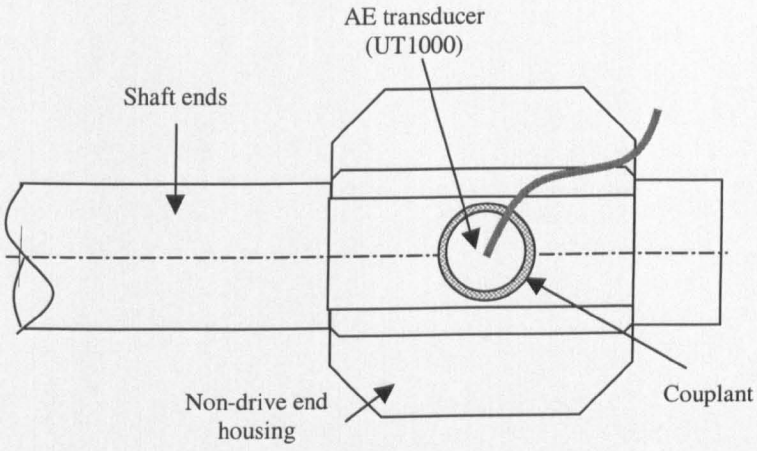


Figure 6.11 Schematic diagram showing the mounting position of an AE transducer (UT1000)

6.4.2.2 Experimental results for loading effects on the STL

Estimated Weibull parameters, characteristic life (θ), shape (γ) and guaranteed life (t_o) together with L_{63} and STL were calculated for the ten recordings obtained from each of the four levels of loads – 0 bar, 40 bars, 60 bars and 80 bars The results are shown in Table 6.5. The shape parameter obtained from the different loading conditions is around 1, which suggests an uniformly random pattern of inter-arrival time distribution. In contrast, the values of the life parameter were a function of the applied load. With no load, the life value, from 0.6 to 0.9 second, was the highest compared with the other loading conditions. With the load at 40 bars, the life values were in the range of 0.3 to 0.4 second, resulting in a slight increase in the STL value. With the load at 60 bars, just above the manufacturer's recommended loading limit of 59 bars, the life values decreased to the range of 0.17 to 0.2 second. When the bearing was loaded at 80 bars, an obvious overload, the shortest life values, from 0.084 to 0.122 second, were observed, resulting in the highest STL values.

Figure 6.12 shows the progression over a 10-hour period of the STL values under the four different levels of load. It is clear that the STL values increased with increasing loading – a pattern observed for all columns of points taken at any given time in Figure 6.12. It is also clear that for the loading of 80 bars, the STL curve showed an increasing trend with time whereas the other curves at lower loads remained more or

less level. The increasing trend at the 80 bar load seems to suggest progression of bearing wear and the horizontal trend at other loads indicates a static wear state.

Table 6.5 Estimated Weibull parameters and STL and L_{63} value at different loads on the low speed heavy duty test rig

Condition	Time (hr)	Life (s)	Shape	t_0 (s)	STL(s^{-1})	L_{63}(s)
No load	1	0.703	1.235	-0.0630	1.756757	0.64
	2	0.749	1.441	-0.0430	1.923899	0.706
	3	0.615	1.094	0.0100	1.778862	0.624986
	4	0.678	1.382	-0.0730	2.038348	0.605
	5	0.594	1.074	-0.0340	1.808081	0.56
	6	0.68	1.013	-0.0210	1.489706	0.659
	7	0.797	0.969	0.0066	1.215809	0.80355
	8	0.67	1.33	-0.0230	1.985075	0.647
	9	0.933	0.968	0.0032	1.037513	0.936209
	10	0.732	1.245	-0.1050	1.70082	0.627
Load at 40 bars	1	0.394	0.75	-0.0090	1.903553	0.385032
	2	0.324	1.088	-0.0310	3.358025	0.293
	3	0.453	1.282	-0.0086	2.830022	0.444405
	4	0.322	1.078	-0.0370	3.347826	0.285
	5	0.403	1.033	0.0097	2.563275	0.412692
	6	0.307	0.796	-0.0065	2.592834	0.300503
	7	0.328	1.042	-0.0250	3.176829	0.303
	8	0.45	1.467	-0.1050	3.26	0.345
	9	0.304	0.966	-0.0210	3.177632	0.283
	10	0.287	0.895	-0.0100	3.118467	0.277
Load at 60 bars	1	0.175	1.144	-0.0007	6.537143	0.174288
	2	0.188	1.04	0.0029	5.531915	0.19094
	3	0.202	1.135	0.0100	5.618812	0.211985
	4	0.186	1.089	0.0011	5.854839	0.187075
	5	0.199	1.072	0.0011	5.386935	0.20011
	6	0.21	1.066	-0.0009	5.07619	0.209141
	7	0.2	1.142	-0.0079	5.71	0.192079
	8	0.176	1.011	0.0028	5.744318	0.17881
	9	0.177	1.205	-0.0004	6.80791	0.176625
	10	0.19	1.059	-0.0005	5.573684	0.189458
Load at 80 bars	1	0.122	1.112	0.0100	9.114754	0.131974
	2	0.115	1.222	0.0100	10.62609	0.124999
	3	0.111	1.168	0.0061	10.52252	0.117124
	4	0.101	1.287	0.0100	12.74257	0.11099
	5	0.104	1.291	0.0008	12.41346	0.104849
	6	0.097	1.361	-0.0004	14.03093	0.096608
	7	0.091	1.418	0.0042	15.58242	0.095153
	8	0.095	1.456	-0.0046	15.32632	0.090439
	9	0.09	1.581	-0.0037	17.56667	0.086336
	10	0.084	1.629	-0.0005	19.39286	0.083453

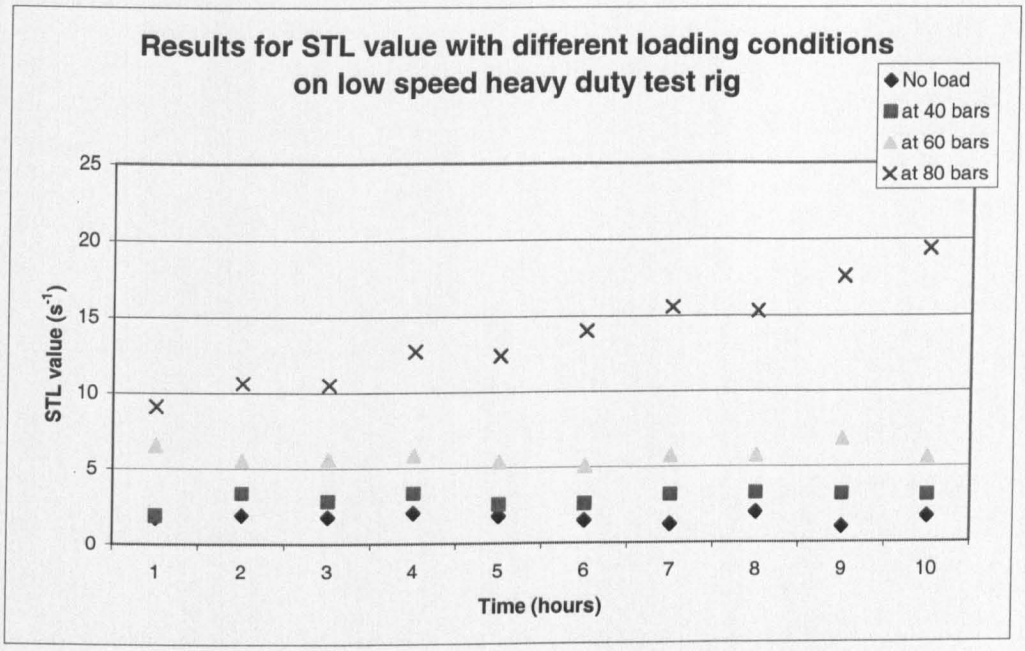


Figure 6.12 Progression of STL with time from all different loading conditions on low-speed heavy duty test rig

Figure 6.13 shows the relationship between the STL and L_{63} at different loads. The graph reveals that the higher the load the shorter is the L_{63} value and the greater the STL value. Not surprisingly, as the load increased the AE events, as an indication of microscopic failure, occurred more frequently, resulting in a shorter L_{63}. Since STL = γ/θ where γ, the shape parameter was observed to be close to unity and θ, the characteristic life, is not much different in value from L_{63}, the graph of STL versus L_{63} is close to a hyperbolic curve, Figure 6.13.

Figure 6.14 shows the progression of STL with times (1-hourly interval) under incremental loading conditions, Table 6.5. It is clear that the STL values are more or less constant when the bearing was subjected under the basic dynamic load rating (recommended load), 59.23 bars. The mean and standard deviation of STL of bearing operated under basic load rating is 2.3031 s^{-1} and 0.7554 s^{-1} respectively.

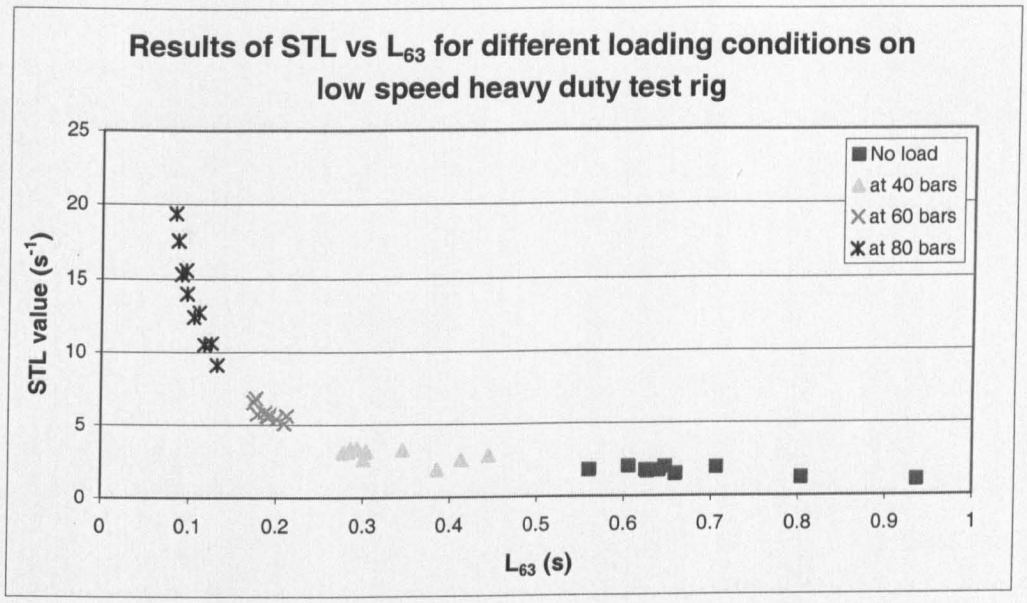


Figure 6.13 STL versus L_{63} values under incremental loading conditions on low speed heavy duty test rig

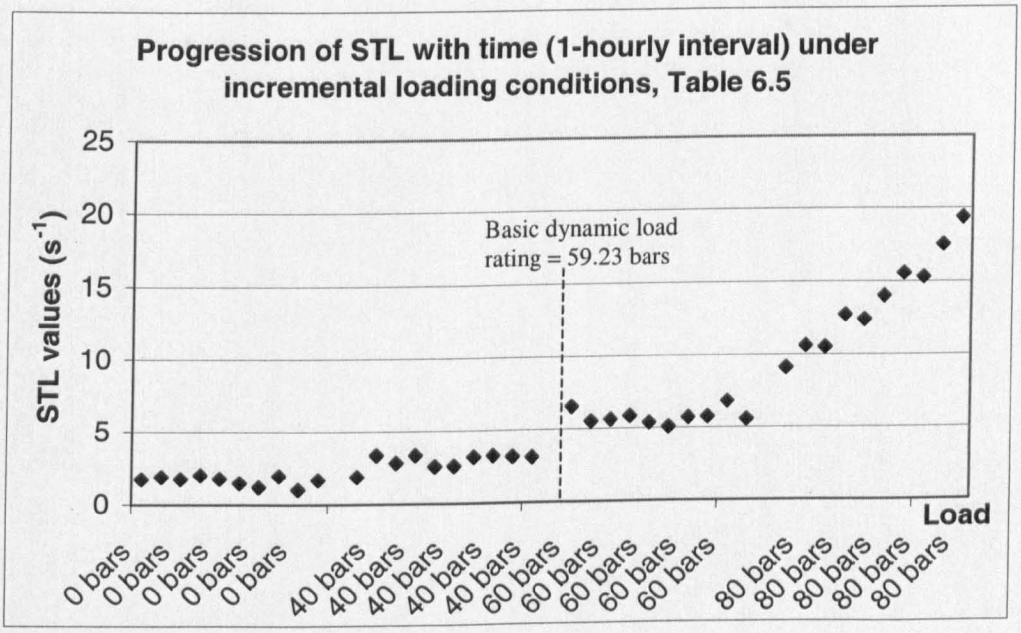


Figure 6.14 Progression of STL with time under incremental loading conditions

Table 6.6 shows the means and standard deviations of the STL values obtained for the four different loads over the 10-hour period. It is noted that the mean increases with increasing load and that the standard deviations for the loads of 0 bar, 40 bars and 60 bars are rather small, being 0.32 s^{-1}, 0.46 s^{-1} and 0.51 s^{-1} the corresponding means. For the load of 80 bars, strictly speaking the standard deviation of 3.27 s^{-1} has very little meaning because of the increasing trend of the STL.

Table 6.6 Means and standard deviations under incremental loading conditions

Loading conditions	STL value (s^{-1})	
	Mean	S.D
No load	1.673487	0.329566
40 bars	2.932846	0.463816
60 bars	5.784175	0.518696
80 bars	13.73186	3.279382

6.4.3 Progressive bearing wear on the low-speed heavy duty rig

An accelerated bearing life test was carried out in order to establish whether the STL method is sensitive even for low speed bearings experiencing progressive wear. The life of a bearing was shortened by subjecting it to a radial load higher than recommended by the manufacturer. This is to facilitate observation of the different stages of bearing wear leading to functional failure within a shorter time scale. The radial load was produced by a hydraulic ram with the accompanying hydraulic circuit that gives an accurate pressure to the ram. For comparison, the root-mean-square value of the acceleration obtained from the bearing was also produced. All tests were performed on the low-speed heavy duty test rig.

6.4.3.1 Experimental set up for progressive bearing wear

The low-speed heavy duty test was run at 0.23 rev/sec. The bearing under test was an SKF 1206E, which is a self-aligned ball bearing with a maximum load capacity of about 137 bars. AE signals were captured using the WD (wideband) transducer mounted on the top of the non-drive end bearing housing. These signals were amplified with a 60 dB gain and filtered with a 100 kHz – 450 kHz band-pass filter. The sampling rate was 1 MHz. An accelerometer (PCB 333A12) was used to measure vibration. It was located next to the AE transducer on the non-drive end housing. The acceleration signals were amplified by 60 dB and filtered using the KEMO filter set to operate as a 1000 Hz low-pass filter. The conditioned signals were then acquired into a PC using the LABVIEW program with a sampling rate of 2000 Hz. Figure 6.15 illustrates the positions of the two transducers.

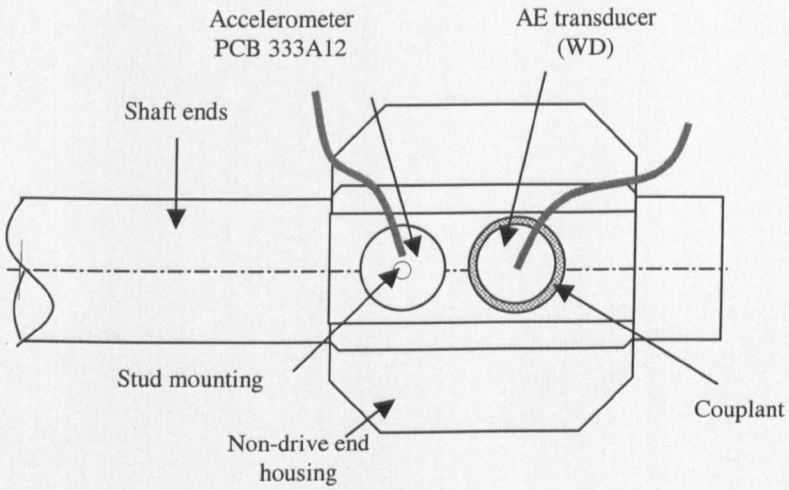


Figure 6.15 Schematic diagram of the mounting positions of the AE transducer and the accelerometer

Measurements started with no radial load applied to the test bearing and the load then was increased at 50-bar steps up to 300 bars. From the loads of 0 to 250 bar, each loading condition was maintained for about 2 hours, thus taking about 12 hours to reach the end of the 250 bars test. Then the load was increased to 300 bars and maintained until the test bearing failed.

6.4.3.2 Experimental results for progressive bearing wear

Tables 6.7 and 6.8 show the computed results from the AE and acceleration signals respectively. It can be seen from the two tables that when a test bearing was run under the load of 200 bars, which is more than the design load capacity (137 bars), the STL values increase abruptly whereas the V_{RMS} remains unchanged.

The total time taken for the test bearing to fail was about 126 hours. This is in good agreement with the expected service life of an SKF 1206E calculated as follows (Eschmann, 1958; Peter, 1998).

Basic dynamic load rating , C = 15600 N

$L = (C/P)^k$ million revolutions

where L is the service life of a bearing, C is the basic dynamic load rating, P (300 bars) is the applied radial load and k, the index, is taken to be 3 for a ball bearing.

$L = 10^6 \times (15600\ N / 34023.3\ N)^3$ revolutions

$L = 0.096 \times 10^6$ revolutions

$L = (0.096 \times 10^6)/(13.8\ rev/min \times 60)$ hours

$L = 107.10$ hours (compared with 126 hours in practice)

This agreement is all the more remarkable if it is recognised that the 126 hours to failure includes 12 hours during which the applied load was less than 300 bars.

Figures 6.17 and 6.18 show the respective trend plots of the STL and V_{RMS} against time. The curve in Figure 6.17 suggests that the STL values increased when the bearing was overloaded, and with the load maintained indefinitely (300 bar load from the 12^{th} to the 120^{th} hour), the STL increased with progressive bearing wear until the final failure. At the 300 bar load, the STL started with a value of 18.9 and increased monotonically to 59.4 when it failed, representing just over a three-fold increase. Over this period of time L_{63}, as shown in Figure 6.19, fell from 0.0546 second down to 0.0177 second, the latter being just under a third of the former. Given that STL = γ/θ, γ is approximately unity and $\theta = L_{63} - t_o$ where t_o is small compared to L_{63}, it is to be expected that the increase in STL and the decrease in L_{63} over the same period of loading are in inverse proportion to each other. Like Figure 6.13, this is yet another demonstration of the hyperbolic relation that exists between STL and L_{63}. In general, it can be asserted that if γ is approximately unity and t_o is small compared to L_{63}, the hyperbolic relationship between STL and L_{63} holds.

In contrast, the V_{RMS} value of the acceleration signal in Figure 6.18 does not show any change either with the load or with the progression of bearing wear. It only rose after the bearing had failed.

Figure 6.20 shows a graph of the STL against L_{63} for all levels of load applied sequentially to the bearing during the accelerated life test. Like Figure 6.10 and 6.13, the hyperbolic relationship between STL and L_{63} is once again demonstrated in Figure 6.20.

At the point of failure, not only did the bearing seize up but the shaft also broke at a plane next to the bearing as shown in Figure 6.16. The shaft failure was characteristic

of fatigue failure caused by the repeated action of a cyclical shear stress created by the radially downward load applied to the shaft.

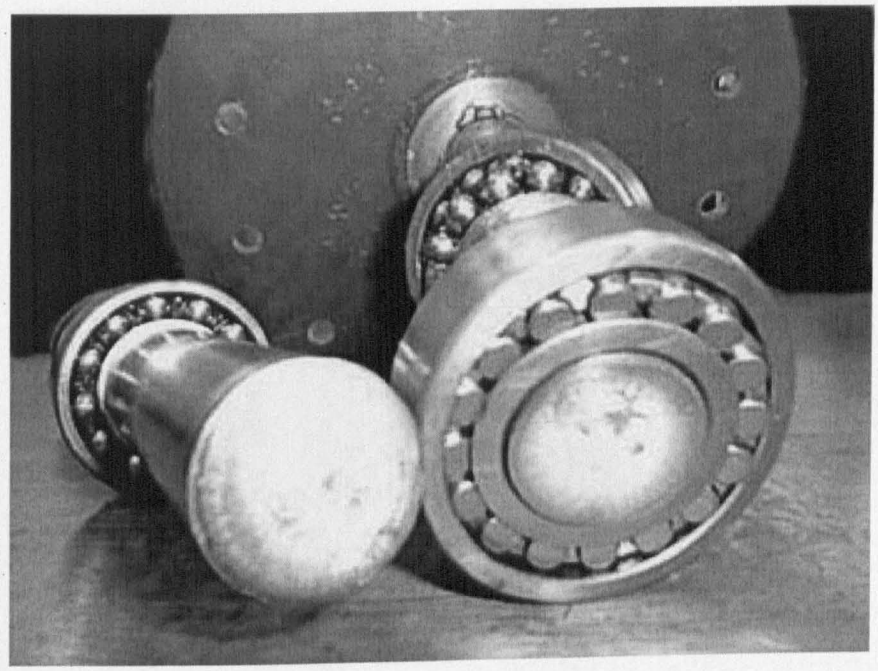


Figure 6.16 Ceased bearing and broken shaft due to fatigue failure

As the shaft broke, the bearing ceased to rotate and hence the number of AE events dropped. Consequently L_{63} increased, bringing down the value of STL.

On the other hand, the V_{RMS} of the acceleration signal increased after the shaft broke because the ensuing vibration on the drive-end bearing grew so large that it shook the base-plate of the whole test rig. Although the accelerometer was attached to the non-drive end bearing housing, the vibration on the base plate was transmitted to this bearing housing and was picked up by the accelerometer.

Table 6.7 STL and L_{63} from progressive bearing wear on the low-speed heavy duty test rig

Load (bars)	Time (hr)	Life (s)	Shape	t_0 (s)	STL(s^{-1})	L_{63}(s)
0	0	0.5340	1.1040	0.0002	2.0674	0.5342
0	1	0.2740	0.9370	0.0001	3.4197	0.2741
50	2	0.8670	1.0890	-0.0051	1.2561	0.8619
50	3	0.8920	1.1900	-0.0720	1.3341	0.8200
100	4	0.4570	0.9480	-0.0460	2.0744	0.4110
100	5	0.3240	0.7740	-0.0023	2.3889	0.3217
150	6	0.4540	1.0850	-0.0330	2.3899	0.4210
150	7	0.3220	1.0620	-0.0170	3.2981	0.3050
200	8	0.0640	1.1990	-0.0029	18.7344	0.0611
200	9	0.0650	0.9990	0.0001	15.3692	0.0651
250	10	0.0400	0.8950	0.0001	22.3750	0.0401
250	11	0.0360	0.9520	-0.0004	26.4444	0.0356
300	12	0.0560	1.0570	-0.0014	18.8750	0.0546
300	24	0.0420	1.0550	0.0001	25.1190	0.0421
300	36	0.0430	1.0130	-0.0010	23.5581	0.0420
300	48	0.0420	1.0190	-0.0007	24.2619	0.0413
300	60	0.0390	1.0860	-0.0016	27.8462	0.0374
300	72	0.0350	1.0670	-0.0021	30.4857	0.0329
300	84	0.0320	0.9980	-0.0004	31.1875	0.0316
300	96	0.0260	0.9870	-0.0009	37.9615	0.0251
300	108	0.0230	1.1070	-0.0017	48.1304	0.0213
300	120	0.0190	1.1290	-0.0013	59.4211	0.0177
300 Damaged	132	0.0980	0.8760	-0.0032	8.9388	0.0948
300 Damaged	138	0.1430	1.0240	-0.0092	7.1608	0.1338
300 Damaged	144	0.1400	0.8910	-0.0017	6.3643	0.1383
300 Damaged	150	0.1250	1.1570	-0.0077	9.2560	0.1173

Table 6.8 V_{RMS} of acceleration signals from progressive bearing wear on the low-speed heavy duty test rig

Load (bars)	Time (hr)	V_{RMS} (V)
0	0	0.557
0	1	0.483
50	2	0.436
50	3	0.427
100	4	0.475
100	5	0.449
150	6	0.415
150	7	0.415
200	8	0.418
200	9	0.419
250	10	0.435
250	11	0.386
300	12	0.408
300	24	0.350
300	36	0.425
300	48	0.427
300	60	0.400
300	72	0.385
300	84	0.395
300	96	0.394
300	108	0.415
300	120	0.406
300 (Damaged)	132	2.998
300 (Damaged)	138	2.911
300 (Damaged)	144	2.914
300 (Damaged)	150	2.927

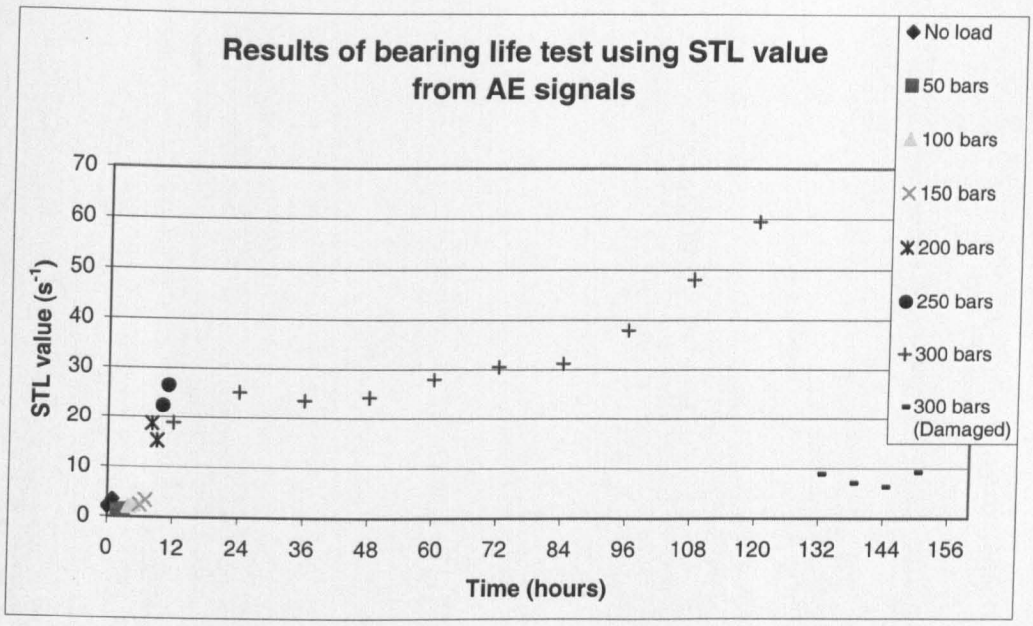


Figure 6.17 Progression of STL with time from the bearing life test on low-speed heavy duty test rig

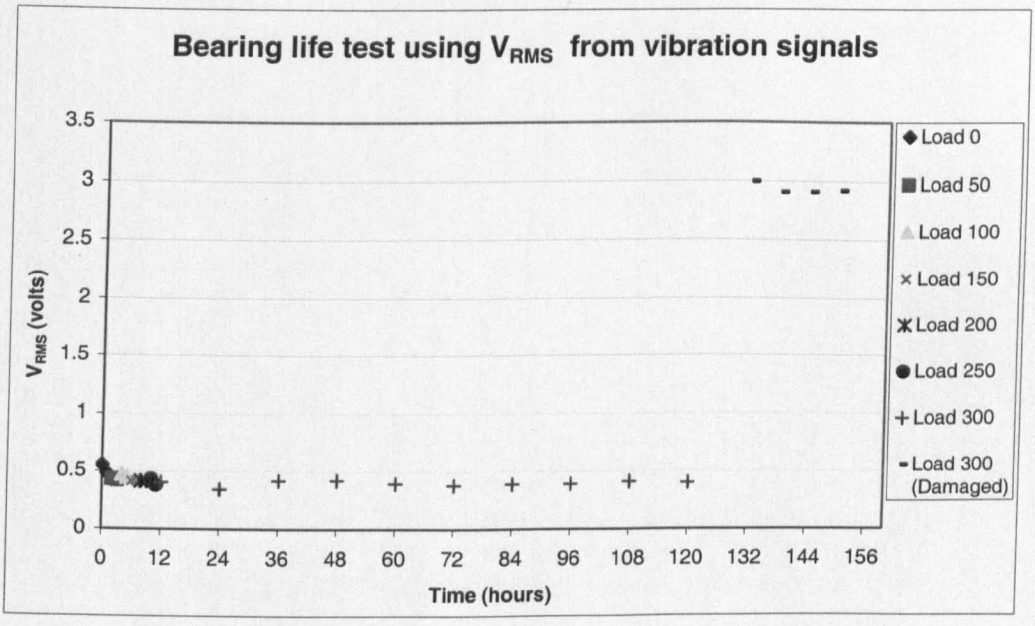


Figure 6.18 Progression of V_{RMS} of acceleration signals with time from bearing life test on low-speed heavy duty test rig

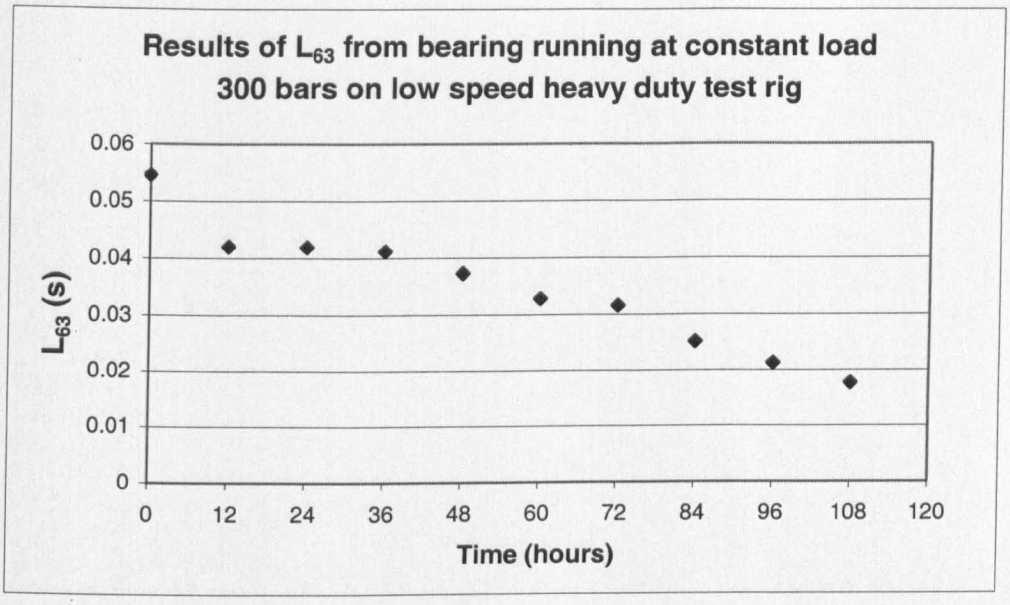


Figure 6.19 L_{63} values from the bearing life test at the constant load 300 bars

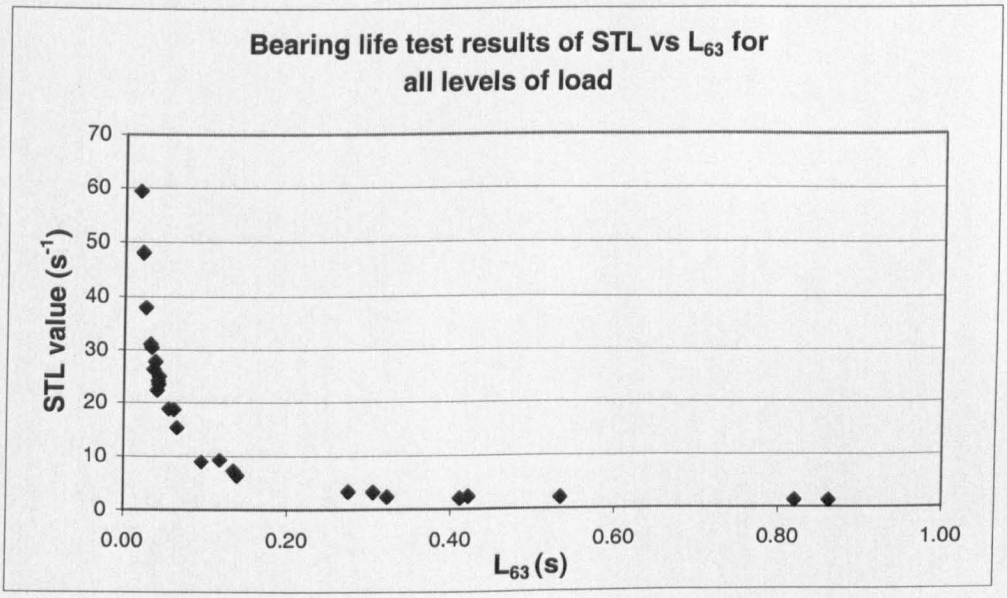


Figure 6.20 STL versus L_{63} in the bearing life test; the graph shows results for all levels of load

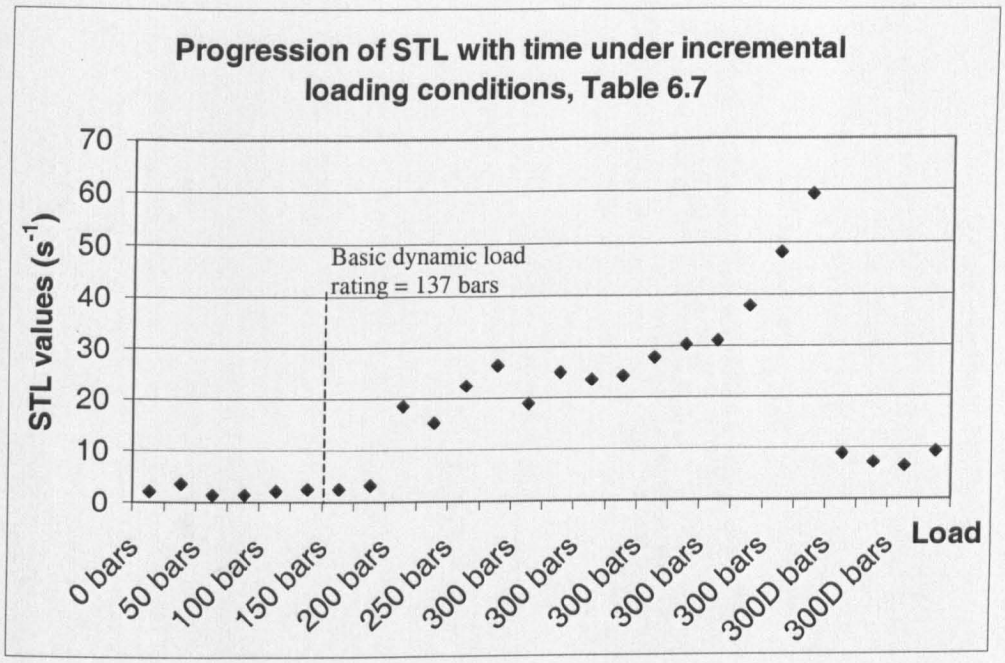


Figure 6.21 Progression of STL with time under incremental loading conditions from bearing life test

Figure 6.21 shows the progression of STL with time under incremental loading conditions from bearing life test, Table 6.7. It is obvious that with the applied load within basic load rating the STL values are more or less level. The mean and standard deviation of STL within basic dynamic load rating are 2.0901 and 0.7905 respectively.

