MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Probabilistic Inter-Disturbance Interval Estimation for Bearing Fault Diagnosis

Kevin Wilson

TR2009-060 October 2009

Abstract

We describe a new method for detecting characteristic bearing fault signatures from accelerometer vibration data based on a probabilistic model of the fault signal generation process. It is common to assume that single-point bearing defects cause periodic disturbances in bearing vibration signals, but this assumption may not be valid in practice. Our new method is less sensitive to departures from periodicity, such as fault disturbance amplitude and timing variations, than standard spectral or autocorrelation-based approaches. We demonstrate the utility of our method by distinguishing among inner race, outer race, and rolling element faults in a bearing fault test rig. Our method is significantly better than standard techniques at detecting rolling element (ball) faults.

IEEE Symposium on Diagnostic for Electrical Machines 2009

Probabilistic inter-disturbance interval estimation for bearing fault diagnosis

Kevin W. Wilson, *Member, IEEE*

***Abstract*—We describe a new method for detecting characteristic bearing fault signatures from accelerometer vibration data based on a probabilistic model of the fault signal generation process. It is common to assume that single-point bearing defects cause periodic disturbances in bearing vibration signals, but this assumption may not be valid in practice. Our new method is less sensitive to departures from periodicity, such as fault disturbance amplitude and timing variations, than standard spectral or autocorrelation-based approaches. We demonstrate the utility of our method by distinguishing among inner race, outer race, and rolling element faults in a bearing fault test rig. Our method is significantly better than standard techniques at detecting rolling element (ball) faults.**

***Index Terms*—bearing fault classification, condition monitoring, fault diagnosis, vibration**

I. Introduction

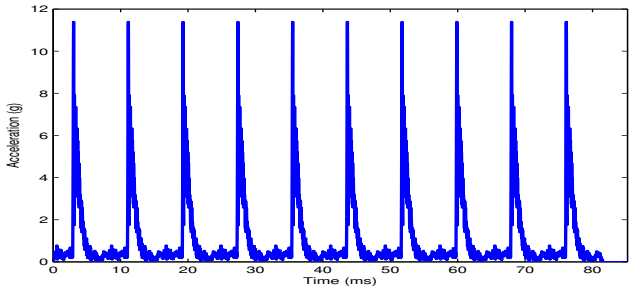
BEARING faults are a common failure mode for electrical machines, accounting for over 40% of the failures observed in induction motors [1]. For this reason, there is strong interest in being able to detect and diagnose such faults, often through the analysis of vibration signals from attached accelerometers [2], [3], [4].

Rolling element bearing faults are typically described by their location (inner race, outer race, cage, or rolling element) and by whether they are localized "single-point defects" or distributed "general roughness" faults [5]. It is commonly assumed that a single-point bearing defect will cause a periodic vibrational disturbance in the bearing and surrounding structures, and that the period at which this disturbance occurs will be determined by the location of the fault and the geometry of the bearing.

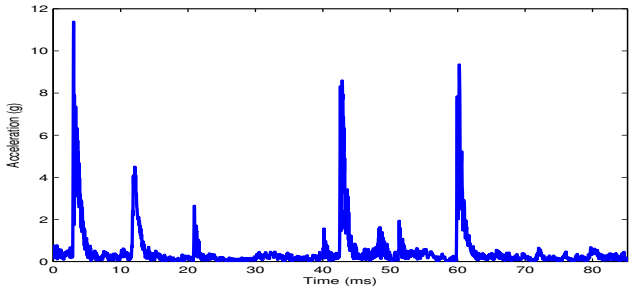
This assumption that the fault-related disturbance will be periodic is not always valid, however. It may be invalid if there is significant bearing slip or if a single-point defect located on a ball varies over time in how it contacts the inner and outer races. In situations such as these, both the inter-disturbance interval and the amplitudes of individual disturbances may vary. These variations can cause techniques that make strong assumptions about periodicity to fail.

This paper presents a new method, which is robust to significant departures from periodicity, for detecting single-point bearing faults using vibration signals. Section II defines the problem and reviews previous approaches. Section III describes our new method, including the simple probabilistic model on which it is based. Section IV presents experimental results from our bearing fault test rig, and Section V concludes.

K. Wilson is with the Mitsubishi Electric Research Lab in Cambridge, MA (email: wilson@merl.com).



(a) Idealized vibration envelope



(b) Actual vibration envelope

Fig. 1. (a) An idealized vibration signal envelope, in which a single point bearing fault generates a periodic impulse-like disturbance. (b) An actual vibration signal envelope recorded from a bearing with a rolling element (ball) fault.

