Automatic Bearing Fault Pattern Recognition using Vibration Signal Analysis

E. Mendel, L. Z. Mariano, I. Drago
S. Loureiro, T. W. Rauber, F. M. Varejão
Department of Computer Science, Technology Center
Federal University of Espírito Santo
Av. Fernando Ferrari s/n, 29060-970 Vitória, ES, Brazil
Email: thomas@inf.ufes.br

R. J. Batista
Espírito Santo Exploration and Production Business Unit
Automation, Inspection and Maintenance Engineering
Petróleo Brasileiro S.A. PETROBRAS, Vitória, ES, Brazil

Abstract—This paper presents vibration analysis techniques for fault detection in rotating machines. Rolling-element bearing defects inside a motor pump are the object of study. A dynamic model of the faults usually found in this context is presented.

Initially a graphic simulation is used to produce the signals. Signal processing techniques, like frequency filters, Hilbert transform and spectral analysis are then used to extract features that will later be used as a base to classify the states of the studied process. After that real data from a centrifugal pump is submitted to the developed methods.

I. Introduction

Detecting or even preventing failures in complex machines usually benefits in terms of economy and security [18]. Continuous technological development contributes to the increase of the lifetime of a rolling bearing. However, defects can occur due to the great number of critical processes where bearings are employed. The precocious diagnosis of possible faults constitutes an important activity to prevent more serious damages.

Predictive maintenance [16], from the analysis of vibration signals produced by the process, allows to monitor and make conclusions about the operational state of the machine, in addition to that allows taking appropriate measures to extend the time of use, and to minimize costs resultant from the machine's downtime.

The objective of the signal analysis is the discovery of discriminative features that allow the identification of problems in their early stages. In particular, bearing problems manifest in alterations of the machine's vibration patterns.

Especially for defects in rolling-element bearings *envelope detection* [9] is an indicated technique because the mechanic defects in components of the bearing manifest themselves in periodic beatings, overlapping the low frequency vibrations of the entire equipment, for instance caused by unbalance of the pump's rotor. The Hilbert transform, [1], [20] plays an important role in the sequence of steps of the analysis. The main idea is the separation of the defect frequency and the natural frequency of the beating by demodulation.

II. Vibration Analysis in Rotating Machines

Motor pumps, due to the rotating nature of their internal pieces, produce vibrations. Accelerometers strategically placed in points next to bearings and motors allows the position, velocity or acceleration of the machine over time to be measured, thus generating a discrete signal of the vibration level. Fig. 1 shows a typical positioning configuration of accelerometers on the equipment. In general, the orientations of the sensors follow the three main axes of the machine, i.e. vertical, horizontal and axial.
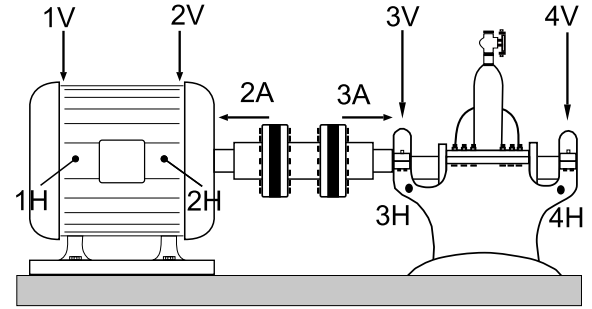

Fig. 1. Motor pump with extended coupling between motor and pump. The accelerometers are placed along the main directions to capture specific vibrations of the main axes. (H=horizontal, A=axial, V=vertical)

A. Fault Models in Bearings

The structure of a rolling bearing allows to establish a model of possible faults. The bearings, when defective, present characteristic frequencies depending on the localization of the defect [13], [10], [14]. Defects in rolling bearings can be foreseen by the analysis of vibrations, detecting spectral components with the frequencies (and their harmonics) typical for the fault.

There are five characteristic frequencies at which faults can occur. They are the shaft rotational frequency F_S, fundamental cage frequency F_C, ball pass inner raceway frequency F_{BPI}, ball pass outer raceway frequency F_{BPO} and the ball spin frequency F_B.

