



A fault detection tool using analysis from an autoregressive model pole trajectory

Suguna Thanagasundram*, Sarah Spurgeon, Fernando Soares Schlindwein

Department of Engineering, University of Leicester, University Road, Leicester LE1 7RH, UK

Received 28 November 2006; received in revised form 17 March 2008; accepted 25 March 2008

Handling Editor: M.P. Cartmell

Available online 13 May 2008

Abstract

A new scheme is proposed that combines autoregressive (AR) modelling techniques and pole-related spectral decomposition for the study of incipient single-point bearing defects for a vibration-based condition monitoring system. Vibration signals obtained from the ball bearings from the high vacuum (HV) and low vacuum (LV) ends of a dry vacuum pump run in normal and faulty conditions are modelled as time-variant AR series. The appearance of spurious peaks in the frequency domain of the vibration signatures translates to the onset of defects in the rolling elements. As the extent of the defects worsens, the amplitudes of the characteristic defect frequencies' spectral peaks increase. This can be seen as the AR poles moving closer to the unit circle as the severity of the defects increase. The number of poles equals the AR model order. Although not all of the poles are of interest to the user. It is only the poles that have angular frequencies close to the characteristic bearing defect frequencies that are termed the 'critical poles' and are tracked for quantification of the main spectral peaks. The time-varying distance, power and frequency components can be monitored by tracking the movement of critical poles. To test the efficacy of the scheme, the proposed method was applied to increasing frame sizes of vibration data captured from a pump in the laboratory. It was found that a sample size of 4000 samples per frame was sufficient for almost perfect detection and classification when the AR poles' distance from the centre of unit circle was used as the fault indicator. The power of the migratory poles was an alternative perfect classifier, which can be used as a fault indicator. The analysis has been validated with actual data obtained from the pump. The proposed method has interesting potential applications in condition monitoring, diagnostic and prognostic-related systems.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Bearing failures are one of the most common reasons for breakdown of rotating machines in industry today. Bearing defects manifest themselves as either excessive wear or damage in the rolling ball elements as well as in the inner/outer races of the bearings [1]. Fault identification of a ball-bearing-related phenomenon using conditional maintenance techniques has been the subject of extensive research for the last two decades [2–4]. One of the possible approaches to fault monitoring of the bearings is the processing of vibration signals obtained from the external housings in which the bearings are mounted for extraction of diagnostic features

*Corresponding author. Tel.: +44 116 252 5053.

E-mail address: st89@le.ac.uk (S. Thanagasundram).

[5]. This technique is more commonly known as vibration signature analysis and there are many conventional procedures based on time harmonic and power spectrum analysis that have shown considerable success in detecting failures in machine components [6,7]. If the bearings are in good condition, the distribution of the demodulated vibration signals is Gaussian-like. When the bearings are damaged, the appearance of cracks and spallings disturbs the signals, modifying its distribution. Wide band impulse-like signals are generated when the bearings pass over the defect points at a frequency determined by shaft speed, bearing geometry and defect location. The defective bearings' vibration signatures are spread across a wide frequency band and are easily masked by noise and frequency components generated by the machine. Theoretical models of single and multiple point defects of the vibration produced by a faulty bearing under constant and varying radial loads, have been established by McFadden and Smith [8,9]. Their model takes into account the impulse series generated by a point defect in a bearing modelled from first principles as a function of the rotation and geometry of the bearing, the modulation of the periodic signal caused by non-uniform bearing load distribution, the transfer function of the vibration transmission from the rolling element bearing to the transducer, as well as the exponential decay of vibration.

Raw vibration signals obtained from rotating machinery are usually complicated and immediate information for fault detection is not available. In diagnostic and prognostic applications, there exist a wide variety of models, algorithms and tools for data pre-processing and analysis to extract the useful information from the raw signals. Amongst these different time- and frequency domain-based diagnostic techniques, vibration analysts often rely on the power spectral densities (PSDs) of vibration data to monitor the health of moving parts of machinery as this method provides the most useful information. The spectral components allow the identification of several types of faults. Common failures such as bearing faults and gear problems can be detected by trending major frequency components and their amplitudes. Most of the frequency domain methods used in industry today are based on the Fast Fourier Transform (FFT) technique [10]. However, the FFT-based spectrum analysis method suffers from some shortcomings. One of its major setbacks is that a large number of frequency components have to be monitored due to the complexity of the system. A standard approach in evaluating an instantaneous frequency implies the computation of the whole spectrum first and then estimation of the amplitude of a particular frequency of interest. For instance, if ball-bearing defect frequencies [3] such as ball pass frequency of outer race (BPFO), ball pass frequency of inner race (BPMFI), ball spin frequency (BSF) and fundamental train frequency (FTF, also known as cage frequency) are to be detected, FFT spectra are computed and then the spectra are filtered to monitor the presence of the fault frequencies. Such a process can firstly be time consuming as whole frames of data have to be estimated. Secondly, it can be power intensive as it involves processing time proportional to $N \log_2 N$ computations where N is the sample size. Another concern of the FFT technique is that a large enough sample size has to be used for the spectral estimation for reasonable resolution capabilities as the resolution of the FFT is inversely proportional to the frame size utilised [11]. This might not be appropriate in real time applications.

An interesting statistical signal processing alternative is to evaluate directly the frequencies of interest. In this case only those frequencies have to be estimated instead of the whole spectrum. This provides a reduction in computing time and effort facilitating real time estimation. In this study, characteristic bearing defect frequencies are extracted from the pole frequencies of a parametric time series AR model [12]. The AR model is used to decompose the signals into a set of poles, which have a correspondence to the peaks of the signals PSDs. Using the AR estimation technique, it is not necessary to obtain the whole spectrum. Instead, the evaluation of the pole frequencies of interest from the derived AR parameters would suffice as AR modelling allows spectral decomposition. This often just involves the calculation of the AR coefficients and the variance of the input vibration signal [13]. Small order AR models can efficiently estimate the pole frequencies which correspond to the poles of the bearing defect frequencies. The AR technique also only requires a fraction of the samples that are required by the FFT method for the same resolution [12]. When compared to the traditional FFT method, the resolution of the AR technique is higher due to its implicit extrapolated autocorrelation sequence. This means that smaller sample sizes can be used for PSD estimation. The positions of the poles, which are the roots of the AR coefficient polynomial, vary for every frame of vibration data and the time-varying behaviour of the spectral components can be monitored by tracking the movement of the AR poles. Faults can be predicted by movement of poles in the complex plane as the pole positions are expected to move closer to the unit circle as the severity of the defect increases. From the position of the poles inside the

unit circle, classification and quantification of the main spectral peaks of defect frequencies can be easily performed, leading to the possibility of having frame-to-frame monitoring of spectral parameters of interest.

In this paper, we present a study of fault identification through differences in the behaviour of the AR poles for vibration signals collected from two similar high speed dry vacuum pumps, one with a healthy set of bearings and another with a ball bearing with an inner race defect. The poles of the AR model are extracted as the feature relating to the characteristic bearing defect frequencies and a classification scheme based upon the position of the coordinates of the autoregressive (AR) poles is proposed and tested with real data.

The organisation of the paper is as follows. The paper first summarises the modelling aspects of the vibration signals based on tracking of AR pole movement on the parametric z domain. A description of the experimental hardware and data acquisition set-up is then given. Experimental tests on vibration data obtained from a pump with normal bearings and damaged bearings are given. Receiver operating characteristics curves are plotted for increasing frame sizes of vibration data and the performance of the classification scheme is tested. Finally, the conclusions are presented.

2. The AR modelling technique

AR modelling belongs to a class of modern spectral analysis techniques, which is more generally known as autoregressive moving average (ARMA) time series modelling [10]. The AR method is the preferred method for this class since it is the best compromise between temporal resolution and speed, efficiency and simplicity of algorithms. An AR process of model order p can be described by Eq. (1) where a_k are the AR parameters, $e[t]$ is white noise with zero mean and variance σ^2 and t is the discrete-time index [12]. The same equation expressed as a linear filter in the z -transform domain is stated as Eq. (2):

$$x[t] = - \sum_{k=1}^p a_k x[t-k] + e[t] \quad (1)$$

$$X[z] = -X[z] \sum_{k=1}^p a_k z^{-k} + E[z] \quad (2)$$

where $X[z]$ and $E[z]$ are the z -transforms of $x[t]$ and $e[t]$, respectively. $H[z]$ is defined as the AR polynomial of the model transfer function relating the input to output and is denoted by Eq. (4):

$$X[z] \left[1 + \sum_{k=1}^p a_k z^{-k} \right] = E[z] \quad (3)$$

$$H[z] = \frac{X[z]}{E[z]} = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (4)$$

The poles, p_k , are obtained by finding the roots of the AR coefficient polynomial in the denominator of $H[z]$. An AR model's transfer function contains poles in the denominator plus only trivial zeros in the numerator at $z = 0$, so it is referred to as an "all-pole" model. Since the coefficients of $H[z]$ are real, the roots must be real or complex conjugate pairs. The number of poles in the z plane equals to p , the AR model order:

$$H[z] = \frac{1}{(z + p_0)(z + p_1) \dots (z + p_p)} = \frac{1}{z^{p+1} \prod_{k=0}^p (1 + p_k z^{-1})} \quad (5)$$

The pole representation of the AR model, Eq. (5), is better than the AR coefficient representation, Eq. (1), for fault detection as the AR filter coefficients a_k are not stable and are highly dependent on the filter order and do not reflect the signal properties. The poles p_k are more closely related to the spectral form and contain important information on the system condition. For a stable AR filter, all $|p_k|$ must be less than 1 and the nearer the pole is to the unit circle, the higher the corresponding peak in the AR spectrum.

