RECOGNITION OF BEARING FAILURES USING WAVELETS AND NEURAL NETWORKS

Alexander Ypma, Ronald Ligteringen, Eduard E. E. Frietman and Robert P. W. Duin*

Pattern Recognition Group
* Computational Physics Group
Faculty of Applied Physics
Delft University of Technology
Lorentzweg 1, 2628 CJ, Delft The Netherlands
email: *ypma@ph.tn.tudelft.nl*

Abstract

Conventional feature extraction methods and a novel variant of the self-organizing map (enabling automatic feature extraction and clustering) are described and investigated for their usefulness in the recognition of bearing failures in a submersible pump. Utilization of spectral features yields adequate results for measurements at one machine running speed, experiments with the ASSOM indicate that it is possible to recognize faults without explicit preprocessing, where also the running speed need not be constant.

1. INTRODUCTION

Predictive maintenance implies that developing faults have to be detected early. Modern approaches focus on the time-frequency characteristics of certain machine faults: incipient failures are often not repetitive in the earliest stages (e.g. an occasional clash occurs), and give rise to nonstationary, impulsive (low-energy broadband) signals, often occuring over several scales[Dow93]. Using a representative set of (spectral and temporal) features, a generic description of a machine's normal behaviour under different cirmcumstances could be made, serving as a reference for the detection of (previously unseen) fault states, i.e. *novelty* [SK95].

2. BEARING FAILURE RECOGNITION

Nonsynchronous high-frequency components are indicative of incipient failure, e.g. fatigue cracks give rise to relative short-time transients in the vibration signal. As wear progresses, the fault becomes repetitive and ultimately related to running speed, e.g. as higher harmonics [Bro84]. Many diagnostic methods rely on being able to identify families of equally spaced components, such as harmonics and sidebands. For this purpose, techniques like cepstrum analysis and envelope detection can be applied. Rolling element bearings introduce specific defect frequency patterns of a spiky nature that can be related to the bearing geometry [Mit93] [Ber91]. For example, the *ball passing frequency inner race* can be computed by

$$BPFI = \frac{N_b}{2}(1 + \frac{B_d}{P_d} cos\theta) \times RPM, \qquad (1)$$

where N_b denotes the number of balls or rollers, B_d is the ball or roller diameter (inches or mm), P_d is bearing pitch diameter and θ is the contact angle between ball and outer race (degrees). Similar formulas can be given for the outer race defect frequency, cage defect frequency and rolling element defect frequency.

Generally, race frequencies will appear first, ball- and cage defect frequencies may appear later as sidebands around the race defect frequencies. Defect frequencies may also appear as a modulation or envelope on a high-frequency signal, which may be resolved by envelope detection [Mit93]. Many bearing components contain natural frequencies already in the range of 500 to 2000 Hz, probably being excited by intermittent impacts of the rolling elements striking flaws on the raceways. Usually, *many bearing frequency harmonics present* in the spectrum are a much better indication of bearing damage than defect frequency amplitudes [Ber91]. The optimum bandwidth for observing harmonics of the defect frequencies sould be limited to the fourth through tenth orders [BB96]. According to [Ran87], presence of sidebands can give additional information, as the fault signal tends to be unmodulated for outer race faults, modulated by the speed shaft for inner race faults, and by the cage speed for rolling element faults.

Unfortunately, many ambiguities are present in the vibration spectrum, since certain components may more or less overlap. First, certain faults may exhibit themselves in different ways in the spectrum. Rolling element bearing failures lead to high nonsynchronous frequencies, and change with variations in contact angle, caused by changes in thrust load or when sliding occurs. Actual failure frequencies may therefore deviate slightly from calculated val-

This work was partially supported by the Dutch Technology Foundation, Project No. DTN-44.3584

ues. Second, high vibration levels may both be caused by forced vibrations due to machine faults and excitation by machine resonances, which clarifies the need for understanding the machines structural behaviour. Third, for detection of high-frequency phenomena, a large dynamic range of the analysis method is required: e.g. in gears and rolling element bearings, vital condition-related characteristics at low amplitudes are mixed with high amplitude components at gear mesh or ball passing frequency. Finally, noise may severely complicate detection of early (low-amplitude high-frequency) failures.

Instead of performing an extensive quest for a large set of possible failure causes using a specific preprocessing of the vibration spectrum for each type of failure, we focus on a more general procedure: we are looking for an approximate characterization of all types of failures simultaneously, using the machine's vibration spectrum and information about its operating state, environment, time-frequency behaviour, etc., denoting a a *generalized machine fingerprint*

3. BASES FOR FEATURE EXTRACTION AND DIMENSION REDUCTION

In order to reduce the dimensionality of the feature vectors characterizing the machine state, an efficient representation using a suitable basis can be advantageous.