Like Figure 6.14, the constant trend of progressive STL, when bearing was subjected under basic dynamic load rating as shown in Figure 6.21, remained. It should be noted that the results shown in Figure 6.14 and 6.21 were obtained from different bearings in terms of their load carrying capacity. Therefore, it is worthwhile considering the test of significance (Moroney, 1951) for STL obtained from the two different bearings running under recommended load.

Denoting the STL from SKF 2206 ETN9 (Figure 6.14) by subscript 1 and the STL from SKF 1206 E (Figure 6.21) the subscript 2, then

N_1=20, μ_1 = 2.3031, σ_1 = 0.7554, and σ_1^2 = 0.5706

N_2=6, μ_2 = 2.0901, σ_2 = 0.7905, and σ_2^2 = 0.6248

where N = sample size,

μ = sample mean

σ = standard deviation (S.D), and

σ^2= variance

The difference between the sample mean is $\mu_1-\mu_2 = 0.213\ s^{-1}$

and the corresponding variance of the difference is

$Var(\mu_1-\mu_2) = \sigma_1^2/N_1 + \sigma_2^2/N_2 = 0.1326$

from which the standard error of the difference is

Std. Error of Diff. = 0.3642

since it is the square root of the variance of the difference.

The observed difference between the sample mean of STL, 0.213 s^{-1}, is not significant, being 0.213/0.3642 = 0.58 times its standard error.

6.4.4 Effects of speed variations on the low-speed heavy duty test rig

The aim of this series of experiments was to investigate the effect of speed variations on the STL value on the low-speed heavy duty test rig. For an AE-producing defect on, for example, the inner race of a bearing, as the shaft rotating speed doubles, the rolling elements in a bearing go twice as fast and so the intervals between the successive AE events become halved. This will affect the STL as well as the L_{63}. The experiments were performed on a good bearing without load in order to eliminate the effects of load and progressive bearing wear on the STL value.

6.4.4.1 Experimental set up for speed variation effects

The low-speed heavy duty test rig was used with the same configuration as that presented in Section 5.4. The test bearing was a SKF 1206E. The rotating speed was varied using an inverter in steps from 9.46, 14.2 to 28.4 rev/min. At each speed, ten recordings, about 30-second each, of the AE signal were made. A wideband AE acoustic transducer (WD) was fixed, using silicon-gel, to the top of the non-drive end housing. An AECL/2100 preamplifier with a 60 dB gain and a band-pass filter at 100k – 450kHz were chosen. The sampling rate used to capture the AE signal into a PC was 1 MHz.

6.4.4.2 Experimental results for speed variation effects

Weibull parameters were estimated from the distribution of the AE inter-arrival times for the three different speeds and are as shown in Table 6.9. The values of the shape parameter for all speeds are around 1. However, the life values are a function of the rotating speed. At 9.46 rev/min, the life values are higher than those obtained from faster speeds, resulting in the smallest STL value of around 2. When the speed is increased to 14.2 rev/min, the life value becomes smaller and the STL rises to around 6. As expected, the STL value from the bearing running at 28.4 rev/min yields the maximum value of around 16.

Figure 6.22 illustrates the relationship between the STL and L_{63} values for all three speeds. It is obvious that at the low speed the L_{63} values are greater and have a wider spread. At the higher speed, the L_{63} becomes smaller and tends to spread less. Figure

6.23 shows the empirical equation that fits the data points in Figure 6.22. The equations is

$$STL = 0.6031 L_{63}^{-1.1843} \qquad \textbf{Equation 6.1}$$

It is noted that the exponent for L_{63} is -1.1843. Had it been -1, then the true hyperbolic relationship between STL and L_{63} could have been claimed.

Table 6.9 STL, L_{63} and estimated Weibull parameters with speed variations

Condition	File no.	Life (s)	Shape	t_0 (s)	STL(s^{-1})	L_{63}(s)
28.4 RPM	1	0.0900	0.9460	-0.0030	10.5111	0.0870
	2	0.0900	1.0140	-0.0043	11.2667	0.0857
	3	0.0980	1.0750	-0.0056	10.9694	0.0924
	4	0.0790	1.1850	-0.0077	15.0000	0.0713
	5	0.0800	0.9440	0.0002	11.8000	0.0802
	6	0.0760	1.2360	-0.0081	16.2632	0.0679
	7	0.0740	0.9610	-0.0050	12.9865	0.0690
	8	0.0770	1.0350	-0.0063	13.4416	0.0707
	9	0.0640	0.8870	-0.0031	13.8594	0.0609
	10	0.0810	0.8370	-0.0016	10.3333	0.0794
14.2 RPM	1	0.1340	1.0610	0.0012	7.9179	0.1352
	2	0.1390	0.9740	0.0001	7.0072	0.1391
	3	0.1410	1.0170	0.0009	7.2128	0.1419
	4	0.1790	1.0270	-0.0084	5.7374	0.1706
	5	0.1520	0.9490	0.0001	6.2434	0.1521
	6	0.1710	0.8220	0.0022	4.8070	0.1732
	7	0.1400	0.8590	0.0001	6.1357	0.1401
	8	0.1160	0.7780	0.0002	6.7069	0.1162
	9	0.1160	0.7040	0.0110	6.0690	0.1270
	10	0.1510	1.0040	-0.0038	6.6490	0.1472
9.46 RPM	1	0.2710	0.8690	-0.0026	3.2066	0.2684
	2	0.2610	1.0470	-0.0055	4.0115	0.2555
	3	0.2590	0.8810	-0.0011	3.4015	0.2579
	4	0.3030	0.8870	-0.0008	2.9274	0.3022
	5	0.2800	0.6280	-0.0020	2.2429	0.2780
	6	0.2670	0.6920	-0.0029	2.5918	0.2641
	7	0.2820	0.6120	-0.0020	2.1702	0.2800
	8	0.3220	0.6150	-0.0078	1.9099	0.3142
	9	0.2750	0.7230	-0.0049	2.6291	0.2701
	10	0.2260	0.6480	-0.0021	2.8673	0.2239

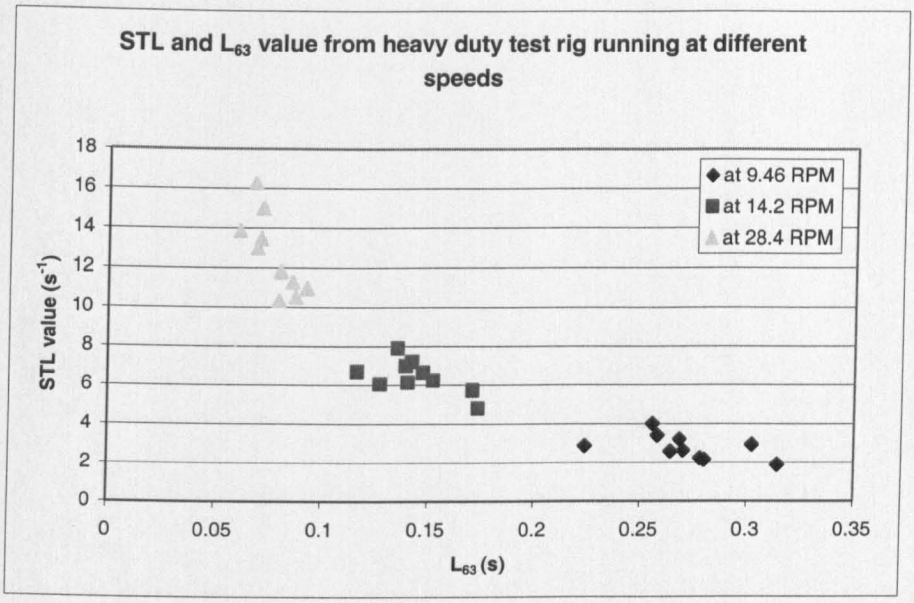


Figure 6.22 STL versus L_{63} values from bearing running at different speeds

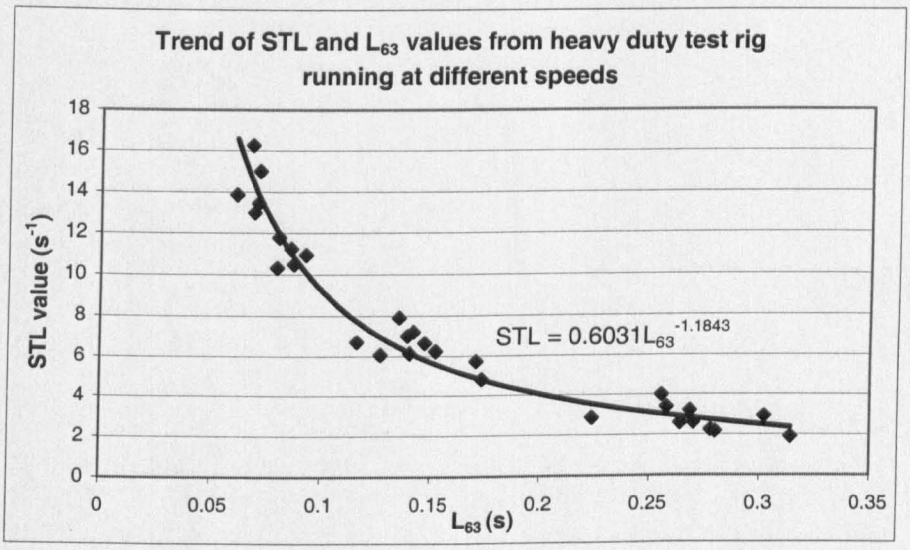


Figure 6.23 Trend of STL versus L_{63} values with bearing running at different speeds

6.4.4.3 Simulation study of speed variation

This section sets out to prove the hypothesis that the rotating speed and inter-arrival time between AE events are inversely proportional to each other. Once this is proven, the STL's of a bearing rotating at two different speeds can be compared.

Consider the bearing in Figure 6.24, where a defect is located on the outer race. An AE event is produced whenever a rolling element passes over the defect, giving an AE signal (top right-hand diagram). If the rotating speed is now halved, the frequency of AE events is doubled (middle right-hand diagram). If the rotating speed is reduced to a third, the frequency is trebled (bottom right-hand diagram).

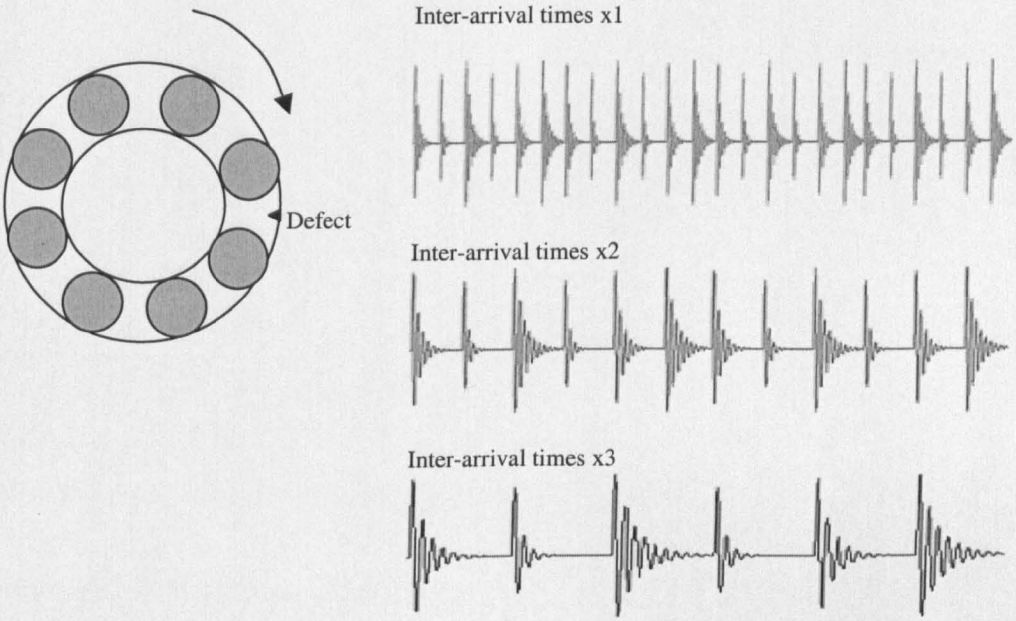


Figure 6.24 Examples of inter-arrival times and its multiplications

From the experiments described in Section 6.4.4.2, it was noted that the STL and L_{63} values are approximately inversely related. If the inter-arrival times of AE events obtained at a particular speed are chosen and their intervals doubled, the resulting Weibull distribution will produce its corresponding shape parameter, guaranteed and characteristic lives. From these, the STL and L_{63} can be computed using Equations 4.15 and 4.16 in Section 4.5.

Similarly, values of STL and L_{63} can be obtained for the case of trebling the inter-arrival time intervals. If these pairs of STL and L_{63} are then plotted on the graph of Figure 6.22 alongside the results of the speed variation experiments, and if they fit in well with each other, then it can be concluded that doubling the speed is equivalent to halving the inter-arrival time. In other words, speed and inter-arrival time are inversely related.

There is a simple relationship in the Weibull distribution between one set of inter-arrival times and the new set created from doubling those in the original set. It is that both distributions have exactly the same value for the shape parameter γ; but the guaranteed life t_o and characteristic life θ will have values twice as large as before the time dilation.

That this is the case is not hard to prove. Consider the cumulative distribution function of Weibull distribution,

$$F(t) = 1 - e^{\left[-\left(\frac{t-t_o}{\theta}\right)^{\gamma}\right]}$$

The doubling of inter-arrival times t can be regarded as a change in the time unit with the new unit being half of the old one. Clearly, changing the unit of measurement of a variate should not affect the shape of the variate's distribution.

A simulation study was conducted in which the inter-arrival times of AE events at the speed of 28.4 rev/min were chosen to be the reference set from which the time dilation was carried out. The inter-arrival times are stretched to twice and three times their original lengths. If the hypothesis is true, then the two stages of stretching are equivalent to the respective rotating speeds of 14.2 rev/min and 9.46 rev/min.

Table 6.10 shows the Weibull parameters obtained from the simulation experiments, each resulting in a data file. These parameters were estimated from the distributions of inter-arrival times at the three simulated speeds. It is noted that the shape parameter, γ, has a value that remains the same for data from the same file irrespective of the simulated speed. For example, File 1 gives $\gamma = 0.9460$ across all three speeds.

It is also noted that the guaranteed life t_o and characteristic life θ increase in ratio to the dilation of the inter-arrival time.

Figure 6.25 is a graphical depiction of the results shown in Table 6.10. Figure 6.26 is similar to Figure 6.25 but in addition, it shows a curve fitted to the data points. The formula for this curve is given as

$$STL = 0.9584 L_{63}^{-0.9948} \qquad \textbf{Equation 6.2}$$

Note that the exponent of L_{63} is very close to -1, thus suggesting a strong hyperbolic relationship between STL and L_{63}.

Table 6.10 Simulated results of STL and L_{63} on speed variation conditions

Inter-arrival time spacing	File number	Estimated Weibull parameters			STL(s^{-1})	L_{63}(s)
		Life (s)	Shape	t_0 (s)		
At 28.4 RPM	1	0.0900	0.9460	-0.0030	10.5111	0.0871
	2	0.0900	1.0140	-0.0043	11.2667	0.0857
	3	0.0980	1.0750	-0.0056	10.9694	0.0924
	4	0.0790	1.1850	-0.0077	15.0000	0.0714
	5	0.0800	0.9440	0.0002	11.8000	0.0802
	6	0.0760	1.2360	-0.0082	16.2632	0.0679
	7	0.0740	0.9610	-0.0050	12.9865	0.0690
	8	0.0770	1.0350	-0.0064	13.4416	0.0707
	9	0.0640	0.8870	-0.0031	13.8594	0.0609
	10	0.0810	0.8370	-0.0016	10.3333	0.0795
At 28.4 RPM x 2	1	0.1800	0.9460	-0.0059	5.2556	0.1741
	2	0.1810	1.0140	-0.0086	5.6022	0.1724
	3	0.1960	1.0750	-0.0110	5.4847	0.1850
	4	0.1570	1.1850	-0.0150	7.5478	0.1420
	5	0.1600	0.9440	0.0003	5.9000	0.1603
	6	0.1530	1.2360	-0.0160	8.0784	0.1370
	7	0.1470	0.9610	-0.0100	6.5374	0.1370
	8	0.1540	1.0350	-0.0130	6.7208	0.1410
	9	0.1270	0.8870	-0.0062	6.9843	0.1208
	10	0.1620	0.8370	-0.0031	5.1667	0.1589
At 28.4 RPM x 3	1	0.2690	0.9460	-0.0089	3.5167	0.2601
	2	0.2710	1.0140	-0.0130	3.7417	0.2580
	3	0.2940	1.0750	-0.0170	3.6565	0.2770
	4	0.2360	1.1850	-0.0230	5.0212	0.2130
	5	0.2400	0.9430	0.0005	3.9292	0.2405
	6	0.2290	1.2360	-0.0240	5.3974	0.2050
	7	0.2210	0.9610	-0.0150	4.3484	0.2060
	8	0.2310	1.0350	-0.0190	4.4805	0.2120
	9	0.1910	0.8870	-0.0093	4.6440	0.1817
	10	0.2430	0.8370	-0.0047	3.4444	0.2383

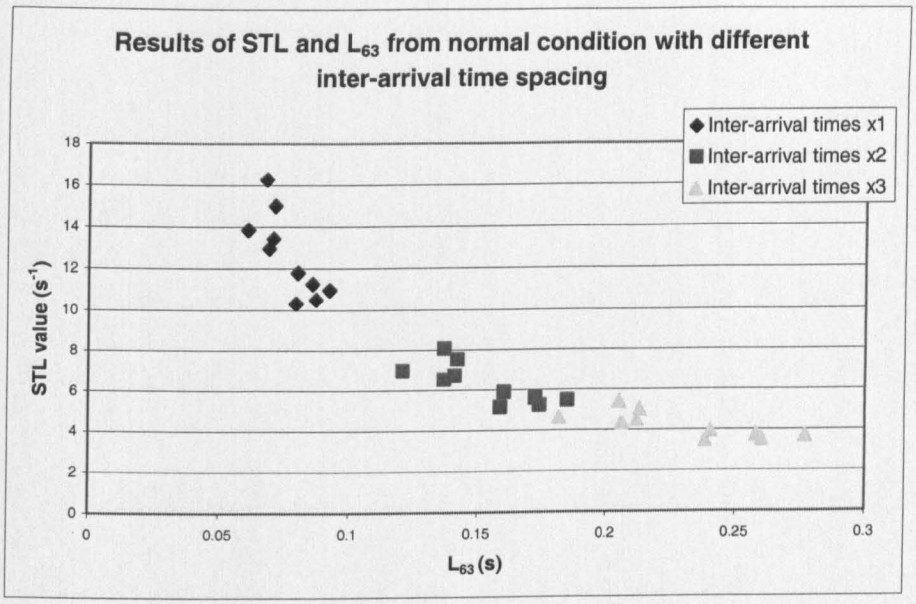


Figure 6.25 STL versus L_{63} from simulated different inter-arrival time spacing

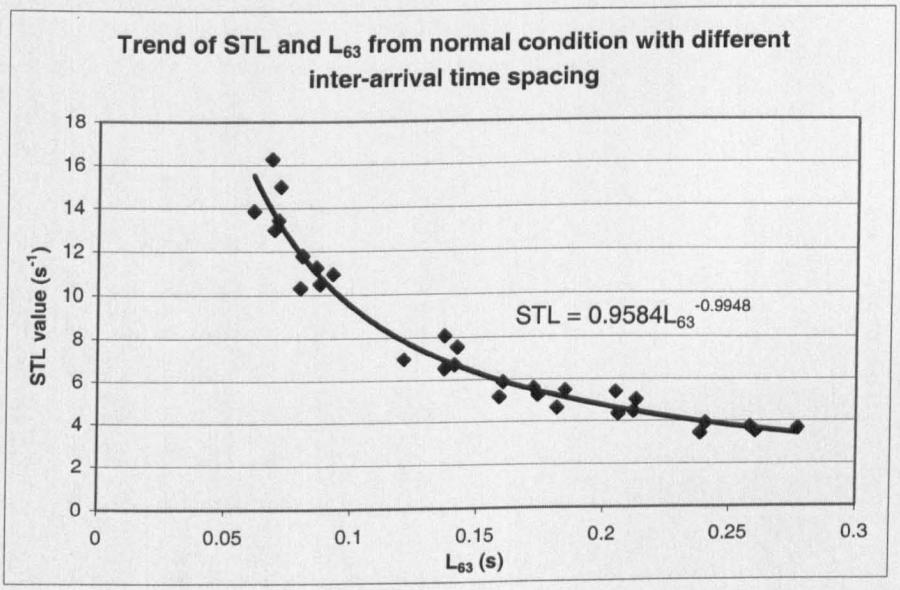


Figure 6.26 Curve of STL versus L_{63} fitted to simulated inter-arrival time spacing

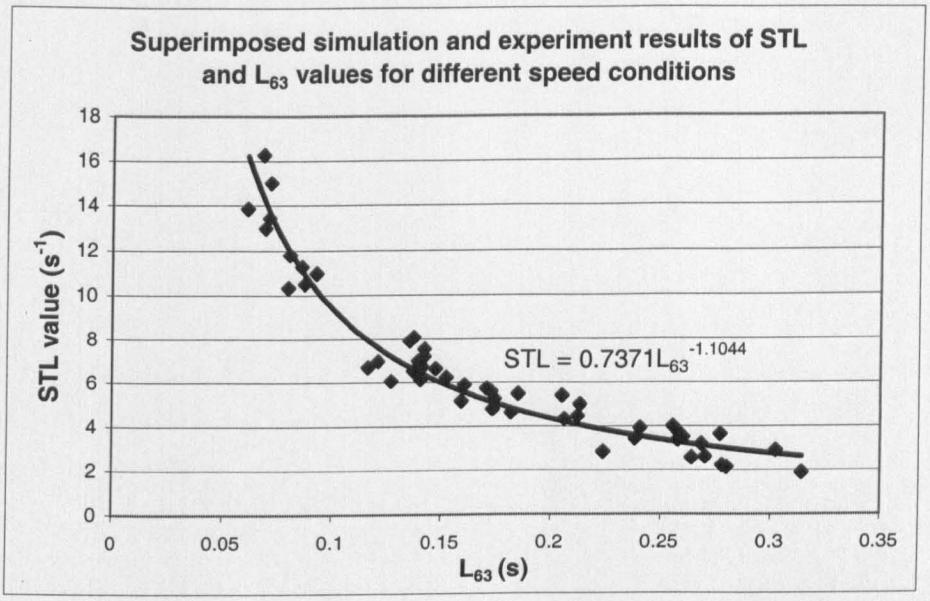


Figure 6.27 Curve of STL as a function of L_{63} for data from simulated inter-arrival time spacing and speed variation experiments.

The verification of the hypothesis that the rotating speed and inter-arrival time between AE events are inversely proportional to each other is given in Figure 6.27. Both sets of data, from speed variation experiments and from inter-arrival time dilation, are shown on the same graph. It is clear that both sets are indistinguishable from each other. The curve fitted to these data points is

$$STL = 0.7371L_{63}^{-1.1044} \qquad \textbf{Equation 6.3}$$

It is noted that the exponent of L_{63} is -1.1044 as compared with -1, the true hyperbolic relation. Equation 6.3 provides the means for comparing STL's across different speeds.

7 Simulation studies

Bearing condition monitoring relies on detection and processing of signals generated from sources of failure such as defects or crack growth or lubrication starvation in a bearing. The characteristics of the sources, such as its location, and the mechanism by which vibration and acoustic emission are produced are generally unknown. It is often desirable to have a repeatable and well characterised vibration or AE source in order to examine the effectiveness of the methods researched. The methods used are: (1) joint time- and frequency- domain analysis on vibration, and (2) the novel STL method on AE. It is also desirable to evaluate the fault prediction capability of the condition monitoring system as a whole. What are required are a simulated acceleration and an acoustic emission source, which are of interest and should be as repeatable as possible. These allow the measurement of sensitivity, reliability and robustness of the proposed multisensor system.

This chapter is divided into two parts. Firstly, the use of simulated vibration sources (producing acceleration signals) for evaluating the relative performance of the joint time- and frequency- domain analysis methods. The performance concerns the ability to discriminate between different types of signal that have transient features in them. Two joint time- and frequency- domain techniques, the Short Time Fourier Transform (STFT) and the Continuous Wavelet Transform (CWT), were studied. Secondly, the sensitivity, reliability and robustness of the STL method were assessed in respect of the variable AE detection level threshold. The AE events were produced by simulation using the real statistical distributions of AE amplitudes and inter-arrival times of AE events, which were obtained from experiments conducted on the heavy-duty test rig. All simulated signals, acceleration and AE, were generated by means of MATLAB programs.

7.1 Simulation study in the joint time- and frequency- domain

When a rolling element comes into contact with a localised defect, it generates a transient signal. While a time-domain method can give accurate time information on the state of the signal, it does not provide the frequency information. On the other hand, a frequency-domain method does not give the time information. It would be most desirable if methods were available that would supply both time and frequency information of a signal and they are said to operate in both the time and frequency

domains. This section presents a comparative study of two such methods, the Short-time Fourier Transform (STFT) and the Continuous Wavelet Transform (CWT).

The signals to which these two methods would be applied were generated artificially. There are four signals with characteristics given as follows:

1) Type 1 (Figure 7.1a) – A damped sinusoidal wave of a single frequency;
2) Type 2 (Figure 7.2a) – A high frequency damped sinusoidal wave followed by a low frequency sine wave;
3) Type 3 (Figure 7.3a) – Two damped sinusoidal waves of different frequencies added together; and
4) Type 4 (Figure 7.4a) – Periodic sinusoidal pulses with sharp rising edges.

It is noted that all four types were corrupted with random noise in other to make them look realistic.

A reference set of signals was created by first generating seven instances of each type and then taking the average. This results in a clearer signal for each type. A test set was also created comprising three instances of each type but no averaging. The MATLAB codes for generating these signals are presented in Appendix C3.

Signals in the reference and test sets were transformed using STFT and CWT as explained in Section 4.2.2 and 4.2.3 respectively. The results were shown as two-dimensional colour-coded images: Figures 7.1b, 7.2b, 7.3b and 7.4b are from the STFT transformation, and Figures 7.1c, 7.2c, 7.3c and 7.4c from the CWT.

By comparing a test image with the corresponding reference image using the method of correlation matching (Section 4.2.5), it is possible to measure the degree of similarity between the two images. Correlation matching produces an output, the correlation coefficient, ranging from +1 to –1. +1 means a perfect match in appearance as well as colour tone. -1 means a perfect match in appearance but the colour tones are completely opposite, like a photograph and its negative. 0 suggests the two images are completely dissimilar.

The results of the comparison are tabulated in Table 7.1 where the groups of three values refer to the correlation coefficients obtained from comparing the reference

signal with each of the three test signals in each type. It is noted that, in each case, correct matching has been achieved as the highest correlation coefficients always appear in the expected row.