There are four main methods for the estimation of the a_k coefficients: Yule–Walker, Burg, Least Squares Forward method and Least Squares Forward Backward method [12]. The AR parameters or the reflection coefficients are estimated in all these methods. Generally all the four estimation methods produce comparable

estimates when using large data samples. However, for moderate sample sizes, differences may arise in the behaviour of the estimation methods, especially the Least Squares Forward method and the Least Squares Forward Backward method that can yield unstable models.

In this study, we have chosen the oldest way of estimating the AR parameters, which is the Yule–Walker approach. The Yule–Walker method calculates the AR parameters recursively, using estimated autocorrelation functions up to order p . The Yule–Walker method with the Levinson–Durbin recursion was used to find the AR coefficients [12] by solving a set of linear equations. However, the Burg method is the most reliable and preferred estimation technique as it estimates the autocorrelation functions by minimising the sum of squares of observed forward and backward prediction errors. The Burg method provides reliable parameter estimates for practical data as it is guaranteed to be stationary and is less biased than the Yule–Walker method. The strong bias of Yule–Walker gives unacceptable errors in many applications and should not be used for processes with unknown characteristics [14]. However, in the current work it nevertheless did give useful results. It might be expected that the Burg method would have given better results [15].

The PSD of the output signal, $P[z]$, is a function of the PSD of the input signal $P_e[z]$, which is noise. $P[z]$ is defined as

$$P[z] = H[z]H^* \left[\frac{1}{z} \right] P_e[z] \tag{6}$$

The PSD of white noise is a constant and is given by its variance σ^2 and Δt , the sampling interval, which is the inverse of the sampling rate f_s :

$$P_e[z] = \sigma^2 \Delta t \tag{7}$$

$P[z]$ can be obtained by replacing z with $\exp(j2\pi f \Delta t)$ and evaluating Eq. (8) along the unit circle for $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ once the a_k coefficients are known:

$$P[z] = \frac{\sigma^2 \Delta t}{\left| 1 + \sum_{k=1}^p a_k z^{-k} \right|^2_{z=e^{j2\pi f \Delta t}}} \tag{8}$$

Each pair of complex conjugate poles in Eq. (8) has a one-to-one relationship with a peak in the AR spectrum, $P[z]$, in the z domain. A P th order AR model with p poles will have a maximum of m peak frequencies where $m = p/2$ when p is even and $m = (p + 1)/2$ when p is odd. Not all poles give rise to sharp peaks in the AR spectrum. Only the poles which are close to the unit circle give rise to sharp peaks in the AR frequency spectrum (see Fig. 1). The other poles are equally distributed around the unit circle to create an equiripple ‘flat’ PSD estimation. The symmetry with respect to the real axis is related to the fact that the signals have real values and this advantage can be conveniently exploited by disregarding the poles in the negative imaginary plane, reducing the redundancy of poles in the pole tracking method.

Each pole p_k has a phase ϕ_k and a magnitude r_k which is the distance of the pole from the origin and $1-r_k$ which is the distance of the pole from the unit circle (see Fig. 2). By knowing the pole position inside the unit circle, the central frequency f_k of each peak can be obtained from the phase ϕ_k of the pole if the sampling

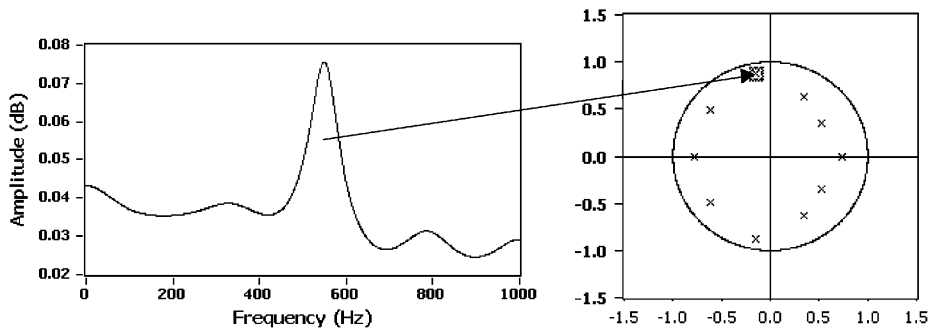


Fig. 1. Monitoring the peaks obtained from the AR model. In this case, AR model order was 8 and sampling frequency was 2000 Hz. Note that it is only the pole nearest to the unit circle that gives rise to a sharp peak in the AR spectrum.

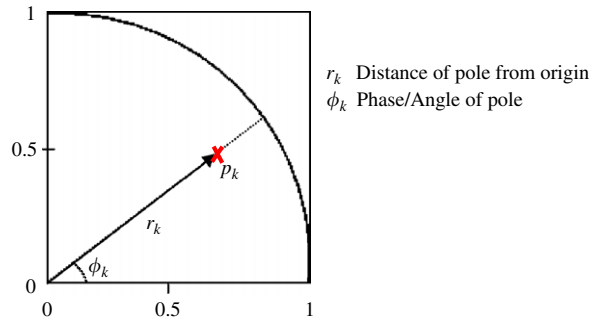


Fig. 2. Defining the parameters of an AR pole.

frequency f_s is known (Eq. (9)). Since the AR time series is quasi-steady stationary, the positions of the poles vary with every frame of data. The trajectory mapped out by the poles can be quantified by finding the area traversed by the poles over a period of time. This aspect can be used as a parameter to indicate the condition of the bearing. The area mapped out by a characteristic bearing frequency pole will be different depending on whether they were obtained from no-fault or faulty conditions. The area mapped out by the migratory poles is given by Eq. (10):

$$f_k = 2\pi\phi_k = \tan^{-1}(\text{Im}(z_k)/\text{Re}(z_k)) \times f_s/2\pi \tag{9}$$

$$\text{Area} = \frac{1}{2}(\Delta r_k)^2 \Delta\phi_k \tag{10}$$

The power associated with each spectral pole p_k is estimated from the residues of the complex poles as proposed in Ref. [16]. The transfer function in Eq. (5) can be rewritten as a partial fraction expansion. The residue (Eq. (11)) is simply the coefficient of the one-pole term $1/(1-p_k z^{-1})$ in the partial fraction expansion of $H[z]$ at $z = p_k$. Each residue r_k provides an estimate of the integrated power in the neighbourhood of the spectral frequency f_k associated with pole p_k . The spectral power P_k of the pole p_k (Eq. (12)) is obtained by multiplying real part of the residue term with the variance of the driving AR time series σ^2 and the scale factor n . $n = 2$ for complex conjugate poles and $n = 1$ for real poles at either 0 Hz or at the Nyquist frequency:

$$r_k = z^{-1}(z - p_k)H(z)|_{z=p_k} \tag{11}$$

$$P_k = \sigma^2 n \text{Re}[r_k]|_{z=p_k} \tag{12}$$

3. Laboratory set-up and data acquisition

This section contains a description of the test equipment and instrumentation used for obtaining the test signals used for the experimental phase of the work. A multistage IGX dry vacuum pump based on the ‘Roots and Claws’ principle [17] was used as the rotating machine. The schematic of the pump, the sensors used for capturing the data and the set-up of the data acquisition system are shown in Fig. 3.

The pump has a single row of deep groove ceramic bearings at both the high vacuum (HV) and low vacuum (LV) ends. Such dry vacuum pumps are the pump of choice in semiconductor and clean room environments where corrosive gases are abundant, temperatures are high and effluents resulting from the processes can cause seizure of the pumps. With constant use, it is possible that the ceramic bearings will fail over time because of their operation in such harsh running conditions. Whilst the design of dry vacuum pumps has remained relatively the same over the last 10 years, more emphasis is being employed in the evolution of intelligent pumping systems with embedded sensors and self-diagnostic capabilities as dry vacuum pumps have become a critical part of the integrated chip (IC) wafer manufacturing process. Pump failure can contribute to the significant loss of valuable products, e.g. loss of IC wafer batches, which produce significant financial losses. Also customers are affected by the inconvenience of down time. Hence high reliability and availability have

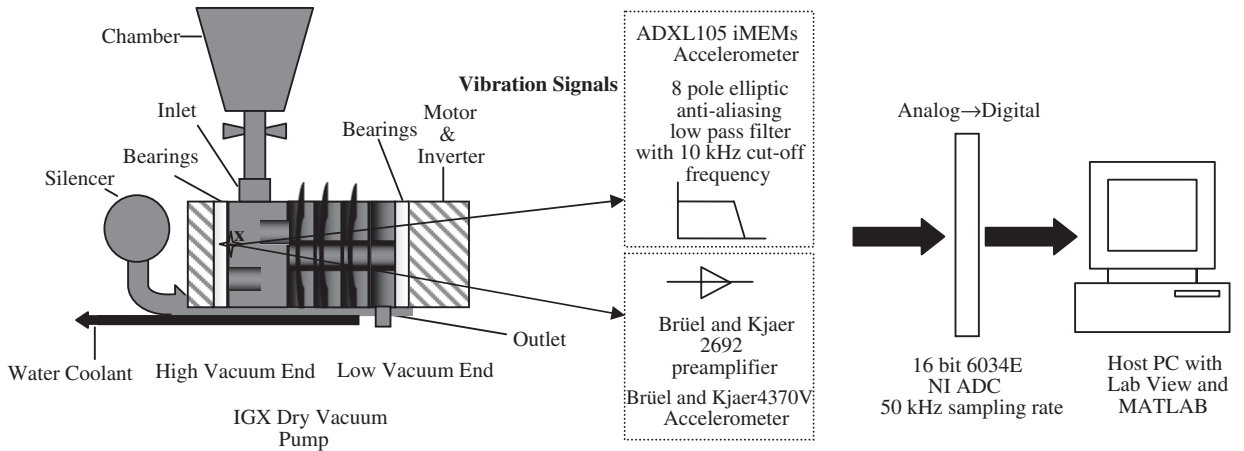


Fig. 3. Schematic of the complete data acquisition system. The ADXL105 and Brüel and Kjær 4370 V accelerometers were mounted radially on point marked X on the dry vacuum pump, near the high vacuum end.

become primary customer requirements and early detection of incipient faults is vital to avert costly failures and unscheduled maintenance.