II. Background

A rolling element bearing consists of an inner race, an outer race, rolling elements that roll between these two races, and optionally a cage around these rolling elements. Faults may occur in any of these parts, and often these faults are single-point defects such as chips or dents. As these elements move past each other, these defects come into periodic contact with other elements in the bearing, and at each contact they can excite a high frequency resonance in the overall structure [2]. This affects the accelerometer signal, although the effect on the raw acceleration signal is often overwhelmed by low frequency vibration from other parts of the motor.

To accentuate the effects of this high-frequency resonance, which in idealized form can be thought of as a high-frequency modulated pulse train, a standard approach is to high-pass filter the signal and then take its envelope [2]. After thus "demodulating" the signal, one can then analyze the fault-related contribution to the vibration signal.

Based on the geometry of the bearing elements (inner race diameter, outer race diameter, number of balls, etc.) and the shaft rotation rate, it is possible to calculate the characteristic frequencies with which defects on different bearing com-

ponents will generate these disturbances [6]. Because these characteristic frequencies are typically different for different components, analysis of the vibration signal allows us to distinguish among defects on these different components.

The simplest approach to detecting disturbances at these frequencies is to look at the energy in the enveloped signal at these frequencies or to auto-correlate the enveloped signal and look for peaks at lags corresponding to the characteristic frequencies. Other more sophisticated approaches are possible when simple high-pass filtering and enveloping is not sufficient.

Two broad classes of approaches to improving vibration-based bearing fault diagnosis are under active investigation in the research community. The first class is to improve the low-level signal processing techniques used to extract fault-related signatures. For example, [7] uses bispectral analysis to simultaneously demodulate the signal and detect energy at characteristic frequencies. Wavelet-based approaches are also being investigated, for example in [8] and [9]. These approaches have the potential to extract additional useful information from the vibration signal beyond what is extracted by a simple short-time Fourier transform, and they typically require no training data. They typically do depend on a model of the motor to define frequencies of interest for fault diagnosis, however. The other broad class of approaches is to apply machine learning techniques to feature vectors derived from the vibration signal. Popular machine learning techniques include neural networks, as used, for example, in [10], [11], and [12], and Hidden Markov Models, as used, for example, in [13] and [14]. Such approaches have the advantage that they do not require a motor model from which to derive fault frequencies of interest. However, they do require training data and may not generalize well to situations not represented in the training data.

Most previous approaches have made the fundamental assumption that these fault-related disturbances are periodic, but this assumption may not always be valid. We have observed that both the inter-disturbance time intervals and the amplitudes of individual disturbances can vary greatly in some cases. Figure 1(a) shows an idealized vibration signal envelope, while Figure 1(b) shows an actual vibration signal envelope from a ball-faulted bearing in our test rig. It is signals like the one in Figure 1(b) that our method is intended to address. These variations in the vibration envelope are presumably due to variations in how the ball defect contacts the inner and outer races and also to the periodic passing of the faulted ball through the load zone induced by the downward radial load.

Our method for dealing with these departures from periodicity requires no training data. We simply analyze the signal assuming a simple model in which we make two basic assumptions – first, that although the disturbance amplitudes may vary, their shapes will remain similar, and second, that although the inter-disturbance time interval may vary, the most common inter-disturbance interval (the mode of the distribution) will still be the one predicted by the geometry of the bearing. We describe our method in detail in the next section.

III. Method

The goal of our method is to detect the characteristic inter-disturbance intervals for different types of bearing faults in spite of some random variation in this interval and in the amplitudes of individual disturbances. Figure 2 illustrates the steps of our method with a synthetic example signal.

A. Motivating example

We motivate our approach with a simple model of the enveloped accelerometer signal, $x(t)$, an example of which is shown in Figure 2(a). In this model, the impulsive disturbances caused by the bearing defect have a known, fixed shape (exponentially decaying, in this example) but varying amplitude (varying according to an exponential distribution in this example). These disturbances occur at regular time intervals (8 ms in this example) except when there is bearing "slip" during that interval, which we assume happens with probability p_{slip} ($p_{slip} = 0.8$ in this example). When a "slip" occurs, the inter-pulse interval is randomly chosen from a uniform distribution between 0 ms and 16 ms. Finally, there is exponentially distributed white additive background noise. We have chosen these particular distributions to yield a signal that looks qualitatively similar to the real bearing data shown in Figure 1(b), but our approach does not depend strongly on these assumed distributions.