3.1. Spectral representations

From the discussion above it is clear that the classical Fourier basis has valuable properties for representation of machinery measurements. Envelope detection prior to Fourier transformation enhances evenly spaced defect fequency sidebands and harmonics. It consists of a two-stage process: (i) bandpass-filtering the data around the frequencies of interest (e.g. the region containing bearing resonances) and (ii) demodulating the data by applying the Hilbert-transform

$$H\{f(t)\} = \tilde{f}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(\tau)(\frac{1}{t-\tau})d\tau \quad (2)$$

where the magnitude $|A(t)|$ of the *analytic signal* $f_a = f(t) + j\tilde{f}(t) = |A(t)|e^{j\phi(t)}$ is the amplitude modulating signal.

3.2. Wavelet representations

Windowing the Fourier transform yields a constant bandwidth analysis of the input data (the *spectrogram*). In order to achieve accurate high-frequency information, the time interval in which the information is being obtained should be short, and vice versa. This can be established by introducing the *Continuous Wavelet Transform*

$$(W_{\psi}f)(b,a) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)}dt, a \in R^+, b \in R \quad (3)$$

where $\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a})$ is a set of functions, called *wavelets*, obtained from a basic wavelet $\psi \in L^2(R)$ by dilation (i.e. scaling) and translation. Here, the width of the frequency window decreases for increasing scale parameter a, and vice versa, implying a *constant percentage bandwidth analysis* [Ran87].

3.3. Principal components analysis

Alternatively, the eigenvectors $u_j, j = 0, \ldots, p-1$ of the covariance matrix G of a n-dimensional dataset $\mathbf{x}, n > p$ can be used as basisvectors for a transformation matrix R, yielding an analysis $\mathbf{x}\prime = R\mathbf{x}$ into principal components

$$x\prime_j = u_j^T\mathbf{x}, j = 0, 1, \ldots, p-1 \quad (4)$$

where $Gu_j = \lambda_j u_j, j = 0, 1, \ldots, p-1$ and the eigenvalues λ_j are arranged in decreasing order $\lambda_0 > \lambda_1 > \ldots > \lambda_{p-1}$. The amount of variance retained in the first p principal components is optimal in a least-squares sense [BD89].

3.4. Adaptive Subspace Self-Organizing Map

In the original Self-Organizing Map architecture, an elastic net of locally connected nodes is trained to represent the input data, employing both a vector quantization and dimension reduction, while retaining the topology of the input space as much as possible. When each node consists of a *subspace* (Adaptive Subspace SOM or ASSOM), a map detecting (certain) invariants of the input data in a self-organized manner can be seen to emerge [KKL97]. Unlike the original SOM, adaptation now takes place w.r.t. *episodes* of the input data, i.e. collections of locally shifted finite segments of the input data that contain the common features of all segments, hence retaining translation-invariant features of the episode. Updating is done analogously to the original SOM: the winning node and its neighbourhood are driven towards the input of the ASSOM, in this case by rotating the target subspace towards the episode subspace

$$\begin{aligned}\mathbf{b}_h^{(i)}(t+1) &= \alpha(t)\mathbf{b}_h^{(i)}(t), \forall \mathbf{x}(t_p), t_p \in S(t),\\ \alpha(t) &= I + \lambda(t)h_c^{(i)}(t)\frac{\mathbf{x}(t_p)\mathbf{x}(t_p)^T}{\|\hat{\mathbf{x}}^{(i)}(t_p)\|\|\mathbf{x}(t_p)\|}\end{aligned} \quad (5)$$

where $\mathbf{b}_h^{(i)}(t)$ denotes basis vector h of SOM node i at time t; $\mathbf{x}(t_p)$ denotes the segment in the episode with time instances $S(t)$, that starts at time t_p; $\|\hat{\mathbf{x}}^{(i)}(t_p)\|$

is the length of the projection of x onto the subspace of node i; and λ and $h_c^{(i)}$ the usual (time-dependent) learning rate and neighbourhood functions, respectively. The winning node is defined as the node that contains the subspace that has maximal sum-of-squared projections of the segments in an episode onto the subspaces of the ASSOM (hence is closest to the subspace spanned by the components of the episode). This approach has the advantage that basisvectors representing features of the input data are formed automatically, i.e. without choosing an a priori basis (like in a wavelet analysis) and the types of features are not restricted to be of a spectral nature, since it relies on a kind of template matching in the time-domain. When a Gaussian window is applied to the data segments before they enter the episode, center symmetry of the resulting basisvectors is enforced. When the input data is of a sinusoidal nature, Gabor wavelet - like basisvectors can be seen to emerge [KKL97].