To simulate a fault condition in the dry vacuum pump, a test bearing which had a single-point defect on the inner race was mounted on the HV end. The acceleration signal was captured on the pump housing directly over the bearing casing on the HV for normal (no-fault) and faulty conditions using two different types of accelerometers, namely, a surface micromachined accelerometer ADXL105 and a Brüel and Kjær (B&K) 4370 V accelerometer which were mounted radially on the pump running at 105 Hz [18]. The signals from the ADXL105 were filtered with an 8th order low pass elliptic anti-aliasing filter with a cut-off frequency of 10 kHz and an attenuation of 70 dB in the stop band. The vibration signals from the Brüel and Kjær 4370 V were conditioned using a Brüel and Kjær 2692 preamplifier that includes a 10 kHz low pass filter. The analogue to digital conversion was performed with a 16-bit NI 6034E analogue to digital converter (ADC) card. The sampling rate was set to 40 kHz and varying lengths of signals were acquired. The data were then downsampled to 2 kHz as it was known that the frequencies of interest lie in the range from 0 to 1 kHz. Prior to downsampling, the vibration signals were also pre-processed by amplitude demodulation [19]. If there is noise or other higher level vibrations generated by the other machine components, there is a chance that the harmonics of the defect frequencies may be buried in the spectrum of the other components. To overcome this limitation, the technique of amplitude demodulation is recommended. This involves two basic steps (1) bandpass filtering around one of the resonant peaks where there is structural resonance and (2) applying the Hilbert transform to the bandpassed signal to obtain the squared envelope [20]. This technique improves the signal-to-noise ratio of the signal of the vibration signatures for a more effective detection of bearing defects.

4. Relating AR pole positions with characteristic bearing defect frequencies

Formulas have been developed to calculate bearing defect frequencies for every bearing geometry, inner raceway, outer raceway and rolling elements [21]. For a bearing with a stationary outer race and inner rotating race, characteristic defect frequencies (Eqs. (13)–(16)) can be obtained for flaws in the outer race, inner race, ball bearings or in the cage as follows [3]:

$$\text{Bearing pass frequency outer race (BPFO)} = f \left(\frac{N_B}{2} \right) \left[1 - \frac{B_D}{P_D} \cos(\alpha) \right] \tag{13}$$

$$\text{Bearing pass frequency inner race (BPFI)} = f \left(\frac{N_B}{2} \right) \left[1 + \frac{B_D}{P_D} \cos(\alpha) \right] \tag{14}$$

$$\text{Ball spin frequency (BSF)} = \frac{f}{2} \left(\frac{P_D}{B_D} \right) \left[1 - \left(\frac{B_D}{P_D} \cos(\alpha) \right)^2 \right] \tag{15}$$

$$\text{Fundamental Train Frequency (FTF)} = \left(\frac{f}{2} \right) \left[1 - \frac{B_D}{P_D} \cos(\alpha) \right] \tag{16}$$

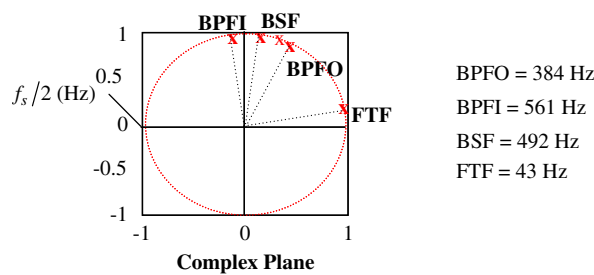
The five bearing parameters that must be known to calculate the bearing defect frequencies are, B_D -ball or roller diameter, P_D -pitch diameter, N_B -number of rolling elements, α -contact angle and f -shaft rotational frequency.

The specifications for the BOC Edwards IGX dry vacuum pump that was used as the test bed in these experiments are: number of balls = 9, pitch diameter = 46.2 mm, ball diameter = 9.5 mm and contact angle = 24.97°. Using the standard reference formulas, the theoretical ball-bearing defect frequencies BSF, BPFO, BPFI and FTF were estimated to be around 492, 384, 561 and 43 Hz, respectively when the pump’s running speed was set to 105 Hz.

The characteristic bearing defect frequencies can be transformed to pole locations in the z domain. As an example, consider the case of a defective bearing with an inner race crack set to rotate at 105 Hz. The theoretical relative AR pole phase angles ϕ_k can be worked out if the sampling rate is known and in this case, the sampling rate f_s used was 2000 Hz. The angle of the theoretical pole locations of the characteristic bearing defect frequencies are then worked out in degrees to be as shown in Fig. 4 using the equation given below:

$$\Delta \text{ angle} = \frac{\text{Characteristic bearing defect frequency}}{f_s/2} \times 180^\circ \tag{17}$$

The angles of the AR pole locations are related to the characteristic bearing defect frequencies. For a fixed rotating speed, the AR pole angles are fixed as they are determined by the geometrical shape of the ball bearings used but the distances of the pole locations from the unit circle are determined by the levels of vibration at that particular frequency. It is known that, as defects appear on the ball bearings and their severity increases with time, the amplitudes of the vibrations of characteristic bearing defect frequencies also increase (see Fig. 5). The appropriate alarm level for the vibration signal can be determined from standards such as the ISO 10816 [22] and ISO 7919 [23]. These can be translated to relative allowable amplitudes and alarm levels and hence corresponding pole displacements for the characteristic ball-bearing defect frequencies.



$$\Delta_{BPFO} = \frac{384 \text{ Hz}}{2000 \text{ Hz}/2} \times 180^\circ = 69.12^\circ$$

$$\Delta_{BPFI} = \frac{561 \text{ Hz}}{2000 \text{ Hz}/2} \times 180^\circ = 100.98^\circ$$

$$\Delta_{BSF} = \frac{492 \text{ Hz}}{2000 \text{ Hz}/2} \times 180^\circ = 88.56^\circ$$

$$\Delta_{FTF} = \frac{43 \text{ Hz}}{2000 \text{ Hz}/2} \times 180^\circ = 7.74^\circ$$

Fig. 4. Theoretical location of characteristic defect frequencies in the z plane can be worked out with standard reference formulas if the bearing dimensions are known. Pump was set to 105 Hz and sampling frequency used was 2000 Hz.

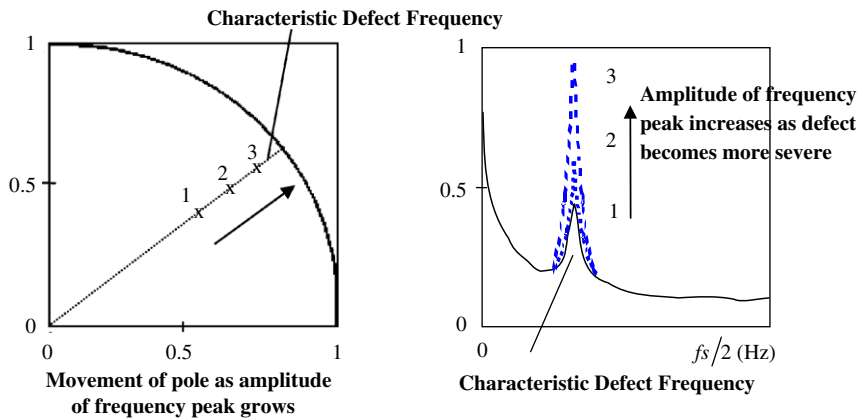


Fig. 5. Typical movement of pole as defect becomes more severe (from 1 to 3) and amplitude of vibration of characteristic defect frequency begins to increase.

The distance parameter of the poles can be used as a useful indicator to classify between faulty and non-faulty conditions. The frequencies of the characteristic bearing defect frequencies also do not remain the same due to variations of the speed of the machines resulting in pole angle variations in the z domain. The area of the region swept by the loci of the migratory poles can be used as another useful fault indicator for diagnostic purposes. The power parameter is another good indicator as it is reflective of the energy content of the bearing signal studied. The vibration signal of a defective bearing is normally much higher than the vibration signal obtained from good bearings as the energy of certain frequency bands are excited due to the presence of faults.

In reality, a bearing with an inner race fault has BPF_I occurring at 555 Hz, which is slightly less than 561 Hz (where 561 Hz is obtained from Eq. (14)). It was observed that the rotating speed of the pump's rotor shaft, on which the bearing case was directly connected to, was often less than the set speed of the pump due to rotor slip. It varied with external running conditions like the ultimate pressure of the inlet of the pump. In this case, the actual speed achieved by the pump was 104.12 Hz (corresponding to a slippage factor of 1%-slippage is typically less than 5%). The running speed of the pump had to be determined accurately for the diagnostics scheme, as this was the frequency that was used in the calculations of the bearing defect frequencies [24]. This effect was taken into consideration when translating the bearing defect frequencies into pole positions.

