A NEW BEARING FAULT DETECTION AND DIAGNOSIS SCHEME BASED ON HIDDEN MARKOV MODELING OF VIBRATION SIGNALS

Hasan OCAK Kenneth A. LOPARO, *Fellow, IEEE*

hxo9@po.cwru.edu kal4@po.cwru.edu

Department of Electrical Engineering and Computer Science, CWRU, Cleveland OHIO

ABSTRACT

This paper introduces a new bearing fault detection and diagnosis scheme based on hidden Markov modeling (HMM) of vibration signals. First features are extracted from amplitude demodulated vibration signals obtained from both normal and faulty bearings. The features are based on the reflection coefficients of the polynomial transfer function of the auto-regressive model of the vibration signal. These features are then used to train HMMs to represent various bearing conditions. The technique allows for online detection of faults by monitoring the probabilities of the pre-trained HMM for the normal case. It also allows for the diagnosis of the fault by the HMM that gives the highest probability. The new scheme was tested with experimental data collected from drive end ball bearing of an induction motor (Reliance Electric 2HP IQPreAlert) driven mechanical system.

1. INTRODUCTION

Induction motors are widely used in industry due to their reliability and relatively low cost. However, diagnosis and isolation of both electrical and mechanical faults of an induction motor is a challenging problem. According to the Motor Reliability Working Group (MRWG) and the investigation carried out by Electric Power Research Institute (EPRI), the most common failure mode of an induction motor is bearing failure followed by stator winding failures and rotor bar failures. A bearing failure increases the rotational friction of the rotor. Thus, detection and diagnosis of mechanical faults in rolling element bearings is very crucial for the reliable operation of an induction motor.

Considerable research has been carried out previously for the development of various algorithms for bearing fault detection and diagnosis. These algorithms can be classified into time domain, frequency domain, time-frequency domain, higher order spectral analysis, neural-network and model based techniques. Various time domain statistical parameters have been used as trend parameters to detect the presence of incipient bearing damage. Kurtosis and skew values of vibration signals are used in [1] for detection of bearing faults at early stages in their development. [2] presents a study on the application of sound pressure and vibration signals to detect the presence of defects in a rolling element bearing using a statistical analysis method. The most important shortcoming of the statistical analysis approach is its inability to detect bearing defects at later stages. In the frequency domain approach the major frequency components of vibration signals and their amplitudes are used for trending purposes. The frequency characteristics of the vibration for a defective bearing subject to various load conditions are investigated in [3]. Envelope analysis, originally known as the high frequency resonance technique, is the most commonly used frequency analysis technique for the detection and diagnosis of bearing faults. The technique is studied in detail in [4]. One of the problems with envelope analysis and the other frequency domain approaches is that, they require the bearing defect frequencies be known or pre-estimated. The other shortcoming is the increasing difficulty in analyzing the vibration spectrum when the signal to noise ratio is low and the vibration spectrum has a large number of frequency components due to the complexity of the system. Time-frequency domain techniques use both time and frequency domain information allowing for the investigation of transient features such as impacts. A number of time-frequency domain techniques have been proposed including Short Time Fourier Transform (STFT), the Wigner-Ville Distribution (WVD), and the Wavelet Transform (WT) [5,6]. Bi-coherence spectra are used in [7] to derive features that relate to the condition of a bearing. The application of bi-spectral and tri-spectral analysis in condition monitoring is discussed in [8]. Neural networks are also applied to bearing fault detection and diagnosis [9,10]. Model based techniques are studied in [3] and [11].

Hidden Markov modeling (HMM) is known as the state of the art technique for speech recognition [12]. HMMs are also successfully applied to machine tool wear monitoring [13,14]. In this paper we present the theory of hidden Markov models and apply it to bearing fault detection and diagnosis problem.

2. THEORETICAL BACKGROUND

2.1. Linear Auto Regressive Modeling

A linear auto-regressive model can be used to predict the value of the next sample of a signal as a linear combination of the previous samples. The next sample of the signal, $\bar{s}_n$, is predicted as the weighted sum of the p previous samples, $s_{n-1}, s_{n-2}, ..., s_{n-p}$ and can be expressed as,

$$\bar{s}_n = a_1 s_{n-1} + a_2 s_{n-2} + ... + a_p s_{n-p} = \sum_{i=1}^{p} a_i s_{n-i} \quad (2.1)$$

The transfer function of the model is given by,

$$H(z) = \frac{\bar{S}(z)}{S(z)} = a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p} = \sum_{i=1}^{p} a_i z^{-i} \quad (2.2)$$

The residual error, e_n, is defined as the difference between the actual and the predicted values of the next sample and can be expressed as,

$$e_n = s_n - \bar{s}_n = s_n - \sum_{i=1}^{p} a_i s_{n-i} \qquad (2.3)$$

The weights can be calculated by minimizing the mean square value of the residual errors over an analysis window.