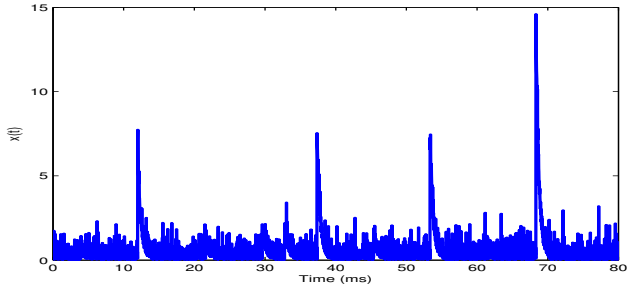
If the disturbance amplitude was constant and if p_{slip} was 0, we would have a signal like that shown in Figure 1(a), and it would be straightforward to detect the fault period either by looking at the autocorrelation of this signal at a lag of 8 ms or by computing the signal spectrum and looking the energy at 125 Hz (= 1/8 ms). This approach fails on our synthetic example signal, however. Figure 2(e) shows the autocorrelation of the example signal for lags from 4 ms-12 ms, which has three large spurious maxima away from the true period of 8 ms. It fails because the autocorrelation is dominated by the few largest-amplitude disturbances, even though these disturbances may suffer from "slip" and thus cause a spurious peak. Spectrum-based approaches (not shown) would suffer from similar problems. Intuitively, we would like to decrease the relative influence of very large disturbances, such as the one near the 70 ms mark in Figure 2(a), and increase the relative influence of small disturbances that are still clearly distinguishable from background noise, such as the one near the 33 ms mark in Figure 2(a). Our approach seeks to do precisely that.

B. Algorithm

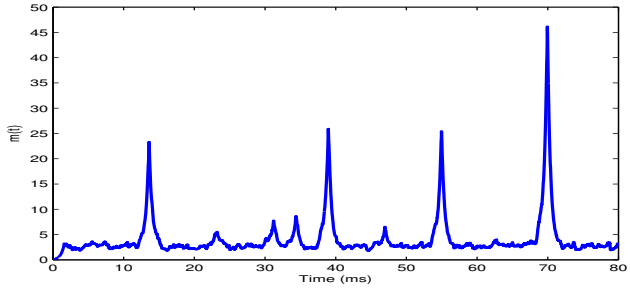
We set as our goal the estimation of the posterior probability of a given inter-disturbance interval, $p_i(\tau)$, for inter-pulse time interval τ. If we know the probability of a disturbance as a function of time, $p_d(t)$, and if disturbances are independent across time, then we have (assuming discrete time):

$$p_i(\tau) \propto P_i(\tau) \sum_t p_d(t) p_d(t-\tau) \tag{1}$$

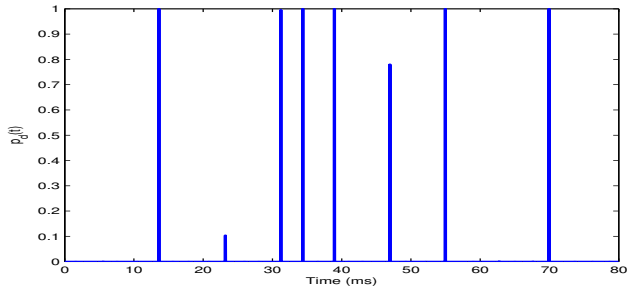
where $P_i(\tau)$ is a prior over possible inter-disturbance intervals, which in practice we will assume is uniform. In other words,