4.1. Fault detection using AR spectra

Fig. 6 shows an example of the spectra for the data with the implanted inner race fault on the HV end for the demodulated vibration signal plotted using both AR (Eq. (8)) and FFT techniques for frame sizes of $N = 4000$ samples. For the AR spectra, a model order of 10 was used. By careful examination, one can see that there is one major sharp spectral peak at the BPF_I line. The dominant peak corresponds to the inner race defect frequency (BPF_I) of 556 Hz and is clearly evident in both FFT and AR spectra.

5. AR pole-based monitoring

The proposed method of monitoring consists of two phases: training and detection. It is assumed that S bearing faults are to be monitored. These bearing faults can be BPF_O, BPF_I, BSF, FTF or even multiples of these depending on the application to be studied. The idea is to compare the frequency, distance, area and power of the poles of the AR model in different time segments. It was required to choose a certain model order that adequately represents the demodulated vibration data for all time segments. Selection of the model order in the AR process is of critical importance. Too low an order produces a smoothed estimate, while too high an order may cause spurious peaks and spectral line splitting. For the purpose of the fault detection tool, the optimum model order was first determined for both the normal and faulty data using order selection criteria (please refer to authors' related work on finding the optimum order for the vibration signals [25]). A 10th

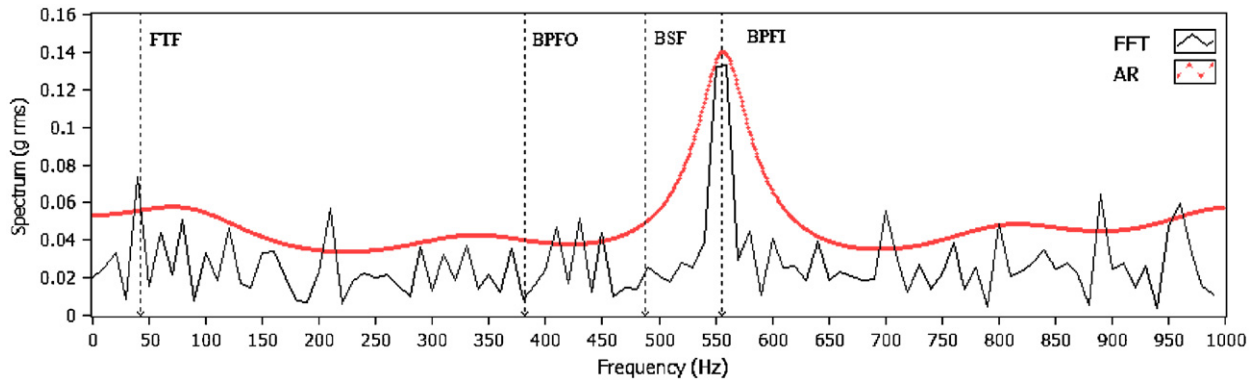


Fig. 6. FFT and AR Spectra for ADXL105 vibration signal obtained from a bearing with a seeded inner race fault. Speed of pump was set to 105 Hz. Sampling rate f_s was 2000 Hz. A sample size of $N = 4000$ samples was used. The ball pass frequency of inner race (BPF1) at 556 Hz is clearly evident in both spectra at 555 Hz (lower than theoretical BPF1 due to slip—see text). A model order of 10 was used for the AR spectrum.

order AR model was found sufficient to study the behaviour of the demodulated ADXL105 vibration signals mounted on the HV end for a pump operating in normal conditions and also for a pump fitted with a bearing with defects. Each bearing fault state is hence modelled by an AR model with a fixed model order of $p = 10$.

Training phase:

1. Each training set of vibration data was then formed constituting an observation of the data when there are no faults in the bearings. First, the AR parameters are found for the bearings in a healthy condition and subsequently, σ^2 , the variance of the input signal in the no-fault condition is also derived.
2. Changes in the condition of the ball bearing and the non-stationarity of the varying signal cause deviations in the AR coefficients which in turn result in deviations in the AR pole positions. The positions of the AR poles vary in every frame. The AR poles, which are the roots of the AR coefficient polynomial, are obtained per frame of data. There are p poles when a model order of p is used.
3. Out of these, the critical poles for each characteristic bearing defect frequency, are found. There is a one-to-one mapping between the characteristic bearing defect frequencies and the angular frequencies of the critical poles (Eq. (17)). The poles of interest, which have angles closest to the angles corresponding to the characteristic bearing defect frequencies, are termed the ‘critical poles’ and are monitored.
4. Next, parameters λ are derived from the critical poles in the normal condition. The parameter λ can be the area of the region swept by the migratory loci of the critical poles, their distances from the centre of the unit circle or the power of the critical poles. The parameters λ reflect the characteristics of the time invariant spectral parameters and comprise important information of the system condition. The impulses produced by ball bearings with defects will be modelled by the changing system dynamics and this in turn will be reflected in the parameters λ . The parameters λ are the conditional indicators that discriminate between the damaged and non-damaged bearings.

Detection phase:

Boundary conditions can be established using the known angles of characteristic defect frequencies and acceptable levels obtained from standards and this can help the optimum threshold for the classification boundary to be determined. Once the boundary conditions for the healthy signal are extracted and normalised for the healthy condition, the procedure is repeated for the detection phase with vibration signals from a test bearing. If the test bearing is a bearing with a defect, the parameter λ of the critical poles will not normally be within the acceptable levels of the established boundary conditions (refer to Fig. 7) or will be above a certain predetermined threshold. A reasonable threshold level can be chosen based on the probabilities calculated from the training data. In this case, the threshold value was determined using receiver operating characteristics (ROC) analysis.

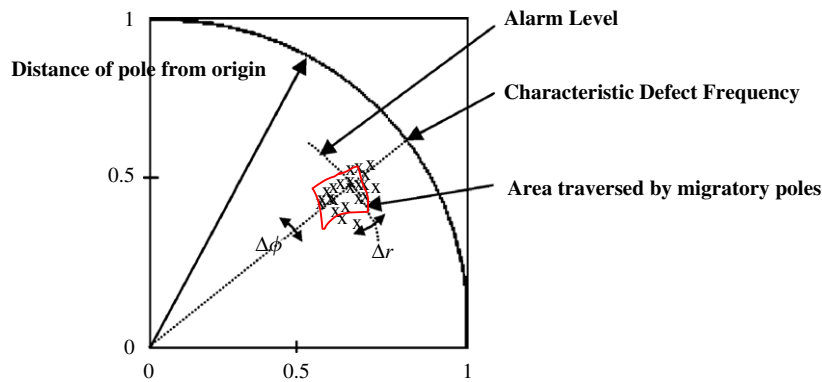


Fig. 7. Locus of a particular pole versus evolution of time. Note that pole is migratory and the area pole traverses can be calculated. Alarm level for the characteristic defect frequencies can be determined from standards such as the ISO 10816 and if pole crosses the alarm level, we can conclude that the defect is causing the characteristic frequency and hence the pole displacement is severe.

One important and useful measure of the performance of any diagnostic test is ROC curve analysis. A threshold is determined for separation of data into two classes based on some decision parameter and depending on whether the subject falls below or above the cut-off level, the subject is termed 'positive' or 'negative'. In reality, there will be some overlap between the two classes of data and depending on the position of the threshold, some subjects will be misclassified as 'false positives' and 'false negatives'. The sensitivity and specificity of the data are defined as

$$\text{Sensitivity} = \frac{\text{Number of true positive decisions}}{\text{Number of actually positive decisions}} \quad (18)$$

$$\text{Specificity} = \frac{\text{Number of true negative decisions}}{\text{Number of actually negative decisions}} \quad (19)$$

ROC curve analysis can be established by taking the parameters λ as the fault indicators, and then working out the condition and fault patterns of the ball bearings. Depending on what defect the bearing has, such as whether it is a defect on the inner race or outer race, the defect can be identified. The above explained diagnosis approach is under the assumption that only one fault is present. However in reality, multiple faults can develop simultaneously. In such a case, the scheme can be easily modified to incorporate the diagnosis of multiple faults.

5.1. A diagrammatic illustration of the procedure

Tracking the parameters λ is performed by processing each raw time domain sample through the stages shown in the diagram of Fig. 8. This diagram summarises all the key steps in processing the data and the major stages in obtaining the fault indicators from the mapped critical poles. The AR coefficients are obtained from each time sample with fixed model order of 10 and subsequently, the AR pole positions per frame of data are also derived. The critical poles corresponding to the bearing defect frequencies are identified. From the positions of the AR critical poles, parameters λ are derived. The parameters λ are compared with those obtained from non-faulty conditions and the bearing is diagnosed to be whether in faulty condition or not.