2.2. Hidden Markov Models

Hidden Markov models (HMM) are extensions of Markov models to include the case where the observations are probabilistic functions of the states rather than the states themselves. A HMM is characterized by several parameters. The first parameter is the transition probability distribution $\boldsymbol{A} = \{a_{ij}\}$, where a_{ij} is the probability of being in state S_j at time $t+1$ provided that the state at time t is S_i, i.e.,

$$a_{ij} = P\{q_{t+1} = S_j \mid q_t = S_i\}, 1 \le i, j \le N \qquad (2.4)$$

where q_t denotes the state at time t and N is the number of states.
The second parameter of a HMM is the observation probability distribution, $\boldsymbol{B} = \{b_j(k)\}$,

$$b_j(k) = P\{o_k \mid q_t = S_j\}, 1 \le j \le N, 1 \le k \le M \qquad (2.5)$$

where o_k is the k^{th} observation and M is the number of distinct observations. If the observations are continuous, a continuous probability density function, generally a weighted sum of several Gaussian distributions, is assigned to each state.
The last parameter is the initial state distribution, $\boldsymbol{\pi} = \{\pi_i\}$,

$$\pi_i = P\{q_1 = S_i\}, 1 \le i \le N \qquad (2.6)$$

which is the probability of S_i being the initial state.
A compact notation $\lambda = [\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}]$ is used to define a HMM. The probability of a given observation sequence, $O = o_1, o_2, ..., o_T$, can be calculated as,

$$P(O \mid \lambda) = \sum_{all\,S} \pi_{S_0} \prod_{t=0}^{T-1} a_{S_t S_{t+1}} b_{S_{t+1}}(o_{S_{t+1}}) \qquad (2.7)$$

The maximum likelihood (ML) method can be used to re-estimate the model parameters, $\lambda = [\boldsymbol{A}, \boldsymbol{B}, \pi]$, as follows:

$$\bar{\pi}_i = expected\ number\ of\ times\ in\ S_i,\ at\ time\ t = 1 \qquad (2.8a)$$

$$\bar{a}_{ij} = \frac{expected\ number\ of\ transitions\ from\ S_i\ to\ S_j}{expected\ number\ of\ transitions\ from\ S_i} \qquad (2.8b)$$

$$\bar{b}_j(k) = \frac{expected\ number\ of\ times\ in\ S_j\ \&\ observation\ k}{expected\ number\ of\ times\ in\ S_j} \qquad (2.8c)$$

Training of a HMM for a given observation sequence can be realized by the so-called Baum-Welch Algorithm. Starting with initial or pre-estimated HMM parameters, the algorithm updates the parameters, by calculating the ML estimates, step by step increasing the probability of the observation sequence in each step. The training procedure along with the other features of hidden Markov models is explained in detail in [12].

3. PREPROCESSING AND FEATURE EXTRACTION

Vibration signals were amplitude demodulated before the feature extraction process. Amplitude demodulation provides a mechanism for effectively extracting out the rolling element fault frequencies from extraneous noise present in the signal.

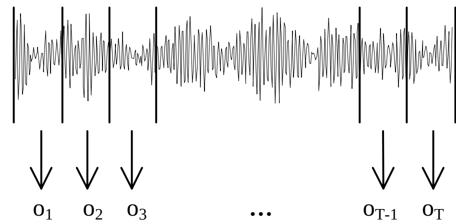

Figure 3.1 - Feature Extraction from Vibration Signals

Amplitude demodulation was carried out by band-pass filtering followed by half-wave rectification. The envelope detection stage was performed by low-pass filtering the half-wave rectified signal. The center frequency and the bandwidth for the band-pass filter were 3 and 2kHz, respectively. The cut-off frequency for the low pass filter was 2kHz.
Figure 3.1 illustrates the feature extraction process. After preprocessing, vibration signals were divided into windows of equal length. A set of features (referred to as a feature vector or a single observation) was extracted from each window. The features for a single window were selected to be the reflection coefficients of the polynomial transfer function of the linear auto-regressive model for that window. The observation sequences were later used in the HMM training process. The effect of the choice of the window length and the order of the model on the performance of the new scheme will be discussed in the Results section.