4. EXPERIMENTAL SETUP

In order to determine a machine's representation in some high-dimensional feature space under various operating conditions (Section 2), an in-house test rig is built up. The rotating machine under investigation is a one-channel impeller submersible centrifugal pump, manufactured by Landustrie, Holland. It rotates at 1500 RPM, has a maximum flow of 1.7 M^3 per minute, and a power consumption of 3 KW. The impeller is attached by the motor via one direcly coupled shaft, which is kept in place by two ball bearings.

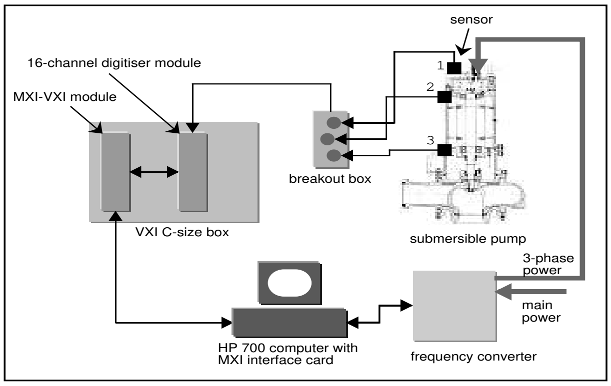

Figure 1: : Experimental setup

Vibration is measured on the casing with three underwater accelerometers. One PCB 324A11 ring accelerometer is placed on top of the pump for axial vibration measurements, another one is mounted close to the double bearing. Near the (uppermost) single bearing (SKF-6305) a PCB 329A11 triaxial accelerometer is used. The sensor data is digitized with a HP E1432B 16 channel 51.2 kSa/s digitizer plus DSP and tachometer input, where an alias protected frequency range from 10 to 23.000 Hz can be measured.

5. EXPERIMENTS

Measurements were obtained from three kinds of machine operating modes: normal operation, imbalance and a bearing failure (an inner race flaw of approx. 1.5 mm), where the running speed of the machine was varied. Measurements from the triaxial accelerometer near the single bearing in the *axial* direction were selected for the experiments, since spectra from this channel differed most significantly for the three operating modes. The defect frequency corresponding to the inner race defect was at 177.8 Hz. The maximum frequency present in the discretized data was 20.000 Hz. Using the heuristics from Section 2 we chose the frequency band of interest as 500 - 3800 Hz. In a wavelet transform this corresponds best to the level 5 detail coefficients (1250 - 2500 Hz). After envelope detection around this subband, a spectrum containing many evenly spaced sidebands was observed in the measurements from the inner race defect, whereas it was absent in the other two operating modes (Figure 2).

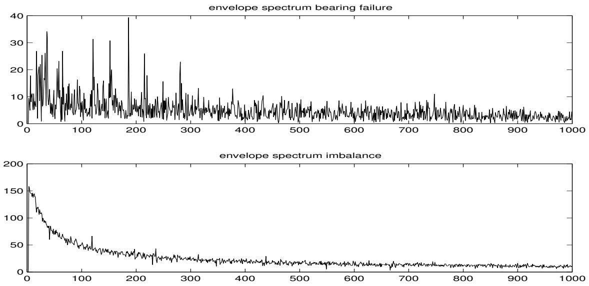

Figure 2: : Envelope spectra

We compared the Fourier and principal component bases w.r.t. their suitability to distinghuish between measurements from three operating classes (normal, imbalance and bearing failure, at the same running speed). A 256-points FFT was applied to (nonoverlapping) 1000-points segments of the (mean detrended) measurement time series, and the logarithm of the power spectral density was determined of each segment. This yielded a set of 300 256-dimensional feature vectors from three classes, which were normalized w.r.t. mean and standarddeviation, and randomly split into two equal-sized train- and testsets. A second set of train- and testsamples was constructed by applying a Principal Component Analysis to the previous feature vectors. We trained neural networks of 5 and 10 hidden units using the Levenberg-Marquard training algorithm, where we varied the dimensionality of the feature vectors in the range of 1 to 256. For comparison, a 10-nearest neighbour classifier was trained on the data. Classification results, denoted by the mean error on each testset (averaged over 5 identical experiments) showed that for already very small dimensionalities of the feature vectors, classification performance is very good (order of 0.01 %), indicating that the spectral features already sufficed for distinghuishing the three classes. This is illustrated by a scatterplot of the first two principal components of the feature vectors from three classes, and the resulting discriminant functions ('–' = 10-nn classifier, '-' = 141x10x3 neural network):