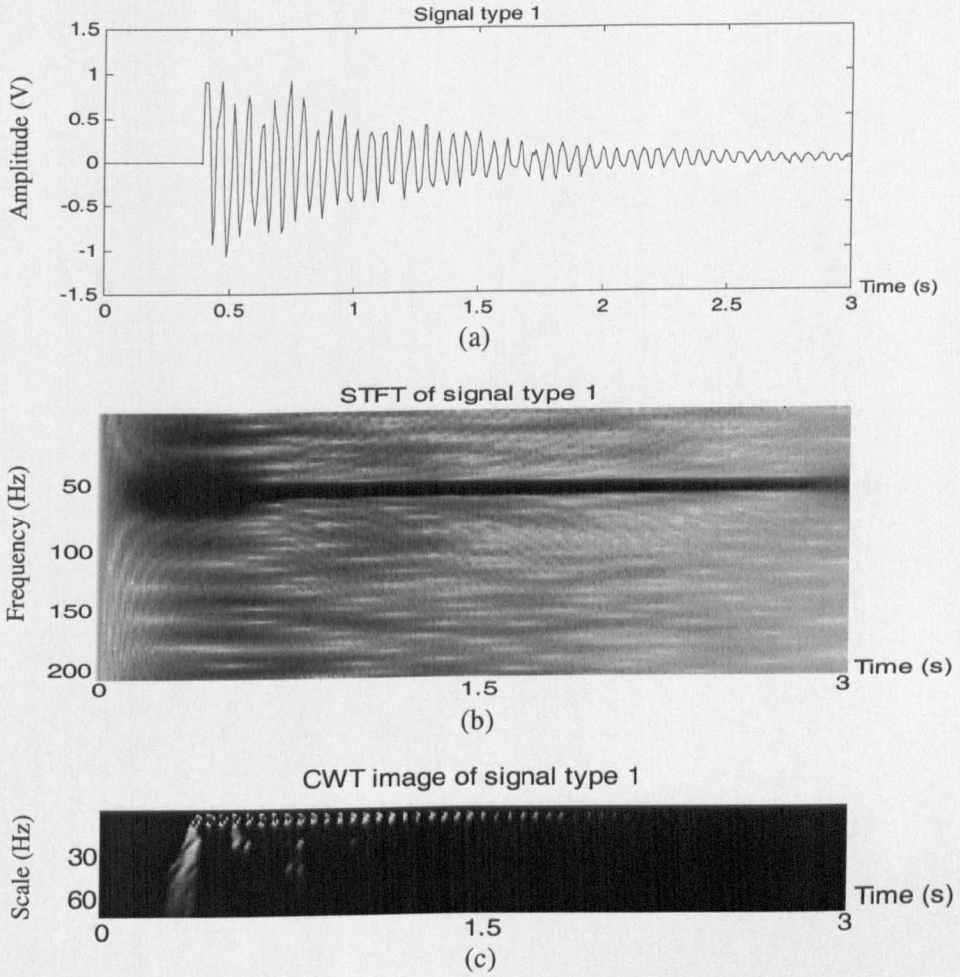


Figure 7.1 Type 1 signal and its STFT and CWT images

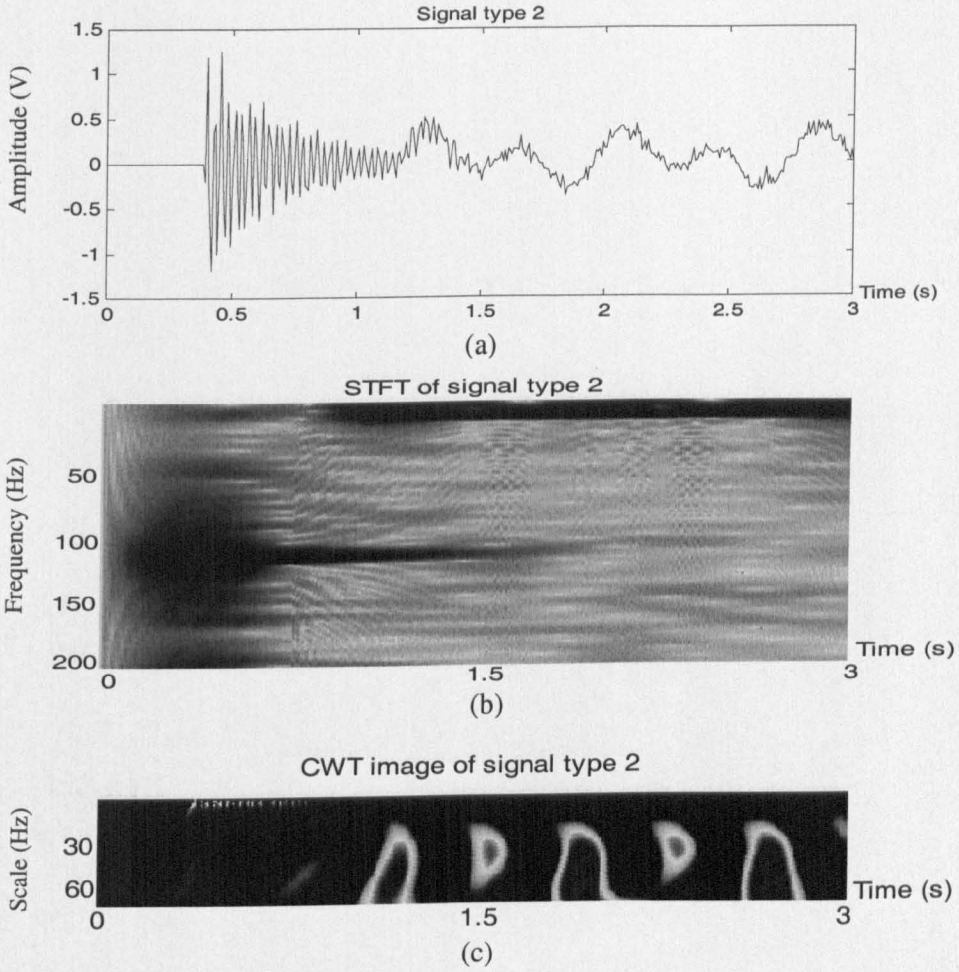


Figure 7.2 Type 2 signal and its STFT and CWT images

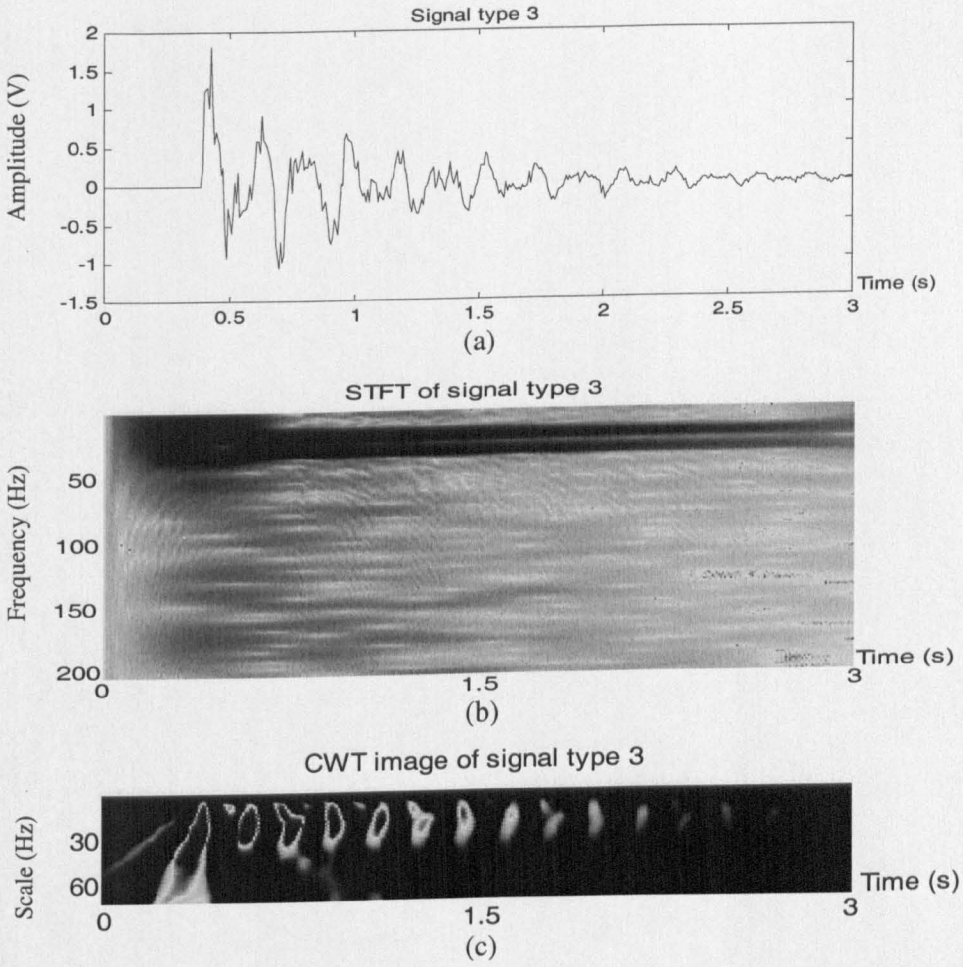


Figure 7.3 Type 3 signal and its STFT and CWT images

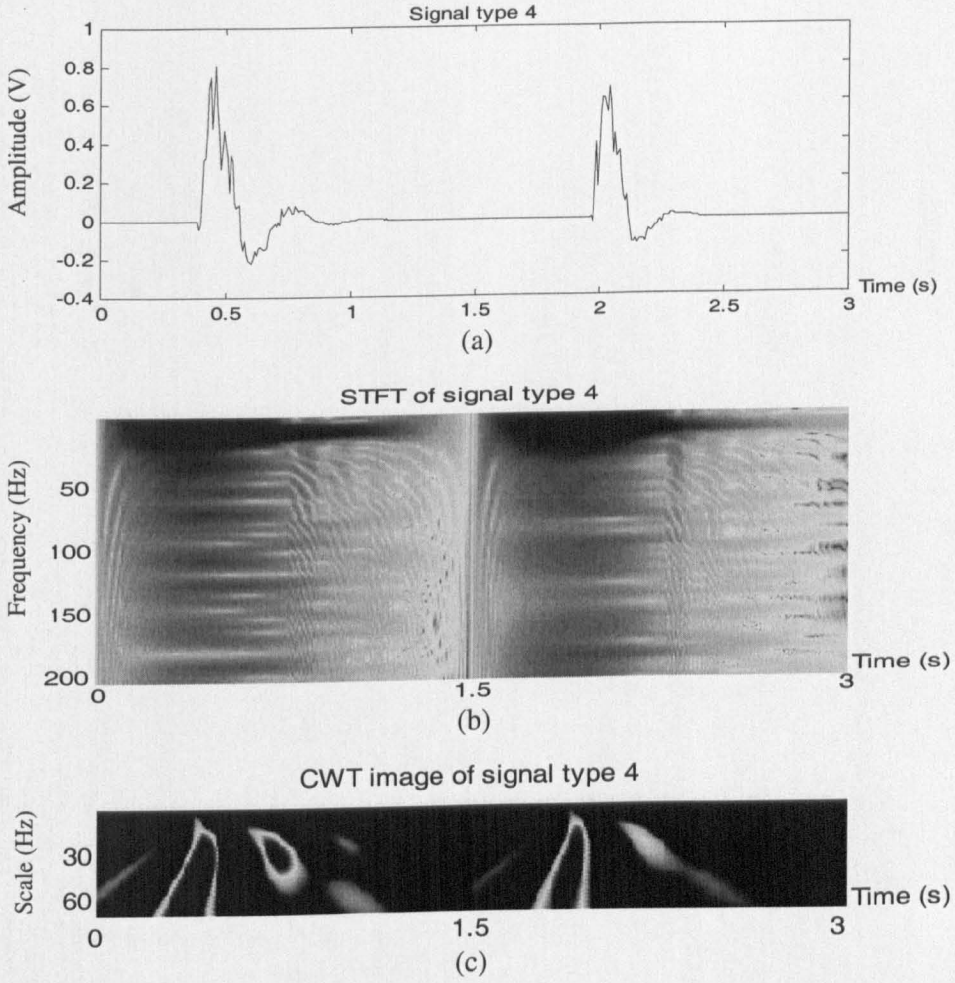


Figure 7.4 Type 4 signal and its STFT and CWT images

Table 7.1 Correlation matching results of STFT and CWT images

Correlation of STFT

Correlation coefficient		Test image (Type 1)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.832	0.8211	0.8347
	Type 2	0.1376	0.1613	0.1753
	Type 3	0.5188	0.525	0.545
	Type 4	0.3578	0.3578	0.343
Correlation coefficient		Test image (Type 2)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.2431	0.2974	0.3129
	Type 2	0.7124	0.6723	0.6562
	Type 3	0.303	0.3328	0.3475
	Type 4	0.3636	0.3636	0.3346
Correlation coefficient		Test image (Type 3)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.5004	0.5298	0.512
	Type 2	0.1838	0.2176	0.1674
	Type 3	0.8053	0.7932	0.7819
	Type 4	0.4588	0.4588	0.4515
Correlation coefficient		Test image (Type 4)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.3423	0.3628	0.3485
	Type 2	0.3232	0.3086	0.2373
	Type 3	0.4127	0.4645	0.4413
	Type 4	0.9057	0.9057	0.9199

Correlation of CWT

Correlation coefficient		Test image (Type 1)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.9041	0.9074	0.8715
	Type 2	-0.0889	-0.0498	-0.0797
	Type 3	0.2964	0.2678	0.3092
	Type 4	0.1157	0.1046	0.1208
Correlation coefficient		Test image (Type 2)		
		No. 1	No. 2	No. 3
Reference image	Type 1	-0.0927	-0.0479	-0.098
	Type 2	0.9909	0.9891	0.9903
	Type 3	-0.0825	-0.0469	-0.1264
	Type 4	0.2166	0.2099	0.2327
Correlation coefficient		Test image (Type 3)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.3469	0.3008	0.2279
	Type 2	-0.1413	-0.1047	-0.1233
	Type 3	0.9739	0.9453	0.9761
	Type 4	0.2718	0.2607	0.2805
Correlation coefficient		Test image (Type 4)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.2449	0.1292	0.0206
	Type 2	0.1432	0.1481	0.1626
	Type 3	0.26	0.3313	0.3112
	Type 4	0.975	0.9738	0.9756

It is noted that in both cases there is correct recognition for each type of signal. However, if the values of the correlation coefficients are examined, the CWT method returns consistently a higher set of values than does the STFT method; in addition, there is better discrimination amongst the different types of signals for the CWT method too.

The superiority of the CWT method is somewhat expected in view of its higher resolutions across the scale and time, as mentioned in Section 4.2.3. CWT is particularly good at detecting discontinuity in the signal, as is evident by comparing, for example, Figures 7.4b and 7.4c, where, in the latter, the abrupt rise in the signal is clearly indicated on the CWT image.

7.2 Simulation study for the shape-to-life (STL) method

The STL method is based on the ability of a monitoring system to detect acoustic emission events. The detection is governed by a threshold primarily designed to eliminate background noise. When the signal of an AE event crosses the threshold, the event is said to have been detected. The higher the threshold, the fewer the AE events captured, and vice versa. In real application, background noise from one machine may be different from the other because their speed, load and other conditions are not the same. It is therefore important to ask: How susceptible is the STL method to the choice of threshold? This is an issue of reliability and robustness of the method. If the method is greatly influenced by the threshold level, not unless some compensation is provided, the usefulness of the method will be severely curtailed. To answer this question, simulation tests were carried out.

Simulated AE events were produced which had inter-arrival times following a Weibull distribution. From the results of the bearing life test performed on the low-speed heavy-duty test rig, Table 6.7, three characteristic life values, 0.5 s, 0.05 s and 0.025 s were chosen and referred to as 'good', 'warning' and 'faulty' conditions in Table 7.2. The guaranteed life t_o, being rather small compared to the characteristic life, was taken to be zero in the simulation.

Since the shape parameter in Table 6.7 has a value of around 1 for all conditions of the bearing, this value was chosen in the simulation. Knowing the shape parameter and the characteristic life, one can then estimate the STL's for the three conditions, as highlighted in the third column of Table 7.2.

Table 7.2 Source simulated signals using different Weibull parameters

Signal Patterns	**Bearing conditions**	**STL value (Shape-to-life) (S^{-1})**	**Weibull distribution**	
			Life parameter (s)	**Shape parameter**
Type 1	Good	2	0.5	1
Type 2	Warning	20	0.05	1
Type 3	Faulty	40	0.025	1

The amplitude of an AE event, being approximately a damped sinusoidal wave and hence a narrow band process, follows a Gaussian distribution with zero mean. According to Newland (1993a), the peak amplitude of this signal has a Rayleigh distribution whose probability density function (pdf) is given by

$$P_p(a) = \frac{a}{\sigma_y^2} e^{\frac{a^2}{2\sigma_y^2}} \quad , 0 \le a \le \infty \qquad \textbf{Equation 7.1}$$

where a = the threshold, and

σ_y = the standard deviation of AE signal amplitude.

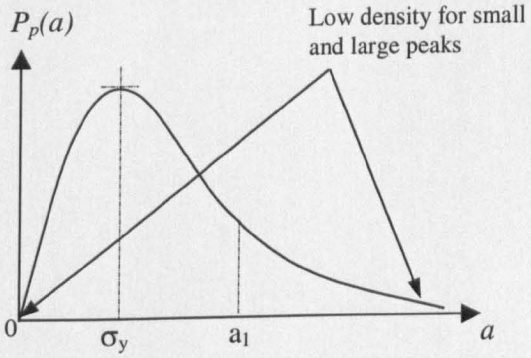


Figure 7.5 Rayleigh distribution of peaks (Newland, 1993a)

Figure 7.5 is a graphical representation of Equation 7.1. Note that the area to the right of the vertical line $a = a_1$ gives the probability that the peak value exceeds a_1. It is also clear from Figure 7.5 that the majority of peaks have about the magnitude of σ_y.

An array of twenty inter-arrival times for each of the three conditions was generated using the Weibull parameters as given in Table 7.2. The MATLAB program for doing this is in Appendix C4. In a similar manner, arrays of peak amplitudes that follow the Rayleigh distribution, Equation 7.1, were generated for the three conditions with the MATLAB program, Appendix C4. The standard deviation, σ_y, of the AE signal amplitude was taken to be unity. This is equivalent to expressing the signal

amplitudes in terms of their standard deviation and so the results of the simulations study can be readily applied to all signals with different gains.

Figures 7.6, 7.7 and 7.8 show the peak amplitudes, following the Rayleigh distribution, erected at such time instants that their inter-arrival times follow the Weibull distribution as explained earlier. These diagrams do not represent the true AE signals or events. But they show the features based on which the inter-arrival times were extracted. As the threshold level is raised, the number of detectable events represented by the peak amplitudes above the threshold is reduced.

The 'good', 'warning' and 'faulty' condition, STL values were computed for each of the twenty simulated signals for a particular threshold. There were four threshold settings, as a percentage of the amplitude standard deviation: 0%, 20%, 30% and 40%. According to Newland (1993b), the probability that the peak value exceeds a threshold a is given by

$$P(Peak > a) = e^{-a^2/2\sigma_y^2} \qquad \textbf{Equation 7.2}$$

By virtue of Equation 7.2, a 0% threshold gives a probability of 1, meaning that all events are counted. For thresholds that are 20%, 30% and 40% of the amplitude standard deviation, the probabilities are respectively 0.980, 0.956 and 0.923. Thus, at the 40% threshold, just under 8% of the events will go undetected. This is unlikely to happen in practice if the threshold is set properly.

Results of the STL values, as a function of the threshold settings of 20%, 30% and 40%, computed for the respective good, warning and faulty conditions are shown in Tables 7.3, 7.4 and 7.5. The life and shape values from which the STL values are calculated are also included.

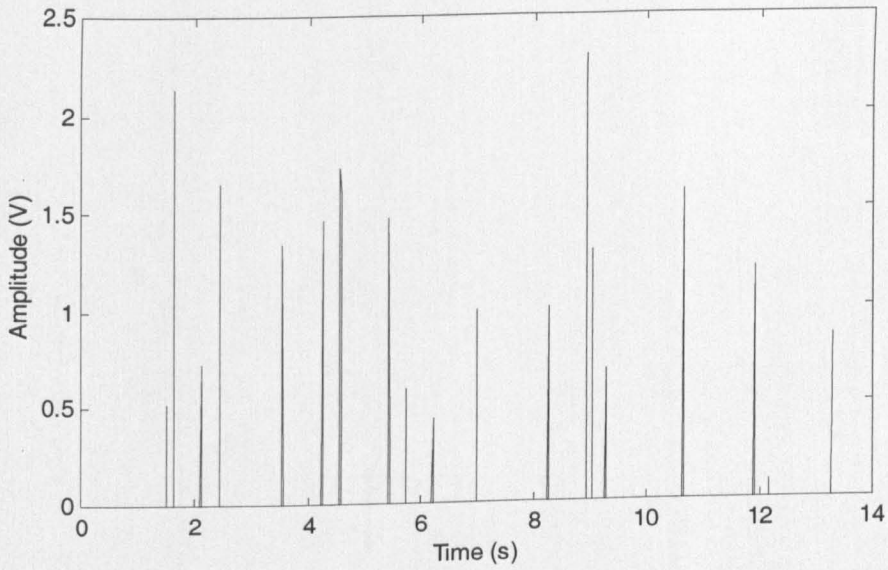


Figure 7.6 Example of simulated signal for good condition (life = 0.5, shape = 1)

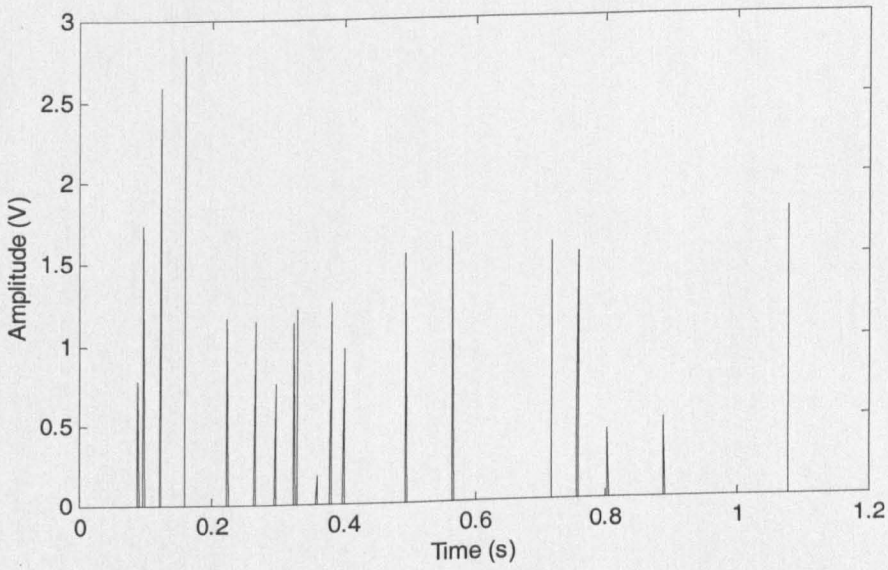


Figure 7.7 Example of simulated signal for warning condition (life = 0.05, shape = 1)

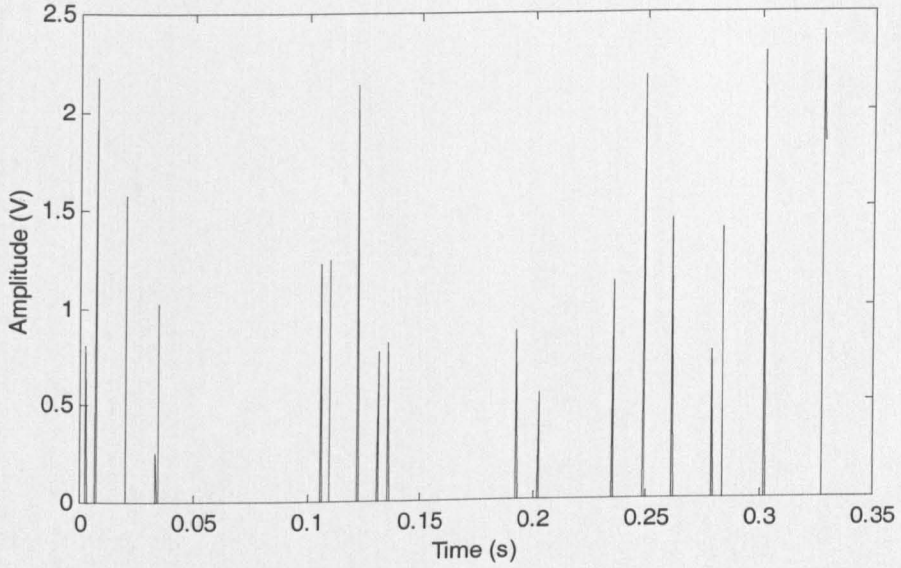


Figure 7.8 Example of simulated signal for faulty condition (life = 0.025, shape = 1)

Table 7.3 STL from simulated signals of GOOD condition with different threshold values

Threshold @ 20%			Threshold @ 30%			Threshold @ 40%		
Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})
0.7351	0.9485	1.2902	0.7351	0.9485	1.2902	0.7351	0.9485	1.2902
0.5704	1.2805	2.2448	0.5916	1.2914	2.183	0.5916	1.2914	2.183
0.4724	1.1983	2.5366	0.4724	1.1983	2.5366	0.5035	1.3136	2.6092
0.397	1.0157	2.5582	0.3979	1.0159	2.5534	0.5016	0.996	1.9856
0.5069	1.2782	2.5215	0.5069	1.2782	2.5215	0.5069	1.2782	2.5215
0.5422	1.0594	1.9538	0.7015	1.0503	1.4971	0.7015	1.0503	1.4971
0.6254	1.1013	1.7608	0.6806	1.0565	1.5522	0.6806	1.0565	1.5522
0.3342	0.9965	2.9821	0.3342	0.9965	2.9821	0.3342	0.9965	2.9821
0.5302	1.2935	2.4399	0.5302	1.2935	2.4399	0.5302	1.2935	2.4399
0.417	1.0109	2.4241	0.3956	1.2062	3.0493	0.3956	1.2062	3.0493
0.6357	1.139	1.7918	0.6357	1.139	1.7918	0.6357	1.139	1.7918
0.6173	0.9649	1.5631	0.6377	0.9887	1.5505	0.6496	0.9437	1.4527
0.5201	0.9251	1.7789	0.5305	1.0049	1.8943	0.5305	1.0049	1.8943
0.4702	1.1711	2.4908	0.4921	1.2329	2.5054	0.533	1.1887	2.2302
0.5468	1.2535	2.2925	0.5468	1.2535	2.2925	0.5741	1.0946	1.9067
0.3644	0.9876	2.7106	0.3644	0.9876	2.7106	0.3644	0.9876	2.7106
0.368	0.9976	2.7108	0.368	0.9976	2.7108	0.368	0.9976	2.7108
0.6025	1.0598	1.7591	0.6675	1.2478	1.8694	0.7135	1.3898	1.9478
0.538	1.1735	2.181	0.538	1.1735	2.181	0.5493	1.183	2.1537
0.4689	0.9801	2.0901	0.4802	0.934	1.9452	0.5039	0.8885	1.7632

Table 7.4 STL from simulated signals of WARNING condition with different threshold values

Threshold @ 20%			Threshold @ 30%			Threshold @ 40%		
Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})
0.0474	0.9342	19.707	0.0474	0.9657	20.3793	0.05	0.9821	19.6564
0.0549	1.0587	19.2894	0.0549	1.0587	19.2894	0.0519	1.0869	20.924
0.0464	0.9565	20.6162	0.0464	0.9565	20.6162	0.0566	0.8945	15.802
0.0403	1.0297	25.5703	0.0403	1.0297	25.5703	0.053	1.0344	19.5166
0.0411	1.0669	25.9549	0.0419	1.0761	25.6644	0.0419	1.0761	25.6644
0.0469	1.0369	22.1037	0.0447	1.0635	23.7867	0.0447	1.0635	23.7867
0.0573	0.9451	16.4964	0.0573	0.9451	16.4964	0.0573	0.9451	16.4964
0.0538	1.1856	22.0191	0.0538	1.1856	22.0191	0.0629	1.1791	18.7387
0.0574	0.944	16.4526	0.0574	0.944	16.4526	0.0615	0.9181	14.9254
0.0414	1.1918	28.759	0.0414	1.1918	28.759	0.0414	1.1918	28.759
0.0397	1.067	26.8685	0.0397	1.067	26.8685	0.0397	1.067	26.8685
0.0525	1.1937	22.7302	0.0525	1.1937	22.7302	0.0525	1.1937	22.7302
0.0588	0.9527	16.1948	0.059	0.9946	16.8585	0.059	0.9946	16.8585
0.0449	0.9545	21.2508	0.0529	0.8959	16.9235	0.0529	0.8959	16.9235
0.0526	0.9499	18.0749	0.0526	0.9499	18.0749	0.0526	0.9499	18.0749
0.0423	1.116	26.356	0.0423	1.116	26.356	0.0423	1.116	26.356
0.0562	1.0182	18.1049	0.0719	0.9198	12.785	0.0719	0.9198	12.785
0.0478	1.0715	22.4319	0.0478	1.0715	22.4319	0.0478	1.0715	22.4319
0.0487	1.1112	22.8213	0.0487	1.1112	22.8213	0.0487	1.1112	22.8213
0.0567	1.0705	18.8821	0.0567	1.0705	18.8821	0.0567	1.0705	18.8821

Table 7.5 STL from simulated signals of FAULTY condition with different threshold values

Threshold @ 20%			Threshold @ 30%			Threshold @ 40%		
Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})	Life (s)	Shape	STL (s^{-1})
0.0219	0.9901	45.2643	0.0257	0.9882	38.4339	0.0225	0.9862	43.9011
0.029	1.0026	34.5448	0.029	1.0026	34.5448	0.029	1.0026	34.5448
0.0221	1.0032	45.451	0.0418	0.8133	19.4436	0.0389	0.8556	21.9772
0.0195	1.041	53.283	0.0195	1.041	53.283	0.0287	0.9284	32.3941
0.0272	0.9877	36.2487	0.0272	0.9877	36.2487	0.0275	1.0013	36.3453
0.0217	1.0644	49.0579	0.0217	1.0644	49.0579	0.0217	1.0644	49.0579
0.0257	1.022	39.7749	0.0283	0.9936	35.0971	0.0321	1.0066	31.3217
0.0228	0.9538	41.7781	0.0228	0.9538	41.7781	0.0228	0.9538	41.7781
0.0278	1.0139	36.4316	0.0278	1.0139	36.4316	0.0197	1.1619	58.9537
0.026	1.0076	38.7192	0.026	1.0076	38.7192	0.026	1.0076	38.7192
0.0239	0.9244	38.7423	0.0267	0.8888	33.2401	0.0267	0.8888	33.2401
0.0271	1.1127	41.1328	0.0271	1.1127	41.1328	0.0306	1.0986	35.9368
0.0261	0.8988	34.443	0.0261	0.8988	34.443	0.0261	0.8988	34.443
0.0249	1.1122	44.5836	0.0259	1.1279	43.542	0.0259	1.1279	43.542
0.0218	1.082	49.5754	0.0213	1.0926	51.3109	0.0213	1.0926	51.3109
0.0237	1.0411	43.8423	0.0237	1.0411	43.8423	0.0237	1.0411	43.8423
0.0221	1.0035	45.4464	0.0221	1.0035	45.4464	0.0221	1.0035	45.4464
0.023	1.0887	47.2727	0.023	1.0887	47.2727	0.035	0.9648	27.5815
0.0255	1.1143	43.7225	0.0255	1.1143	43.7225	0.0255	1.1143	43.7225
0.0243	1.0162	41.8581	0.0243	1.0162	41.8581	0.0243	1.0162	41.8581

Table 7.6 Statistics of STL values of all conditions with different threshold values

	STL (Mean) (s^{-1})			STL (S.D) (s^{-1})		
	Good	Warning	Faulty	Good	Warning	Faulty
Threshold @20%	2.204035	21.5342	42.55863	0.442393	0.510605	0.515267
Threshold @30%	2.20284	21.18827	40.44244	3.727281	4.237344	4.369682
Threshold @40%	2.133595	20.45008	39.49584	5.168055	7.601809	8.604922

Table 7.6 gives the means and standard deviations of the STL values. It is noted that despite the threshold having an effect on the means STL, it is far smaller than that due to the bearing conditions. Consider the possibility of misdiagnosis. There are two cases of interest: 1) Missed detection - mistaking 'faulty' for 'warning'; and 2) False alarm – mistaking 'warning' for 'faulty'. Whilst under certain situations, one case may be preferred to the other as being the lesser of the two evils, the evaluation, presented below, of the probability that such cases do happen does not make any distinction between them.

An inspection of Table 7.6 shows that the two cases that can cause problem of misdiagnosis are the warning STL of 21.5342 and the faulty STL of 39.49584, as circled in the table. A test of significance is performed based on the procedure detailed by Moroney (1951). Denoting the faulty condition at the 40% threshold by the subscript 1 and the warning condition at the 20% threshold the subscript 2, then

N_1=20, μ_1 = 39.4958, σ_1 = 8.6049, and σ_1^2 = 74.0446

N_2=20, μ_2 = 21.5342, σ_2 = 0.5106, and σ_2^2 = 0.2607

where N = sample size,

μ = sample mean

σ = standard deviation (S.D), and

σ^2= variance

The difference between the sample mean is μ_1-μ_2 = 17.9616

and the corresponding variance of the difference is

$Var(\mu_1-\mu_2) = \sigma_1^2/N_1 + \sigma_2^2/N_2 = 3.7152$

from which the standard error of the difference is

Std. Error of Diff. = 1.9275

since it is the square root of the variance of the difference.

The observed difference between the sample mean of STL, 17.9616, is highly significant, being 17.9616/1.9275 = 9.32 times its standard error and so the probability of misdiagnosis (mistaking warning for faulty and vice versa) is very remote indeed.

Figure 7.9 shows a bar graph of the mean STL values with their associated error bars of ± 1 S.D for all three conditions with different threshold settings. It can be seen that, as the threshold setting increases, the mean STL falls slightly. The reason is that as the threshold is raised, due to the fact that some events are missed, the inter-arrival times on average become longer. This in turn means that the characteristic life has become longer, resulting in a smaller STL (STL = shape parameter/characteristic life) since the value of the shape parameter remains roughly at unity.

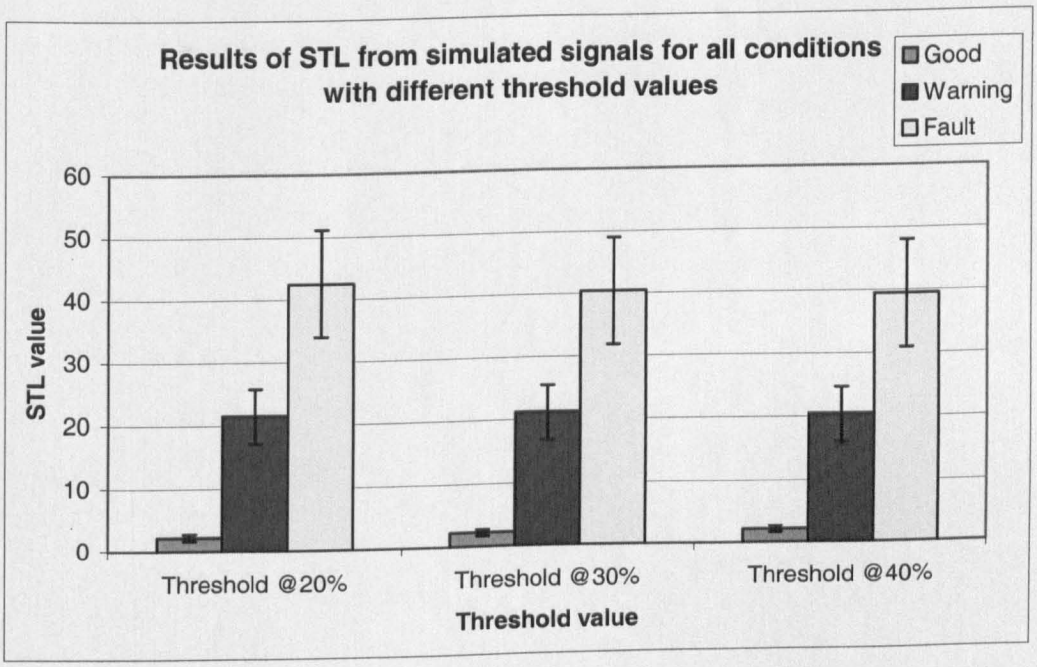


Figure 7.9 Bar plots of mean STL showing ± 1 SD error bars at different threshold values

In conclusion from this simulation study, the ability of the STL method to discriminate between the warning and faulty condition is not affected by the choice of threshold settings up to 40% of the standard deviation of the amplitude of AE signals. The 40% threshold represents just under 8% of the AE events not being detected, an unlikely situation if the threshold is set judiciously.

8 Proposed monitoring scheme using the STL method

Section 6.4 has shown that there are three factors that can affect the values of STL and L_{63}:

1. Speed
2. Load
3. Wear

In order to use the STL method for condition monitoring of bearings, an important question is how the effects of speed and load can be compensated. This chapter attempts to answer this question.

8.1 Establishing the initial STL value for a good bearing

With a new bearing, the initial values for the STL and L_{63} are obtained. These will serve as a reference for monitoring the bearing condition because, from experiments, it has been established that as the bearing wears progressively, it causes the:

1. STL to increase; and
2. L_{63} to decrease

such that they follow a hyperbolic relation (Figure 6.20).

8.2 Speed compensation for STL and L_{63}

When the rotating speed of a bearing changes, both STL and L_{63} will change (Figure 6.23). In particular, as shown in Section 6.4.4, L_{63} is inversely proportional to rotating speed. For example, if the rotating speed is doubled, its corresponding L_{63} is halved.

Using Equation (6.3),

$$STL = 0.7371 L_{63}^{-1.1044}$$

a new value for STL can be produced, adjusted to a reference rotating speed. The reference speed chosen can be any speed but logically it is the typical running speed for the bearing in question.

8.3 Load compensation for STL and L_{63}

The effect of a greater bearing load is to increase the STL and decrease the L_{63}. This can be seen from Figure 6.13 and the part of the curve in Figure 6.20 with values of STL less than about 25 s^{-1}. If the load acting on the bearing is measurable, then adjustment to STL and L_{63} can be made in the same way as speed compensation is made using the fact that STL and L_{63} are hyperbolically related. However, more often than not, knowledge of bearing load is not as easy to gain as of speed and so a different method has to be found.

8.4 Setting of STL threshold for bearing condition alert

The choice of STL threshold level is governed by past experience of the same or similar bearings and by the criticality of the process in which the bearing is a component. The more critical the process, the lower the threshold should be set.

It has been observed that bearing loading below the bearing's load carrying capacity produces a rather small STL value, typically not more than 8 s^{-1}. The bearing that produces Figure 6.14 has a load carrying capacity of 59 bars and the bearing in Figure 6.21 a capacity of 137 bars. This low value of 8 s^{-1}, compared to the range of increase in STL of (59-19) = 40 s^{-1} due to the effect of bearing wear alone for the bearing subjected to a constant load of 300 bars (Figure 6.17), is only a fifth. Another piece of evidence can be found from Figure 6.10 and Table 6.4 for a bearing running under no load. In this case, the ratio of STL of balanced shaft to defective bearing is 5.2:70.9 = 1:13.6. It is remembered that a defective bearing condition means a 1-mm diameter small indentation produced with an electric-arc scriber. It therefore seems to suggest that lower bearing loading makes the STL value even more sensitive to bearing wear.