6. Results

The loci of the critical poles (BPFIs) on the z plane were tracked in different time segments, considering the mean and standard deviations of the parameters λ , seeking differences between results for damaged and undamaged bearings (see Fig. 9). Frame sizes of 4000 samples were used and 100 frames of data each of length 2 s were subsequently processed to monitor the movement of the BPFIs pole corresponding to the inner race defect frequency. A 10th order AR model was used to estimate the parameters for successive segments. The

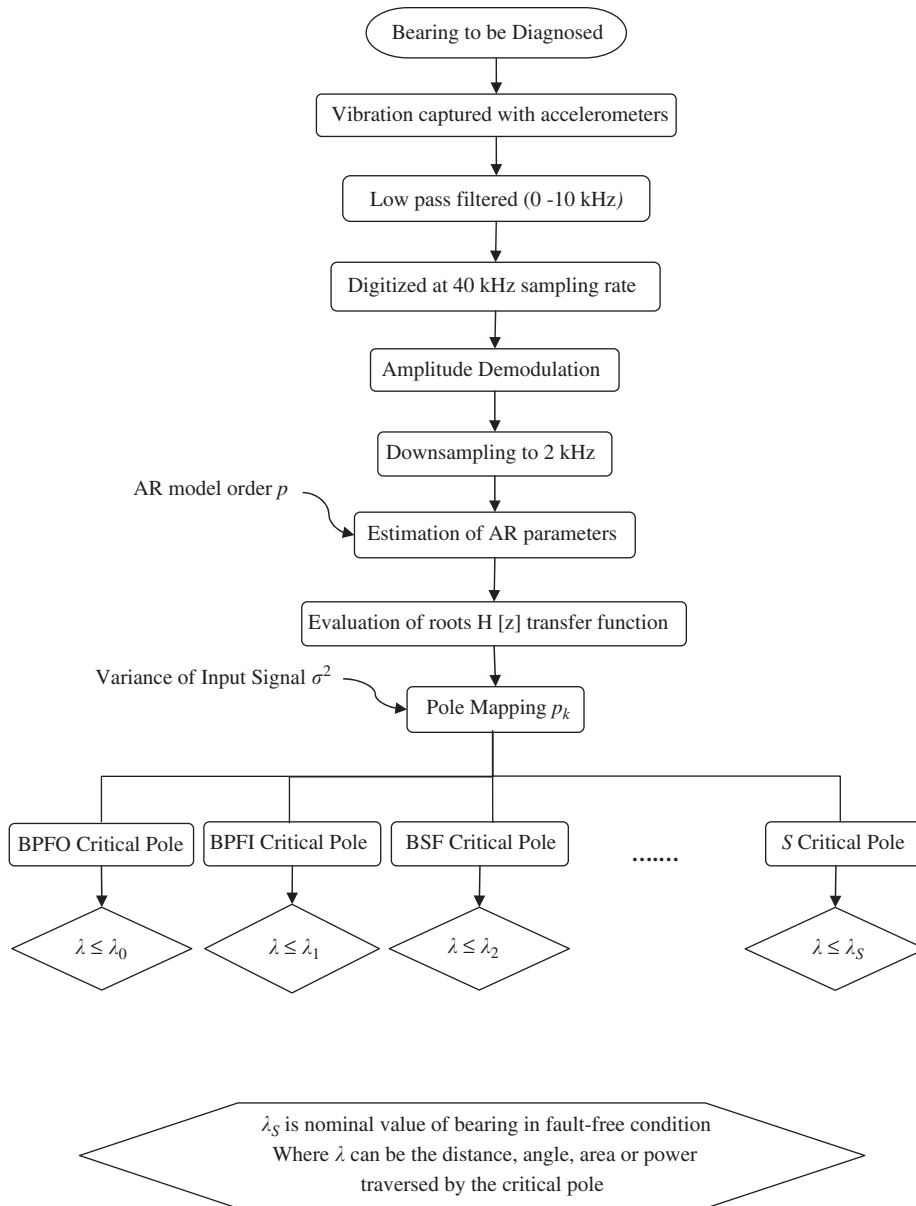


Fig. 8. Block diagram of the procedure for bearing fault detection using AR pole tracking.

distance of the BPFI pole from the origin, power of the BPFI pole, the angle mapped by the BPFI pole and the cumulative area traversed by the BPFI pole were plotted for the vibration data for both normal and faulty conditions. The mean \pm standard deviation distance of the BPFI poles from the origin in the non-faulty case was 0.8005 ± 0.0580 . The mean \pm standard deviation distance of BPFI poles from the origin in the faulty case was 0.9303 ± 0.0148 . It can be seen that the mean distance of the BPFI poles was much larger for the faulty condition than in the normal case. This will be seen as bigger amplitude of spectral peaks associated with critical poles (amplitude of spectral peaks is proportional to the inverse of the distance of a pole to the unit circle) when the AR spectrum is plotted for the faulty condition.

The mean \pm standard deviation of the power of the BPFI poles in the non-faulty case was $9.570 \times 10^{-4} \pm 2.831 \times 10^{-4}$ and in the faulty case 0.0369 ± 0.0086 , respectively. Clearly the power of the faulty poles is much larger than for the non-faulty case. The power and distance of the poles from the origin

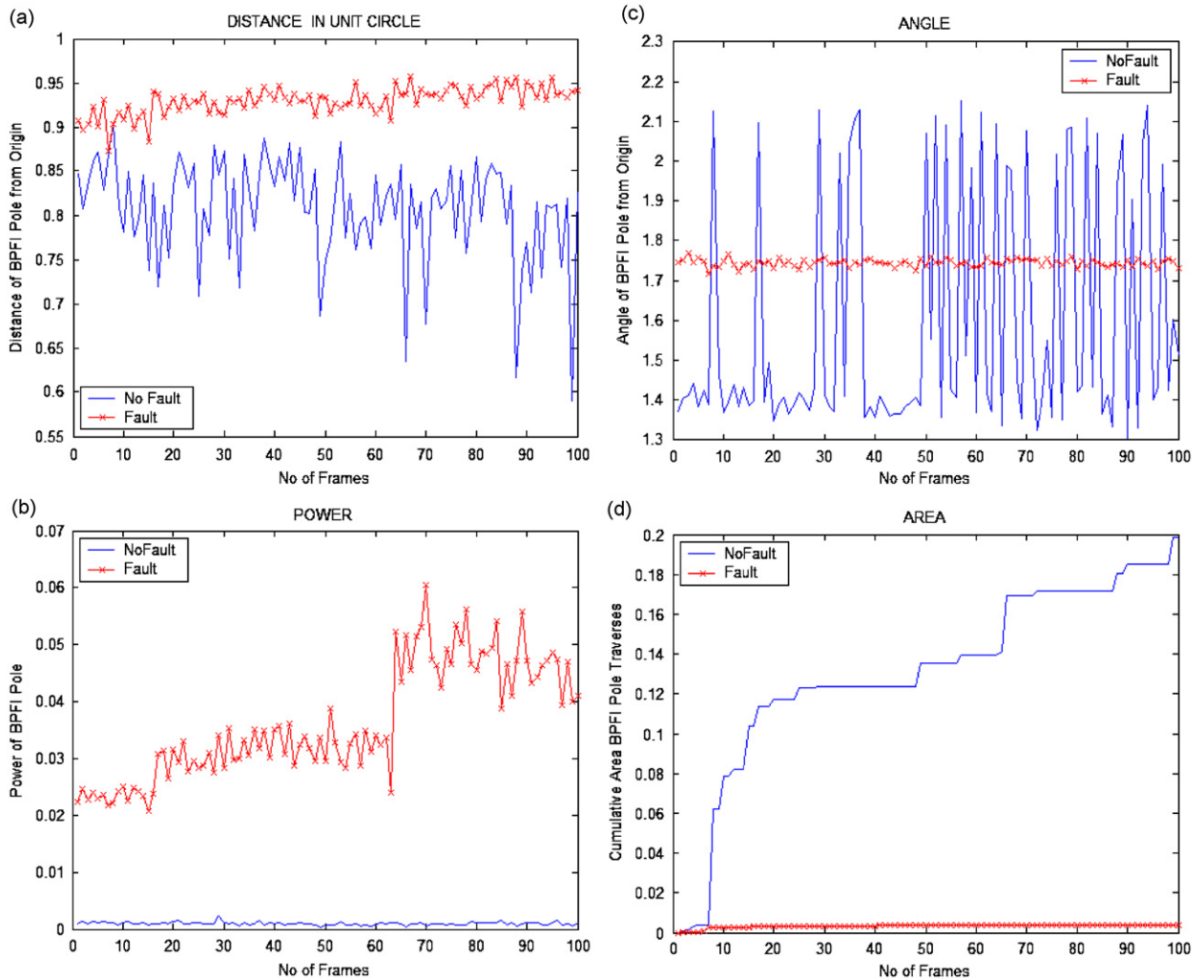


Fig. 9. Monitoring parameters λ . Pump running at 105 Hz. Hundred frames of ADXL105 vibration data used each of 2 s length. A 10th order AR model was used: (a) Distance of BPF1 pole from origin. The distance of the BPF1 pole for faulty conditions marked much higher than in normal conditions. (b) Power of BPF1 poles. The power of faulty poles was much larger than the non-faulty poles. (c) Angle BPF1 pole traverses. For normal condition, there was much variation in the BPF1 angles but for faulty condition the variance was much less. (d) Cumulative area traversed by the BPF1 poles. The area mapped out for normal condition was much larger than that for faulty condition.

are related and are dependent on each other. Though the distance and power of the BPF1 critical poles can be used as effective indicators for fault classification, the power is a better classifier than the distance since it incorporates the variance term of the signal (which is equal to the energy of the signal and energy of the faulty signal is much more than the non-faulty one). In the example shown, the power seemed to increase with time as the number of frames monitored increased. This is not surprising as the amplitude of vibrations also increased with time due to the pump heating up and increased friction of the bearings when it was running for a longer time.