For the ball bearings with angular contact with the cage, the outer ring is static and the inner ring rotates at the shaft speed. The characteristic fault frequencies can be calculated by the following equations:

$$F_{\mathrm{C}} = \frac{1}{2}F_{\mathrm{S}}\left(1 - \frac{D_b\cos(\theta)}{D_c}\right) \tag{1}$$

$$F_{\mathrm{BPI}} = \frac{N_B}{2}F_{\mathrm{S}}\left(1 + \frac{D_b\cos(\theta)}{D_c}\right) \tag{2}$$

$$F_{\mathrm{BPO}} = \frac{N_B}{2}F_{\mathrm{S}}\left(1 - \frac{D_b\cos(\theta)}{D_c}\right) \tag{3}$$

$$F_{\mathrm{B}} = \frac{D_c}{2D_b}F_{\mathrm{S}}\left(1 - \frac{D_b^2\cos^2(\theta)}{D_c^2}\right) \tag{4}$$

where D_b is the ball diameter, θ is the contact angle between the balls and the cage, D_c is the cage diameter and N_b is the number of balls in the bearing. These equations consider that the rolling elements do not slide, but roll over the race's surfaces.

For bearings where the balls do not have an angular contact with the cage, when there are defects in a rolling element, the fault vibration frequency appears as twice the frequency F_{B}, because the defect will collide on both races at each ball rotation.

These frequencies stem, in fact, from defects. They will only be present in the vibration spectrum when the bearings are really defective or, at least, when their components are subject to excessive tensions and deformations that can induce a fault.

Fig. 2 illustrates a basic model of a bearing with the rolling elements, the inner and outer raceways and the cage.
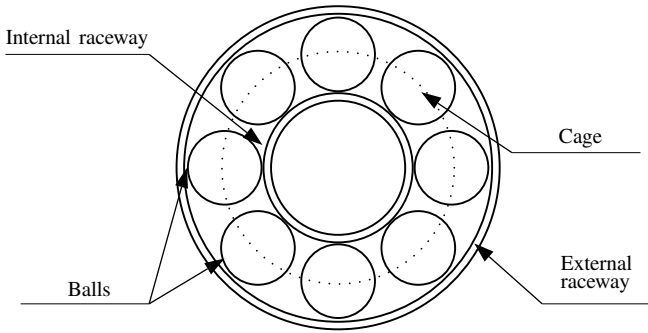

Fig. 2.   Sectional view of a bearing model [10].

B. Spectral Composition

In the presence of bearing defects there are vibrations that overlap the normal functioning signals. Besides that, faults from other problems of the machinery can also occur. An example are the lower frequency vibrations which typically occur in case of unbalance of the rotating parts of the pump.

Whenever a collision between a defect and some bearing element happens, a short duration pulse is produced. This pulse excites the natural frequency of the bearing, resulting in an increase of the vibrational energy. We consider three basic frequency bands that are relevant for the defect analysis: the lower unbalance frequencies, the higher frequencies of the

mechanic shocks of the balls with the cage (resonance) and one or more of the frequencies defined in the equations (1) to (4).

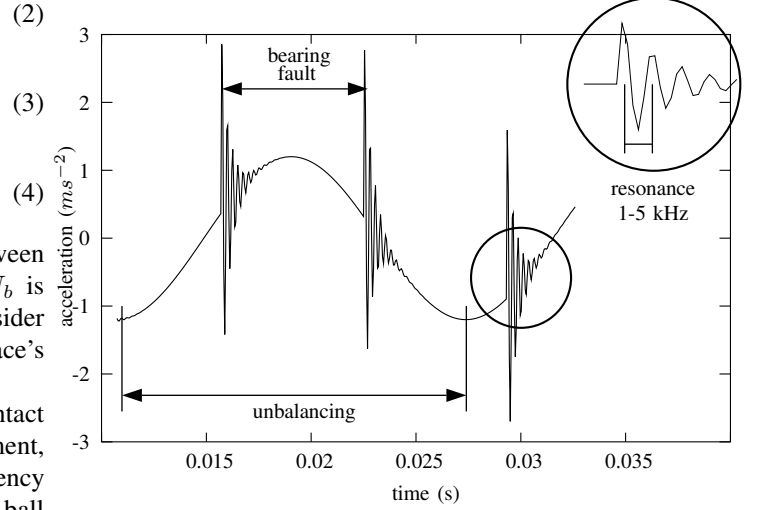Fig. 3 shows some of the involved frequencies.