4. HMM BASED FAULT DETECTION AND DIAGNOSIS

4.1. Fault Detection

For the detection of the presence of a fault, it is sufficient to train a single HMM to represent the normal condition. A HMM is trained with features obtained from the vibration data collected from a normal bearing using the Baum-Welch Algorithm. To determine whether a given vibration signal is from a normal or a faulty bearing, first the signal is preprocessed. Then the feature vectors are extracted. Given the set of feature vectors, the probability of the previously trained HMM for the normal bearing is calculated. If the probability is above a pre-determined threshold, then the vibration signal is from a normal bearing. The vibration is from a faulty bearing, otherwise.

4.2. Fault Diagnosis

For the diagnosis of the fault, it is necessary that a separate HMM be trained for all the possible fault types in addition to the HMM for normal condition. To diagnose the condition of a bearing, vibration signals gathered from the bearing are preprocessed followed by feature extraction. Then, the probability of the set of feature vectors is calculated given all the HMMs in the previously constructed database. The HMM for which the probability is maximum, determines the condition of the bearing. The recognition stage of a HMM-based fault diagnosis scheme is illustrated in Figure 4.1.

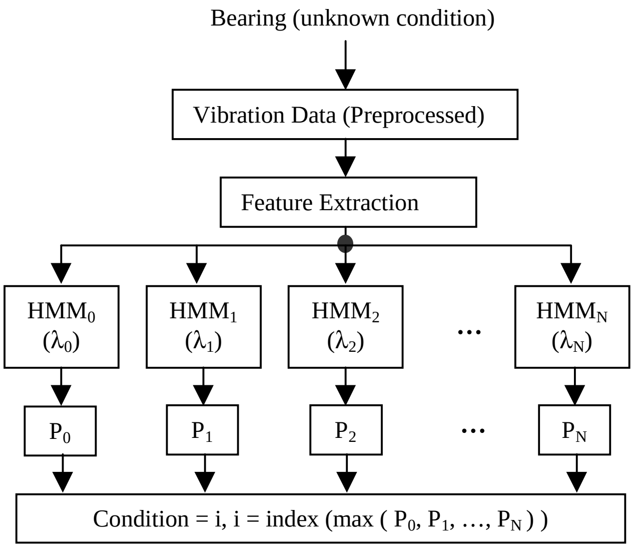

Figure 4.1 – HMM based Fault Diagnosis

5. EXPERIMENTAL RESULTS

5.1. Experimental Setup

Experimental data were collected from drive end ball bearing of an induction motor (Reliance Electric 2HP IQPreAlert) driven mechanical system. The accelerometer was mounted on the motor housing at the drive end. Data was gathered for four different conditions: (i) normal (N); (ii) inner race fault (IRF); (iii) outer race fault (ORF); (iv) ball fault (BF). Faults were introduced into the drive end bearing by EDM method. For inner race and ball fault cases, vibration data for three severity levels (0.007, 0.014 and 0.021 inches) was collected. For outer race fault case, vibration data for two different severity levels (0.007 and 0.021 inches) was collected. All the experiments were repeated for four different load conditions (O, 1, 2 and 3HP). Therefore, experimental data consisted of 4 vibration signals for normal condition and 12 vibration signals for the inner race and ball fault conditions. For the outer race faulty case there were 8 vibration signals. The sampling period was 12 kHz and the duration of each vibration signal was 10 seconds.

5.2. Results

All the vibration signals were divided into ten equal sections (one second duration each). The first sections were used for training, and the remaining nine sections along with the first sections were used to test the performance of the new scheme. Four states left-to-right continuous density HMMs were used to model the bearing conditions. Two Gaussian distributions were used in each state.

Table 5.1 shows the diagnosis accuracies for various model orders for a fixed window size and for various window sizes for a fixed model order. The best diagnosis accuracies were achieved when the window size and model order were 0.25 seconds and 25, respectively. Increasing the model order increased the diagnosis accuracies. However, changing the window size as long as it was above 0.05 seconds did not significantly affect the performance.