In summary, bearing wear exerts a much stronger influence on STL than does the load as long as the latter is kept below the bearing's recommended load carrying capacity. Since it is not good practice to drive a bearing beyond its recommended load and hence this situation will seldom be encountered on a properly designed machine or process, load compensation for STL and L_{63} can often be safely ignored.

From the examples given above for the two extremes of loading conditions – at 300 bars which is way above the bearing's load carrying capacity and at 0 bar, a guideline on setting an alarm threshold for STL can be formulated. The rule is that the alarm level should be set no higher than five times the initial STL value of the bearing. Evidently, this is a somewhat conservative recommendation and it should be reviewed and updated in light of further experimental results obtained from the field.

9 Conclusions and future work

9.1 Validity of systems approach

A systems approach has been adopted in the development of a condition monitoring system for rotating machines. For various operating conditions of changing loads and speeds, the particular cases studied are defined as follows:

1. balanced rotating shaft;
2. unbalanced rotating shaft;
3. misalignment of bearing supports; and
4. artificial and natural wear on bearings

The monitoring system consists of two complementary sensing methods: vibration and acoustic emission (AE). Employing both methods based on the different physical principles, the dynamic range of shaft rotating speeds in monitoring can be increased significantly. In addition, a multi-sensor system increases fault tolerance and hence its reliability. This is because if one sensor fails to operate or perform at its best, the other can still provide useful information.

Figure 9.1 provides a summary of the proposed multi-sensor system for bearing condition monitoring in block diagram form. This diagram is similar to the one as shown in Figure 4.1, but additional information on the implementation and the success rates obtained from experiments are included. The following sections describe the flow of the block diagram.

9.1.1 Vibration monitoring using CWT and BPNN

Vibration monitoring has been proven to be an effective technique for monitoring rotating machines running at high speed by various researchers. The Continuous Wavelet Transform (CWT), joint time- and frequency- domain, has been used successfully in this research to provide a greater discrimination than the short time Fourier Transform (STFT) between good (balanced), unbalanced, misaligned and defective bearing conditions. This work is based on the transformation of an acceleration signal using CWT into a three-dimensional image map, which consists of time, frequency (scale) and signal intensity axes. Then classification of the various conditions is performed using a back-propagation neural network (BPNN). Results show that this method is effective, gaining a recognition rate of 90 %, for detection and diagnosis of the different failures that may occur in rotating machines running at

the high speed of 20 rev/sec. However, if the machine operates at low speed, vibrations generated can barely be detected resulting in a poor signal-to-noise ratio, thus making the signal difficult to be interpreted.

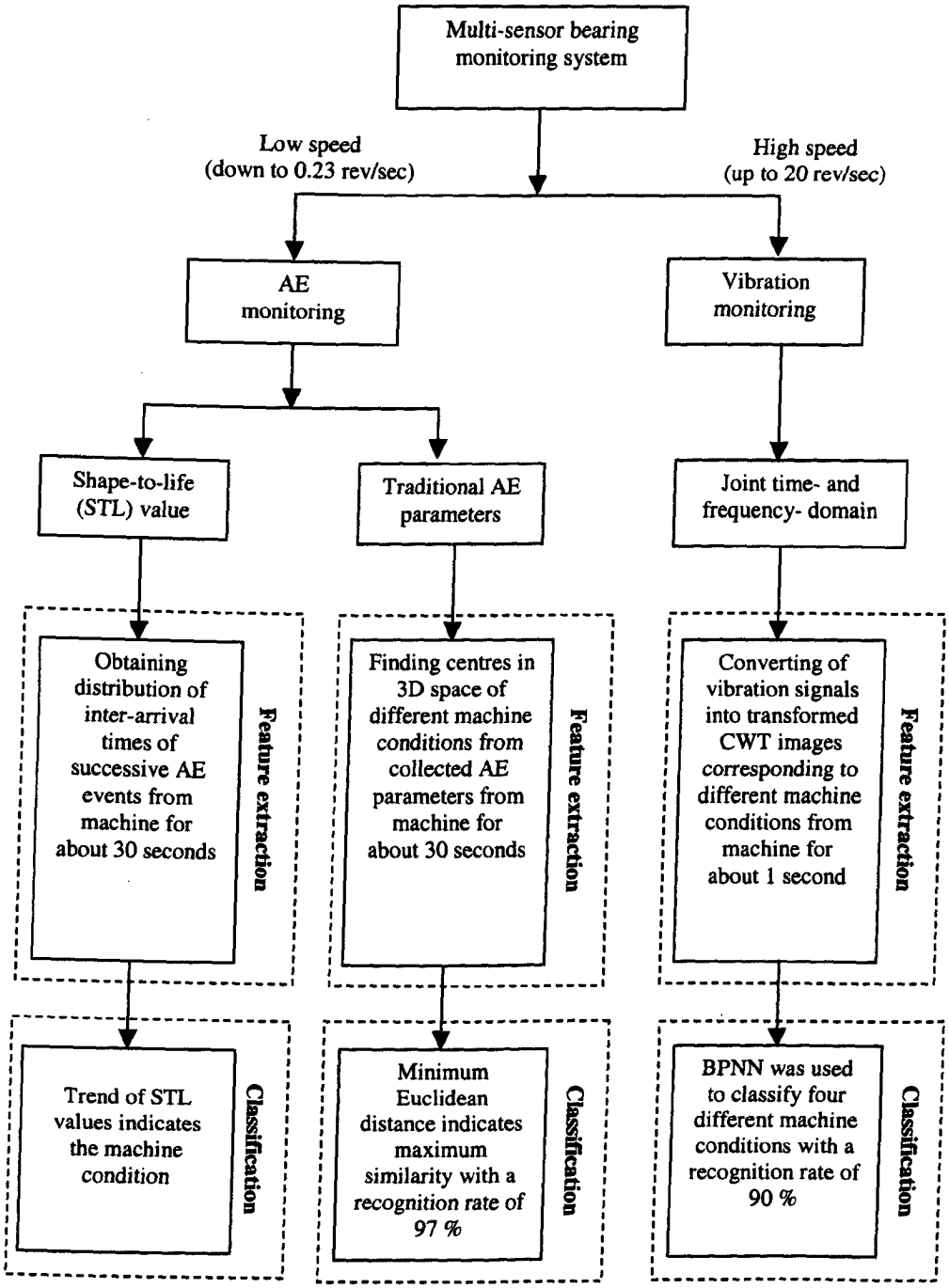


Figure 9.1 Summary block diagram of the proposed multi-sensor system for bearing condition monitoring

Vibration can also be significantly influenced by the dynamics of machine structure and it tends to be more easily detected when it is close to or at the resonant frequencies of the structure.

On the other hand, the usefulness of an AE technique extends well into low rotating speed. The technique essentially strives to detect material deformation arising from crack initiation and growth or from the friction process, all being independent of the machine structure dynamics because of the different operating frequencies.

Two AE methods have been explored in detail in this research: 1) the fuzzy c-mean clustering technique applied to traditional AE parameters, and 2) the STL method.

9.1.2 AE monitoring using fuzzy C-mean

Three traditional AE parameters, peak amplitude, AE event duration and energy, have been demonstrated to be effective in the fuzzy c-mean clustering technique as signatures for different machine operating conditions. The fuzzy c-mean is used to provide a centre of a cluster corresponding to each machine condition. Then the minimum distance classifier is performed to calculate the Euclidean distance between the unknown and the centre of each machine condition, thus making it possible to assign the unknown to the nearest cluster. The four different machine conditions are the balanced shaft, the unbalanced shaft, the misaligned shaft and the defective bearing at the high rotating speed of 20 rev/sec. The results show that centres from all conditions are distinctively clustered (Figure 6.6) with a recognition rate of 97 %.

Moreover, the technique has also demonstrated its capability to discriminate the different bearing lubrication conditions. In this case, AE parameters were collected, at Corus, Middlesborough, on bearings running at 0.5 rev/min with two types of grease, clean and contaminated. Results show that both lubrication conditions have been classified correctly using the fuzzy-c mean technique (Figure 6.8).

9.1.3 AE monitoring using the novel STL method

An innovative method, known as the STL method, has been developed in this project for monitoring the condition of rotating machines. It is based on the statistical modelling of inter-arrival times of AE events with the Weibull distribution. Once the

collection of inter-arrival times is made, the STL can then be calculated as the ratio between the estimated shape and characteristic life parameters (Equation 4.15) which follows Weibull distribution. The trend of STL can be used to indicate machine status as progressive bearing wear leading to its eventual failure.

The STL method was studied thoroughly via experimental and simulation work. Experiments were conducted in order to verify the method in terms of its sensitivity and discrimination capability on different machine conditions such as rotating speed and loading.

At the high speed of 20 rev/sec, the STL method can clearly differentiate between the four different machine conditions: balanced, unbalanced, misaligned and defective bearing. The 'balanced' condition yields the minimum STL values, whilst the other conditions give greater STL values (Figure 6.9).

At the lower speed of 4.7 rev/sec, the STL values were measured over time when the bearing was subjected to stepwise loading increments from below the recommended bearing load to overload (with a high capacity hydraulic ram). The STL values obtained at normal load were at minimum and remained constant with time; whereas at overload the STL values started at a higher level and increased monotonically with time (Figure 6.12).

A bearing life test was performed at the low speed of 0.23 rev/sec. The bearing was run from new to failure. After the initial part of the test under normal bearing loading, the test was accelerated by intentionally subjecting the bearing to a load of 300 bars, above the recommended 137 bars. The results show that the AE method using the STL is far superior to the vibration method using the V_{RMS} of acceleration. The STL values are far more sensitive to progressive wear as well as imminent bearing failure than does the acceleration V_{RMS} (Figure 6.17 and Figure 6.18).

The rotating speed and inter-arrival time are related by the empirical Equation 6.3, as repeated below:

$$STL = 0.7371 L_{63}^{-1.1044}$$

9.2 Summary of findings

9.2.1 Recognition of a visualisation of mechanical vibration

Condition monitoring of bearing conditions via visualisation of mechanical vibration using transformed CWT image has been explored in this study (Section 6.1.2). Binarising, an image processing technique, was employed to convert the colour-map image into a black and white image, with the advantage that computational effort for classification becomes much simpler. An artificial back-propagation neural network (BPNN) was used as machine classifier. The results show that the proposed technique is able to distinguish different machine conditions with a recognition rate of 90 % (Table 6.1).

9.2.2 Discrimination capability of CWT and STFT images

From the simulation study in joint time- and frequency- domain (Section 7.1), it has been shown that using Continuous Wavelet Transform (CWT) images for classification is superior to Short Time Fourier Transform (STFT) images in its higher correlation coefficient and discrimination capability (Table 7.1). This is because CWT images have higher resolutions across the scale and time axes (Section 4.2.3).

9.2.3 Utilising the trigger for acquiring vibration signal

The main problem with bearing condition monitoring using visualisation of vibration signal is that two transformed images cannot be compared unless both signals start at the same position in the time domain. To overcome this problem, it is important to assure that all signal recordings start at the same before they are converted into image maps. The trigger, produced by a proximity sensor, was used to facilitate all vibration signals to be acquired at the same shaft rotating reference point.

9.2.4 Traditional AE parameters for machine classification

Three traditional AE parameters: AE event duration, peak amplitude and energy have been found to be effective signatures for representing different machine conditions: balanced, unbalanced, misaligned and defective bearing (Section 6.2.1). The fuzzy C-mean was applied to the three AE traditional parameters, resulting in representing centre clusters as for machine classification (Figure 6.6).

9.2.5 Lubrication conditions and traditional AE parameters

From the experiment (Section 6.3.1), the relationship between the three traditional AE parameters and lubrication conditions has been found as shown in Figure 6.8. It is concluded that the two lubrication conditions - clean grease and contaminated grease - can be clearly distinguished from each other.

9.2.6 Novel STL method for bearing condition monitoring

The STL and L_{63} values have been explored in this project and demonstrated as sensitive condition monitoring AE parameters. The STL method is based on the modelling of inter-arrival times of AE events with Weibull distribution. The STL is defined as the ratio of two estimated Weibull parameters, shape to characteristic life (Equation 4.15). The L_{63} can be defined as the summation of the estimated guaranteed life and characteristic life (Equation 4.16).

9.2.7 Effect of load on STL

For the two different bearings studied in the experiment, it has been found that the STL values remain more or less level if the bearing is subjected to the applied load within basic dynamic load rating. When the applied load becomes greater, the STL is increased whilst the L_{63} is decreased as can be seen in Figure 6.13 and the part of the curve in Figure 6.20.

9.2.8 Effect of speed on STL

The change in speed of rotating machines has been found to affect the STL values. This is because the rotating speeds and inter-arrival times are approximately inversely related. The relationship between speed variations and STL values are as shown in Figure 6.23. The effect of speed variations on STL can be compensated using the hyperbolic relationship between STL and L_{63} (Section 9.2.10).

9.2.9 Effect of wear on STL

Similar to load and speed, STL has been found to be influenced by wear. From simulated bearing wear test (Section 6.4.1.2), a mere 1 mm indentation on the outer race bearing yields an STL value 13.6 times that from a good bearing (Figure 6.9). From natural progressive bearing wear test (Section 6.4.3.2), it is clear that the progression of wear results in a monotonically increasing trend of STL. When the

bearing failed to operate, the STL was increased to about 30 times the value for the initial bearing condition (Figure 6.17). In order to set the threshold for STL as bearing condition alert, the choice of threshold level is governed by the rule that the alarm level should be set no higher than five times the initial STL value of the bearing (Section 8.4).

9.2.10 Hyperbolic relationship between STL and L_{63}

It has been found from both simulation and empirical study that the rotating speed and inter-arrival time between AE events are inversely proportional to each other, resulting in the hyperbolic relationship between STL and L_{63} (Figure 6.27). This hyperbolic curve provides the basis for adjusting STL and L_{63} for speed compensation (Section 8.2) and load compensation (Section 8.3)

9.2.11 Effect of threshold for AE detection on STL

From simulation study of STL (Section 7.2), the STL method is insensitive to the variation of threshold values, which lie in the range from 20 %, 30 % and 40 % of the signal amplitude standard deviation. The probability of misdiagnosis was calculated using the test of significance (Table 7.6). The result shows that the probability of misdiagnosis is very remote indeed.

9.2.12 Limitation of V_{RMS} from vibration for low speed bearing monitoring

As can be seen from the progressive bearing wear test (Section 6.4.3), the traditional vibration V_{RMS} has been found to be insensitive to a bearing's incipient failure at the low speed of 0.23 rev/sec. The level of V_{RMS} only rose after the bearing had failed to operate (Figure 6.18), which is rather pointless for condition monitoring.

9.2.13 Cyclical shear stress causing shaft failure

From the progressive bearing wear test (Section 6.4.3), it was seen that when the bearing had seized up, the shaft broke at the plane next to the bearing (Figure 6.16). The shaft failure has the characteristic of fatigue failure. This fatigue failure is a result of the cyclical shear stress created by the excessive radially downward-acting load applied to the shaft using the hydraulic ram.

9.3 Recommendation for future work

The work on bearing condition monitoring and diagnosis based on AE and vibration techniques as an attempt for systems approach has been demonstrated throughout this thesis. Experimental and simulation work has been carried out and the results show the effectiveness of the system in fault detection and identification. The thesis has addressed all major issues concerning the creation of a monitoring system using both acceleration and AE signals. However, there are still some related questions that should be investigated before a truly robust monitoring system can be built.

9.3.1 Failure types

The types of abnormal conditions for rotodynamic machines investigated are rather limited. Not only can bearings be found on such machines, gearboxes are also a very common component. Gearbox problems, including cracked or missing gear teeth, should be a useful area of study. In order to add new failure types into an artificial neural network or a minimum distance classifier, there is need for training data.

9.3.2 Utilisation of other traditional AE parameters

The current work on traditional AE parameters concerns four different bearing conditions: balanced (normal), unbalanced, misaligned and defective bearing. In this case, it is found that the three AE parameters of event duration, peak amplitude and AE energy are good discriminatory features. However, when other failure types are considered, it is likely that other traditional AE parameters (such as rise time and ringdown count) may be more appropriate. Consequently further investigation should be conducted to explore the possibility.

9.3.3 Natural progressive bearing wear test with realistic load using STL

In the interest of time, the progressive bearing wear test was an accelerated test in that the applied load was set intentionally high. However, natural bearing wear test under realistic loading conditions should be performed. This will establish more accurately the STL warning threshold level.

9.3.4 Automatic threshold setting for AE detection

The STL method is based on the detection of AE events during the machine operation. In real applications, the huge variety of rotating machines with vastly different operating conditions poses a great challenge to the setting of AE detection threshold, made even more difficult by the presence of the ubiquitous background noise. Whilst the simulation study (Section 7.2) has answered the fundamental question from a theoretical viewpoint, it is prudent to verify this by empirical study. .

9.3.5 Implementation of STL method as portable device

The implementation of STL method as a portable device, which can be used in industry, should be developed in the future. This is because in industry, there are needs for condition monitoring devices, which are mobile on the shop floor. With today's digital signal processing technology, it is possible to develop such a device using the principle of the STL method.

References

Arvid P. (1945) Ball and Roller bearing engineering. England: SKF Industries, Inc.

Au J. H. and Butler C. (1990) Manufacturing measurement and quality systems 2, Great Britain: Brunel University. 18-1.

Barron R. (1996) Engineering condition monitoring: practice, methods and applications. England: Addison Wesley.

Barwell F. T. (1979) Bearing systems: Principles and practice. England. Oxford University Press.

Bashir I., Bannister R. and Robert D. (1999) Release of acoustic energy during the fatiguing of a rolling element bearing. Applied Energy. 62 97-111.

Boulahbal D., Golnaraghi M. F. and Ismail F. (1999) Amplitude and phase wavelet maps for the detection of cracks in geared systems. Mechanical Systems and Signal Processing. 13(3) 423-436.

Broch J. T. (1980) Mechanical Vibration and Shock measurements. Denmark: Bruel & Kjaer.

Butler D. E. (1973) The shock pulse method for the detection of damaged rolling bearings. Non-Destructive Testing. 92-95.

Charles W. R. (1998) The vibration monitoring handbook, Oxford: Coxmoor.

Choudhury A. and Tandon N. (2000) Application of acoustic emission technique for the detection of defects in rolling element bearings. Tribology International. 33 39-45.

Collacott R. A. (1977) Mechanical fault diagnosis and condition monitoring. London: Chapman and Hall.

Conforto S. and D'Alessio T. (1999) Spectral analysis for non-stationary signals from mechanical measurements: A parametric approach. Mechanical Systems and Signal Processing. 13(3) 395-411.

Dalianis S. A. and Hammond J. K. (1997) Time-frequency spectra for frequency-modulated processes. Mechanical Systems and Signal Processing. 11(4) 621-635.

Dalpiaz G. and Rivola A. (1997) Condition monitoring and diagnostics in automatic machines: Comparison of vibration analysis techniques. Mechanical Systems and Signal Processing. 11(1) 53-73.

Dalpiaz G., Rivola A. and Rubini R. (2000) Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. Mechanical Systems and Signal Processing. 14(3) 387-412.

Daubechies I. (1990) The wavelet transform, Time – frequency localisation and signal analysis. IEEE Transaction on Information Theory. 36 (5) 961-1005.

Donald F. W. and Richard E. B. (1957) Bearing design and application. London: McGraw-Hill.

Elsayed E. A. (1996) Reliability Engineering. Canada: Addison Wesley.

Eschmann, Hasbargen and Weigand (1958) Ball and Roller bearings: Their theory, design, and application. Germany: Kugelfischer Georg Schafer & Co.

Esteban J., Starr A. G. and Cross P. B. (1999) Data fusion: Models and procedures. Integrating Dynamics, Condition Monitoring and Control for the 21st Century. Rotterdam: A. A. Balkema. 427-432.

Gaul L. and Hurlebaus S. (1997) Identification of the impact location on a plate using wavelets. Mechanical Systems and Signal Processing. 12(6) 783-795.

Geropp B. (1997) Envelope analysis: A signal analysis technique for early detection and isolation of machine faults. IFAC Symposium on Fault Detection. 268-288.

Gertsbakh I. B. (1989) Statistical reliability theory. New York: Marcel Dekker, INC.

Gonzales R. C. and Woods R. E. (1993) Digital Image Processing. United States of America: Addison Wesley.

Han Y. S. (1999) Directional Wigner distribution for order analysis in rotating/reciprocating machines. Mechanical Systems and Signal Processing 13(5) 723-737.

Heng R. B. W. and Nor M. J. M. (1998) Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. Applied Acoustics. 53(3) 211-226.

Holroyd T. J. and Randall N. (1993) Use of acoustic emission for machine condition monitoring. British Journal of Non-Destructive Testing. 35 (2) 75-78.

Hu W., Starr A. G., Leung A. Y. T. and Qiu J. (1999a) Multisensor-based condition monitoring of CNC machine tools. Integrating Dynamics, Condition Monitoring and Control for the 21st Century. Rotterdam: A. A. Balkema. 139-144.

Hu W., Starr A. G., Leung A. Y. T. and Zhou Z. (1999b) Integrated hierarchical diagnostic reasoning for FMS's. Integrating Dynamics, Condition Monitoring and Control for the 21st Century. Rotterdam: A. A. Balkema. 125-130.

James L. C. and Li S. Y. (1995) Acoustic emission analysis for bearing condition monitoring. Wear. 185 (1) 67-74.

Jeffrey D. (1991) Principles of machine operation and maintenance. Oxford: Newnes.

John E. G. and Grosvenor R. I. (2000) Failure data analysis: The application of excel spreadsheet add-ins as an aid to curve fitting. Quality and Reliability Engineering International. 16 81-89.

Joshi R. and Sanderson A. C. (1999) Multisensor fusion – A minimal representation framework. Singapore: World Scientific Publishing.

Jun M. (1995) Detection of localised defects in rolling element bearings via composite hypothesis test. Mechanical Systems and Signal Processing. 9(1) 63-75.

Karl V. B. (1975) Statistical Models in Applied Science. America: John Wiley & Sons.

Kelly A. and Harris M. J. (1978) Management of Industrial Maintenance. London: Butterworths.

Kennedy J. B. and Neville A. M. (1986) Basic Statistical Methods for Engineers and Scientists. London: Happer and Row.

Lee S. K. and White P. R. (1997) Higher-order time-frequency analysis and its application to fault detection in rotating machinery. Mechanical Systems and Signal Processing. 11(4) 637-650.

Li C. J. and Ma J. (1996) Wavelet decomposition of vibrations for detection of bearing-localised defects. NDT&E International. 30(3) 143-149.

Li Y. Billington S., Zhang C., Kurfess T., Danyluk S. and Liang S. (1999) Adaptive prognostics for rolling element bearing condition. Mechanical Systems and Signal Processing. 13(1) 103-113.

Lin J. (2001) Feature extraction of machine sound using wavelet and its application in fault diagnosis. NDT&E International. 34 25-30.

Liu B. and Ling S. F. (1999) On the selection of informative wavelets for machinery diagnosis. Mechanical Systems and Signal Processing. 13(1) 145-162.

Liu T. I. (1996) Detection of roller bearing defects using expert system and fuzzy logic. Mechanical Systems and Signal Processing. 10(5) 595-614.

McCormick A. C. and Nandi A. K. (1997) Classification of the rotating machine condition using artificial neural networks. Proceeding of Instrumentation and Mechanical Engineering. 211 439-450.

McFadden P. D. and Smith J. D. (1984) Acoustic emission transducers for vibration monitoring of bearings at low speeds. Proceeding of Instrument and Mechanical Engineering. 198 (8) 127-130.

McFadden P. D., Cook J. G. and Forster L. M. (1999) Decomposition of gear vibration signals by the generalised S transform. Mechanical Systems and Signal Processing.13(5) 691-707.

Mechefske C. K. (1998) Objective machinery fault diagnosis using fuzzy logic. Mechanical Systems and Signal Processing. 12(6) 855-862.

Mechefske C. K. and Mathew J. (1995) Fault detection and diagnosis in low speed rolling element bearings using inductive inference classification. Mechanical Systems and Signal Processing. 9(3) 275-286.

Mellor D. J. (1988) Rolling bearing fault diagnosis. Bearing Conference. 5 1-6.

Michael N. (1979) A guide to the condition monitoring of machinery, London: Her Majesty's Stationary Office.

Miettinen J. and Andersson P. (2000) Acoustic emission of rolling bearings lubricated with contaminated grease. Tribology International. 33 777-787.

Mori K., Kasashima N., Yoshioka T. and Ueno Y. (1996) Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals. Wear. 195 162-168.

Moroney M. J. (1951) Facts from figures. Penguin Books Ltd: England. 220-221.

Newland D.E. (1993a) An introduction to random vibrations, spectral & wavelet analysis. London: Longman Scientific & Technical. 90-91.

Newland D.E. (1993b) An introduction to random vibrations, spectral & wavelet analysis. London: Longman Scientific & Technical. 92.

Nisbet T. S. and Mullett G. W. (1978) Rolling bearings in service: Interpretation of types of damage. London: Hutchison Benham.

O'Connor P. D. T. (1991) Probability plotting. Practical Reliability Engineering, 3rd Edition, Chichester: Wiley. Chapter 3.

Oehlmann H., Brie D., Tomczak M. and Richard A. (1997) A method for analysing gearbox faults using time-frequency representations. Mechanical Systems and Signal Processing. 11(4) 529-545.

Pachaud C., Salvetat R. and Fray C. (1997) Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces. Mechanical Systems and Signal Processing. 11 (6) 903-916.

Paya B. A. and Esat I. I. (1997) Artificial neural network based fault diagnostics of rotating machinery using wavelet transform as a pre-processor. Mechanical Systems and Signal Processing. 11(5) 751-765.

Peter R. N. C. (1998). Mechanical Design, Great Britain: Arnold.

Pineyro J., Klempnow A. and Lescano V. (2000) Effectiveness of new spectral tools in anomaly detection of rolling element bearings. Journal of Alloys and Compounds. 310 276-279.

Pollock A. (1973) Acoustic emission 2 - Acoustic emission amplitudes. Journal of Non-Destructive Testing. 10 264-269.

Qu L., Chen Y. and Liu X. (1989) A new approach to computer added vibration surveillance of rotating machinery. International Journal of Computer Applications in Technology, 2(2) 108-112.

Qu L., Liu X., Peyronne G. and Chen Y. (1989) The holospectrum: A new method for rotor surveillance and diagnosis. Mechanical Systems and Signal Processing 3(3) 255-267.

Randall R. B. (1977) Application of B&K equipment to frequency analysis. Denmark: Bruel & Kjaer.

Rioul O. and Vetterli M. (1991) Wavelets and signal processing. IEEE Signal Processing magazine. 14-37.

Rogers L. M. (1979) The application of vibration signature analysis and acoustic emission source location to on-line condition monitoring of anti-friction bearings. Tribology International. 9 51-59.

Ross T. J. (1995) Fuzzy logic with engineering applications. United States of America: McGraw-Hill, Inc.

Safizadeh M. S., Lakis A. A. and Thomas M. (2000) Using short time Fourier transform in machinery fault diagnosis. International Journal of COMADEM. 3(1) 5-16.

Sato I, Yoneyama T., Sato K., Tanaka T., Yanagibashi M. and Takikawa K. (1991) Applications of acoustic emission techniques for diagnosis of large rotating machinery and mass production products. Acoustic Emission: Current Practice and Future Directions, ASTM STP. 1077 287-304.

Scruby C. B. (1987) An introduction to acoustic emission. Journal of Physics Instrument Science and Technology. 20 946-953.

Shibata K., Takahashi A. and Shirai T. (2000) Fault diagnosis of rotating machinery through visualisation of sound signals. Mechanical Systems and Signal Processing. 14(2) 229-241.

Shiroishi J., Li Y., Liang S., Kurfess T., and Danyluk S. (1997) Bearing condition diagnostics via vibration and acoustic emission measurements. Mechanical Systems and Signal Processing. 11(5) 693-705.

SKF (1994) Bearing failures and their causes PI 401 E, Palmeblads Tryckeri AB.

Smith J. D. (1989) Vibration and measurement analysis. England: Butterworths.

SPM Instrument AB. (1970) SPM Technical information. Sweden: Strangnas.

Staszewski W. J. and Tomlinson G. R. (1997a) Local tooth fault detection in gearboxes using a moving window procedure. Mechanical Systems and Signal Processing. 11(3) 331-350.

Staszewski W. J., Worden K. and Tomlinson G. R. (1997b) Time-frequency analysis in gearbox fault detection using the Wigner-Wille distribution and pattern recognition. Mechanical Systems and Signal Processing. 11(5) 673-692.

Subrahmanyam M. and Sujatha C. (1997) Using neural networks for the diagnosis of localised defects in ball bearings. Tribology International. 30(10) 739-752.

Suzuki H., Kinjo T., Hayshi Y., Takemoto M. and Ono K. (1996) Wavelet transform of acoustic emission signals. Journal of Acoustic Emission. 14(2) 69-84.

Swindlehurst W. (1973) A series on acoustic emission - Introduction. Non-Destructive Testing. 152-158.

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

Tandon N. and Nakra B. C. (1992) Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. Tribology International. 25(3) 205-212.

Tang H., Cha J. Z., Wang Y., and Zhang C. (1991) The principle of Cepstrum and its application in quantitative fault diagnostics of gears. Modal Analysis, Modelling, Diagnostics, and Control - Analytical and Experimental ASME. 38 141-144.

Thomas F. D. (1996) A history of acoustic emission. Journal of Acoustic Emission. 14 (1) 1-34.

Tou J. T. (1974) Pattern recognition principles. United States of America: Addison-Wesley Publishing company.

Vyas N. S. and Satishkumar D. (2001) Artificial neural network design for fault identification in a rotor-bearing system. Mechanism and Machine Theory 36 157-175.

Wallace R. H. (1970) Understanding and measuring vibrations. London: Wykeham publications LTD.

Wang W. (2000) An investigation into the effects of load on several gear fault diagnostic techniques. International Journal of COMADEM. 3(3) 29-36.

Wang W. J. and McFadden P. D. (1993) Early detection of gear failure by vibration analysis – I. Calculation of the time-frequency distribution. Mechanical Systems and Signal Processing. 7(3) 193-203.

Wang W. J. and McFadden P. D. (1993) Early detection of gear failure by vibration analysis – II. Interpretation of time-frequency distribution using image processing technique. Mechanical Systems and Signal Processing. 7(3) 205-215.

Williams J. H., Davies A. and Drake P. R. (1994) Condition-based maintenance and machine diagnostics. London: Chapman & Hall.

Yoon D. J., Kwon O. Y., Chung M. H. and Kim K. W. (1995) Early detection of damages in journal bearings by acoustic emission monitoring. Journal of Acoustic Emission. 13 1-10.

Yoshioka T. and Fujiwara T. (1982) A new acoustic emission source locating system for the study of rolling contact fatigue. Wear. 81 183-186.

Ypma A., Ligteringen R., Frietma E. and Duin R. (1997) Recognition of bearing failures using wavelets and neural networks. Netherlands: Physics group.

Zhang C., Qiu J., Kurfess T. R., Danyluk S. and Liang S. Y. (2000) Impact dynamics modelling of bearing vibration for defect size estimation. International Journal of COMADEM. 3(3) 37-42.

Zhang S., Ganesan R., and Xistris G. D. (1996) Self-organising neural networks for automated machinery monitoring systems. Mechanical Systems and Signal Processing. 10(5) 517-532.

Appendix – A

Bearing defect frequencies calculation

Bearing characteristic defect frequencies

Characteristic defect frequencies are effectively the roller passing frequencies for defect on different locations of a bearing. Roller passing frequencies can be calculated from the geometry and speed of a bearing (Collacott, 1977; Charles, 1998). Figure A1 illustrates a cross sectional view of a ball bearing in which the inner race, a ball and the outer are shown. The ball-passing frequencies can be calculated as the equations shown below.