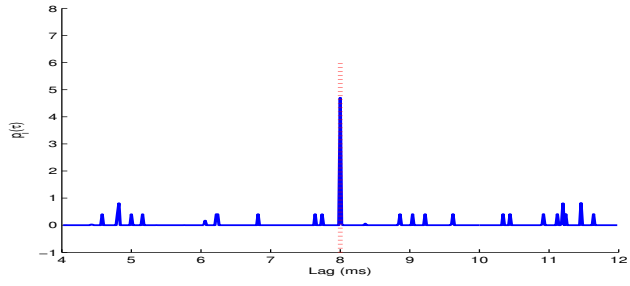
(a) Synthetic example waveform



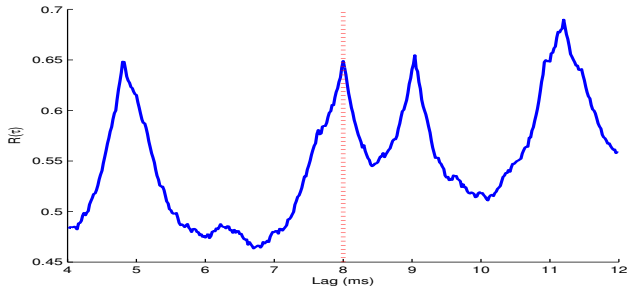
(b) Matched filter output



(c) Probability of disturbance



(d) Probability of inter-disturbance interval



(e) Autocorrelation

Fig. 2. Synthetic example of an accelerometer signal envelope and subsequent processing steps for our method. To show disturbances at a reasonable temporal resolution, (a-c) show only the first 80 ms of the signal. However (d-e) are calculated from a full 500 ms signal to be consistent with our experimental results. See Section III for details.

the posterior probability of an inter-disturbance interval τ is proportional to the prior probability times the autocorrelation of $p_d(t)$. To make use of this relationship, we must find an expression for $p_d(t)$.

The optimal filter to detect a fixed waveform in white noise is to use a matched filter, and this is basis of the approach we take. In the case of our simulated example, we know the pulse shape and can therefore apply a true matched filter. (In practice, we will estimate the disturbance waveform shape as described in the following subsection and use this estimate for the matched filter.) Applying this matched filter to $x(t)$ yields $m(t)$, shown for our example in Figure 2(b). Note how the matched filtering operation brings out even the lower-amplitude disturbances from the background noise. To get $p_d(t)$ from $m(t)$, we apply a pointwise transformation $p_d(t) = g(m(t))$, where $g(m(t))$ is the posterior probability of a disturbance at time t after observing $m(t)$:

$$g(m) = \frac{P_d \cdot p(m|d=1)}{P_d \cdot p(m|d=1) + (1-P_d) \cdot p(m|d=0)} \qquad (2)$$

where P_d is the prior probability of a disturbance, $p(m|d=1)$ is the probability of observing a given value of $m(t)$ given that a disturbance exists at time t, and $p(m|d=0)$ is the probability of observing a given value of $m(t)$ given that no disturbance exists at time t. All of these probabilities are straightforward to compute for our synthetic example. We describe our approach for determining these probabilities in practice in the following subsection.

In practice, a single disturbance will result in large values of $m(t)$ for a range of t in the neighborhood of the disturbance time. Because of this, we preserve only large local maxima of $p_d(t)$ and zero out $p_d(t)$ in a small neighborhood surrounding each large local maximum. At the times of moderate to large disturbances, like the one near 40 ms in Figure 2, $p_d(t)$ (shown in Figure 2(c)) will be nearly 1, and at times with no obvious disturbances present, for example around 60 ms, $p_d(t)$ will be nearly 0. Small disturbances, such as the one near 47 ms, may produce borderline matched filter outputs, resulting in an intermediate value for $p_d(t)$.

Figure 2(d) shows $p_i(t)$, the result of applying our method to the example signal. Note the peak at the correct lag of 8 ms and the absence of spurious peaks and compare this to the spurious peaks resulting from applying standard autocorrelation in Figure 2(e).

C. Practicalities

The above approach works well assuming we can accurately determine the relevant probability distributions. In practice, these distributions are unknown, so in this section we describe our heuristics for determining them. Our goal is to do some simple preprocessing and estimation on a short segment of a vibrations signal envelope (500 ms in duration in our experiments) that will allow us to apply the technique described in the previous section to that signal segment. No separate training data is required.

First we assume that at least a few of the impulsive disturbances will have large amplitude (in fact, we assume

that the largest-amplitude peaks in the signal will all be due to fault-related disturbances), and we create a matched filter by averaging the waveforms in windows around the N highest peaks in the envelope waveform. (For $N = 1$, this would correspond to assuming that the waveform in a small interval about the global maximum of the signal is a prototypical example of the disturbance waveform shape.)