		N	**IRF**	**ORF**	**BF**
W_n=0.1 sec	n = 5	100%	89%	76%	78%
	n = 10	100%	96%	96%	88%
	n = 15	100%	98%	94%	100%
	n = 20	100%	98%	98%	100%
	n = 25	100%	98%	100%	100%
	n = 30	100%	98%	100%	100%
n = 25	W_n = 0.05 sec	100%	100%	99%	100%
	W_n = 0.10 sec	100%	98%	100%	100%
	W_n = 0.20 sec	100%	98%	100%	100%
	W_n = 0.25 sec	100%	100%	100%	100%

Table 5.1 – Diagnosis Accuracies

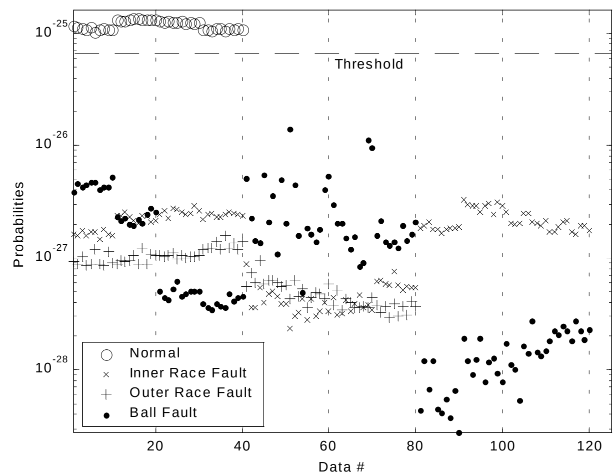

Figure 5.1 – Probabilities of Normal and Faulty Data given the HMM for Normal Condition

Faults were detected with 100% accuracy. Figure 5.1 shows the probabilities of the vibration signals for the normal, inner race, outer race and ball fault conditions given the HMM for the normal condition. The probabilities of the normal data are clearly separable from the probabilities of the faulty data by thresholding. Window size and the model order were 0.25 seconds and 25, respectively.

Figures 5.2, 5.3 and 5.4 show the probabilities of the inner race, outer race and ball fault data, respectively. P1, P2, P3 and P4 refer to the probabilities given the models for normal, inner race, outer race and ball fault conditions, respectively. As seen from the figures, the probability of a given data set is largest given the HMM that represents the condition of the data.

The results verify that the new scheme is able to detect and diagnose faults with 100 % accuracy independent of the load condition, the type and severity of the fault.