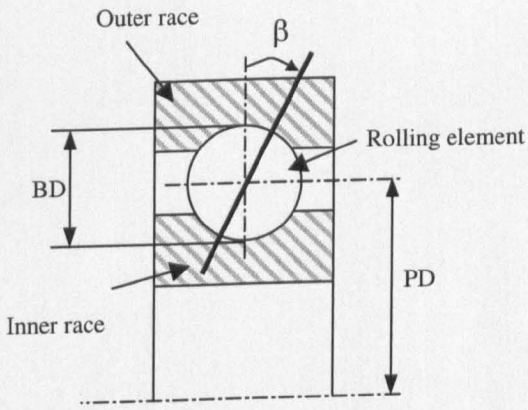


Figure A-1 Cross-sectional view of a rolling element bearing

$$f_{or} = \frac{n}{2}.\frac{N}{60}\left(1-\frac{BD}{PD}\cos\beta\right)$$

$$f_{ir} = \frac{n}{2}.\frac{N}{60}\left(1+\frac{BD}{PD}\cos\beta\right)$$

$$f_{ro} = \frac{PD}{BD}.\frac{N}{60}\left(1-\left(\frac{BD}{PD}\right)^{2}\cos\beta\right)$$

$$f_{c} = \frac{N}{120}\left(1-\frac{BD}{PD}\cos\beta\right)$$

where f_{or} = characteristic defect frequency for defect at outer race

f_{ir} = characteristic defect frequency for defect at inner race

f_{ro} = characteristic defect frequency for defect at rolling element

f_c = characteristic defect frequency for defect at cage

N = rotational speed (rev/min)

BD = rolling element diameter

PD = bearing pitch diameter

β = contact angle between raceway and rolling element (degree)

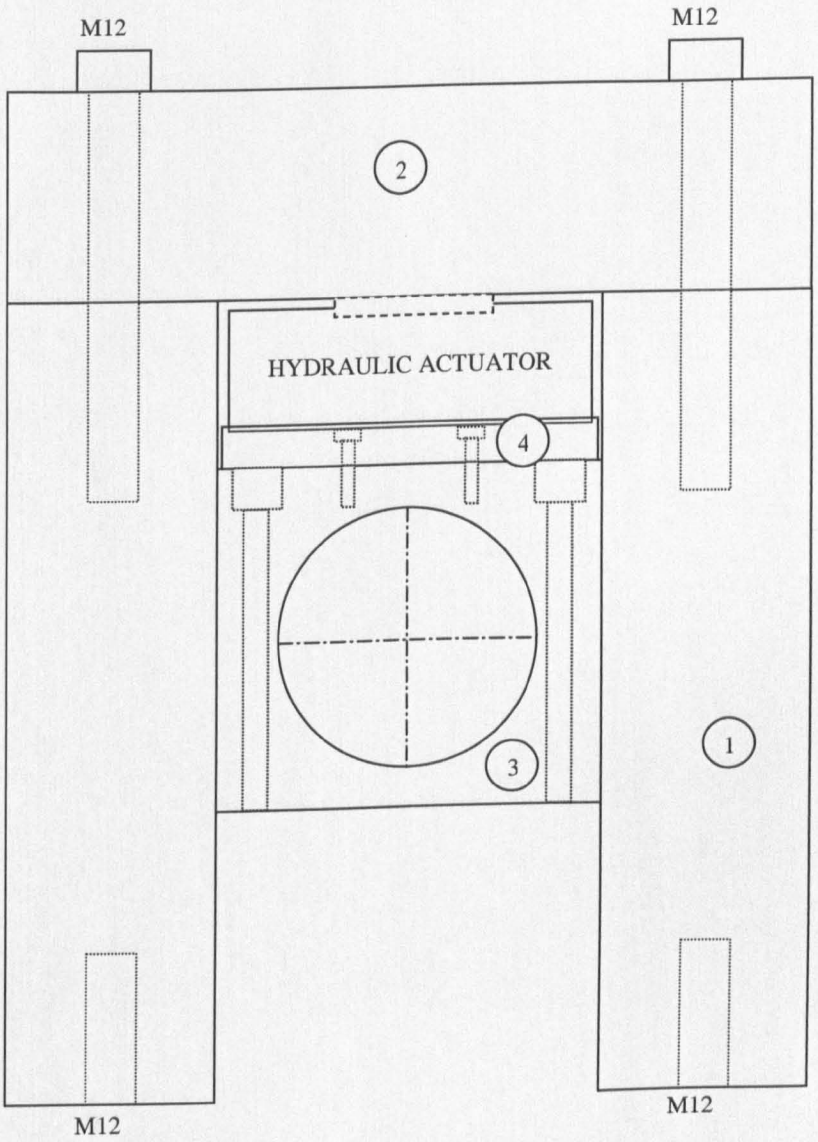
n = number of rolling elements

Appendix – B

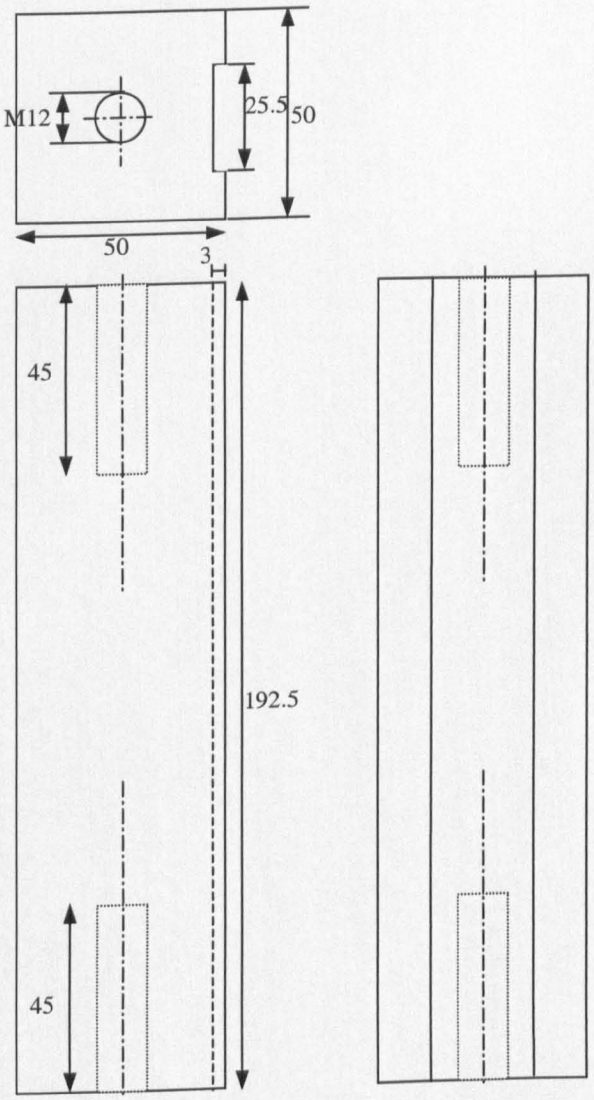
Technical Drawings

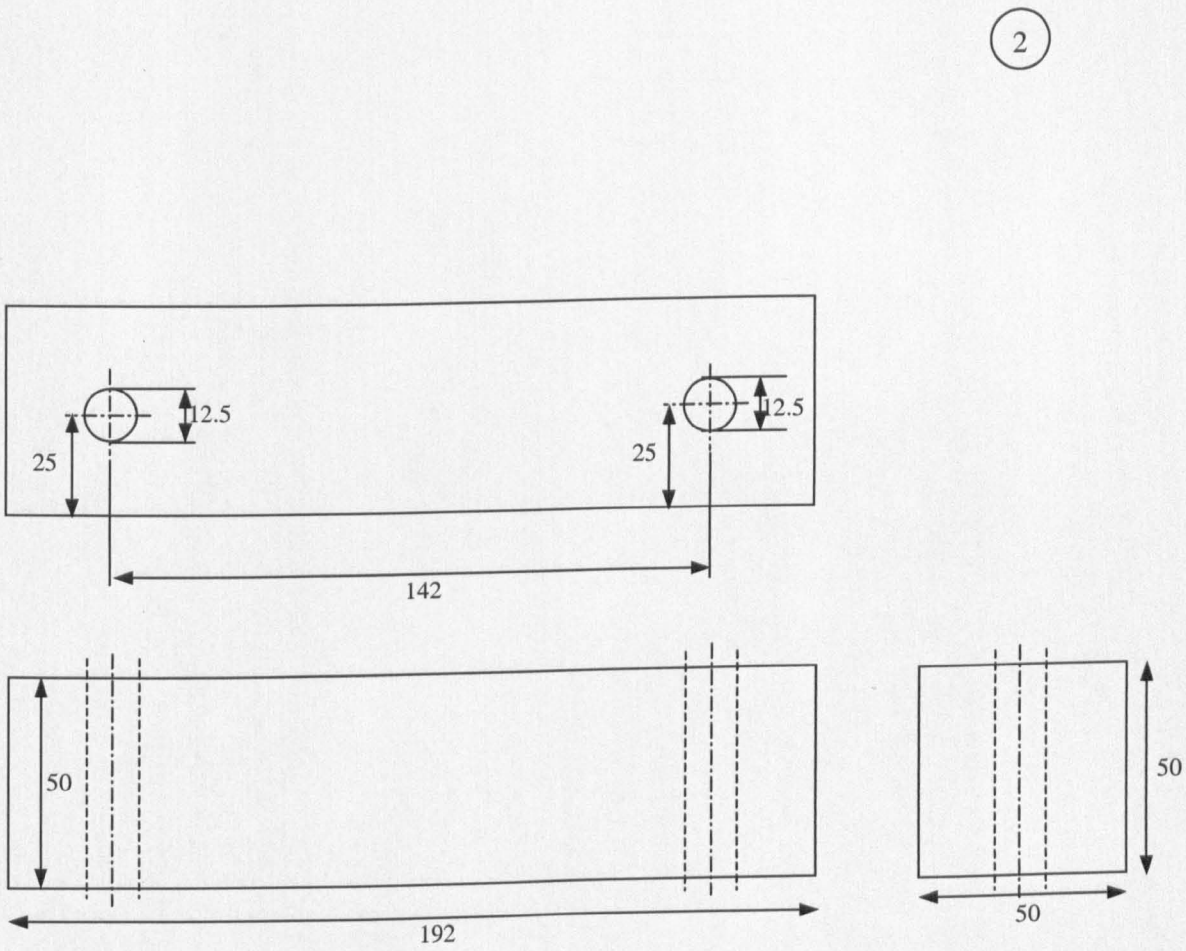
Modified hydraulic loading parts list

Part no.	Description	Dimension	Material	Quantity
1	Supported frame	50x50x192.5	AL	2
2	Cover frame	50x50x192	AL	1
3	Bearing supported frame	25x92x82	AL	1
4	Hydraulic supported fixture	90x90x10	AL	1
Project name: Modified hydraulic loading system				
By: Tonphong Kaewkongka **Supervisor**: Dr Y H Joe Au				

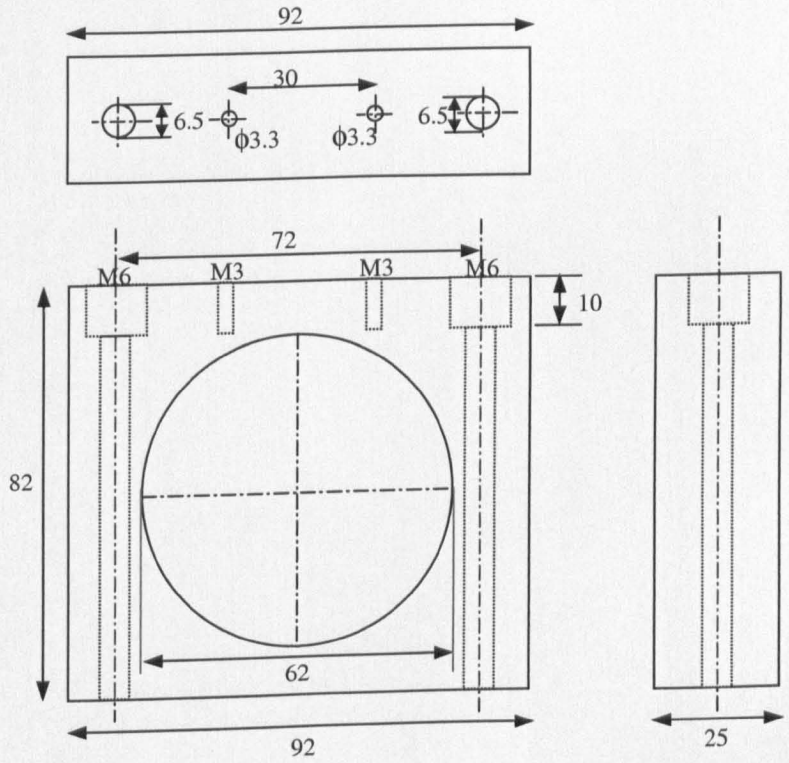


1

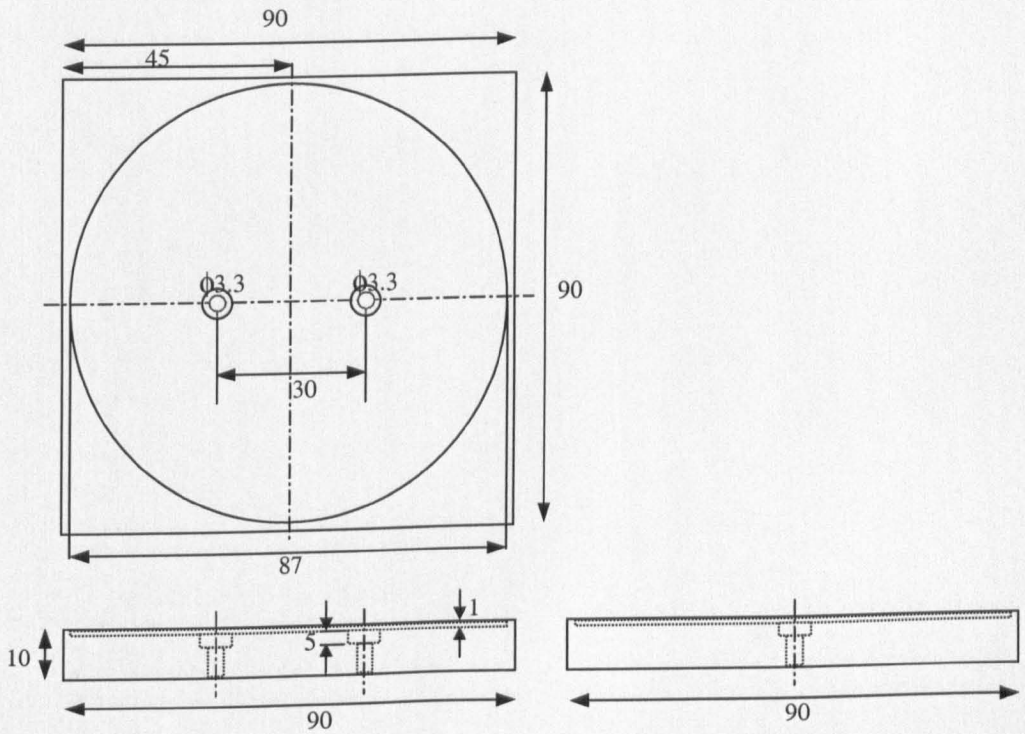



2
12.5
25
12.5
25
142
50
192
50
50

3



4



Appendix – C

Computer Programs

C1: MATLAB source code for fuzzy C-mean clustering technique

```
clear all
clc
load data.dat
fcmdata = data;

figure,plot3(fcmdata(:,1),fcmdata(:,2),fcmdata(:,3),'.','MarkerSize',15)
xlabel('Event duration')
ylabel('Energy')
zlabel('Peak amplitude')
figure

[center, U, obj_fcn] = fcm(fcmdata, 4);
maxU = max(U);
index1 = find(U(1, :, :) == maxU);
index2 = find(U(2, :, :) == maxU);
index3 = find(U(3, :, :) == maxU);
index4 = find(U(4, :, :) == maxU);
line(fcmdata(index1, 1), fcmdata(index1, 2), fcmdata(index1, 3), 'linestyle',...
'none','marker', 'o','color','g');
line(fcmdata(index2,1),fcmdata(index2,2), fcmdata(index2, 3),'linestyle',...
'none','marker', 'x','color','r');
line(fcmdata(index3,1),fcmdata(index3,2), fcmdata(index3, 3),'linestyle',...
'none','marker', '+','color','m');
line(fcmdata(index4,1),fcmdata(index4,2), fcmdata(index4, 3), 'linestyle',...
'none','marker', '*','color','b');

hold on
legend('Balanced shaft','Misaligned shaft','Unbalanced shaft','Defective bearing',4);


title('Clustering of balanced, unbalanced, misaligned and defective bearing');
xlabel('Event duration')
ylabel('Energy')
zlabel('Peak amplitude')
plot3(center(1,1),center(1,2),center(1,3),'kx','markersize',15,'LineWidth',2)
plot3(center(2,1),center(2,2),center(2,3),'ko','markersize',15,'LineWidth',2)
plot3(center(3,1),center(3,2),center(3,3),'ks','markersize',15,'LineWidth',2)
plot3(center(4,1),center(4,2),center(4,3),'kd','markersize',15,'LineWidth',2)
```

C2: MATLAB source code for minimum distance classification

```
% Minimum distance classification
% -- > Calculating Euclidean distance
% For user:
% --> vary test# as a testing set
% --> vary ref(#,..) as its condition
% when ref(1,n) Balanced
%      ref(2,n) Misaligned
%      ref(3,n) Unbalanced
%      ref(4,n) Defective bearing

for i = 1:4,
  for j = 1:4,
    result(i,j)=0;
  end
end

for i = 1:4,
  for j = 1:4,
    result(i,j) = sqrt((test(j,1)-ref(i,1))^2+(test(j,2)-ref(i,2))^2+(test(j,3)-ref(i,3))^2);
    j = j+1;
  end
  i = i+1;
end

out = result'

for i = 1:4,
  for j = 1:4,
    out2(i,j) = out(i,j)/max(out(:,j));
    j = j+1;
  end
  i = i + 1;
end
```

C3: MATLAB source code for simulated signal for comparison between CWT and STFT

The source codes of the four types of waveform created in Matlab are shown below:

Type 1 Waveform:

```
t = 1:1:350;
s = sin(100.*linspace(0,pi,350)) + 0.2*randn(1,350);
x = exp(-0.01*t);
y = (s.*x);
Signal = [zeros(1,50) y];
figure,plot(Signal);
```

Type 2 Waveform:

```
t = 1:1:350;
s = sin(200.*linspace(0,pi,350)) + 0.2*randn(1,350);
x = exp(-0.02*t);
y = (s.*x);
s2 = 0.2.*sin(10.*linspace(0,pi,250)) + 0.08*rand(1,250);
s3 = 0.15.*sin(5.*linspace(0,pi,250))+ 0.05*randn(1,250);
s_ = s2 + s3;
ss = [zeros(1,150) s_];
Sig = [zeros(1,50) y];
Signal = Sig + ss;
figure,plot(Signal);
```

Type 3 Waveform:

```
t = 1:1:350;
x = exp(-0.01*t);
s1 = 0.8.*sin(30.*linspace(0,pi,350)) + 0.2*randn(1,350);
s2 = 0.5.*sin(50.*linspace(0,pi,350))+ 0.2*randn(1,350);
s_ = s1+s2;
s = (x.*s_);
ss = [zeros(1,50) s];
Signal = ss;
figure,plot(Signal);
```

Type 4 Waveform:

```
t1 = 1:1:100;
s1 = sin(5.*linspace(0,pi,100)) + 0.2*randn(1,100);
x1 = exp(-0.06*t1);
y1 = 1.2*(s1.*x1);
s2 = sin(8.*linspace(0,pi,150)) + 0.2*randn(1,150);
t2 = 1:1:150;
x2 = exp(-0.09*t2);
y2 = 1.2*(s2.*x2);
s = [zeros(1,50) y1 zeros(1,100) y2];
figure,plot(s);
```

C4: MATLAB source code for simulated signal for different inter-arrival times of successive AE events followed Weibull distribution

```
clear all;

% Generating different times of arrival followed Weibull distribution
% Define parameters: shape and characteristic life

shape = 1;
life = 0.5

% Good condition      -> life = 0.5
% Warning condition  -> life = 0.05
% Faulty condition    -> life = 0.025

in = weibrnd(life,shape,[1 50]);
in2 = sort(in)';
in3 = cumsum(in2);
t_in = cumsum(in)';
t = linspace(0,in3(length(in3)),1000);
delta = t(2)-t(1);
t_start = t_in;

for i = 1:100;
  temp(i) = 0;
end

N = length(t_start);
M = length(t);
t1 = linspace(0,in3(length(in3))/100,1000);

% Generating distribution of peak follows Rayleigh distribution
% with standard deviation = 1

sigma_y = 1;
a = raylrnd(sigma_y,1,N);
```

C5: MATHCAD source code for 3-parameters estimation of Weibull distribution

TC := input file

N := rows(TC) N = 149

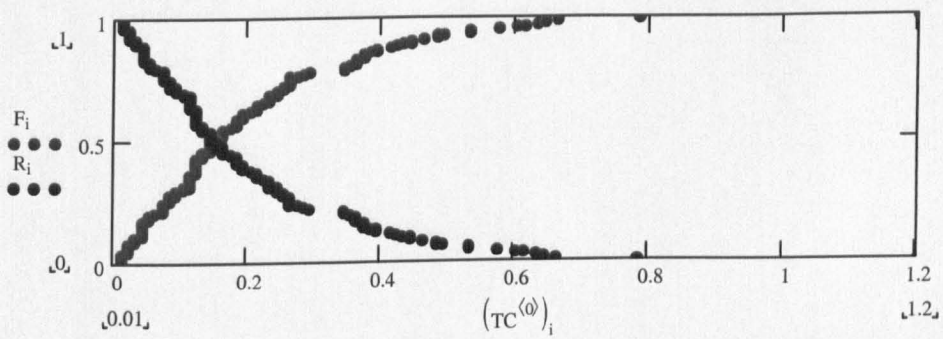
Ne := $TC_{(N-1),1}$ Ne = 149

$$F := \frac{TC^{\langle 1 \rangle}}{Ne}$$

R := 1 − F

The graphs of F and R are plotted below as:

i := 0.. rows (TC) − 1



$$R(t) = \exp\left[-\left(\frac{t - t_o}{\eta}\right)^{\beta}\right]$$

where η is the scale parameter and β the shape parameter.

$\eta := 3$ $\beta := 1$ $t_o := 0$

Given

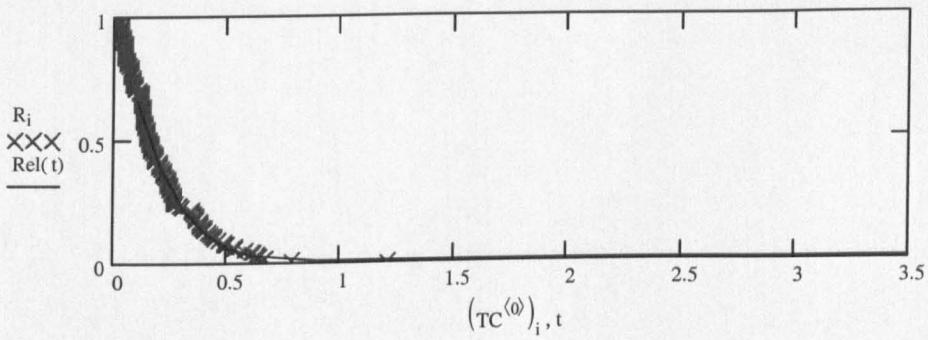
$$\sum_{j=0}^{\text{rows}(TC)-1}\left[R_j - \exp\left[-\left[\frac{\left(TC^{\langle 0\rangle}\right)_j - t_o}{\eta}\right]^{\beta}\right]\right]^2 = 0$$

$$\begin{pmatrix}\eta\\ \beta\\ t_o\end{pmatrix} := \text{Minerr}\left(\eta, \beta, t_o\right) \qquad \text{ERR} = 0.052$$

$\eta = 0.209$ $\beta = 1.165$ $t_o = 9.017\times 10^{-4}$

$t := 0, 0.1 .. 3.5$

$$\text{Rel}(t) := \exp\left[-\left(\frac{t - t_o}{\eta}\right)^{\beta}\right]$$



Appendix – D

Published papers

Doctoral Research Conference
DRC2000, Runnymede.

Continuous wavelet transform and neural network for condition monitoring of rotodynamic machinery

T Kaewkongka, Y H Joe Au, R T Rakowski and B E Jones

Department of Systems Engineering, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom
Telephone: +44 1895 274000, E-mail : empgttk@brunel.ac.uk

Abstract

This paper describes a novel method, Figure 1, of rotodynamic machine condition monitoring and recognition using wavelet and neural network. Since in the real machinery environment, there can be transients or abrupt changes in the measured signal, traditional analyses in either time- or frequency-domain are not always suitable for revealing machine faults. The proposed approach is to utilise the joint time-frequency domain method by applying wavelet transform to the measured signal. The transformed signals are represented as grey-scale images, simplified to binary images, which may contain characteristic features relating to the various types of faults. The act of binarising the image has the effect of removing the broadband noise from the signal before the final stage of classification is performed using a back-propagation neural network.

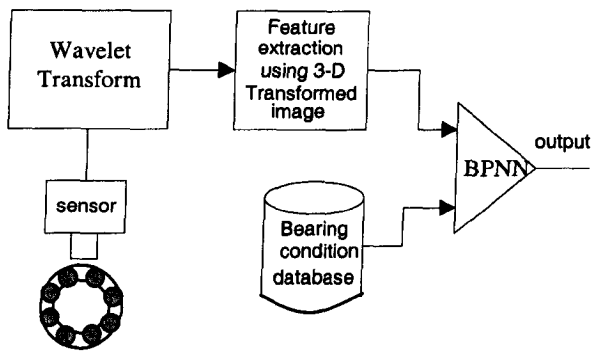


Figure 1 The proposed condition monitoring of rotodynamic machinery block diagram.

1. Introduction

With ever growing competition in industry, the need to automate machine condition monitoring has become more and more acute. A reliable condition monitoring system will significantly reduce failure and unplanned maintenance, and hence the huge attendant cost due to machine downtime. Often, the system is used with an operator who assists in the interpretation of the machine signals for early failure detection and fault diagnosis.

Nowadays there are two basic methods available for bearing maintenance: statistical bearing life estimation, and bearing condition monitoring and diagnostics [1]. Statistical bearing life estimation predicts the fatigue life of a bearing. However, its application has many limitations. Since unusual operating conditions can severely decrease a bearing's life, bearing life estimates become unreliable leading to unexpected breakdown. On the other hand, bearing condition and diagnostics can be a very reliable method because it gives up-to-date information about the condition of a bearing. The more popular techniques used for bearing condition monitoring are vibration and acoustic emission analyses.

This paper is organized as follows. After the introduction in Section 1, time-domain and frequency-domain methods are briefly reviewed in Section 2, followed by a description of the methodology of the proposed method in Section 3. The experiments are presented in Section 4 with results and discussions in Section 5; finally, conclusions are given in Section 6.

2. Previous Work

Currently, there are many conventional methods for identifying and diagnosing bearing faults. Based on the representation of a signal during its processing, a method can be referred to as time-domain or frequency-domain.

Time domain methods are usually sensitive to impulsive oscillations. Characteristics arising from the defects being monitored, also known as features, of the raw time signal can be extracted from a machine. Typical features are the r.m.s. value, peak value [2], crest factor, kurtosis [3] and the shape, size and orientation of a bearing locus derived from orbital analysis. These features, once established to be related to the defect being monitored, often yield satisfactory results. However, if the signal generation mechanism is complex, time-domain methods are often not refined enough.

Frequency domain methods assume that the signal being analyzed have components that are periodic. Thus, a defect produces a periodic signal at the characteristic defect frequency. Examples of frequency-domain methods include spectrum analysis, cepstrum analysis, high frequency resonance technique (HFRT) [4] and holospectrum [5]. Among them, spectrum analysis seems to dominate the fault diagnosis scene. The main limitation of spectrum analysis is that although a local transient will contribute towards the overall frequency spectrum, its location on the time axis is lost. There is no way of knowing whether a particular frequency component exists throughout the life of the time signal or

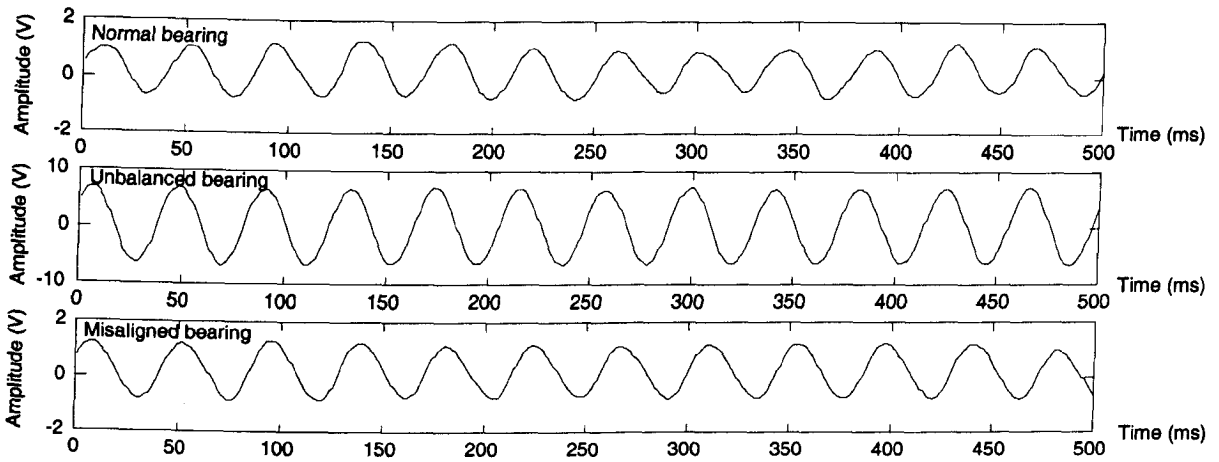


Figure 2 Acceleration signals from the three bearing conditions

during just one or a few selected periods. Unfortunately, many monitoring situations demand knowledge of not just the frequency composition of a transient but also its time of occurrence. When a rolling element passes a localized defect in a bearing, it generates a transient in the measured signal, as does the contact of a damaged tooth with other teeth in a gearbox. A machine with rapidly varying speed is another example of transient events.

Continuous Wavelet Transform (CWT), a joint time- and frequency-domain technique, is proposed in this paper. CWT is capable of indicating abrupt changes in machine conditions [6]. In addition, it can give a better representation of the signal than conventional methods, providing fuller information on the machine operating condition. CWT is used here to produce a 3-D image from the measured signal. Features are then extracted from the image to be used as inputs to a back propagation neural network, a tool for assessing the discriminatory power of CWT.

3. Methodology

The proposed method of machine bearing faults recognition, Figure 1, using continuous wavelet transform and back propagation neural network consists of three steps.

Step 1 Applying continuous wavelet transform

Continuous wavelet transform of a time-signal $f(t)$ is defined by [7]

$$F_{\Psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt$$

The quantity $\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right)$ given in the definition is referred to as the wavelet function. The position variable b shifts the wavelet function along the time axis t of $f(t)$ while the scale variable a expands or compresses the wavelet function $\Psi_{a,b}(t)$. Compared to Fourier transform, the scale variable a is equivalent to the function $\Psi_{a,b}(t)$ and the time-signal $f(t)$ at the scale a and position b of $\Psi_{a,b}(t)$.

Figure 2 shows the acceleration signals obtained from the three machine conditions: normal bearing, unbalanced bearing and misaligned bearing. CWT was applied to these signals to calculate the coefficients $F_{\Psi}(a,b)$, which were then displayed as a grey-scale map with the vertical and horizontal axes denoting the scale a and position b, which is equivalent to time t, as shown in Figure 3 for the three different conditions. Although there is general similarity between them, subtle differences are still noticeable.

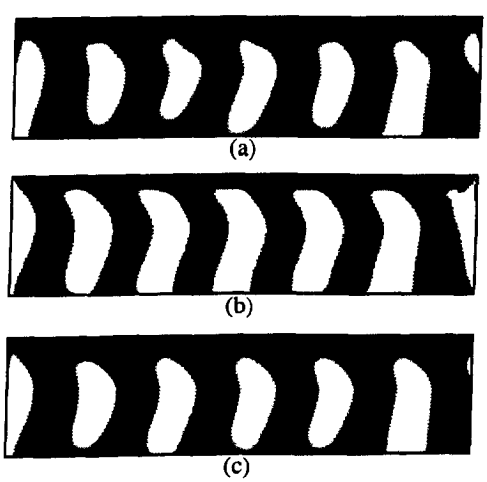


Figure 3 Transformed images of (a) normal,(b) unbalanced and (c) misaligned bearing

Step 2 Preprocessing

Image preprocessing was performed using the method of thresholding or binarising [8]. It was applied to the grey-scale CWT image to convert the colour of each pixel into either black or white. A binary image has the obvious advantage that when classification using neural networks is done, the computation time will be much shortened, as multiplication involving a 0 or 1 is much easier to perform. Figures 4a and 4b show a sample image before and after binarising.

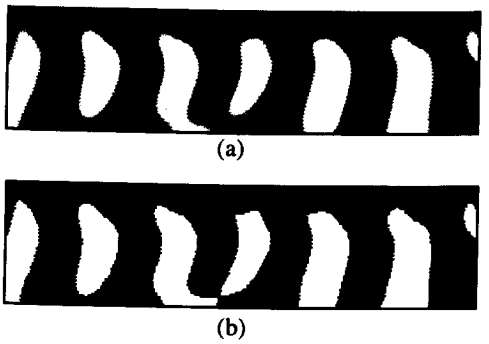


Figure 4 The original image (a) and its binary image (b)

Step 3 Classifying

The back propagation neural network [9], with the architecture as shown in Figure 5, was used to classify bearing faults. The network has an input layer, a hidden layer and an output layer. The values of the pixels composing the CWT binary image provided the inputs to the neural network; and 3 nodes, representing the bearing conditions of normal, unbalanced and misaligned, were available as outputs. Through a process of trial and error based on minimizing the mean square error (MSE), the choice of 5 nodes in the hidden layer was determined to be optimal.

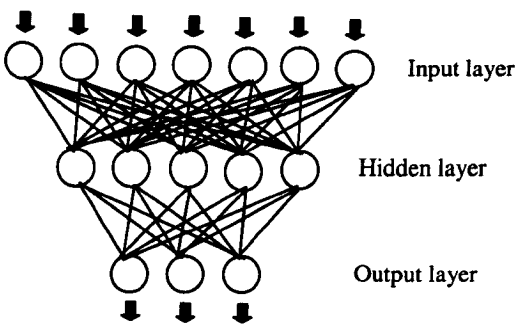


Figure 5 BPNN architecture

The value *NET* appearing at a node in a layer is computed by summing the products of all inputs leading to that node with their corresponding weights plus a bias. This value then forms the argument of a transfer function f that produces an output for the node. In vector notation, given the input vector X_i, the weight vector W_i, and the bias vector θ_i, the output is given by

$$OUT = f(NET) = f(\sum_{i}^{n} X_i W_i + \theta_i)$$

The transfer function f used is the commonly used sigmoid function defined as

$$f = \frac{1}{1 + e^{(-NET)}}$$

The sigmoid function acts as an output gate that can be opened (0) or closed (1). The computing process as described is schematically shown in Figure 6.

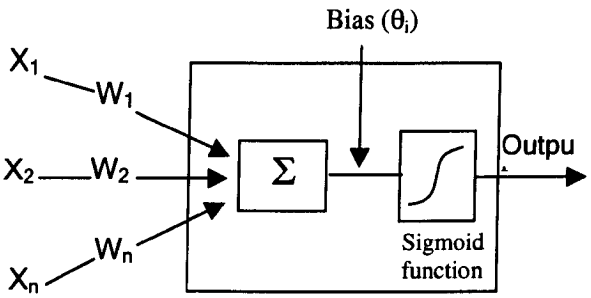


Figure 6 Schematic process of BPNN

The back propagation algorithm is used to obtain the correct weights and biases in a training process. A set of training data with known outputs is fed into the network. The weights are initially set to random values; the biases are fixed at unity. The input data are presented to the network; outputs are calculated and compared with the desired outputs. The normalized mean square error (MSE) is then calculated and propagated back to adjust the weights on the neural connection. This process is repeated for a large number of epochs until the error is relatively low and acceptable which allows the network to classify the test set correctly.