Fig. 3.   Time domain signal with low frequency unbalance, resonance of the bearing collisions and intervals between the defects.

III. ENVELOPE DETECTION

The defect detection based on the frequencies of eqs. (1) to (4) follows a set of consecutive stages usually denominated as envelope detection [2], [6], [9]. The envelope detection is an important signal processing technique that helps in the identification of the bearing defects, extracting characteristic frequencies from the vibration signal of the defective bearing. The objective is the isolation of these frequencies and their harmonics, previously demodulated by the Hilbert transform. With this analysis it is possible to identify not only the occurrence of faults in bearings, but also identify possible sources, like faults in the inner and outer race, or in the rolling elements.

A. Spectral Filtering

The first step in amplitude demodulation is signal filtering with a band-pass filter to eliminate the frequencies associated with low frequencies defects (for instance unbalance and misalignment) and eliminating noise.

The frequency band of interest is extracted from the original signal using a FIR filter, [8], [6], [17], [11] in the time domain.

The response to the impulse $b(n)$, i.e. the coefficients, of the used band-pass FIR filter, with ideal response $H_{\mathrm{PB}}(e^{i\omega})$, is given by [11]

$$\begin{aligned} b(n) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_{\mathrm{PB}}e^{in\omega}d\omega \\ &= \frac{1}{\pi n}\left[\sin(n\omega_{\mathrm{c}_2}) - \sin(n\omega_{\mathrm{c}_1})\right]. \end{aligned} \tag{5}$$

$$b(n) = \begin{cases} (\omega_{\mathrm{c}_2} - \omega_{\mathrm{c}_1})/\pi, & n = 0 \\ \\ \frac{1}{\pi n}\left[\sin(n\omega_{\mathrm{c}_2}) - \sin(n\omega_{\mathrm{c}_1})\right], & |n| > 0 \end{cases} \tag{6}$$

where the frequencies ω_{c_1} and ω_{c_2} are the normalized cut-off frequencies.

An impulse response of finite length is obtained by a truncation, $b'(n) = b(n) \cdot w(n)$. The effects of the *Gibbs phenomenon*, [8], [17], caused by the abrupt discontinuity (or truncation) of the impulse response in the time domain, are reduced by the utilization of a window, $w(n)$, with small lateral lobes like the *Hamming window*.

A delay in $b'(n)$, in order to obtain a causal filter, is introduced by left shifting the origin and re-indexing the coefficients, that is, $b'(n) = b'(n - M); \quad n = 0, 1, \ldots, 2M$.

The spectral filtering in the time domain is concluded by a convolution of the input signal $x(n)$ with the coefficients, i.e., $y(n) = \sum_{k=0}^{N} b'(k) \cdot x(n-k)$, where N is the filter order and $y(n)$ is the filtered signal.

B. Hilbert Transform

The vibration signals of interest have repetitive high frequency manifestations as a consequence of the excitation of high frequency resonances in regular intervals (see Fig. 3). These free vibrations generated by the bearing defects are modulated in amplitude by the sequence of repetitive impacts and by the damping effect.

The direct frequency analysis of the signals does not provides much information [6], because in the high frequency bands there is noise and other defects mixed with the characteristic frequencies of bearing faults. These repeating frequencies are, however, easily measured in the signal envelope. The envelope detection method (or amplitude demodulation) provides an important and effective approximation to analyze fault signals in high frequency vibrations.