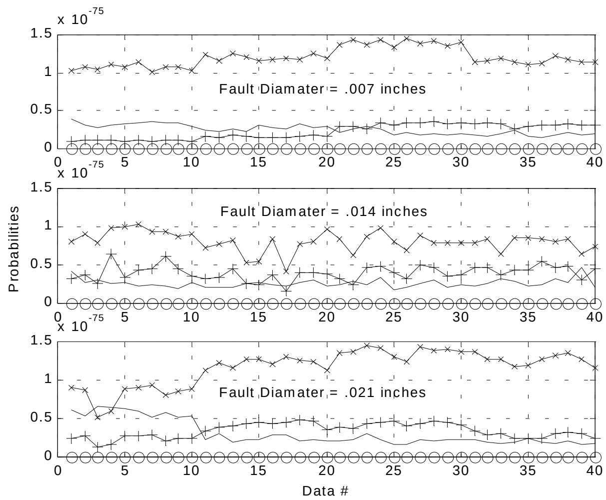

Figure 5.2 – HMM Probabilities for the Inner Race Faulty Data

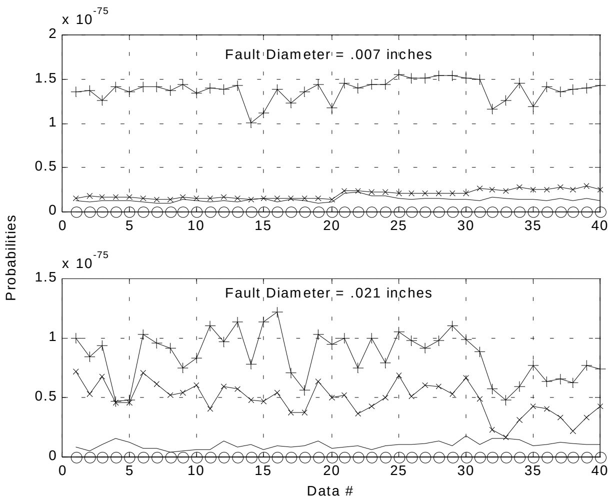

Figure 5.3 - HMM Probabilities for the Outer Race Faulty Data

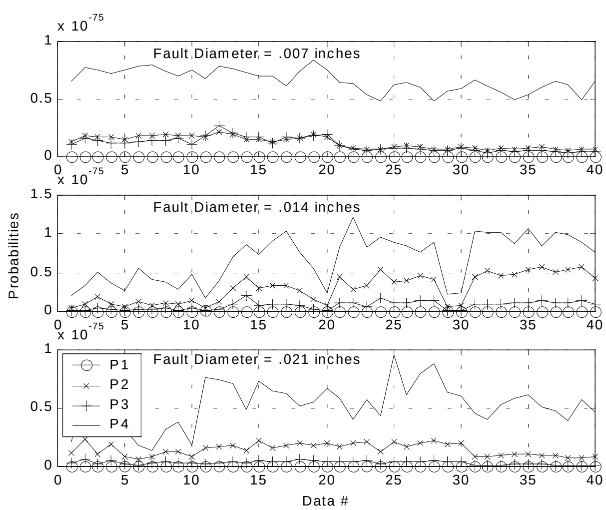

Figure 5.4 – HMM Probabilities for the Ball Faulty Data

6. REFERENCES

[1] H. R. Martin, F. Honarvar, "Application of Statistical Moments to Bearing Failure Detection", Applied Acoustics, Vol. 44, pp. 67-77, 1995

[2] R. B. W. Heng, M. J. M. Nor, "Statistical Analysis of Sound and Vibration Signals for Monitoring Rolling Element Bearing Condition", Applied Acoustics, Vol. 53, No. 1-3, pp. 211-226, 1998

[3] Y. T. Su, S. J. Lin, "On Initial Fault Detection of a Tapered Roller Bearing: Frequency Domain 75
Analysis", Journal of Sound and Vibration, 155(1), pp. 75-84, 1992

[4] P. D. McFadden, J. D. Smith, "Vibration Monitoring of Rolling Element Bearings by High Freq. Resonance Technique – A Review", Tribology International, Vol. 77, pp. 3-10, 1984

[5] C. James Li, Jun Ma, "Wavelet Decomposition of Vibrations for Detection of Bearing-Localized Defects", NDT&E International, Vol. 30, No. 3, pp. 143-149, 1997

[6] Gary Y. Yen, Kuo-Chung Lin, "Wavelet Packet Feature Extraction for Vibration Monitoring", Proceedings of the IEEE International Conference on Control Applications, pp. 1573-1578, August 1999

[7] C. James Li, J. Ma, B. Hwang, "Bearing Localized Defect Detection by Bicoherence Analysis of Vibrations", Journal of Engineering for Industry, Vol. 117, pp. 625-629, November 1995

[8] A. C. McCormick, A. K. Nandi, "Bispectral and Trispectral Features for Machine Condition Diagnosis", IEEE Proceedings of Visual Image Signal Processing, Vol. 146, No. 5, pp. 229-234, October 1999

[9] M. Subrahmanyam, C. Sujatha, "Using Neural Networks for the Diagnosis of Localized Defects in Ball Bearings", Tribology International, Vol. 30, No. 10, pp. 739-752, 1997

[10] Bo Li, Gregory Godi, Mo-Yuen Chow, "Detection of Common Motor Bearing Faults Using Frequency Domain Vibration Signals and a Neural Network Based Approach", Proceedings of the American Control Conference, pp. 2032-2036, June 1998

[11] Nader Afshari, Kenneth A. Loparo, "A Model Based Technique for the Fault Detection of Rolling Element Bearings Using Filter Design and Sliding Mode Techniques", Proceedings of the 37th IEEE Conference on Decision & Control, pp. 2593-2598, December 1998

[12] Lawrance R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol. 77, No. 2, February 1989

[13] L. Atlas, M. Ostendorf, G.D. Bernard, "Hidden Markov Models for Monitoring Machining Tool-Wear", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, pp. 3887 –3890, 2000

[14] L.P. Heck, J.H. McClellan, "Mechanical System Monitoring using HMMs", IEEE International Conference on Acoustics, Speech, and Signal Processing Vol.3, pp. 1697-1700, 1991

Acknowledgments: This research was supported in part by the US Office of Naval Research under agreement number: N00014-98-3-0012 and the National Science Foundation under grant number ECS-9906218.