The mean angle of the BPF1 pole was 1.6117 radians (513 Hz) and 1.7441 radians (555 Hz) for normal and faulty condition, respectively. The average mean angles of the BPF1 poles in both cases were nearly the same. However the mean standard deviation of the angles of the BPF1 poles for the normal condition ($\Delta\phi_k = 0.3094$) was much larger than for the faulty case ($\Delta\phi_k = 0.097$). The mean standard deviation of the BPF1 angles can be used as an indicator for fault classification to distinguish between the no-fault and defective cases as the difference between them is significant. The changes in the frequency content of the

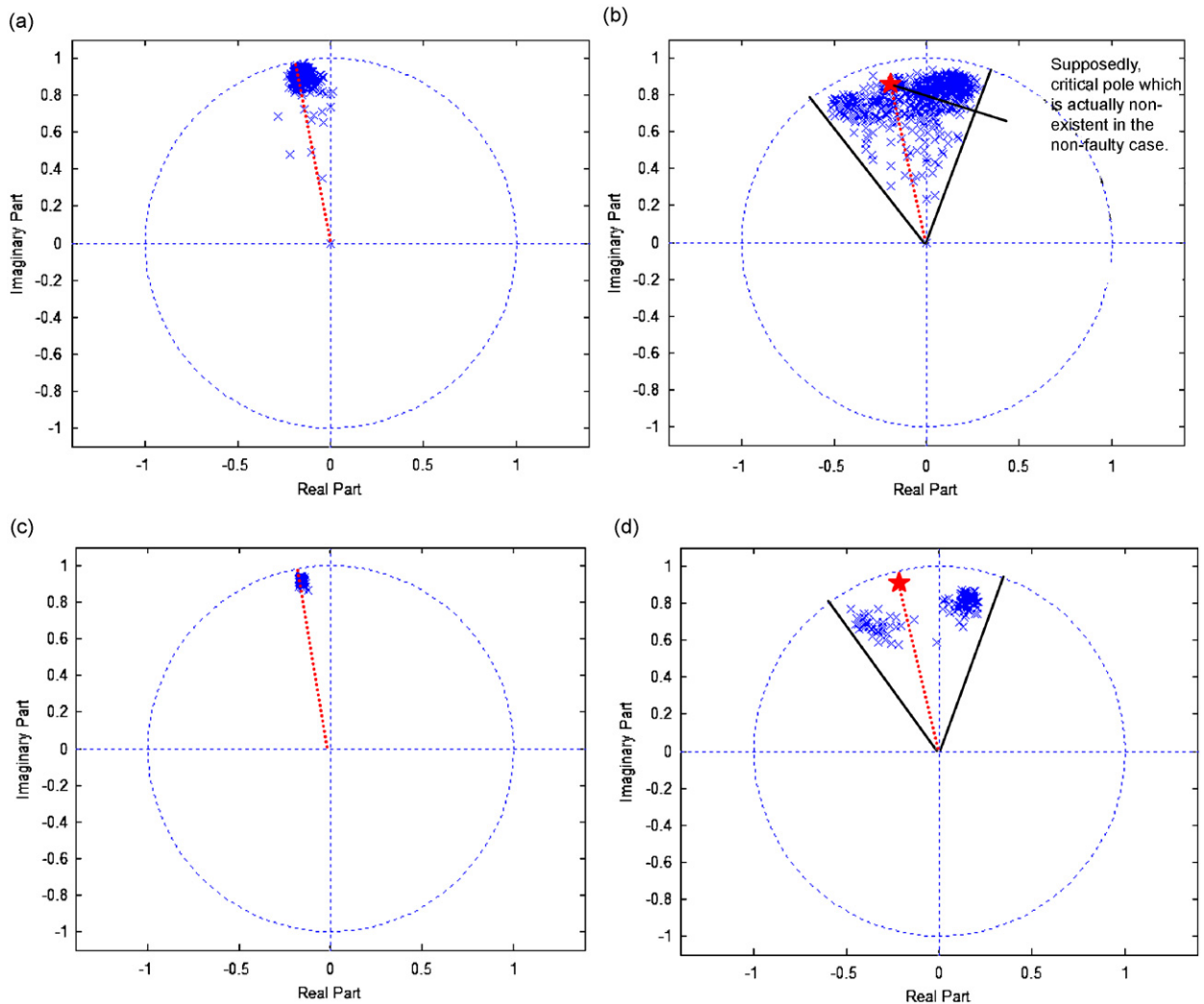


Fig. 10. Distribution of poles from a pump running in normal conditions (right) and for a case with a bearing that has an inner race fault (left). Dotted red line denotes BPF frequency of 555 Hz. (a) Frame size of 2000 samples—faulty poles. (b) Frame size of 2000 samples—non-faulty poles. (c) Frame size of 5000 samples—faulty poles. (d) Frame size of 5000 samples—non-faulty poles. Note concentrated clustering of poles near unit circle for BPF frequency at approximately 555 Hz for faulty data. Spread of poles is much larger for normal conditions.

characteristic defect frequencies can be quantified by using AR modelling by looking at the corresponding changes in the angles of the BPF poles. The cumulative area traversed by the migratory faulty BPF poles (less than 0.01) was also markedly much smaller than the case for the non-faulty condition (0.2). This is because when there is a BPF fault the pole corresponding to the BPF frequency remains at almost the same position, since there is a vibration at BPF. When there is no BPF fault the so-called ‘critical pole’ is simply the pole closest to the BPF angular frequency, but since there is no strong vibration at that frequency the non-faulty BPF poles move considerably from frame to frame.

In order to model the changes in the vibratory signatures, the critical poles are obtained for 2 different frame sizes for both normal and faulty conditions. Fig. 10 (a and b) shows the BPF poles plotted for a frame size of 2000 samples. Fig. 10 (c and d) shows the BPF poles plotted for a frame size of 5000 samples. The BPF pole was chosen as the dominant pole as the fault condition studied was the case of a pump fitted with a bearing with an inner race fault. If other bearing defects are to be monitored, the corresponding critical pole

positions relating to the characteristic defect frequencies of significance can be tracked. The AR identification method placed the faulty BPF critical poles centred around 555 Hz. There was a significant difference in the spread of the poles for the cases of damaged and undamaged bearings, respectively. For the faulty condition, the spread of the poles was more concentrated and nearer to the unit circle for both frame sizes. For the non-faulty BPF poles, the spread of the poles was much wider. When one compares Fig. 10 (b and d), they were plotted using the same data but it can be seen the frame size used can have an effect on the location of the coordinates of the AR poles. The difference obtained is probably due to the fact that when a bigger frame size was used, the mapping of the poles seems to be more deterministic and definite. For Fig. 10 (d), the location of the BPF poles seems to be either on the right or left of the BPF characteristic defect frequency. In reality, the demodulated vibration data from a good bearing looks more like white noise and there is no actual significant peak and hence no corresponding BPF pole in that sense. Because of our definition of the critical pole been the closest pole to the characteristic defect frequency and hence the BPF pole in this case, the method has located the poles that happen to be closest in angular frequency to the BPF frequency. The frequency variation of the characteristic bearing defect frequency (BPF) was much less for the faulty data than the normal data. The complex conjugate poles fell closer to the unit circle for vibration data obtained from faulty conditions. For normal data, the amplitudes of the dominant poles (moduli) were considerably smaller than their counterparts for the faulty case.

6.1. Sensitivity, specificity and ROC curve analysis

ROC curve analysis was used to determine suitable threshold levels for the fault indicators such as the AR poles distances from the centre of unit circle and the AR poles power for faulty and normal cases. The distribution of the normal and faulty AR poles power and distance is shown in Figs. 11 and 12 as histograms. A frame size of 2000 samples was used for these. In terms of power, a near perfect classification can be seen. When the power (using Eq. (12)) of the AR poles of the vibration data of the good bearings were tracked, it can be seen that they all had power values of less than 0.005. But when the power of the AR poles of the vibration data of the bearing with an inner race defect was tracked, the faulty poles' power had a wide distribution. They had values ranging from 0.015 to 0.065. Although there was no overlapping of the values. Hence the power indicator of the AR poles can be used as a 'perfect' fault classifier. For the distance of the AR poles from the origin of the unit circle, the normal poles had a distribution ranging from 0.4 to 0.95. The faulty poles had a distribution ranging from 0.81 to 0.97. There was some overlapping between the two classes. For instance, if 0.9 was chosen as the threshold level to determine if the poles corresponded to the case of damaged or non-damaged bearings, there is a chance that some normal poles with a distance of more than 0.9 are misclassified as faulty poles and vice versa. However it was noted that the frame size used did have an effect on the amount of overlapping between the two classes of data and hence the sensitivity and specificity of the analysis. Fig. 13 shows the distribution of normal and faulty poles for the distance indicator but for a bigger

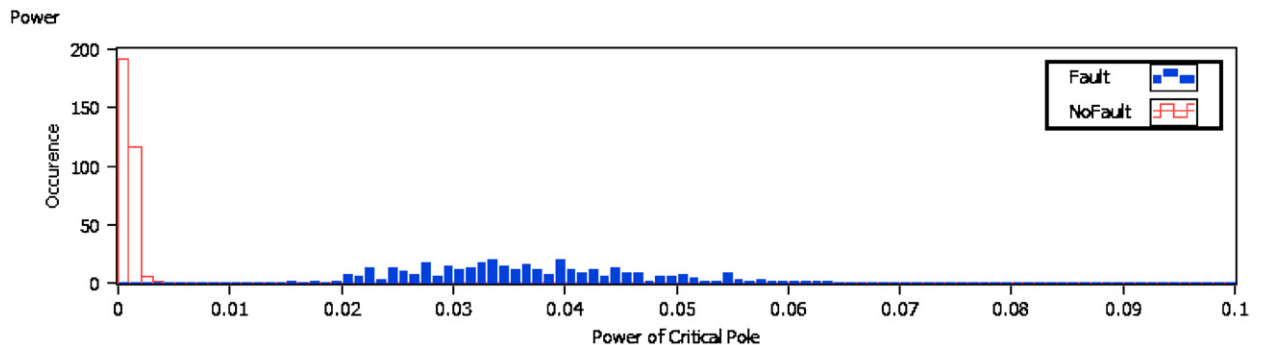


Fig. 11. Distribution of normal and faulty poles power parameter. Pole distribution is separated into two distinct regions. No overlapping of normal and faulty poles. Note these were plotted for frame sizes of 2000 samples. No ROC analysis necessary.