We apply this matched filter estimate to get $m(t)$. We assume that most disturbances will generate noticeable peaks in $m(t)$, so we use the following heuristic to make a first guess at which times correspond to disturbances. We start with the global maximum of $m(t)$ and assume that this time, t_0, is the time of a disturbance. We then assume that no additional disturbances exist in the region around t_0 for which $m(t) > 0.5m(t_0)$ (the "full width half maximum" region around t_0). We subsequently find the largest value of $m(t)$ outside of this region, at time t_1, and do the same thing. We continue this until the entire signal is covered by these "full width half maximum" regions, at which point we have a set, $\{t_0, ..., t_M\}$ of assumed disturbance times and a the corresponding set $\{m(t_0), ..., m(t_M)\}$ of matched filter responses. We set $p(m|d = 1)$ as the maximum likelihood fit of a log-normal distribution to the $\{m(t_0), ..., m(t_M)\}$. (We choose the log-normal distribution because it is a simple parametric form for modeling positive-valued data and has worked well in practice.)

Given this estimated $p(m|d = 1)$, we fix this distribution as one component of a two-component mixture of log-normal distributions, and we use expectation-maximization (EM) [15] to fit this mixture to the full set of values $m(t)$ (for all time, not just peak values). After doing EM, we assume the other log-normal mixture component corresponds to $p(m|d = 0)$. EM also gives us an estimate of P_d.

Our approach makes a number of assumptions, such as the assumption of white noise and the assumption that the probability of a disturbance is independent across time, and in practice we use heuristic approaches to estimate some of the relevant distributions. We will show in the next section that these assumptions and heuristics lead to good results in practice.

We have implemented our algorithm in Matlab. An overview of our algorithm, with parameter settings as used for our results (below), is shown in Algorithm 1.

Algorithm 1 Overview of the method as implemented.

1) Acquire vibration signal at sampling rate 50 kHz.
2) Band pass from 9-20 kHz and extract signal envelope, $x(t)$.
3) Define an approximate matched filter for the disturbance pulse, $h(t)$, by averaging together time-reversed segments of length 3.2 ms centered on the locations of the $N = 10$ largest peaks in $x(t)$.
4) Compute the matched filter output, $m(t) = h(t) * x(t)$.
5) Find $\{t_{max_i}\}$, the set of times of all local maxima of $m(t)$, discarding any local maxima that are within the full-width half-max (FWHM) neighborhood of a larger local max.
6) Fit a lognormal distribution, parameterized by μ_1 and σ_1, to $\{m(t_{max_i})\}$, the set of local maximum signal values. (This lognormal distribution, $p(m|d = 1)$, is assumed to model $m(t)$ for times when a disturbance is occurring.)
7) Use expectation-maximization (EM) to fit a mixture of two lognormal distributions to all of the values $\{m(t)\}$. Keep one of the mixture components, specified by μ_1 and σ_1, fixed, and allow the EM procedure to modify the other mixture component, specified by μ_0 and σ_0. ($p(m|d = 0)$ is now assumed to be parameterized by μ_0 and σ_0; it models $m(t)$ for times when no disturbance is occuring.)
8) Using the lognormal distributions determined in the previous steps, compute the posterior probability of disturbance, $p_d(t) = g(m(t))$, with $g(m)$ as defined in Equation 2. Set $p_d(t) = 0$ for all times not in $\{t_{max_i}\}$.
9) Compute $p_i(\tau)$, the probability of inter-disturbance interval τ, as described in Equation 1.

IV. Results

For our experimental results, we collect data from a "Bearing/Balancing Fault Simulator" from Spectraquest, Inc. This simulator, pictured in Figure 3, consists of a variable-speed driven 1/3 horsepower induction motor connected to a 5/8" shaft. At the center of the shaft is a radially symmetric load weighing a few kilograms, and at either end of the shaft is a ball bearing (type ER-10K). The simulator comes with a set of faulted bearings, and we use the inner race, outer race, and ball faulted bearings in our experiments. For all experiments, the faulted bearing was installed on the far end of the shaft from the motor, and an accelerometer was bolted to the support structure for this bearing as shown. Our accelerometer has a 0.2-10kHz bandwidth, and our data was sampled at 50000 samples/second.

We collected data at three different speeds (1250 RPM, 1750 RPM, and 2400 RPM), and with three different types of faulted bearings (inner race, outer race, and ball-faulted). Our dataset is roughly 5 minutes of vibration data in total, and it is roughly evenly balanced across all speeds and fault types. Figure 4 shows ROC curve results for our technique, autocorrelation, and spectral energy approaches. Each ROC curve shows the detection performance for distinguishing one type of fault from the other two types of faults based on a measurement at the characteristic fault frequency for the fault of interest. Each trial consists of 0.5 seconds of data, which is a very short by some standards. However, rapid and accurate fault detection can be critical, for example for fast testing immediately after assembly or for situations where the motor is only run intermittently.