4. Experimental verification of the proposed method

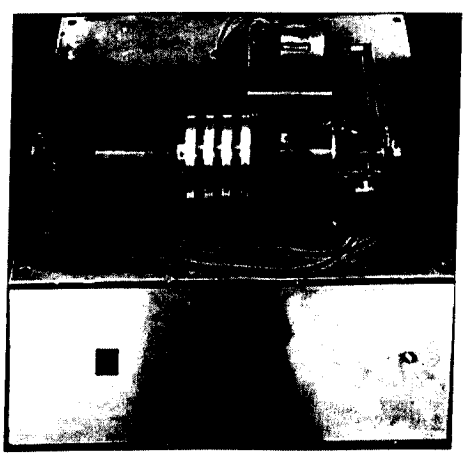


Figure 7 Test rig set-up

Experiments were conducted on a rotodynamic test rig, Figure 7, consisting of a rotating shaft driven by a DC motor at 20 rev/sec. The shaft, supported in two bearings, carried 4 discs with attachable masses in order to produce rotating unbalance. The two bearings were an FAG 20205K.T.C3 self-aligning single row taper-bore bearing and an FAG 6304 ball bearing. Radial vibration at the two bearings was measured using PCB 333A12 accelerometers, one on top of the non-

drive end housing while the other on top of the drive-end housing labeled as #1 and #2 respectively in Figure 7. The acceleration signals, having been low-pass filtered at 1 kHz for anti-aliasing, were sampled into a Schlumberger FFT analyser.

Measurements were obtained from three different machine conditions: normal, unbalanced and misaligned. For each condition, five signals were collected which were divided into two sets of three and two signals respectively. The first set was used for training the neural network while the second set was used for testing it.

As mentioned in Step 2, it was the binarised CWT image of a signal that was being used in either training or testing. The image consisted of 64 scales and 252 time intervals giving a resolution of 64x252=16128 pixels.

5. Results and discussion

Table 1 shows the correlation results obtained using the two signals from each of the three test sets. It should be noted that a good match returns a correlation coefficient of around 1 while a poor match returns a value of around 0.

Table 1 Correlation coefficient results

Bearing condition	Training sets		
Test sets	Normal	Unbalanced	Misaligned
Normal	0.97402	0.02054	0.01047
	0.98377	0.01019	0.01108
Unbalanced	0.02120	0.97891	0.00314
	-0.00087	1.00059	0.00029
Misaligned	0.49346	-0.03897	0.61209
	0.11262	-0.07624	1.00747

With the threshold value set at 0.5, it can be seen from Table 1 that perfect recognition has been achieved in every case.

6. Conclusion

The experiments indicated that continuous wavelet transform of acceleration signals from three different rotodynamic machine conditions – normal, unbalanced and misaligned – produced distinct correlation coefficient images that could be discriminated using the back propagation neural network. The image used was a simplified binarised image in which each pixel assumed a value of either 0 or 1. This allows faster computation in the neural network. The neural network had as many inputs as there were pixels of the image, that is, 16128 pixels. It had one hidden layer with five nodes and three outputs corresponding to the three discriminating conditions of normal, unbalanced and misaligned. The recognition rate achieved with the limited test sets was 100%.

Wavelet transform has the main advantage that it provides information of the signal on scale (frequency) as well as on time, compared to Fourier transform where only frequency information is available. Wavelet transform is therefore most suited to analyzing signals that are not stationary, a situation often encountered in condition monitoring.

Acknowledgements

The authors wish to thank Corrus, Middlesborough, who kindly provided material help and advice for experimentation and the Thai government for its support.

References

[1] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk and S. Liang, "Adaptive prognostics for rolling element bearing condition," Mechanical systems and signal processing, Vol.13(1), pp.103-113, 1999.

[2] S. Zhang, R. Ganesan, and G. D. Xistris, "Self-organizing neural networks for automated machinery monitoring systems," Mechanical systems and signal processing, Vol.10(5), pp. 517-532,1996.

[3] C. Pachaud, R. Salvetat and C. Fray, "Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces," Mechanical systems and signal processing, Vol.11(6), pp. 903-916, 1997.

[4] J. Shiroishi, Y. Li, S. Liang, T. Kurfess and S. Danyluk, "Bearing condition diagnostics via vibration and acoustic emission measurement," Mechanical systems and signal processing,Vol.11(5), pp. 693-705, 1997.

[5] L. Qu, X. Liu, G. Peyronne and Y. Chen, "The Holospectrum: A new method for rotor surveillance and diagnosis," Mechanical systems and signal processing, Vol.3(3), pp. 255-267, 1989.

[6] H. Suzuki, T. Kinjo, Y. Hayshi, M. Takemoto and K. Ono, "Wavelet transform of acoustic emission signals," Journal of acoustic emission, Vol.14(2), pp. 69-84, 1996.

[7] L. Gaul and S. Hurlebaus, "Identification of the impact location on a plate using wavelets," Mechanical systems and signal processing, Vol.12(6), pp. 783-795, 1997.

[8] Rafael C. Gonzales, Richard E. Woods, Digital Image Processing, Addison Wesley, 1993.

[9] B. A. Paya and I. I. Esat, "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor," Mechanical systems and signal processing, Vol.11(5), pp. 751-765, 1997.

IEEE Instrumentation and Measurement
Technology Conference
Budapest, Hungary, May 21-23, 2001.

Continuous wavelet transform and neural network for condition monitoring of rotodynamic machinery

T Kaewkongka, Y H Joe Au, R T Rakowski and B E Jones
The Brunel Centre for Manufacturing Metrology,
Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Phone: (+44) 01895 274000 ext. 2608, E-mail: joe.au@brunel.ac.uk

***Abstract** - This paper describes a novel method of rotodynamic machine condition monitoring using a wavelet transform and a neural network. A continuous wavelet transform is applied to the signals collected from accelerometer. The transformed images are then extracted as unique characteristic features relating to the various types of machine conditions. In the experiment, four types of machine operating conditions have been investigated: a balanced shaft, an unbalanced shaft, a misaligned shaft and a defective bearing. The back propagation neural network (BPNN) is used as a tool to evaluate the performance of the proposed method. The experimental results result in a recognition rate of 90 percent.*

***Keywords** – Wavelet transform, neural network, rotodynamic machinery.*

I. INTRODUCTION

With ever growing competition in industry, the need for machine condition monitoring has become more important. A reliable condition monitoring system will significantly reduce failure and unplanned maintenance, and hence the huge attendant cost due to machine downtime. Often, the system is used with an operator who assists in the interpretation of the machine signals for early failure detection and fault diagnosis.

Nowadays there are two kinds of methods available for bearing maintenance: statistical bearing life estimation and bearing condition monitoring and diagnostics [1]. Statistical bearing life estimation predicts the fatigue life of a bearing. However, its application has many limitations, since unusual operating conditions can severely decrease a bearing's life. Bearing life estimates become unreliable leading to unexpected breakdown. On the other hand, bearing condition and diagnostics can be a very reliable method because it gives up-to-date information about the condition of a bearing. The more popular techniques used for bearing condition monitoring are vibration and acoustic emission analyses.

II. PREVIOUS WORK

Currently, there are many conventional methods for identifying and diagnosing bearing faults. Based on the representation of a signal during its processing, a method can be referred to as time-domain or frequency-domain.

Time domain methods are usually sensitive to impulsive oscillations. Characteristics arising from the defects being monitored, also known as features, of the raw time signal can be extracted from a machine. Typical features are the r.m.s. value, peak value [2], crest factor, kurtosis [3] and the shape, size and orientation of a bearing locus derived from orbital analysis. These features, once established to be related to the defect being monitored, often yield satisfactory results. However, if the signal generation mechanism is complex, time-domain methods are often not refined enough.

Frequency domain methods assume that the signal being analyzed has components that are periodic. Thus, a defect produces a periodic signal at the characteristic defect frequency. Examples of the frequency-domain methods include spectrum analysis, cepstrum analysis, high frequency resonance technique (HFRT) [4] and holospectrum [5]. Among them, spectrum analysis seems to dominate the fault diagnosis scene. The main limitation of spectrum analysis is that although a local transient will contribute towards the overall frequency spectrum, its location on the time axis is lost. There is no way of knowing whether a particular frequency component exists throughout the life of the time signal or during just one or a few selected periods. Unfortunately, many monitoring situations demand knowledge of not just the frequency composition of a transient but also its time of occurrence. For instance, when a rolling element passes a localized defect in a bearing, it generates a transient in the measured signal, as does the contact of a damaged tooth with other teeth in a gearbox. A machine with rapidly varying speed is another example of transient events.

The continuous wavelet transform (CWT), a joint time- and frequency-domain technique, is proposed in this paper. CWT is capable of indicating abrupt changes in machine conditions [6]. In addition, it can give a better representation of the signal than conventional methods, providing fuller information on the machine operating condition. CWT is used here to produce a 3-D image from the measured signal. Features are then extracted from the image to be used as

inputs to a back propagation neural network, the tool used for assessing the discriminatory power of CWT.

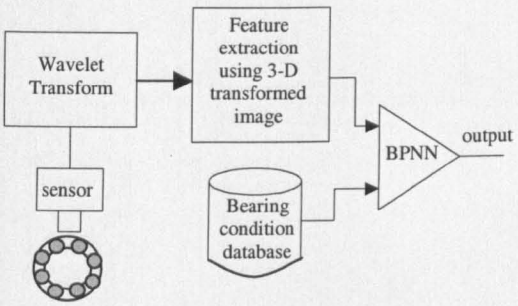


Fig. 1. The proposed condition monitoring of rotodynamic machinery block diagram.

III. METHODOLOGY

The proposed method of machine bearing faults recognition (Fig. 1), using continuous wavelet transform and back-propagation neural network consists of three steps.

Step 1 Applying continuous wavelet transform

Continuous wavelet transform of a time-signal $f(t)$ is defined by [7]

$$F_{\Psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt \qquad (1)$$

The quantity $\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right)$ given in the definition is referred to as the wavelet function. The position variable b shifts the wavelet function along the time axis t of $f(t)$ while the scale variable a expands or compresses the wavelet function $\Psi_{a,b}(t)$. Compared to Fourier transform, the scale variable a is equivalent to the inverse of the frequency. The definition also suggests that $F_{\Psi}(a,b)$ is the correlation coefficient between the wavelet function $\Psi_{a,b}(t)$ and the time-signal $f(t)$ at the scale a and position b of $\Psi_{a,b}(t)$.

The continuous wavelet transform was applied to the acceleration signals from the four machine conditions: (a) balanced shaft, (b) unbalanced shaft, (c) misaligned shaft and (d) defective bearing. Samples of these signal are as shown in Fig. 2. CWT was applied to these signals to calculate the coefficients $F_{\Psi}(a,b)$, which were then displayed as a grey-scale map with the vertical and horizontal axes denoting respectively the scale a and position b, which is equivalent to time t, as shown in Fig. 3 for the four different conditions. Although there is general similarity between them, subtle differences are clearly visible.

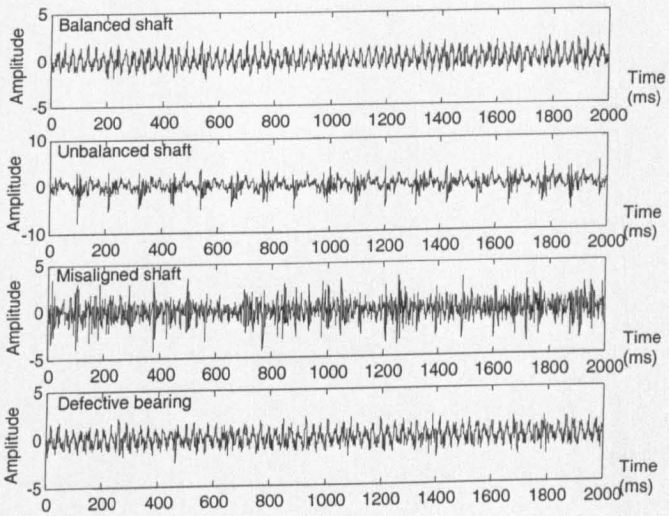


Fig. 2. Acceleration signals from the four bearing conditions

Step 2 Preprocessing

Image preprocessing was performed using the method of thresholding or binarising [8]. It was applied to the grey-scale CWT image to convert the colour of each pixel into either black or white.

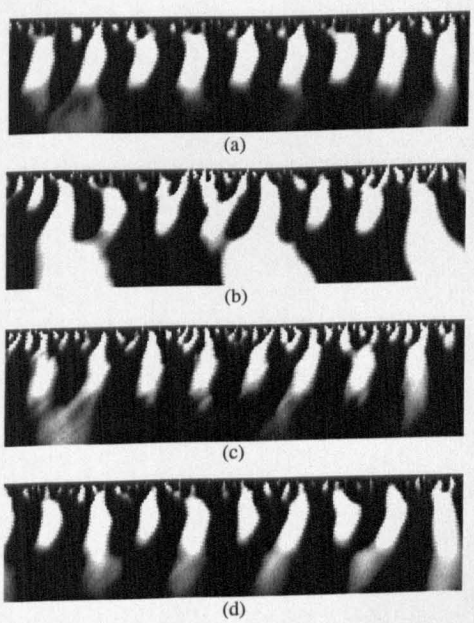


Fig. 3. Transformed images of (a) Balanced shaft, (b) Unbalanced shaft, (c) Misaligned shaft and (d) Defective bearing

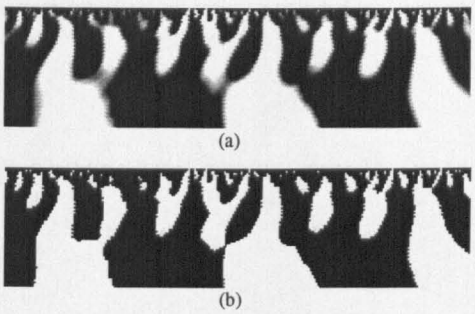


Fig. 4. The original image (a) and its binary image (b)

A binary image has the obvious advantage that when classification using neural networks is done, the computation time will be much shorter, as multiplication involving a 0 or 1 is much easier to perform. Fig. 4a and 4b show a sample image before and after binarising.

Step 3 Classifying

The back-propagation neural network (BPNN) [9], with the architecture as shown in Fig. 5, was used to classify bearing faults. The network has an input layer, a hidden layer and an output layer. The values of the pixels composing the CWT binary image provided the inputs to the neural network; and 4 nodes, representing the bearing conditions of normal, unbalanced, misaligned and defective bearing were available as outputs.

Through a process of trial and error based on minimizing the mean square error (MSE), the choice of 12 nodes in the first hidden layer and 18 nodes in the second hidden layer were determined to be optimal.

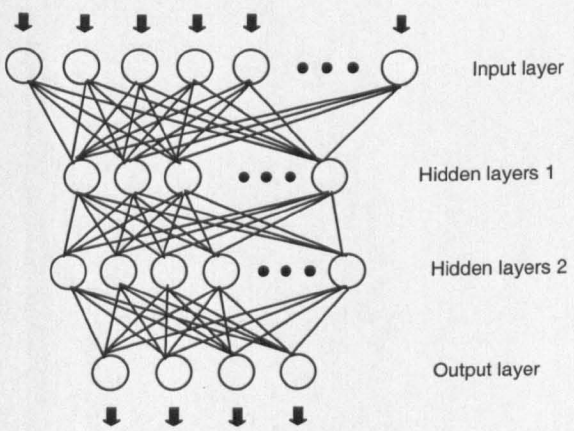


Fig. 5. BPNN architecture

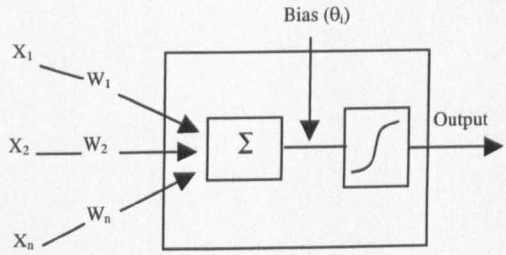


Fig. 6. Schematic process of BPNN

The value *NET* appearing at a node in a layer is computed by summing the products of all inputs leading to that node with their corresponding weights plus a bias. This value then forms the argument of a transfer function f that produces an output for the node. In vector notation, given the input vector X_i, the weight vector W_i, and the bias vector θ_i, the output is given by

$$OUT = f(NET) = f(\sum_{i}^{n} X_i W_i + \theta_i) \tag{2}$$

The transfer function f is the commonly used sigmoid function defined as

$$f = \frac{1}{1 + e^{(-NET)}} \tag{3}$$

The sigmoid function acts as an output gate that can be opened (0) or closed (1). The computing process as described is schematically shown in Fig. 6.

The Back-propagation algorithm is used to obtain the correct weights and biases in a training process. A set of training data with known outputs is fed into the network. The weights are initially set to random values; the biases are fixed at unity. The input data are presented to the network; outputs are calculated and compared with the desired outputs. The normalized mean square error (MSE) is then calculated and propagated back to adjust the weights on the neural connection. This process is repeated for a large number of epochs until the error is relatively low and acceptable, which allows the network to classify the test set correctly.

IV. EXPERIMENTAL VERIFICATION OF THE PROPOSED METHOD

Experiments were conducted on a rotodynamic test rig, Fig. 7, consisting of a rotating shaft driven by a DC motor at 20 rev/sec. The shaft, supported in two bearings, carried 4 discs with attachable masses in order to produce rotating unbalance. The two bearings were a FAG 20205K.T.C3 self-

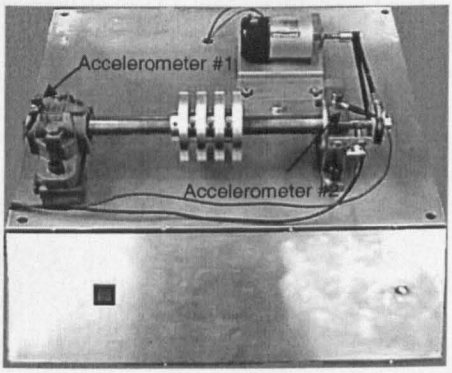


Fig. 7. Test rig set-up

aligning single row taper-bore bearing and an FAG 6304 ball bearing. Radial vibration at the two bearings was measured using two PCB 333A12 accelerometers, one on top of the non-drive end housing while the other on top of the drive-end housing labeled as #1 and #2 respectively in Fig. 7. The acceleration signals, having been low-pass filtered at 1 kHz for anti-aliasing, were sampled into a Labview data acquisition system.

Measurements were obtained from four different machine conditions: balanced shaft, unbalanced shaft, misaligned shaft and defective bearing. For each condition, twenty signals were collected which were divided into two equal sets of ten each. The first set was used for training the neural network while the second set was used for testing it.

As mentioned in Step 2, it was the binarised CWT image of a signal that was being used in training and testing. The image consisted of 64 scales and 250 time intervals giving a resolution of 64x250=16000 pixels.

V. RESULTS AND DISCUSSION

Classification results of 40 events from the accelerometer signals of all four types of machine conditions are given in Table I. It shows the output value from the back propagation neural network after the network has been trained. For the similar pattern, the output value is around unity, rather than zero. In contrast, for a poor match, the value returned is around zero.

To examine the results, the maximum output is used to identify the bearing condition. The symbols in the second column from the left indicate the following:

R = classified correctly
W = misclassified

Table I
Correlation coefficient results

Bearing conditions		Training sets			
Test sets		Balance shaft	Unbalanced shaft	Misaligned shaft	Defective bearing
Balance shaft	W	0.22	-0.01	0.00	0.79
	R	0.94	0.00	0.00	0.06
	R	0.36	0.19	0.00	0.34
	R	0.92	-0.08	0.00	0.23
	R	0.98	0.00	0.00	0.03
	R	0.96	0.03	0.03	-0.03
	R	0.86	-0.02	0.00	0.18
	R	0.73	0.00	0.00	0.27
	W	0.20	-0.01	0.00	0.81
	R	0.98	0.05	0.01	-0.02
Unbalanced shaft	R	0.04	0.35	0.22	0.26
	R	0.04	1.00	0.00	-0.04
	R	-0.12	0.84	0.00	0.45
	R	0.01	0.89	0.00	0.08
	R	-0.01	0.97	0.00	0.04
	W	-0.01	0.19	0.29	0.47
	R	0.00	1.01	0.00	-0.01
	R	-0.08	0.92	0.01	0.18
	R	0.1	0.70	0.04	0.09
	R	-0.01	1.05	0.00	-0.05
Misaligned shaft	R	0.00	0.00	0.99	0.00
	R	0.00	0.00	1.00	0.00
	R	0.01	0.07	0.95	0.12
	R	0.00	0.02	1.00	-0.02
	R	0.00	0.00	1.00	0.01
	R	0.01	-0.03	0.95	0.05
	R	0.00	0.00	1.00	0.00
	R	0.00	0.00	0.99	0.00
	R	0.00	0.00	0.99	0.00
	R	0.00	0.04	1.00	-0.04
Defective bearing	R	-0.01	0.08	0.33	0.51
	R	0.02	0.01	0.00	0.94
	R	-0.08	-0.10	0.10	1.10
	W	0.30	0.34	-0.02	0.30
	R	-0.02	0.13	0.26	0.60
	R	-0.05	0.00	0.42	0.67
	R	0.24	-0.02	0.00	0.78
	R	-0.09	0.32	0.29	0.52
	R	0.00	-0.05	0.44	0.64
	R	0.38	-0.10	-0.02	0.88

The signals were classified correctly 36 times out of 40 with a recognition rate of 90 percent. However, due to the fact that an intensity pixel being less than the threshold value had to be omitted, this brought down the recognition rate. Therefore, if an extracted image consists of many light intensity pixels, it may not find enough significant intensity pixels to represent the characteristic features efficiently. In other words, the shape of the extracted image is distorted by the lack of pixels which contain the unique pattern for each bearing condition. Therefore, this may lead to incorrect results.

VI. CONCLUSIONS

Condition monitoring of rotodynamic machinery has been investigated by using the continuous wavelet transformed image as a characteristic feature of each signal condition and the back propagation neural network as a classification tool. Results are summarised below:

1) A major contribution of this work is the introduction of a new feature extraction method and feature representation for condition monitoring using continuous wavelet transform. Image processing techniques have been employed to remove the broadband noise from the signal before the final stage of classification is performed using a back-propagation neural network.

2) The wavelet transform has the main advantage that it provides information of the signal on scale (frequency) as well as on time. The transformed image therefore contains information in both time and frequency which enhances the ability to discriminate properly between the four types of bearing condition.

3) The classification error is due to the omission of some pixels by binarising and it is dependent on the threshold value. It is believed that if the setting of the threshold level is optimised, a possible higher recognition rate would be achieved.

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge the Royal Thai government for its support, also thank Corus, Middlesborough UK, which kindly provided material help and advice for experimentation, the support of EPSRC (Grant GR/M44200) and the nine industrial collaborators including the UK National Physical Laboratory within the INTErSECT Faraday Partnership Flagship Project (1998-2002) entitled "Acoustic Emission Traceable Sensing and Signature Diagnostics(AESAD)"(Proect website: http://www.brunel.ac.uk/research/bcmm/aesad/).

REFERENCES

[1] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk and S. Liang, "Adaptive prognostics for rolling element bearing condition," Mechanical Systems and Signal Processing, Vol.13(1), pp.103-113, 1999.

[2] S. Zhang, R. Ganesan, and G. D. Xistris, "Self-organizing neural net works for automated machinery monitoring systems," Mechanical Systems and Signal Processing, Vol.10(5), pp. 517-532,1996.

[3] C. Pachaud, R. Salvetat and C. Fray, "Crest factor and kurtosis contribu tions to identify defects inducing periodical impulsive forces," Mechanical Systems and Signal Processing, Vol.11(6), pp. 903-916, 1997.

[4] J. Shiroishi, Y. Li, S. Liang, T. Kurfess and S. Danyluk, "Bearing con dition diagnostics via vibration and acoustic emission measurement," Mechanical Systems and Signal Processing,Vol.11(5), pp. 693-705, 1997.

[5] L. Qu, X. Liu, G. Peyronne and Y. Chen, "The Holospectrum: A new method for rotor surveillance and diagnosis," Mechanical Systems and Signal Processing, Vol.3(3), pp. 255-267, 1989.

[6] H. Suzuki, T. Kinjo, Y. Hayshi, M. Takemoto and K. Ono, "Wavelet transform of acoustic emission signals," Journal of Acoustic Emission, Vol.14(2), pp. 69-84, 1996.

[7] L. Gaul and S. Hurlebaus, "Identification of the impact location on a plate using wavelets," Mechanical Systems and Signal Processing, Vol.12(6), pp. 783-795, 1997.

[8] Rafael C. Gonzales, Richard E. Woods, Digital Image Processing, (Addison Wesley), 1993.

[9] B. A. Paya and I. I. Esat, "Artificial neural network based fault diagnos tics of rotating machinery using wavelet transforms as a preprocessor," Mechanical Systems and Signal Processing, Vol.11(5), pp. 751-765, 1997.

Proceeding of Conference
COMADEM 2001, Manchester.

Condition monitoring of rotodynamic machinery using Acoustic Emission and fuzzy c-mean clustering technique

T Kaewkongka, Y H Joe Au, R T Rakowski and B E Jones

The Brunel Centre for Manufacturing Metrology,
Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Phone: (+44) 01895 274000 ext. 2608, E-mail: joe.au@brunel.ac.uk

ABSTRACT

This paper describes a method of bearing condition monitoring using the fuzzy c-mean clustering technique applied to the pre-processed acoustic emission parameters. Acoustic emission (AE) events were detected by a piezoelectric transducer mounted on the bearing housing of a test rig. AE parameters were extracted from the events and used as characteristic features to represent a machine operating condition. In this experiment, four machine conditions that may happen to a rotodynamic machine were investigated and they corresponded to (a) a balanced shaft, (b) an unbalanced shaft, (c) a shaft with misaligned supportive bearings and (d) a shaft running in a defective bearing. During training, the fuzzy c-mean clustering technique was applied to establish the centres of the four clusters. For testing, a minimum distance classifier was used to classify an AE event from an unknown condition into one of the four conditions. The recognition rate was 97.22 percent.

KEYWORDS

Acoustic emission, condition monitoring, fuzzy c-mean clustering technique, minimum distance classifier, rotodynamic machinery.

INTRODUCTION

Rolling element bearings are perhaps the most ubiquitous machine elements in engineering as they can be found in almost all-rotating machines. With ever growing competition in industry, these bearings are considered critical components because any malfunction, if not detected in time, leads to catastrophic failure and hence losses due to machine downtime and other forms of damage. It is evident that a reliable condition monitoring system is highly desirable so that it will reduce the cost of these consequences and enhance the overall equipment effectiveness.

Basically there are two approaches to bearing maintenance: (1) statistical bearing life estimation and (2) bearing condition monitoring and diagnostics [1]. Statistical bearing life estimation predicts the fatigue life of a bearing. However, its application has limitations, since unusual operating conditions often occur and can severely decrease a bearing's life. In this situation, estimating a bearing's life based on standard operating conditions is unrealistic. The other approach - bearing condition and diagnostics - can be more reliable because it gives up-to-date information about the condition of a bearing. The more popular condition monitoring techniques for bearings are based on vibration and acoustic emission analyses.

Previous research [2,3,4] has demonstrated that AE monitoring is superior to vibration monitoring in that the former can detect subsurface crack growth whereas the latter can at best detect a defect only when it emerges on the surface of a structure.

Acoustic emission (AE) is a natural phenomenon of sound generation in a material under stress. If the material is subjected to stress, a sudden release of strain energy takes place in the form of elastic wave. Each release of energy results in an AE event which often lasts no longer than a millisecond. A rotation bearing can produce AE events each time a surface defect comes into contact with other elements. These AE events are high-frequency transients with frequency components typically in the range from 100 kHz to 1 MHz.

An AE event is characterized using parameters such as ring-down count, rise time, event duration, energy and peak amplitude. A threshold is used in order to eliminate 'noise' and only events that rise above the threshold are counted. Evidently, the threshold level affects the value of some of these parameters. A typical example is the event duration. By definition, it is the time that the envelope of an AE event is above the threshold. When the threshold level is high, the time will be shorter.

The peak amplitude of an AE event is the maximum excursion of the corresponding voltage signal from the zero level. The energy of an AE event is the energy contained in the corresponding voltage signal and, strictly speaking, is not the true energy of the event itself. Energy is calculated using the formula

$$Energy \propto \int_0^T V^2(t)dt \qquad (1)$$

The objectives of the reported work are:

1. To represent the AE events from the four different machine operating conditions (detailed below) in terms of their event duration, peak amplitude and energy in a three-dimensional space;

2. To establish the centres of the clusters for the four conditions in this three-dimensional space using the fuzzy c-mean clustering technique; and

3. To classify an AE event from an unknown condition by computing the minimum Euclidean distance of this event from the respective centres.

APPARATUS AND EXPERIMENTS

Experiments were conducted on a rotodynamic test rig consisting of a rotating shaft driven by a DC motor at 20 rev/sec. The two bearings were a FAG 20205K.T.C3 which is a self-aligning single-row taper-bore roller bearing and a FAG 6304 ball bearing. They were mounted in bearing housings which in turn were attached to a base plate. The test rig provides facilities to produce the four machine operating conditions characterised by:

a) the rotating shaft dynamically balanced (referred to as 'balanced shaft'),
b) the rotating shaft dynamically unbalanced in one plane to the extent of 65×10^{-5} kg.m at mid-span of the shaft (referred to as 'unbalanced shaft'),
c) the shaft with misalignment achieved from moving one bearing laterally by 1 mm relative to the other (referred to as 'misaligned shaft') and
d) the roller bearing seeded with a defect on the outer raceway of 1mm diameter produced with an electric discharge pen (referred to as 'defective bearing').

The methodology for machine condition monitoring and recognition is as shown schematically in Figure 1.

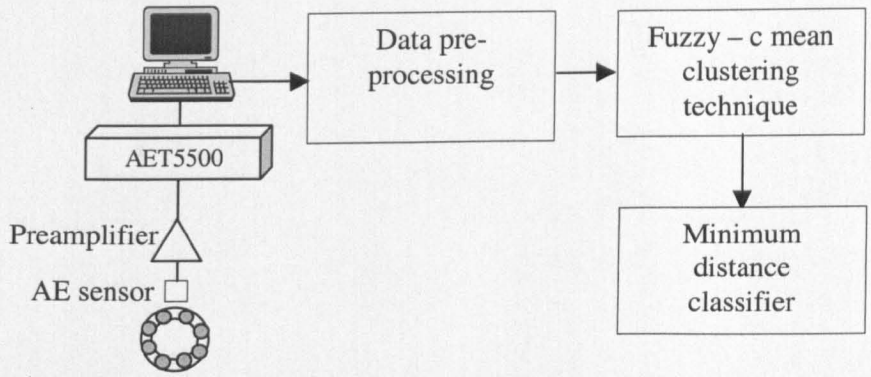


Figure 1: The methodology for machine condition monitoring and recognition.

Acoustic emission instrumentation

Acoustic emission was measured with a microprocessor-based system AET5500 from Acoustic Emission Technology Corp. (AET), USA. A wide band transducer (of model WD, PAC) was attached to the side of the roller bearing housing via a silicone-gel couplant. The AE signal from the transducer was amplified to 60 dB and bandpass filtered - 100kHz to 1 MHz - with a PAC preamplifier before entering the AET5500 for AE parameter extraction.

The threshold level on the AET5500 was chosen to be 1V (floating). With a floating threshold, it can adjust itself automatically such that background noise is excluded. The peak amplitude of the AE event was expressed in dB with 0 dB corresponding to 1 mV at the preamplifier output.

Pre-processing of data

For each condition, twelve recordings each of about 30-second duration were made and they captured the AE parameters of event duration, peak amplitude and energy. These recordings were then divided into two sets of 3 and 9 each. The first set served as the reference generated from the training exercise whereas the second set provided the test samples for validating the classifier obtained from the training process.

Both the training sets and the testing sets were processed as follows:

1. Sort the AE events in each set in the descending order of the event duration.
2. Discard the first ten AE events in the sorted list as they may contain outliers which, if included, would distort the characteristics the AE events in a sub-set
3. Select the next five AE events from the remaining list for the subsequent clustering analysis.
4. Normalise each feature in a unit vector.

Fuzzy c-mean clustering

Fuzzy c-mean is an iterative technique for data clustering. The user decides on the number of clusters that a data set is to be separated into, initialises a proximity matrix and defines an error threshold for the stop condition of the iteration [5].

The five AE events partitioned from the sorted list can be displayed as points on a three-dimensional graph using the event duration, peak amplitude and energy as the three orthogonal axes. As there are four different conditions, it is expected that there will be four clusters taking up different regions in the three-dimensional space.

The proximity matrix contains the membership values of an individual AE event that it does or does not belong to a particular machine condition. The initial values assigned to this matrix is arbitrary and binary logic values are generally used. Thus full membership is represented by 1 whereas non-membership by 0. As there are four different machine conditions and five AE events, the proximity matrix is a 4-by-20 matrix.

The first iteration generates the estimated locations of the cluster centres and a refined proximity matrix in which the membership values become fuzzified, that is they now have values between 0 and 1. With each subsequent iteration, the estimate for the cluster centre locations will be more and more accurate and the proximity matrix will be updated. The iteration will stop when the change in the norm of the proximity matrix from its previous iteration becomes less than the designated error threshold. The cluster centres returned from the last iteration are taken to be the 'best' estimates.