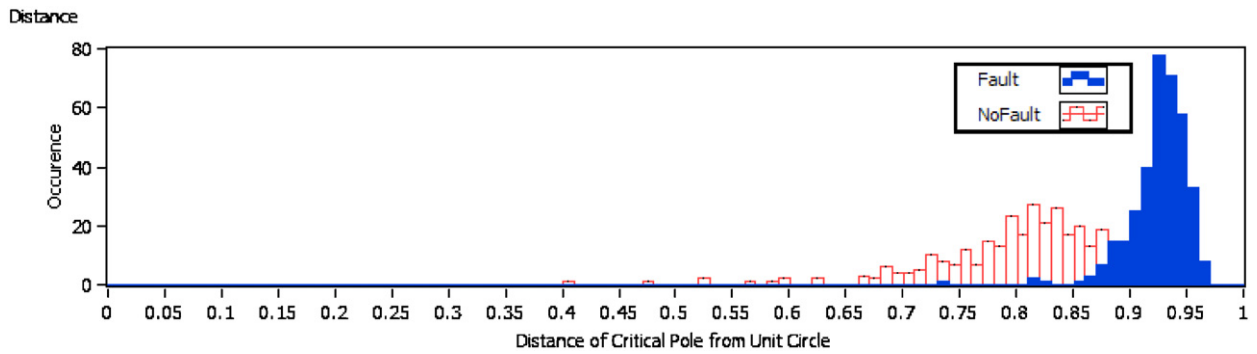


Fig. 12. Distribution of normal and faulty poles using distance from the centre of unit circle as the decision parameter. Some overlapping of normal and faulty poles. Note these were plotted for frame sizes of 2000 samples. If bigger sample sizes are used, less overlapping will occur.

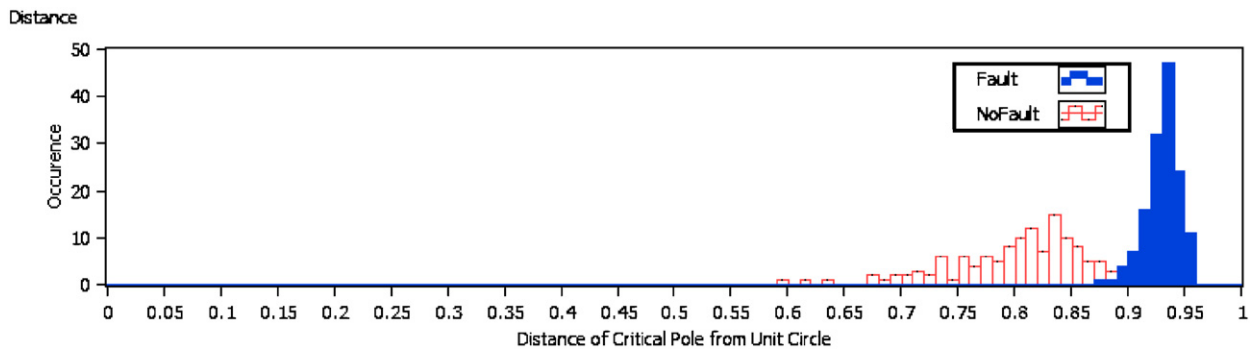


Fig. 13. Distribution of normal and faulty poles using distance from the centre of unit circle as the decision parameter. Less overlapping of normal and faulty poles. Note these were plotted for frame sizes of 5000 samples.

frame size of 5000 samples. It can be seen that there is less overlapping between the classes compared to Fig. 12 and hence less chance of misclassification.

6.2. Frame size versus increase in accuracy of classification

The diagnostic effectiveness of the decision parameter (the distance of the critical poles from the origin in the unit circle) was estimated by constructing a curve of sensitivity against specificity for windows of increasing frame sizes of 1100–5000 samples. The poles with a distance from the origin smaller than the threshold were termed the ‘true negative’ (non-faulty data). The poles with a distance from the origin greater than the threshold were termed the ‘true positive’ (faulty data). The task then becomes one of discriminating between the faulty and non-faulty data. The upper ROC curve is for a frame size of 5000 samples and the lower curve is for 1100 samples per frame. The corresponding ROC curve for the AR BPF I poles distance indicator is shown in Fig. 14. Each point on the curve represents the combination of true negatives against false positives estimated for a given threshold of the parameter. The aim is to maximise detection probability while minimising false alarm rates. A good test curve is one for which sensitivity rises rapidly and 1-specificity hardly increases at all until the sensitivity is high. Using a frame size of 2000 samples gives a sensitivity of 92% and a 1-specificity of 8%. If it is desired to increase both sensitivity and specificity, a bigger frame size can be used. Frames with 4000 samples (corresponding to 2 s of data collection) produced excellent results with sensitivity and specificity of 99%. This means that the undamaged and damaged bearing with the inner race defect can be distinguished with almost 100% accuracy. It is obvious that as the frame size increases the improvement in the classification also increases. The sensitivity and specificity versus distance curves for a frame size of 4000

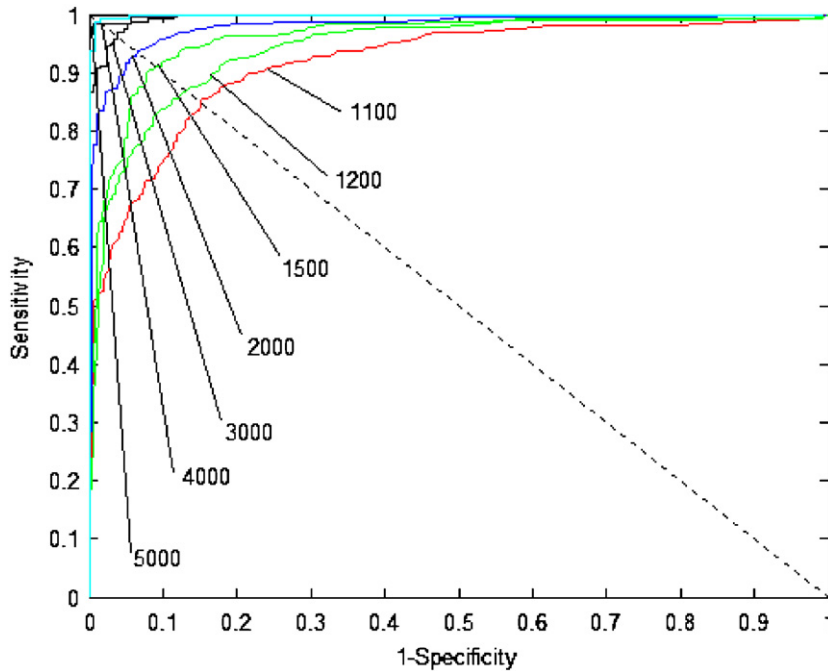


Fig. 14. ROC curves for increasing frame sizes for the distance of the poles from the centre of the unit circle as a decision parameter.

samples are shown in Fig. 15. For example, selecting a distance value of 0.885 as threshold, leads to a sensitivity of 99% and a specificity of 99%.

7. Conclusions

A formal framework for automated assessment of bearing faults using AR pole positions has been established. Metrics in the form of area, angle, power and distance of the AR pole positions (parameters λ) were defined for fault detection purposes. By quantifying the values of these parameters λ , bearing faults can be identified. The characteristic bearing frequencies were modelled and mapped using the AR method. The critical poles are defined as those poles, which are the closest in terms of angular displacements from the characteristic bearing frequencies poles' positions. It is proposed that the parameters λ are then compared with the baseline figures obtained from healthy bearings to detect the presence of faults. A higher power and a bigger distance from the origin of the AR poles will be indicative behaviours of faulty bearings.

The AR pole representation allows a more straightforward quantitative estimation of the spectral parameters and facilitates the understanding of the intrinsic spectral characteristics of the process. There is a one-to-one correspondence between the position of the AR poles and the spectral peaks. By studying the movement of poles in the complex z plane, the development and progression of bearing faults can be monitored. It was shown, with experimental results that the AR poles move closer to the unit circle as the severity of the defects increase. The application of the proposed scheme enabled the detection of the damaged bearings with an inner race defect with a 100% detection rate when a frame size of more than 4000 samples was used when the distance of the critical poles was used as the fault indicator. With a smaller sample size of 2000 samples, a lower detection rate of 92% was achieved. When power was used as the fault indicator, a 100% detection rate was obtained even for small sample sizes such as a sample size of 2000 samples, indicating that this processing methodology is very effective even for short data records. The simplicity of the use of the classification parameters λ , the short length of data needed for the diagnostics (2 s) and the obtained specificity and sensitivity of the classification scheme reveal the relevance and usefulness of model-based diagnostics system in the classification of normal and faulty bearings. Results show that tracking the movement of the AR critical poles as a maintenance decision-making criterion, this method can be used effectively to detect changes