In Figure 4, "Prob disturbance" is the technique described in the previous section, and the measurement used for the ROC curve is $p_i(\tau_f)$, where τ_f is the period corresponding to the expected fault frequency as determined by the shaft speed and fault location. "Autocorrelation" is the autocorrelation of the vibration envelope evaluated at lag τ_f. "Spectral Energy" is

Fig. 3. Photograph of our experimental setup. The faulted bearings were installed on the far end of the shaft, and the accelerometer was bolted to the top of the far bearing support as shown.

computing a modified periodogram of the vibration envelope and looking at the energy at frequency $F_f = 1/\tau_f$. (Because the overall signal energy can vary significantly as motor speed changes, in practice we normalize all of our measurements by the total signal energy. This improves the performance of all methods.)

All ROC curves are based on the full range of motor speeds, so to achieve good ROC operating points, a technique must provide consistent measurement values independent of motor speed. Figures 4(b) and 4(c) show that detecting outer- and inner-race faults is relatively easy, and all techniques perform well. However, as was shown in Figure 1(b), the ball-faulted bearing signal has large deviations from periodicity, and as Figure 4(a) shows, only our probabilistic disturbance modeling technique reliably detects ball faults.

Our unoptimized Matlab implementation of our technique takes less than one second to process a 0.5 second data segment on a standard PC, so we expect that an optimized implementation could easily achieve real-time performance.

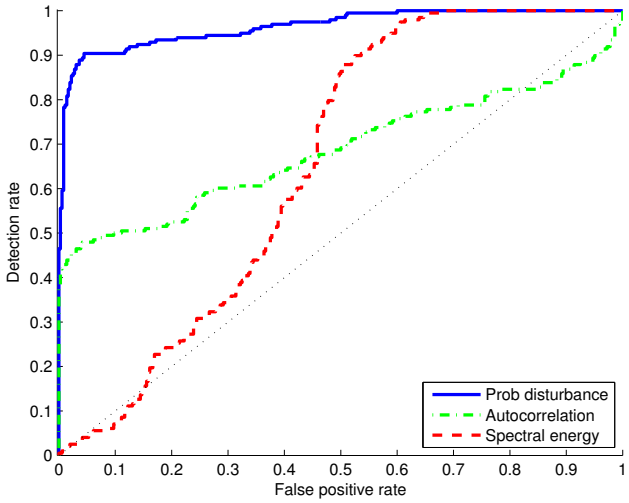
V. Conclusion

We presented a novel approach for detecting the characteristic repetition rates of fault-related disturbances in bearings with single-point defects. Our approach is robust to the significant deviations from periodicity that can be present in some practical situations.

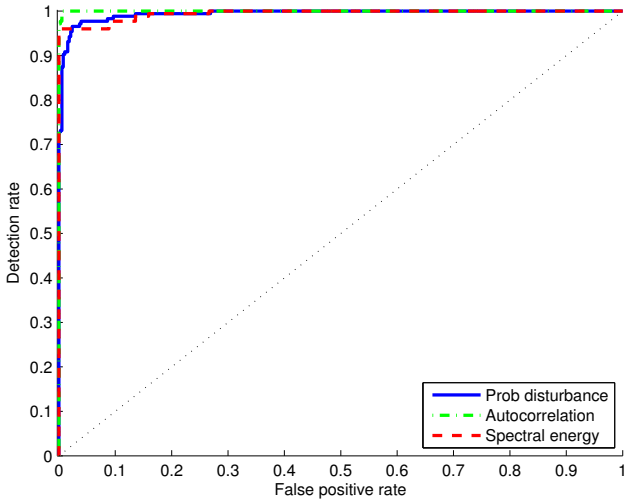
Our technique is based on a simple probabilistic model of fault-related vibration disturbances. We have made a number of simplifying assumptions to arrive at a practically useful technique and have shown that these assumptions work well for data from our experimental setup. We have shown that our technique, in contrast to other techniques, can differentiate ball-faults from other types of faults reliably across a range of motor speeds and given only a short (0.5 second) observation.