Specifically, the fuzzy – c mean algorithm consists of the following step [5]:

1. Fix the number of c-cluster centres and a threshold ε for the stop condition in step 4. Initialise the proximity matrix $U^{(0)}$.
2. Update the c-cluster centres $\left\{v_i^{(r)}\right\}$ according to the current proximity matrix, using

$$v_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} \cdot x_{kj}}{\sum_{k=1}^{n} \mu_{ik}^{m'}} \tag{2}$$

   where μ_{ik} is the membership of the k^{th} data point in the ith class, and m' is the weighting parameter (the arbitrary value of $m' = 2$ was used).
3. Update the proximity matrix for the r^{th} iteration, $U^{(r)}$ according to previous cluster centres, using

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{2/(m'-1)} \right]^{-1} \tag{3}$$

   where $d_{ik} = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2}$ is the distance measured.

4. If the objective function, as defined below, is less than the threshold ε, then stop; else, go to step 2.

$$\left\| U^{(r+1)} - U^{(r)} \right\| \leq \varepsilon \text{ otherwise set r = r+1} \tag{4}$$

The results that emerge from the application of the fuzzy c-mean clustering technique are shown as three-dimensional graphs in Figures 2 and 3. Figure 2 shows the AE events, 20 in total, for the four different machine conditions from the training sets. Figure 3 shows the same AE events but this time with clusters identified and their centres computed.

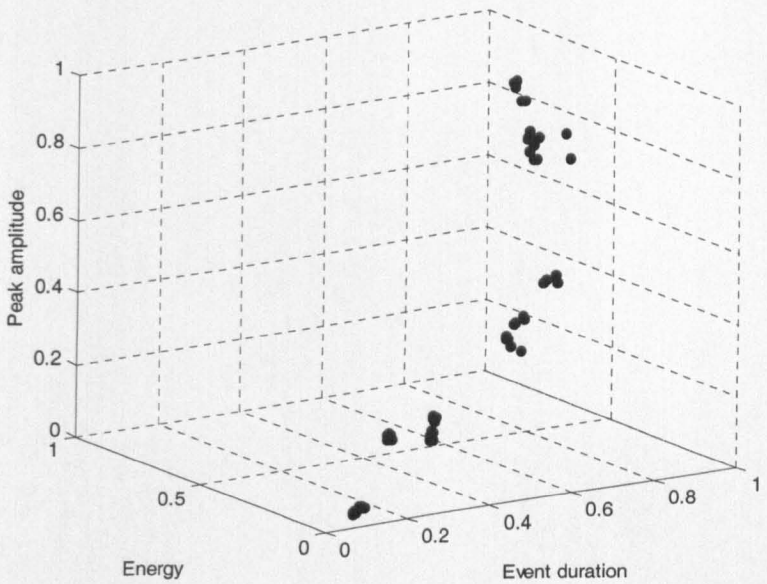


Figure 2: AE events for the four different machine conditions from the training sets

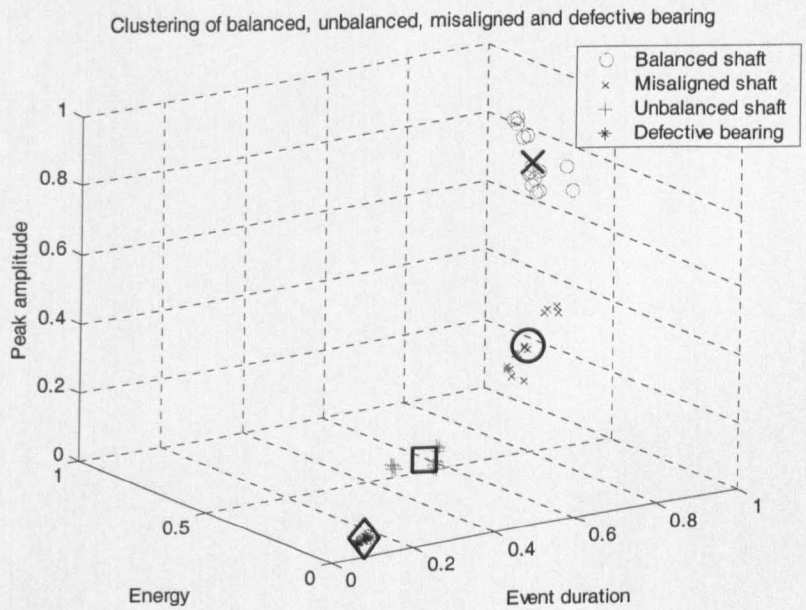


Figure 3: Established cluster centres using training data set from four different operating conditions

Minimum distance classification

A minimum distance classifier [6] was used to determine the machine condition to which a particular AE event belongs. It works by computing the Euclidean distances of the AE event (expressed as a point in the three-dimensional space of event duration, peak amplitude and energy) from the centres of the clusters for the four machine conditions. The AE event is considered to belong to the cluster whose centre is closest. In other words, the distance is defined as an index of similarity so that the minimum distance is equivalent to the maximum similarity.

Thus if the four centres have co-ordinates (x_{i1}, x_{i2}, x_{i3}) where i = 1, 2, 3, 4, and the AE event (y1,y2,y3), the minimum Euclidean distance is then

$$D = \min\left(\sqrt{(x_{i1} - y_1)^2 + (x_{i2} - y_2)^2 + (x_{i3} - y_2)^2}\right) \qquad (5)$$

RESULTS AND DISCUSSION

Classification results of 36 AE events from the test samples of all four types of machine conditions are given in Table 1. It shows the output values from the minimum distance classifier using Eq(5). The events were classified correctly 35 times out of 36, a recognition rate of 97.22 percent. The only error occurs when the unbalanced shaft condition was misclassified as that due to a defective bearing (55.139 in row 4, column 2).

It can also be noted that the normal machine condition of 'balanced shaft' is very distinctive from the other three abnormal conditions, as their distance values are all very small in comparison with others. For example, in the column headed 'Balanced shaft', the first value in each sub-section always turns out to be significantly smaller than the rest. This means that using this approach there is very little risk of raising a false alarm.

CONCLUSION

The methodology described in this paper has been shown to work well for discriminating the four different operating conditions on the rotodynamic test rig. The recognition rate achieved was 97.22 percent. In addition, the probability of misclassifying a good balanced shaft condition is extremely low.

The method involves sorting the AE events according to event duration, identifying clusters and locating their centres in the three-dimensional space of event duration, peak amplitude and energy using the fuzzy c-mean technique, and classifying using a Euclidean minimum distance classifier.

Overall, the method has the advantage that it is simple and efficient to implement and can be readily adapted to include other abnormal operating conditions as may be identified on a rotodynamic machine.

TABLE 1
MINIMUM DISTANCE CLASSIFICATION RESULTS

Bearing conditions	Balanced shaft	Unbalanced shaft	Misaligned shaft	Defective bearing
Testing set #1				
Balanced shaft	**1.4701**	807.9	507.8	1162.3
Unbalanced shaft	670.6	**136.28**	164.1	490.6
Misaligned shaft	478.6	328.4	**28.076**	682.74
Defective bearing	861.81	**55.139**	355.41	299.32
Testing set #2				
Balanced shaft	**4.3499**	810.9	510.81	1165.3
Unbalanced shaft	821.35	**15.163**	314.9	339.8
Misaligned shaft	516.47	290.31	**10.529**	644.71
Defective bearing	1003.8	197.11	497.41	**157.34**
Testing set #3				
Balanced shaft	**28.284**	778.76	478.58	1133.2
Unbalanced shaft	839.19	**32.51**	332.83	321.93
Misaligned shaft	468.16	338.71	**38.416**	693.09
Defective bearing	1002.3	195.59	495.8	**158.9**
Testing set #4				
Balanced shaft	**5.7054**	802.85	502.82	1157.3
Unbalanced shaft	841.6	**35.472**	335.11	319.61
Misaligned shaft	400.91	405.84	**105.69**	760.26
Defective bearing	1038.9	232.27	532.5	**122.2**
Testing set #5				
Balanced shaft	**70.657**	736.04	435.92	1090.5
Unbalanced shaft	773.04	**33.841**	266.61	388.11
Misaligned shaft	421.84	384.97	**84.738**	739.38
Defective bearing	1031.2	224.54	524.8	**129.9**
Testing set #6				
Balanced shaft	**15.982**	790.71	490.61	1145.1
Unbalanced shaft	761.76	**44.958**	255.5	399.37
Misaligned shaft	372.86	433.85	**133.77**	788.28
Defective bearing	1028.2	221.56	521.71	**133.04**
Testing set #7				
Balanced shaft	**40.244**	766.44	466.34	1120.9
Unbalanced shaft	868.07	**61.403**	361.74	293.05
Misaligned shaft	391.57	415.12	**115.4**	769.55
Defective bearing	1222.8	416.14	716.4	**61.706**
Testing set #8				
Balanced shaft	**4.2853**	804.99	504.82	1159.4
Unbalanced shaft	1046.9	**240.23**	540.5	114.21
Misaligned shaft	461.38	345.31	**45.768**	699.74
Defective bearing	1108.1	301.4	601.72	**53.219**
Testing set #9				
Balanced shaft	**2.7767**	809.05	508.92	1163.5
Unbalanced shaft	905.94	**99.297**	399.5	255.2
Misaligned shaft	477.31	329.39	**29.992**	683.82
Defective bearing	1081.2	274.57	574.8	**79.909**

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge the award of a Ph.D. scholarship by the Royal Thai government for its support, also thank Corus, Middlesborough UK, which kindly provided material help and advice for experimentation.

REFERENCES

[1] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk and S. Liang. (1999). Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing* **13:1,** 103-113.

[2] T. Kaewkongka, J. H. Au, R. Rakowski and B. E. Jones. (2001) Continuous wavelet transform and neural network for condition monitoring of rotodynamic machinery *IEEE Instrumentation and Measurement Technology* **3:3,** 1962-1966.

[3] C. James Li and S. Y. Li. (1995). Acoustic emission analysis for bearing condition monitoring. *Wear* **185,** 67-64.

[4] N. Tandon and B. C. Nakra. (1992). Comparison of vibration and acoustic measurement t echniques for the condition monitoring of rolling element bearings *Tribology International* **25:3,** 205-212.

[5] Timoty J. Ross. (1995). *Fuzzy logic with engineering applications*, MaGraw-Hill, Inc.

[6] Julius T. Tou. (1974). *Pattern recognition principles*, Addison-Wesley Publishing Company.

Application of acoustic emission to condition monitoring of rolling element bearings

T Kaewkongka and Y H Joe Au

The Brunel Centre for Manufacturing Metrology, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

Executive summary

This paper describes the application of acoustic emission (AE) for bearing condition monitoring using pre-processed AE parameters and fuzzy c-mean clustering to establish characteristic features relating to each bearing condition. In the experiments, four types of bearing operating conditions were examined: a balanced shaft, an unbalanced shaft, a misaligned shaft and a defective bearing. A minimum distance classifier is then used as a tool to classify bearing operating conditions. The results are very promising with a recognition rate of 97 %.

Introduction

Rolling element bearings are omnipresent in almost all kind of rotating machines. Its condition monitoring has obtained considerable attention for many years because majority of problems in rotating machines are caused by faulty bearings. A reliable condition monitoring system will significantly reduce failure and unplanned maintenance, and hence the huge attendant cost due to machine downtime. Often, the system is used with an operator who assists in the interpretation of the machine signals for early failure detection and fault diagnosis.

Nowadays there are two kinds of methods available for bearing maintenance: statistical bearing life estimation and bearing condition monitoring and diagnostics[1]. Statistical bearing life estimation predicts the fatigue life of a bearing. However, its application has

many limitations, since unusual operating conditions can severely decrease a bearing's life. Bearing life estimates become unreliable leading to unexpected breakdown. On the other hand, bearing condition and diagnostics can be a very reliable method because it gives up-to-date information about the condition of a bearing. The more popular techniques used for bearing condition monitoring are vibration and acoustic emission analyses.

The popular bearing condition monitoring approaches generally are vibration and acoustic emission (AE) techniques. The vibration techniques are practically useful only when the abnormal condition of structures developed. The latter one, acoustic emission techniques, are claimed to be superior to the former because it is an effective tool for early detection of damages due to metallic contact or wear action[2].

Acoustic emission is a natural phenomenon of sound generation applied to the spontaneously generated elastic wave produced within a material under stress[3]. Plastic deformation and growth cracks are the primary sources of acoustic emission in metals. The acoustic signal can be detected by a piezoelectric sensor, which converts the mechanical energy carried by the elastic wave into an electrical signal.

In the case of rotating element bearing, when a defective roller surface comes into contact with other elements, it can then produce AE event because of the release of strain energy picked up by a piezoelectric transducer.

Conventional AE parameters extracted from acoustic signal are ring-down count, rise time, event duration, energy and peak amplitude. To characterise those parameters and also to eliminate 'noise', a threshold is selected and only events that rise above the

threshold are counted. Apparently, the threshold level affects the value of some of these parameters. A typical example is the event duration. By definition, it is the time that the envelope of an AE event is above the threshold. When the threshold level is high, the time will be shorter.

The peak amplitude of an AE event is the maximum arising of the corresponding voltage signal from the zero level. The energy of an AE event is the energy contained in the corresponding voltage signal and, strictly speaking, is not the true energy of the event itself.

Experimental setup

The experimental setup consisted of rotodynamic test rig which can produce multi fault operating conditions. The spindle is driven by a variable speed motor running at 20 rev/sec. The two bearings were a FAG 20205K.T.C3 which is a self-aligning single-row taper-bore roller bearing and a FAG 6304 ball bearing. They were mounted in bearing housings which in turn were attached to a base plate. The test rig provides facilities to produce the four machine operating conditions characterised by:

1. the rotating shaft dynamically balanced (referred to as 'balanced shaft'),
2. the rotating shaft dynamically unbalanced in one plane to the extent of 65×10^{-5} kg.m at mid-span of the shaft (referred to as 'unbalanced shaft'),
3. the shaft with misalignment achieved from moving one bearing laterally by 1 mm relative to the other (referred to as 'misaligned shaft') and
4. the roller bearing seeded with a defect on the outer raceway of 1mm diameter produced with an electric discharge pen (referred to as 'defective bearing').

Figure 1 illustrates a systematic diagram of the proposed bearing condition monitoring technique.

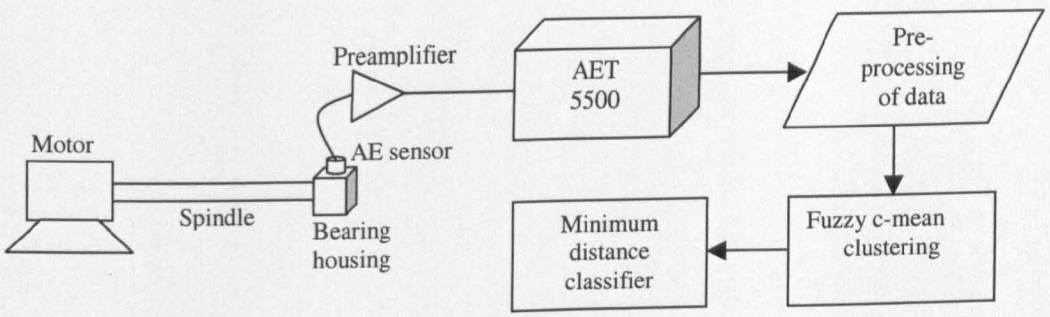


Figure 1 AE monitoring systematic diagram.

Acoustic monitoring system

The acoustic emission transducer, wide band transducer from Physical Acoustic Corporation (PAC) is mounted on the bearing housing with a silicone gel couplant. The AE signal from the transducer was amplified to 60 dB and bandpass filtered - 100kHz to 1 MHz - with a PAC preamplifier before entering an AET5500. The micro-processor based system AET5500, from Acoustic Emission Technology Corp. (AET), USA, is used to process the acoustic signal into AE parameters.

As described earlier, the selection of threshold is important in order to obtain the high signal-to-noise ratio. In this experiment, the floating threshold of 1V is chosen because, with floating threshold, it can adjust the level of threshold to eliminate the background noise.

Pre-processing of data

For each condition, twelve recordings each of about 30-second duration were made and they captured the AE parameters of event duration, peak amplitude and energy. These

recordings were then divided into two sets of 3 and 9 each. The first set served as the reference generated from the training exercise whereas the second set provided the test samples for validating the classifier obtained from the training process.

Both the training sets and the testing sets were processed as follows:

1. Sort the AE events in each set in the descending order of the event duration.
2. Discard the first ten AE events in the sorted list as they may contain outliers which, if included, would distort the characteristics the AE events in a sub-set.
3. Select the next ten AE events from the remaining list for the subsequent clustering analysis.
4. Normalise each feature in a unit vector.

The new data sets are now much shorter, and the corresponding parameters of event duration, peak amplitude, and energy will be further analysed using the fuzzy c-mean clustering technique.

Fuzzy c-mean clustering technique

Fuzzy c-mean is an iterative technique for clustering analysis and classification. The user decides on the number of clusters that the data set is to be separated into, sets up an initial proximity matrix according to some rule and defines an error threshold for the stop condition of the iteration. The first iteration will generate the estimated locations of the cluster centres and a refined proximity matrix. With subsequent iterations, the estimates of the cluster centre locations will be more and more accurate and the proximity matrix will be updated. The iteration will stop when the change in the norm of the proximity matrix from its previous iteration becomes less than the designated error threshold. The cluster centres returned from the last iteration are taken to be the 'best' estimates.

The results that emerge from the application of the fuzzy c-mean clustering technique are shown as three-dimensional graphs in Figures 2 and 3. Figure 2 shows the AE events, 20 in total, for the four different machine conditions from the training sets. Figure 3 shows the same AE events but this time with clusters identified and their centres computed.

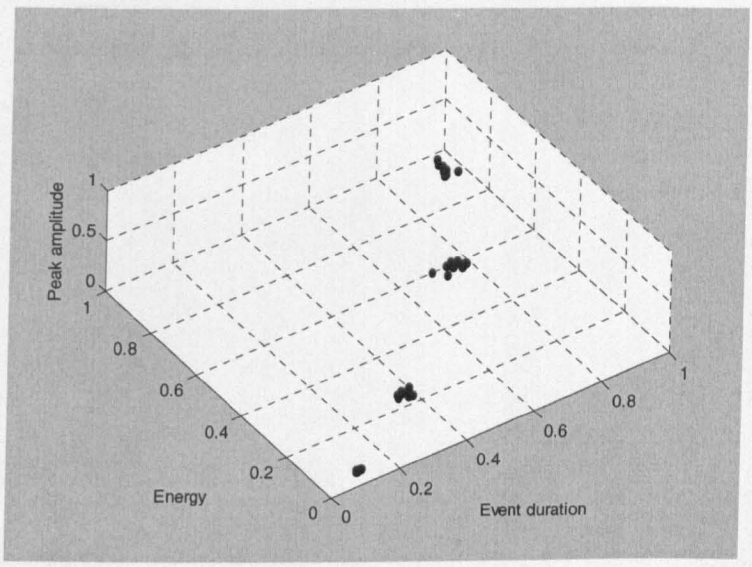


Figure 2 AE events for the four different machine conditions from the training sets.

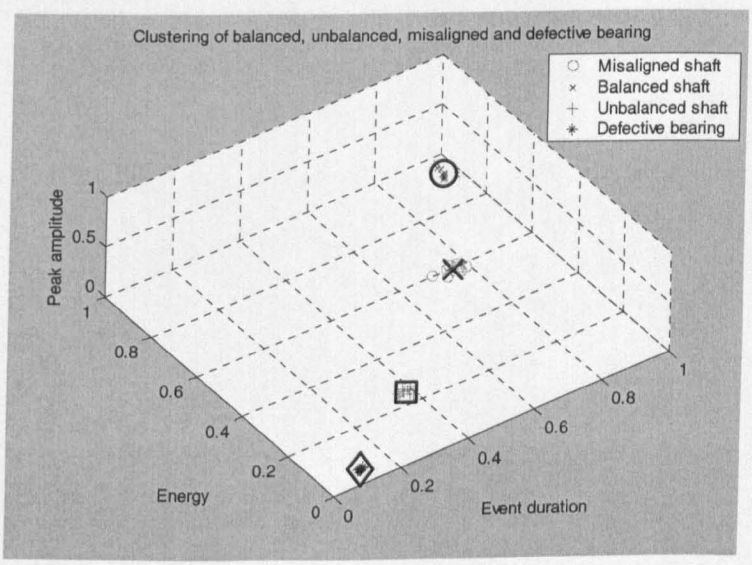


Figure 3 Established cluster centres using training data set from four different operating conditions.

Minimum distance classification

A minimum distance classifier [5] was used to determine the machine operating condition to which a particular AE event belongs. It works by computing the Euclidean distances of the AE event (expressed as a point in the three-dimensional space of event duration, peak amplitude and energy) from the centres of the clusters for the four machine conditions. The AE event is considered to belong to the cluster whose centre is nearest. In other words, the distance is defined as an index of similarity so that the minimum distance is equivalent to the maximum similarity.

Results and discussions

Classification results of 36 AE events from the test samples of all four types of machine conditions are given in Table 1. It shows the output values from the minimum distance classifier. The events were classified correctly 35 times out of 36, a recognition rate of 97 percent. The only error occurs when the unbalanced shaft condition was misclassified as that due to a defective bearing.

It can also be noted that the normal machine condition of 'balanced shaft' is very distinctive from the other three abnormal conditions, as their distance values are all very small in comparison with others. In other words, this approach has very little risk of raising a false alarm.

Table 1 Minimum distance classification results

Bearing conditions	Balanced shaft	Unbalanced shaft	Misaligned shaft	Defective bearing
Testing set #1				
Balanced shaft	1.47	807.90	507.80	1162.30
Unbalanced shaft	670.60	136.28	164.10	490.60
Misaligned shaft	478.60	328.40	28.07	682.74
Defective bearing	861.81	55.13	355.41	299.32
Testing set #2				
Balanced shaft	4.34	810.90	510.81	1165.30
Unbalanced shaft	821.35	15.16	314.90	339.80
Misaligned shaft	516.47	290.31	10.52	644.71
Defective bearing	1003.8	197.11	497.41	157.34
Testing set #3				
Balanced shaft	28.28	778.76	478.58	1133.20
Unbalanced shaft	839.19	32.51	332.83	321.93
Misaligned shaft	468.16	338.71	38.41	693.09
Defective bearing	1002.30	195.59	495.80	158.90
Testing set #4				
Balanced shaft	5.70	802.85	502.82	1157.30
Unbalanced shaft	841.6	35.47	335.11	319.61
Misaligned shaft	400.91	405.84	105.69	760.26
Defective bearing	1038.90	232.27	532.50	122.20
Testing set #5				
Balanced shaft	70.65	736.04	435.92	1090.50
Unbalanced shaft	773.04	33.84	266.61	388.11
Misaligned shaft	421.84	384.97	84.73	739.38
Defective bearing	1031.20	224.54	524.80	129.90
Testing set #6				
Balanced shaft	15.98	790.71	490.61	1145.10
Unbalanced shaft	761.76	44.95	255.50	399.37
Misaligned shaft	372.86	433.85	133.77	788.28
Defective bearing	1028.20	221.56	521.71	133.04
Testing set #7				
Balanced shaft	40.24	766.44	466.34	1120.90
Unbalanced shaft	868.07	61.40	361.74	293.05
Misaligned shaft	391.57	415.12	115.40	769.55
Defective bearing	1222.80	416.14	716.40	61.70
Testing set #8				
Balanced shaft	4.28	804.99	504.82	1159.40
Unbalanced shaft	1046.90	240.23	540.50	114.21
Misaligned shaft	461.38	345.31	45.76	699.74
Defective bearing	1108.10	301.40	601.72	53.21
Testing set #9				
Balanced shaft	2.77	809.05	508.92	1163.50
Unbalanced shaft	905.94	99.29	399.50	255.20
Misaligned shaft	477.31	329.39	29.99	683.82
Defective bearing	1081.20	274.57	574.80	79.90

Conclusions

The high sensitivity of the acoustic emission measurement has been shown to work well for discriminating the four different operating conditions on the rotodynamic test rig. The recognition rate achieved was 97 percent. In addition, the probability of misclassifying a good balanced shaft condition is extremely low.

The method involves sorting the AE events according to event duration, identifying clusters and locating their centres in the three-dimensional space of event duration, peak amplitude and energy using the fuzzy c-mean technique, and classifying using a Euclidean minimum distance classifier.

In summary, the method has the advantage that it is simple and efficient to implement and can be readily adapted to include other abnormal operating conditions as may be identified on a rotodynamic machine.

Acknowledgements

The authors gratefully acknowledge the support of the Royal Thai government and the Department of Physics, Faculty of Science, Chulalongkorn University in awarding a Ph.D. scholarship. The authors also wish to thank Corus, Middlesborough UK, which has kindly provided advice and equipment for experimentation, the support of EPSRC (Grant GR/M44200) and the nine industrial collaborators including the UK National Physical Laboratory within the INTErSECT Faraday Partnership Flagship Project (1998-2002) entitled "Acoustic Emission Traceable Sensing and Signature Diagnostics (AESAD)"(Project website: http://www.brunel.ac.uk/research/bcmm/aesad/).

References

1. Li Y., Billington S., Zhang C., Kurfess T., Danyluk S. and Liang S., *Adaptive prognostics for rolling element bearing condition*, Mechanical Systems and Signal Processing, **13(1)**, 103-113, 1999.
2. Yoon D. J., Kwon O. Y., Chung M. H. and Kim K. W., *Early detection of damages in journal bearings by acoustic emission monitoring,* Journal of acoustic emission, **13**, 1-10, 1995.
3. Rogers L. M., *The application of vibration signature analysis and acoustic emission source location to on-line condition monitoring of anti-friction bearings,* Tribology international, **9**,51-59, 1979.
4. Ross T. J., *Fuzzy logic with engineering applications*, MaGraw-Hill, Inc., 1995.
5. Tou J. T., *Pattern recognition principles,* Addison-Wesley Publishing Company, 1974.

IOM Communications: Instrumentation to innovation
Applications and Developments in Metal Production and Use, London.

Fuzzy C-Mean for Bearing Lubrication Condition Monitoring using Acoustic Emission

T. Kaewkongka, Y. H. Joe Au, R. T. Rakowski and B. E. Jones

The Brunel Centre for Manufacturing Metrology,
Department of Systems Engineering,
Brunel University, Uxbridge, Middlesex UB8 3PH
Telephone: 01895 274000 ext.2637, Fax: 01895 812556, E-mail:
barry.jones@brunel.ac.uk

Abstract

One of the most important aspects of bearing operation is to assure that a film of lubricant such as grease will be introduced between bearing surfaces, so that friction and surface wear can be minimised. Contaminated grease or the lack of lubricant may lead to an ineffective bearing operating condition or indeed malfunction of the machinery. Therefore, there is a significant need for reliable and robust condition monitoring of bearings in order to avoid the unexpected breakdown. This paper describes a novel method for bearing condition monitoring at low speed using acoustic emission (AE) and a fuzzy c-mean clustering technique. The experimental results produce a recognition rate of approaching 100 percent.

The work is being undertaken as part of a INTErSECT Faraday Partnership Flagship Research Project (AESAD, 1998-2002) [1,2,3,4,5]. The objective of this study has been to categorise two bearing lubrication operating conditions according to lubrication conditions: clean grease and contaminated grease which is slag from the steel making process. The experiments were performed on a bearing running at speed of 0.5 revolution per minute at Corus. The bearing was instrumented with AE sensors in order to capture an elastic wave generated producing an acoustic wave by the material when to subjected stress or friction as is the case when a rolling element comes into contact with wear debris.

The equipment that used for the bearing Acoustic emission trials is PAC Ltd., LOCAN AT acoustic emission system. The AE transducers are R15I integral preamplifier type with a resonant frequency of 150kHz and a range of 60 – 600 kHz. The transducers are attached to the bearing using magnetic clamps and silicone grease is used for the couplant between the transducer and the bearing housing. The bearing is FAG 515518K and the lubricating grease is Molub Alloy type.

Conventional AE parameters, such as 'event duration', 'peak amplitude' and 'energy' were extracted from the raw AE signal by a threshold level method. 'Event duration' is the time duration between the first crossing point of the signal which rises above the threshold and the last crossing point when the signal has fallen below the threshold. 'Peak amplitude' is the maximum level of the signal compared to the ground reference. 'Energy' is related to the area of the signal in time envelope.

To recognise the bearing operating conditions, feature extraction has been performed on the AE parameters by using a data processing method, which produced a set of characteristic features for the bearing running condition. The data

processing method included sorting and normalising of the AE parameters. The proposed method for bearing lubrication monitoring is shown in Figure 1.

Fuzzy c-mean was then used for signal classification [6]. The process starts with a user deciding on the number of clusters that the data-set is to be separated into, and then sets up an initial proximity matrix according to some rule and defines an error threshold for the stop condition of the iteration. The first iteration will generate the estimated locations of the cluster centres and a refined proximity matrix. With subsequent iterations, the estimates of the cluster centre locations will be more and more accurate and the proximity matrix will be updated. The cluster centres returned from the last iteration are taken to be the 'best' estimates of the location of the two separated clusters for the different lubrication conditions. The results that emerge from the application of the fuzzy c-mean clustering technique are shown as a three-dimensional graph in Figure 2.

Work is currently being undertaken to try and apply these advanced techniques to the main bearing of the British Airways London Eye to provide long-term condition monitoring for the purpose of preventative maintenance.

Acknowledgements
The authors wish to gratefully acknowledge the award of a PhD scholarship to the first-named author by the Royal Thai Government and also thank Mr Winfield Stewart from Corus, UK, who kindly provided data, material help and advice for experimentation. Acknowledgement is given for the support of EPSRC (Grant GR/M44200) and the nine industrial collaborators including the UK National Physical Laboratory within the INTErSECT Faraday Partnership Flagship Research Project (1998-2002) entitled "Acoustic Emission Traceable Sensing and Signature Diagnostics (AESAD)"
(Project website: http://www.brunel.ac.uk/research/bcmm/aesad/).

References

[1] T. Kaewkongka, Y. H. Joe Au, R. T. Rakowski and B. E. Jones, Continuous wavelet tansform and neural network for condition monitoring of rotodynamic machinery. Proc of IEEE Conference IMTC 2001, Budapest, May 2001. pp 1962-1966.
[2] T. Kaewkongka, Y. H. Joe Au, R. T. Rakowski and B. E. Jones, Condition monitoring of rotodynamic machinery using acoustic emission and fuzzy c-mean clustering technique. Proc of Conference COMADEM 2001, Manchester, September 2001. pp 49-56.
[3] T. Kaewkongka, Y. H. Joe Au, R. T. Rakowski and B. E. Jones, Application of acoustic emission to condition monitoring of rolling element bearings. Measurement and Control, 34, October 2001. pp 245-247.
[4] A. Terchi and Y. H. Joe Au, Acoustic emission signal processing. Measurement and Control, 34, October 2001. pp 240-244.
[5] T. Yan and B. E. Jones, Establishment of absolute measurement standards for in-situ calibration of acoustic emission energy. Measurement and Control, 34, December 2001. pp 300-305.
[6] T. J. Ross, Fuzzy logic with engineering applications, McGraw-Hill, Inc.,1995.

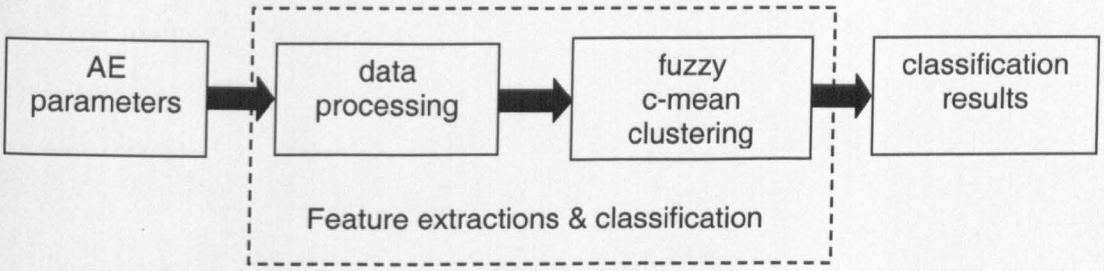


Figure 1 The proposed method for bearing lubrication condition monitoring

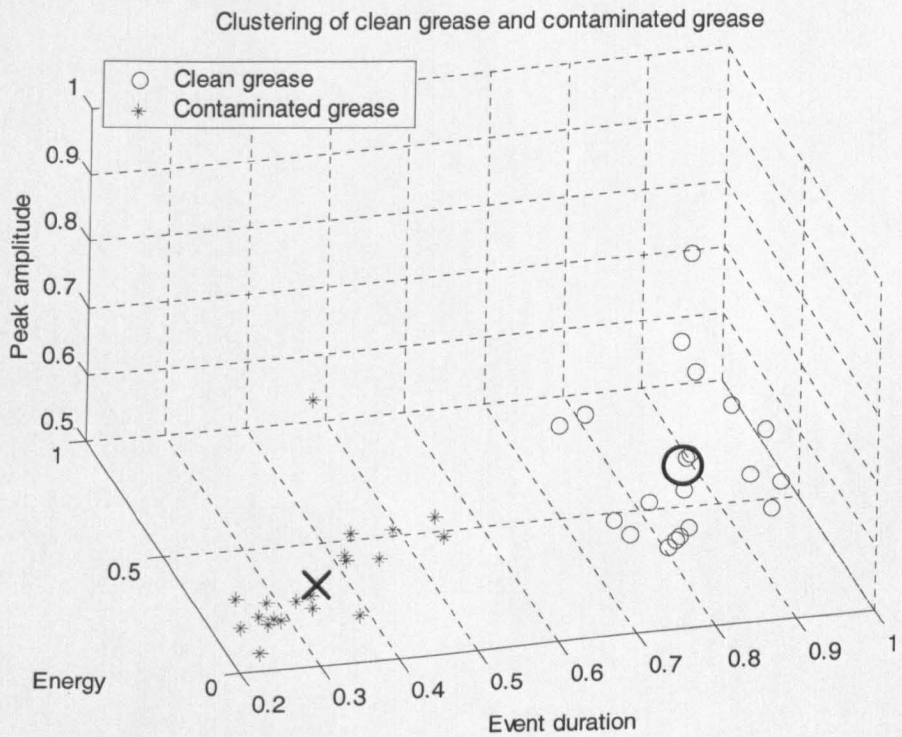


Figure 2 Results after applying the fuzzy c-mean clustering technique to the three dimensional graph of AE parameters

A Comparative Study of Short Time Fourier Transform and Continuous Wavelet Transform for Bearing Condition Monitoring

Tonphong Kaewkongka, Y H Joe Au, Richard T Rakowski and Barry E Jones

The Brunel Centre for Manufacturing Metrology, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom
Telephone: +44 1895 274000, E-mail : joe.au@brunel.ac.uk

Abstract

Rolling-element bearings can be found on almost all rotating machines and their failure is one of the major causes of machine breakdown. This paper provides a review on joint time-frequency domain analysis. Since in the real machine monitoring environment, the monitored signal, such as from vibration measurement, can be transient events with abrupt changes in the waveform, traditional analyses conducted solely in either the time or frequency domain are not capbable of revealing the occurrence of bearing faults. An approach is to utilise joint time-frequency domain method by applying the wavelet transform to the measured signal. The transformed signals are represented as 3-D images which may contain unique characteristic features relating to the various types of bearing faults. The simulation is performed to compare the performance between STFT and WT for signal classification. The broadband noise, by virtue of its low correlation with the main signal, is then removed from the image before the final stage of classification is performed. Classification is, in essence, pattern recognition of the refined image. The correlation matching is then applied to the transformed images for classification. This diagnostic process will discriminate between different bearing defects and also provide a quantitative measure of the defect size.