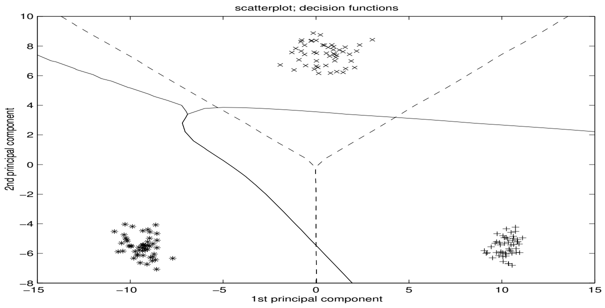

Figure 3: : Scatterplot and discriminant functions

Next, we trained an ASSOM with 8x1 nodes, each consisting of 5 basisvectors of length 32, for 10.000 cycles. In order to check whether certain parts of the map were being tuned to invariants from a certain class, we calibrated the map by constructing episodes out of the measurement time series used to train the map, and measurements at different machine running speeds that were not included in the ASSOM train data. The histograms of the units that were matching best the episodes from the time series are shown in Figure 4. Upper left denotes imbalance, upper right normal operation, lower left bearing failure and lower right bearing failure at a different running speed. Similar behaviour was observed with other time series, indicative of different running speeds (previously unseen by the ASSOM). A majority voting mechanism can hence be used to calibrate the map, and classification may be performed on the basis of a representative set of episodes from a raw measurement series.

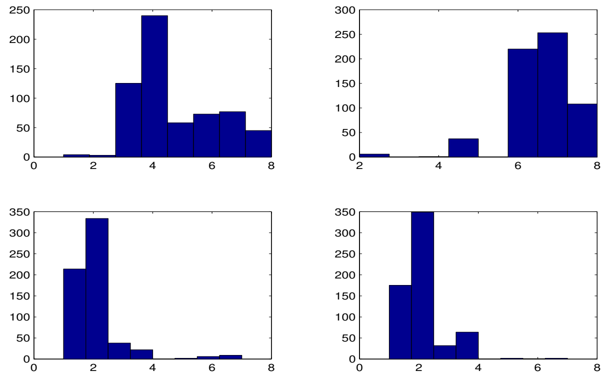

Figure 4: : Histograms of winning nodes for measurements from several machine conditions

6. CONCLUSION

We investigated some characteristics of time series indicative of bearing failures in rotating machines and proposed some methods for the extraction of temporal and spectral features for diagnosis of bearing failure. Measurements from a single machine running speed could already be classified as normal, imbalance or bearing failure very accurately using conventional Fourier techniques. Automatic extraction of invariants of the time series by the ASSOM showed that different map units learned features from separate classes, not limited to a certain running speed. Future work will include the investigation of wavelet networks for feature extraction from vibration measurements and the application of Learning Vector Quantization to the ASSOM, probably enhancing separate localization of classes on the map.

7. REFERENCES

[BB96] A. Barkov and N. Barkova. Condition assessment and life prediction of rolling element bearings - part 1. Technical report, VibroAcoustical Systems and Technologies, St. Petersburg, Russia, 1996.

[BD89] E. Backer and R. P. W. Duin. *Statistische patroonherkenning (in Dutch)*. Delftse Uitgevers Maatschappij, b.v., 1989.

[Ber91] J. E. Berry. How to track rolling element bearing health with vibration signature analysis. *Sound and Vibration*, pages 24–35, november 1991.

[Bro84] J. T. Broch. *Mechanical vibration and shock measurement*. Brüel & Kjær, Denmark, 1984.

[Dow93] M. J. Dowling. Application of nonstationary analysis to machinery monitoring. *IEEE*, 1993.

[KKL97] T. Kohonen, S. Kaski, and H. Lappalainen. Self-organized formation of various invariant-feature filters in the adaptive subspace som. *Neural computation*, 1997.

[Mit93] J. S. Mitchell. *An introduction to machinery analysis and monitoring - 2nd ed.* PennWell Publ. Comp., 1993.

[Ran87] R. B. Randall. *Frequency analysis*. Brüel & Kjaer, Denmark, 1987.

[SK95] O. Simula and J. Kangas. *Process monitoring and visualization using Self-Organizing Maps*, chapter 14, pages 377–390. Elsevier Science B.V., 1995.