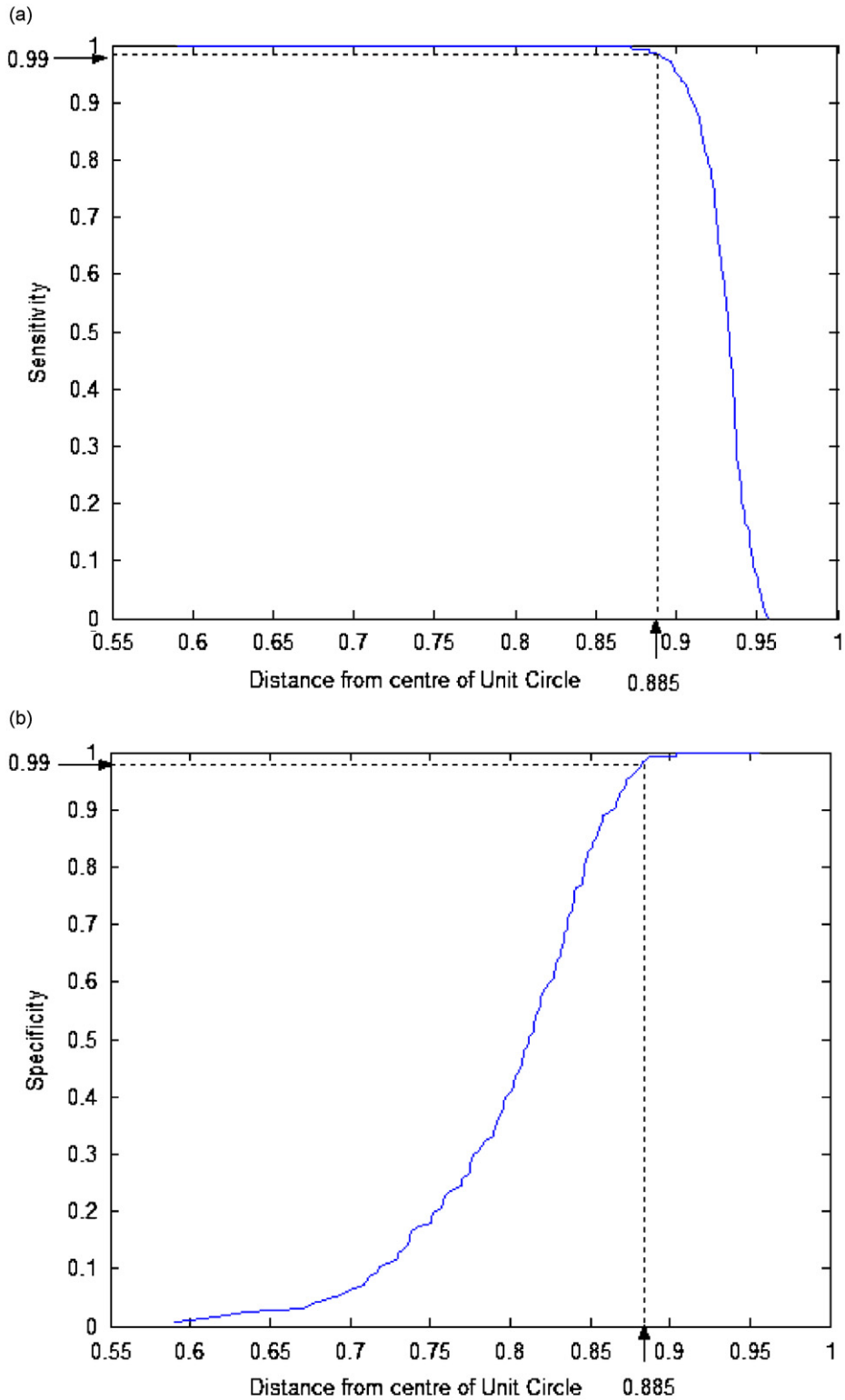


Fig. 15. Sensitivity (a) and specificity (b) curves for a frame size of 4000 samples. Note that distance from centre of unit circle was used as a decision parameter. Even when using the distance as the diagnostic parameter, one can obtain 99% sensitivity and 99% specificity for frames of 2 s duration.

in the frequency and power of the faulty bearings and consequently differentiate pumps running in normal and defective working environments.

It is speculated that the proposed technique would also be usable for other types of machines or machine elements as long as a model for the system can be obtained and the frequencies of likely localised faults can also be calculated. It is also speculated that this technique would be useful even if they have no a priori knowledge of the behaviour of non-faulty pumps as when there is a localised fault, this can be seen as the concentrated clustering of poles around a localised frequency in the unit circle and this behaviour is characteristic of the existence of a propagating fault as can be seen in Fig. 10 (c). Hence with the use of the proposed detection tool, the current state of a structure or machine element can also be detected without any information about its undamaged or initial state.

Acknowledgements

The authors would like to thank EPSRC (Grant GR/S42866/0), which funded Suguna Thanagasundram's Ph.D. studies, BOC Edwards for the supply of the IGX dry vacuum pumps, Barden Bearings for providing us the faulty ceramic ball bearings on which our experimental investigations were based on and Mr. Kalu Gurung for conducting some of the preliminary investigations on the pole tracking method as part of his final year project in 2005.

References

- [1] S.A. McNerny, Y. Dai, Basic vibration signal processing for bearing fault detection, *IEEE Transactions on Education* 46 (1) (2003) 149–156.
- [2] W. Wang, A.K. Wong, Some new signal processing approaches for gear fault diagnosis, *Proceedings of the Fifth International Symposium on Signal Processing and its Applications*, vol. 2, 1999, pp. 587–590.
- [3] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, *Tribology International* 32 (8) (1999) 469–480.
- [4] J.R. Stack, T.G. Habetler, R.G. Harley, Fault classification and fault signature production for rolling element bearings in electric machines, *Proceedings of the IEEE International Symposium on Diagnostics for Electric Machines*, 2003, pp. 172–176.
- [5] K.Z. Tang, K.K. Tan, C.W. de Silva, T.H. Lee, K.C. Tan, S.Y. Soh, Application of vibration sensing in monitoring and control of machine health, *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, vol. 1, Como, Italy 2001, pp. 377–382.
- [6] R.J. Alfredson, J. Mathew, Time domain methods for monitoring the condition of rolling element bearings, *Mechanical Engineering Transactions—Institution of Engineers, Australia* 10 (2) (1985) 102–107.
- [7] R.J. Alfredson, J. Mathew, Frequency domain methods for monitoring the condition of rolling element bearings, *Mechanical Engineering Transactions—Institution of Engineers, Australia* 10 (2) (1985) 108–112.
- [8] P.D. McFadden, J.D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, *Journal of Sound and Vibration* 96 (1) (1984) 69–82.
- [9] P.D. McFadden, J.D. Smith, The vibration produced by multiple point defects in a rolling element bearing, *Journal of Sound and Vibration* 98 (2) (1985) 263–273.
- [10] S.M. Kay, *Modern Spectral Estimation—Theory and Application. Signal Processing Series*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [11] S. Lawrence Marple Jr., *Digital Spectral Analysis with Applications*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.
- [12] S.M. Kay, S.L. Marple, Spectrum analysis—a modern perspective, *Proceedings of the IEEE* 69 (11) (1981) 1380–1419.
- [13] L.T. Mainardi, A.M. Bianchi, G. Baselli, S. Cerutti, Pole-tracking algorithms for the extraction of time-variant heart rate variability spectral parameters, *IEEE Transactions on Biomedical Engineering* 42 (3) (1995) 250–259.
- [14] M.J.L. de Hoon, T.H.J.J. van der Hagen, H. Schoonewelle, H. Van-Dam, Why Yule–Walker should not be used for autoregressive modelling, *Ann. Nucl. Energy* 23 (1996) 1219–1228.
- [15] P.M.T. Broersen, *Automatic Autocorrelation and Spectral Analysis*, Springer, Berlin, 2006.
- [16] S.J. Johnsen, N. Andersen, On power estimation in maximum entropy spectral analysis, *Geophysics* 43 (4) (1978) 681–690.
- [17] P.A. Lessard, Dry vacuum pumps for semiconductor processes: guidelines for primary pump selection, *Proceedings of the the 46th International Symposium of the American Vacuum Society*, vol. 18, 2000, pp. 1777–1781.
- [18] S. Thanagasundram, F.S. Schlindwein, Comparison of integrated micro-electrical–mechanical system and piezoelectric accelerometers for machine condition monitoring, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 220 (8) (2006) 1135–1146.
- [19] D.R. Harting, Demodulated resonance analysis—a powerful incipient failure detection technique, *ISA Transactions* 17 (1) (1978) 35–40.

- [20] R.B. Randall, J. Antoni, S. Chobsaard, A comparison of cyclostationary and envelope analysis in the diagnostics of rolling element bearings, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, 2000, pp. 3882–3885.
- [21] N. Tandon, A. Choudhury, An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect, *Journal of Sound and Vibration* 205 (3) (1997) 275–292.
- [22] International Standard of Organisation ISO 10816, Mechanical vibration—evaluation of machine vibration by measurements on non-rotating parts, *ISO Standard*, 1995.
- [23] International Standard of Organisation ISO 7919, Mechanical vibration of non-reciprocating machines, *ISO Standard*, 1996.
- [24] J.R. Stack, T.G. Habetler, R.G. Harley, Effects of machine speed on the development and detection of rolling element bearing faults, *Power Electronics Letters, IEEE* 1 (1) (2003) 19–21.
- [25] S. Thanagasundram, Y. Feng, I.S.M.A. Rayan, F.S. Schlindwein, A case study of Auto-Regressive Modelling and Order Selection for a Dry Vacuum Pump, *Proceedings of the ICSV12-Twelfth International Congress on Sound and Vibration*, Lisbon, Portugal, 2005.