References

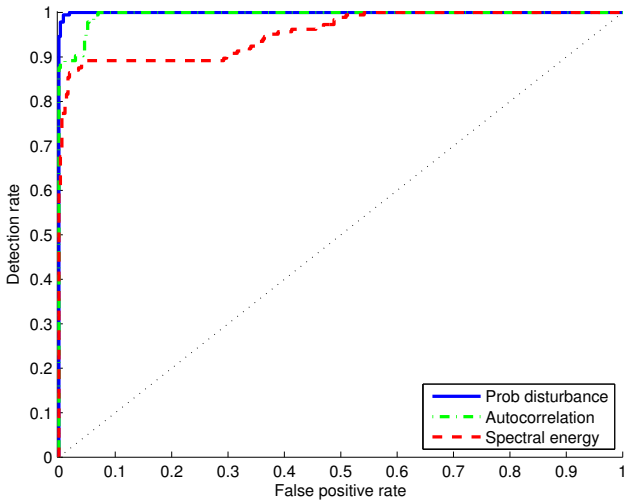
[1] G. K. Singh, "Induction machine drive condition monitoring and diagnostic research-a survey," *Electric Power Systems Research*, vol. 64, pp. 145–158(14), February 2003.



(a) Ball fault



(b) Outer race fault



(c) Inner race fault

Fig. 4. ROC curve showing performance for detecting faults in a dataset consisting of ball faults, inner race faults, and outer race faults measured at a range of different speeds. All methods perform well on outer and inner race faults. Our method is superior for ball faults.

[2] P. D. McFadden and J. D. Smith, "The vibration monitoring of rolling element bearings by the high-frequency resonance technique - a review," *Tribology International*, vol. 17, no. 1, pp. 3–10, 1984.

[3] S. McInerny and Y. Dai, "Basic vibration signal processing for bearing fault detection," *Education, IEEE Transactions on*, vol. 46, no. 1, pp. 149–156, Feb 2003.

[4] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detectio n of defects in rolling element bearings," *Tribology International*, vol. 32, no. 8, pp. 469 – 480, 1999.

[5] J. Stack, T. Habetler, and R. Harley, "Fault classification and fault signature production for rolling element bearings in electric machines," *Industry Applications, IEEE Transactions on*, vol. 40, no. 3, pp. 735–739, May-June 2004.

[6] T. A. Harris and M. N. Kotzalas, *Essential Concepts of Bearing Technology, Fifth Edition*. CRC Press, 2006.

[7] J. Stack, R. Harley, and T. Habetler, "An amplitude modulation detector for fault diagnosis in rolling element bearings," *IECON 02 [Industrial Electronics Society, IEEE 2002 28th Annual Conference of the]*, vol. 4, pp. 3377–3382 vol.4, Nov. 2002.

[8] D.-M. Yang, "Induction motor bearing fault detection with non-stationary signal analysis," in *Mechatronics, ICM2007 4th IEEE International Conference on*, May 2007, pp. 1–6.

[9] S. Seker, S. Gullulu, and E. Ayaz, "Transfer function approach based upon wavelet transform for bearing damage detection in electric motors," in *Industrial Electronics, 2008. ISIE 2008. IEEE International Symposium on*, 30 2008-July 2 2008, pp. 749–752.

[10] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *Industrial Electronics, IEEE Transactions on*, vol. 47, no. 5, pp. 1060–1069, Oct 2000.

[11] H. Su and K. T. Chong, "Induction machine condition monitoring using neural network modeling," *Industrial Electronics, IEEE Transactions on*, vol. 54, no. 1, pp. 241–249, Feb. 2007.

[12] A. Mahamad and T. Hiyama, "Development of artificial neural network based fault diagnosis of induction motor dearing," in *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, Dec. 2008, pp. 1387–1392.

[13] H. Ocak and K. Loparo, "A new bearing fault detection and diagnosis scheme based on hidden markov modeling of vibration signals," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 5, 2001, pp. 3141–3144 vol.5.

[14] A. Lebaroud and G. Clerc, "Classification of induction machine faults by optimal timefrequency representations," *Industrial Electronics, IEEE Transactions on*, vol. 55, no. 12, pp. 4290–4298, Dec. 2008.

[15] J. A. Bilmes, "A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Tech. Rep., 1997.

Kevin W. Wilson received B.S., M.Eng, and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology in 1999, 2000, and 2006, respectively.

He is currently a member of the technical staff at the Mitsubishi Electric Research Laboratories in Cambridge, MA, where he works on a variety of applied signal processing and machine learning problems.