1. Introduction

Bearings can be found on almost all machines. Their failure invariably has production consequences and sometimes even health and safety consequences. A reliable condition monitoring system is therefore highly desirable for it will alleviate the cost of these consequences and enhance the overall equipment effectiveness.

For bearing maintenance, two methods have been used, namely the statistical bearing life estimation and the bearing condition monitoring and diagnostics [1]. The first method relies on a model of the bearing survival probability in terms of the dynamic load rating and the equivalent load to give a prediction of the fatigue life of a bearing. However, since operating conditions can vary significantly from one machine to another, the prediction based on the assumption of normal duty on a bearing can be in serious error. The second method, in theory, is superior to the first if the signals monitored have useful features that can reliably indicate a potential failure well ahead of the occurrence of the corresponding functional failure. Signals that have been studied for bearing condition monitoring include vibration and acoustic emission.

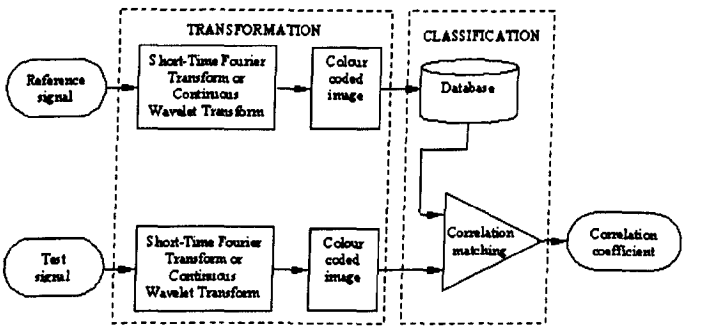


Figure 1 The proposed bearing condition monitoring block diagram.

When a bearing is damaged, the defect may occur on the rolling element, inner race or outer race. The signal generated from a defective bearing contains information not just about the defect but also about other factors such as the location of the defect relative to the loaded zone of the bearing and other mechanical noise sources that are present in the machine. The difficulty lies in being able to identify correctly the key elements in the monitored signal that are related to the defect.

2. Previous Work

Bearing monitoring methods can be categorized based on the domain of the signal in which the analysis takes place. Thus, monitoring methods can be regarded as time-domain or frequency-domain.

Time-domain methods are usually sensitive to impulsive oscillations. Examples of features that have been successfully extracted and used for monitoring purposes are the rms value, peak value [2], crest factor and kurtosis [3]. These features can produce satisfactory results when the time signal contains predominantly the frequency bands that are sensitive to the defect. However, if the machine dynamics is complex generating correspondingly complex signals, these time-domain methods may become ineffective.

Frequency-domain methods work best when the signals being monitored are periodic in nature, as is the case for vibration on a rotating shaft supported in bearings. Since a bearing defect produces a signal that is periodic at the characteristic defect frequency, frequency-domain analysis appears to be a suitable tool for condition monitoring. Examples of frequency-domain techniques include the spectral analysis, cepstrum analysis, high frequency resonance

technique (HFRT) [4] and holospectrum [5]. Among them, spectral analysis using Fourier Transform seems to be the most common and dominate the fault diagnostics.

A major disadvantage of Fourier Transform used in spectral analysis is that the frequency information provided is an average over the whole length of the analyzed signal. Thus, if there is a localized event at some point in time of the signal, it will appear in the Fourier transform as a frequency component but its location on the time axis is lost. On the other hand, although time-domain methods can reveal the time of occurrence of a local event, they do not provide any frequency information of the event.

Examples of localized events in condition monitoring abound. When a rolling element passes over a localized defect in a bearing, it generates a transient oscillation in the measured signal, so does the contact of a damaged tooth with other teeth in the gearbox. A machine with rapidly varying speed is another example of transient events.

In this paper, a comparative study was reported on two joint time- and frequency-domain techniques, namely the Short Time Fourier Transform (STFT) and the Continuous Wavelet Transform (CWT). They were assessed in terms of their relative ability to discriminate between different types of signals that had transient features in them. These signals were generated artificially in Matlab so that their features could be well defined. The better method arising from this assessment was then used for a condition-monitoring test on a bearing that had an outer race defect. The assessment procedure adopted for both tests was based on the principle of correlation matching.

3. Theoretical background

Figure 1 shows a block diagram of the proposed condition monitoring procedure. Both the reference signal and the test signal are first filtered to remove noise and then processed by either STFT or CWT to produce colour-coded two-dimensional images. For STFT, the image has a horizontal axis representing time and a vertical axis representing frequency whilst the colour at a point in the image represents the intensity of the frequency component (see, for example, Figure 4b). For CWT, the horizontal and vertical axis denote the time and the scale respectively whilst the colour at a particular combination of time and scale denotes the correlation of the signal with the mother wavelet at that combination.

Images from the reference and test signals are then compared using the method of correlation matching to determine the degree of similarity between them. Whether an STFT image is superior to the corresponding CWT image for a given signal in their relative ability to discriminate depends on the values of the correlation coefficients obtained from correlation matching.

3.1 Short-Time Fourier transform

The Fourier transform of a signal x(t) is defined as

$$F[x(t)] = X(f) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}\,dt \qquad (1)$$

For Equation (1) to be valid, the signal x(t) being transformed must be stationary, which means that its amplitude distribution does not depend on absolute time. In other words, the moments of the distribution – for example, mean, variance, and so on - are stationary. A signal with localized events is clearly not stationary.

To overcome this difficulty, the signal is divided into segments such that within each segment the stationarity property is approximately true and so Equation (1) can be applied. The Short-Time Fourier Transform (STFT), also known as the Windowed Fourier transform (WFT), does exactly that [7]. STFT uses a constant-width time window, and the segment of the signal exposed is transformed into the frequency domain. As the window slides to a new position along the time axis, the Fourier transform is again computed. This is repeated until the whole duration of the signal is covered. Mathematically, STFT can be expressed as a function of the frequency ω and the position b along the time axis; thus,

$$F(\omega,b) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\tau)g(\tau-b)e^{-j\omega\tau}\,d\tau \qquad (2)$$

This is the Fourier transform of the function f(t) windowed by the function g(t) for all b. A disadvantage of this method is that the time and frequency resolutions are constant as determined by the constant size of the window used and the fixed number of points of the signal exposed by window. Consequently, when the signal has a wide frequency bandwidth, STFT cannot give the high resolutions for frequency and time simultaneously.

The function $F(\omega, b)$ in Equation (2) is depicted as a two-dimensional colour-coded map where the x-axis denote b, which, for a time signal, is time itself, and the y-axis denotes ω, the frequency. The colour at a point (ω, b) represents $F(\omega, b)$.

3.2 Wavelet transform

Continuous wavelet transform (CWT) converts a time signal f(t) into the time- and scale- domain. The basis function used is a mother wavelet, ψ(t), analogous to the sine or cosine function used in Fourier Transform. CWT is defined by [8,9]

$$F_{\Psi}(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\Psi\left(\frac{t-b}{a}\right)dt \qquad (3)$$

The quantity in Equation (3)

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \qquad (4)$$

is the wavelet function. The position variable b is the translation parameter, same as b in the STFT definition, Equation (2). Whereas STFT uses a frequency variable ω,

CWT involves a scale variable *a*. Scaling is a primary characteristic of wavelet analysis. The mother wavelet function ψ((t) in Equation (3), besides being a basis function, also plays the role of a window function, equivalent to g(t) in Equation (3). The scale variable *a* is loosely the reciprocal of the frequency variable ω; changing *a* controls the frequency band. Variable *b* controls the size of the window. It is therefore possible to adjust the resolutions of time and scale (frequency) independently by changing *a* and *b*.

The function $F_\psi(a, b)$ in Equation (2) is depicted as a two-dimensional colour-coded map where the x-axis denote *b*, which, for a time signal, is time itself, and the y-axis denotes *b*, the scale. The colour at a point *(a, b)* represents $F_\psi(a, b)$.

3.3 Correlation matching

Correlation matching is used for pattern recognition in the proposed method. The two-dimensional transformed images obtained from either STFT or CWT is used to compute the similarity coefficient between the reference and test images. The method works as follows.

Given an image f(x,y) of size M by N pixels and a reference image w(x,y) of size JxK pixels, one starts by placing the geometric centre of the reference image w(x,y) on the pixel on the top-left corner of the image f(x,y) and then compute the correlation coefficient based on the pixels that overlap between the two images. One then shifts the geometric centre of w(x,y) one pixel to the right along the first row of f(x,y) at which another correlation coefficient is calculated. Each pixel in the first row is visited in this fashion; and when the first row is completed, other rows are visited in turn until the whole image f(x,y) is covered. The result is an MxN correlation coefficient matrix. The correlation coefficient in row *s* and column *t* of the matrix is given by [10]

$$\gamma(s,t) = \frac{\sum_x \sum_y \left[f(x,y) - \bar{f}(x,y)\right]\left[w(x-s,y-t) - \bar{w}\right]}{\left\{\sum_x \sum_y \left[f(x,y) - \bar{f}(x,y)\right]^2 \sum_x \sum_y \left[w(x-s,y-t) - \bar{w}\right]^2\right\}^{1/2}}$$

where

s = 0,1,2,...,M-1, t = 0,1,2,..,N-1, $\bar{w}$ =is the average value of pixels in w(x,y) (computed only once), and $\bar{f}(x,y)$ = the average value of f(x,y) in the region coincident with the w(x,y). Furthermore, it should be noted that the summation is taken over the image coordinates common to both f and w.

Figure 2 illustrates the procedure. The correlation coefficient γ(s,t) is in the range of –1 to 1. For perfect correlation,

$$f(x,y) = w(x,y), \Rightarrow \gamma(s,t) = 1$$

and for imperfect correlation,

$$1 \geq \gamma(s,t) \geq -1$$

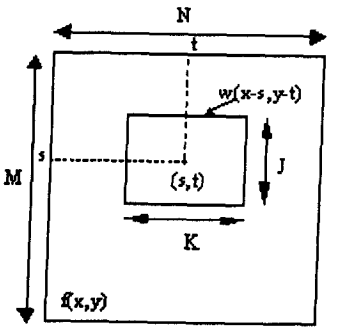


Figure 2 Correlation matching of f(x,y) and w(x,y) at point (s,t).

4. Simulation study

To determine the relative efficiency of discrimination between STFT and CWT, the procedure of Figure 1 was adopted. Both the reference and test signals are transformed by the same technique – either STFT or CWT; the images are then compared using correlation matching producing a correlation value between –1 and +1.

Four artificial signals with characteristic wave shapes were generated in Matlab using the following formulae:

Type 1 waveform

$$y(t) = e^{-0.01t}\sin(120\pi) \quad (6)$$

Type 2 waveform

$$y(t) = e^{-0.02t}\sin(240\pi) + \sin(20\pi) + \sin(10\pi) \quad (7)$$

Type 3 waveform

$$y(t) = e^{-0.01t}[\sin(50\pi) + \sin(30\pi)] \quad (8)$$

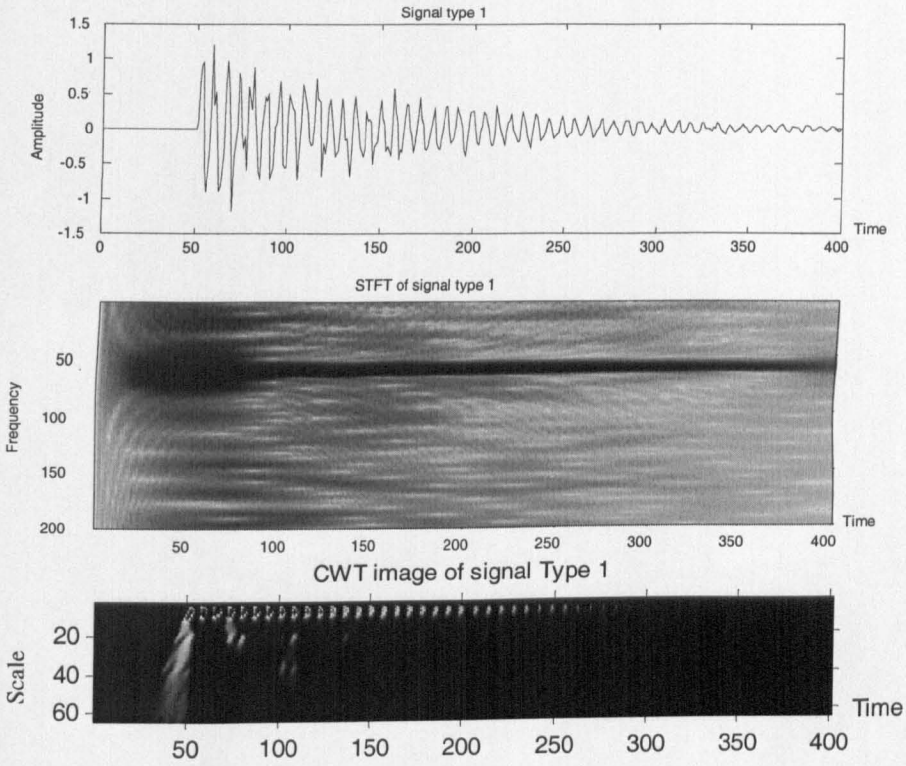
Type 4 waveform

$$y(t) = e^{-0.06t}\sin(20\pi) + e^{-0.09t}\sin(10t) \quad (9)$$

The signals thus generated were intentionally corrupted with noise that followed a normal distribution (produced using the Matlab function, 0.2*randn) to make the signals look more realistic. They are shown as in Figures 3a, 4a, 5a and 6a.

There is a distinction between a reference signal and a test signal. A reference signal of a particular type was produced from an average of seven instances of the signal of that type whereas a test signal is merely one single instance.

STFT was applied to the reference and test signals for each of the four signal types and a correlation coefficient matrix then computed from the corresponding pair of two-dimensional colour-coded maps shown as Figures 3b, 4b, 5b and 6b. The maximum value was then taken from the entries in the correlation coefficient matrix and would be used to indicate the extent of similarity between the reference and test signals. Three test signals were used for the comparison and the results are shown in Table 1. It is noted that in each case, correct matching has been achieved as the highest correlation coefficients always appear in the expected row.

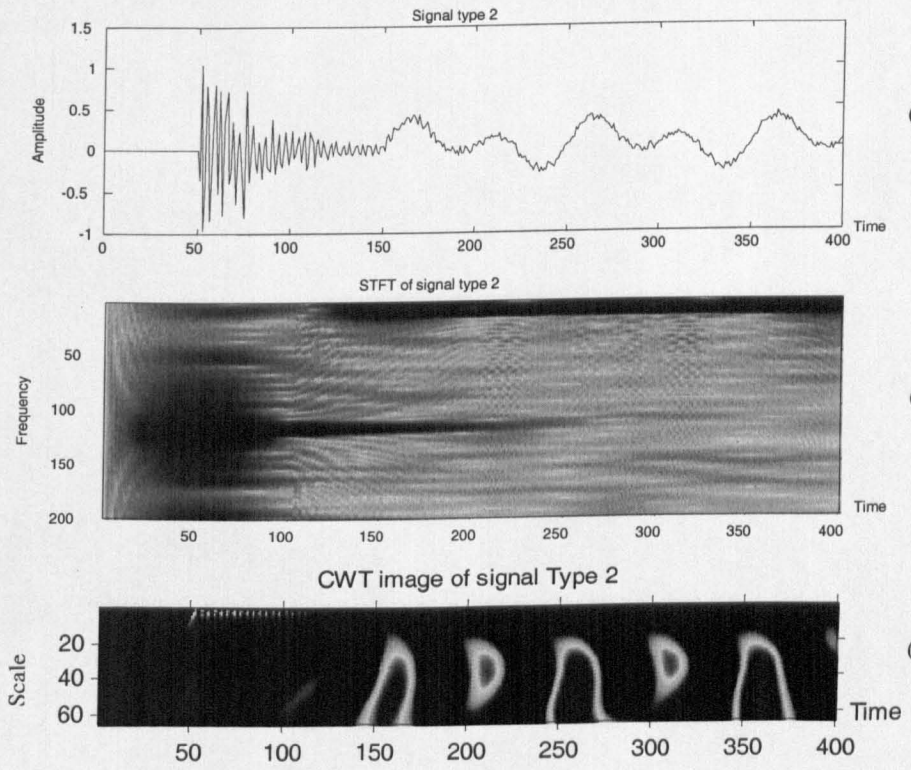


(3a)

(3b)

(3c)

Figure 3 Source simulated signals and its STFT and CWT images of signal Type 1

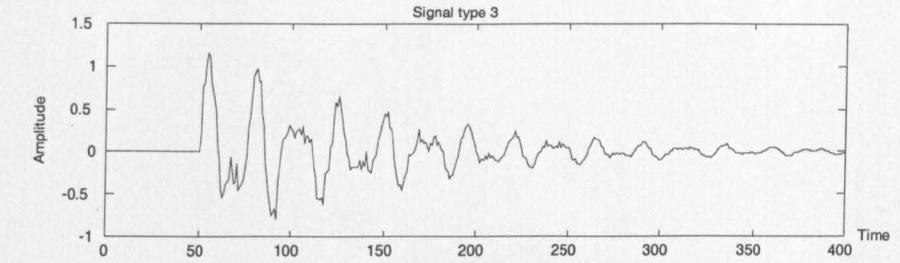


(4a)

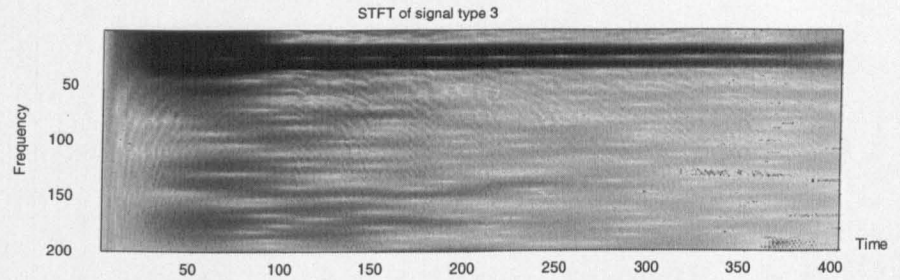
(4b)

(4c)

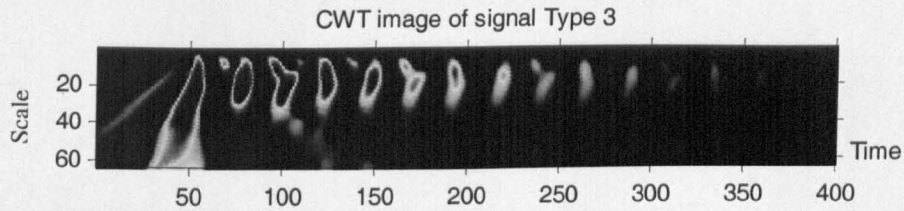
Figure 4 Source simulated signals and its STFT and CWT images of signal Type 2



(5a)

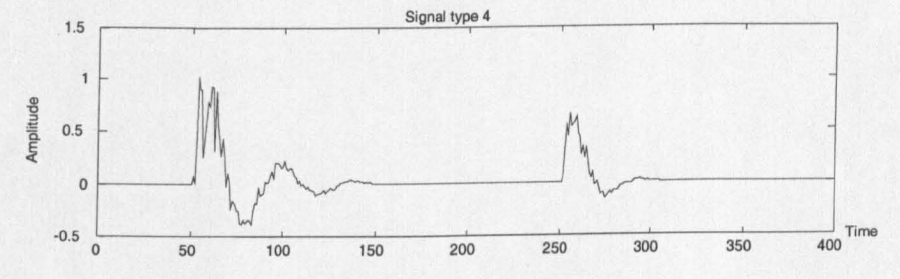


(5b)

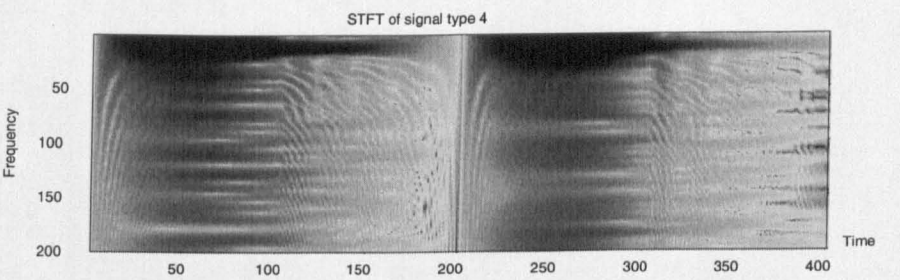


(5c)

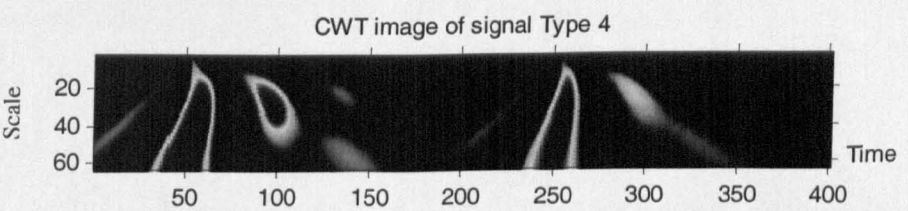
Figure 5 Source simulated signals and its STFT and CWT images of signal Type 3



(6a)



(6b)



(6c)

Figure 6 Source simulated signals and its STFT and CWT images of signal Type 4

Table1 Correlation matching results of STFT images

Correlation coefficient		Test image (Type 1)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.832	0.8211	0.8347
	Type 2	0.1376	0.1613	0.1753
	Type 3	0.5188	0.525	0.545
	Type 4	0.3578	0.3578	0.343
Correlation coefficient		Test image (Type 2)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.2431	0.2974	0.3129
	Type 2	0.7124	0.6723	0.6562
	Type 3	0.303	0.3328	0.3475
	Type 4	0.3636	0.3636	0.3346
Correlation coefficient		Test image (Type 3)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.5004	0.5298	0.512
	Type 2	0.1838	0.2176	0.1674
	Type 3	0.8053	0.7932	0.7819
	Type 4	0.4588	0.4588	0.4515
Correlation coefficient		Test image (Type 4)		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.3423	0.3628	0.3485
	Type 2	0.3232	0.3086	0.2373
	Type 3	0.4127	0.4645	0.4413
	Type 4	0.9057	0.9057	0.9199

Table2 Correlation matching results of CWT images

Type 1		Test image		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.9041	0.9074	0.8715
	Type 2	-0.0889	-0.0498	-0.0797
	Type 3	0.2964	0.2678	0.3092
	Type 4	0.1157	0.1046	0.1208
Type 2		Test image		
		No. 1	No. 2	No. 3
Reference image	Type 1	-0.0927	-0.0479	-0.098
	Type 2	0.9909	0.9891	0.9903
	Type 3	-0.0825	-0.0469	-0.1264
	Type 4	0.2166	0.2099	0.2327
Type 3		Test image		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.3469	0.3008	0.2279
	Type 2	-0.1413	-0.1047	-0.1233
	Type 3	0.9739	0.9453	0.9761
	Type 4	0.2718	0.2607	0.2805
Type 4		Test image		
		No. 1	No. 2	No. 3
Reference image	Type 1	0.2449	0.1292	0.0206
	Type 2	0.1432	0.1481	0.1626
	Type 3	0.26	0.3313	0.3112
	Type 4	0.975	0.9738	0.9756

time and scale (obtained by dilation parameter). Figure 5 compares both transformed images obtained from the same source of simulated signal.

Correlation coefficient is then computed by applying correlation matching to reference and test image. This can give the similarity between two images in order to classify test image into its category. The highest correlation coefficient is used to identify the class of test signal in which belong to. The results of correlation coefficient of STFT images and WT images are shown in table 1 and 2 respectively.

To examine how 3-D images are classified, there are three sets of unknown (test) signals and reference signals. The correlation coefficient value of the test signal of either STFT or WT gives highest correlation value for all unknown signal in its class which it belong to. Therefore, the recognition rate of the source simulated signal is 100 percent. However, discrimination capability of using STFT image is poorer than using STFT image, whilst the correlation coefficient of WT image gives the higher dissimilarity than the other classes. In addition, the WT image can also reflect the discontinuity characteristic of the signal.

5. Experimental application of the proposed method

After verifying the program by using computer simulated source signal, the actual vibration signals from a bearing test rig is performed. The application of the WT image that appears to be superior to STFT is then utilised in order to classify bearing faults. The experimental setup consists of a spindle driven by a DC motor. The rotation speed can be controlled by using potentiometer to control the speed of the motor. The electric motor provided 20Hz rotation for the bearing shaft. The test was performed in a laboratory using the test setup shown in Figure 7.

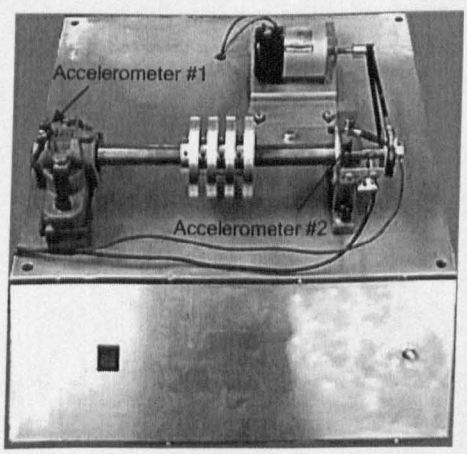


Figure 7 Test rig set-up

A rotating shaft is supported by two bearings. Tapered bored FAG 20205K.T.C3 (self-aligning single row) and ball FAG 6304 bearings were used in this study. Vibration signals were measured in the radial direction by using PCB 333A12 accelerometer. One is placed on the top of the non-drive end housing, another one is mounted on the drive-end housing labeled #1 and #2 respectively. The sensor data is digitised with a National Instrument with 3 channels. The output signals of the sensors were low-pass filtered at 1 kHz for anti-aliasing before sampling. The sampling rate used in the experiments was 2,000 samples/second.

6. Results and discussion

Measurements were obtained from two kinds of machine conditions: normal bearing and a defected bearing. In order to accelerate a defect progress, an artificial defect is generated by scratching an outer raceway of bearing with an electrical discharge machine. Since it is impossible to access the inner race of the bearing. Figure 8 shows the signal measured on normal and defected bearing.

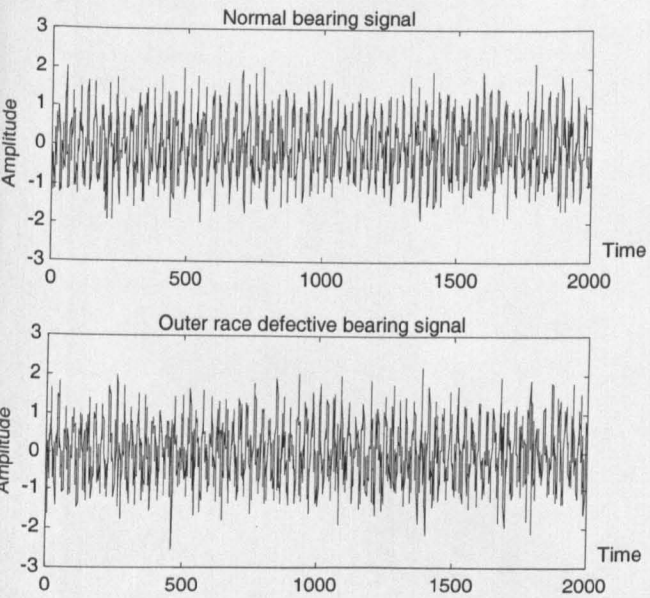


Figure 8 The signal measured from normal and outer race defective bearing.

In the experiment, the signals are collected separately as reference and test (unknown) sets. Selection of appropriate images is important in the proposed method of signal classification. To represent a range of the signals, 7 waveforms were selected for each class. Then, a reference signal can be obtained from averaging the entire signal of reference set. The 3-D reference image is obtained by applying Wavelet transform. The test set consists of three unknown vibration signals. To achieve 3-D image of test signal, this can be performed similarly as reference set.

Figure 9 shows the 16000 pixels used for the correlation matching classification. The image consists of 64 pixels in frequency scale and 250 pixels in time scale.

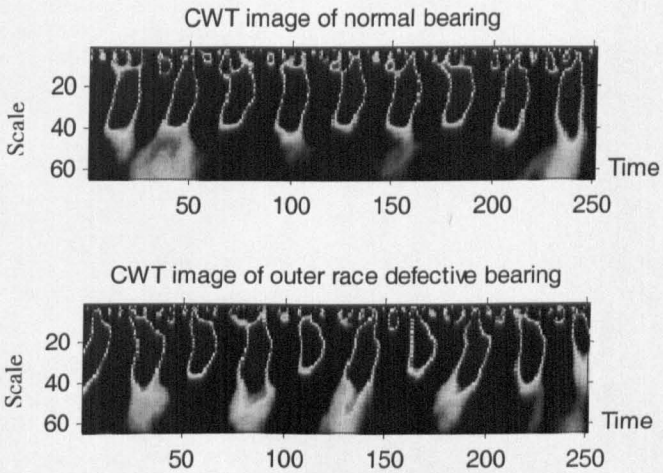


Figure 9 The 3-D WT image obtained from normal and defected bearing signal.

The classification is performed by using the 3-D image as characteristic feature of the signal. The correlation coefficient results of normal bearing which are computed by using equation (5) are shown in table 3.

Table 3 Correlation coefficient results

Condition of reference image	Test image			
	Condition	No. 1	No. 2	No. 3
Normal bearing	Normal	0.4660	0.8710	0.7204
	Defected	-0.3619	-0.2655	-0.1806
Defective bearing	Normal	-0.3427	-0.3642	-0.4186
	Defected	0.6655	0.3160	0.3102

From the results, the recognition is promising as it can be seen by using highest correlation value obtained from WT image to identify its machine condition. Intuitively, another set which has lower correlation coefficient is belonged to defected bearing category. However, the correlation coefficient results achieved by real signals are less than the one obtained from source simulated signals. This may caused by the wide band noise spread over the measured signal. Therefore, it results the change of the shape of the 3-D image. Nevertheless, even the distortion of the 3-D image does exist, the characteristic feature of the signal is significant enough for signal classification.

7. Conclusion

In order to classify the machine condition, there are many conventional methods yet proven to reveal neither transient event nor simultaneous representation of time and frequency domain. Therefore, joint time and frequency domain analysis is investigated in this study. Short-time Fourier transform and Wavelet transform is effective method of time-frequency analysis and powerful tool for machine condition monitoring. Image processing technique has been employed to classify the vibration signature of the bearing failure. The correlation matching is applied to the transformed image and similarity coefficient is then computed.

From the simulation result, it has been shown that using WT images for classification is superior to STFT images in its higher correlation coefficient and discrimination capability. Therefore, WT image is used as characteristic features for machine condition classification. Experimental application from a machine bearing shows that the proposed method can be used successfully to detect a bearing failure.

Acknowledgements

The authors wish to thank the British Steel, who kindly provided us their test rig. The authors are also grateful to Thai government for their support.

References

[1] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk and S. Liang Adaptive prognostics for rolling element bearing condition. Mechanical systems and signal processing, 1999,13(1),103-113.

[2] S. Zhang, R. Ganesan, and G.D. Xistris Self-organising neural networks for automated machinery monitoring systems. Mechanical systems and signal processing, 1996,10(5),517-532.

[3] C. Pachaud, R. Salvetat and C. Fray Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces. Mechanical systems and signal processing, 1997,11(6),903-916.

[4] J. Shiroishi, Y. Li, S. Liang, T. Kurfess and S. Danyluk Bearing condition diagnostics via vibration and acoustic emission measurement, Mechanical systems and signal processing, 1997,11(5),693-705.

[5] L. Qu, X. Liu, G. Peyronne and Y. Chen The Holospectrum: A new method for rotor surveiliance and diagnosis, Mechanical systems and signal processing 1989, 3(3),255-267.

[6] W. J. Wang and P. D. McFadden Early detection of gear failure by vibration analysis-I. Calculation of the time-frequency distribution. Mechanical systems and signal processing,1993,7(3),193-203.

[7] M. S. Safizadeh, A. A. Lakis and M. Thomas Using short-time fourier transform in machinery fault diagnosis, International journal of COMADEM, 2000,3(1),5-16.

[8] L. Gaul and S. Hurlebaus Identification of the impact location on a plate using wavelets, Mechanical systems and signal processing, 1997,12(6),783-795.

[9] H. Suzuki, T. Kinjo, Y. Hayshi, M. Takemoto and K. Ono Wavelet transform of acoustic emission signals, Journal of acoustic emission,14(2),1996,69-84.

[10] Rafael C. Gonzales, Richard E. Woods, Digital Image Processing, Addison wesley,1993.583-586.