The signal envelope can be calculated by the Hilbert transform [1], [20]. Given a signal $h(t)$ in the time domain, the Hilbert transform is the convolution of $h(t)$ with the signal $\frac{1}{\pi t}$, producing a new signal in the time domain.

$$\widetilde{h}(t) := \mathscr{H}\{h(t)\} := h(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} h(t) \frac{d\tau}{t - \tau}. \quad (7)$$

Considering the spectrum of $h(t)$ and $\widetilde{h}(t)$, knowing that the convolution in the time domain is equivalent to a multiplication in the frequency domain, and that $\text{sgn}(\omega)$ is the sign function, we get

$$\mathscr{F}\{\widetilde{h}(t)\} = -i \cdot \text{sgn}(\omega) \cdot \mathscr{F}\{h(t)\}, \quad (8)$$

i.e. the Hilbert transform causes a shift of $+90°$ for the positive frequencies and of $-90°$ for the negative frequencies, leaving the amplitudes unmodified.

The *analytic signal*, $h_a(t)$, a complex signal composed by the original signal $h(t)$ and its Hilbert transform $\widetilde{h}(t)$ in quadrature, defined as $h_a(t) := h(t) + i\widetilde{h}(t)$, has a spectrum with positive frequencies only.

It is possible to obtain the analytic signal from the equation (8)

$$\mathscr{F}\{h(t) + i\widetilde{h}(t)\} = \mathscr{F}\{h(t)\} \cdot (1 + \text{sgn}(\omega)) \quad (9)$$

$$\mathscr{F}\{h_a(t)\} = \mathscr{F}\{h(t)\} \cdot \begin{cases} 2, & \omega > 0 \\ 1, & \omega = 0 \\ 0, & \omega < 0 \end{cases} \quad (10)$$

Considering the signal (original and analytic) as a modulation by the (complex) signal $e^{i\omega_c t}$ of a carrier frequency of angular frequency ω_c, the magnitude of the Fourier transform of the analytic signal $|\mathscr{F}\{h_a(t)\}|$ is a (scaled) version of the magnitude of the Fourier transform of the demodulated original signal $|\mathscr{F}\{h(t)\}(\omega - \omega_c)|$, i.e. relocated to the low frequencies $\omega - \omega_c$. In this way it is possible to isolate the bearing defects.

In the discrete form, utilizing the DFT (discrete Fourier transform), the equation (10) can be represented in the following way, [20] and [12]

$$\text{DFT}\{h_a[k]\} = \text{DFT}\{h[k]\} \cdot \begin{cases} 2, & k = 1, \ldots, N/2 - 1 \\ 1, & k = 0, N/2 \\ 0, & k = N/2, \ldots, N - 1 \end{cases}$$
$$(11)$$

The inverse transform of the equation (11) is the analytic signal $h_a[k]$, which imaginary part is the Hilbert transform, by which it is possible to extract the envelope of the signal, i.e., the magnitude of $h_a[k]$

$$\mathscr{E}[k] = \|h_a[k]\| = \sqrt{h^2[k] + \widetilde{h}^2[k]} \quad (12)$$

The analysis steps for the calculus of the bearing defect frequencies spectrum are then resumed: 1º) Low frequency filtering to eliminate the influence of slow vibrations, 2º) Calculus of the analytic signal $h_a(t)$ of the original signal $h(t)$, 3º) Fourier transform of the analytic signal, 4º) Analysis of the magnitude of the spectrum.

After the calculus of the spectrum, with the knowledge of the bearing properties, a classification module is responsible for the diagnosis of the possible fault.

IV. SIMULATION

A dynamic simulator with a graphical interface for synthetic signal generation was developed. Fig. 4 shows the graphical model of the simulator's bearing, without the cage representation.

Fig. 4. Graphical model of the simulator's bearing.

The simulator was implemented in C, with the OpenGL graphical interface library and Gnuplot for graphics generation in real time. The objective of the simulator is to generate signals of defects in bearings to facilitate the learning and training of the discussed signal processing techniques. With the simulated signals, all the techniques presented here can be applied to extract necessary information in order to diagnose if the bearing is defective, which is the possible defect and what is the level of degradation.

It is possible to simulate defects in the inner and outer raceways, fissures in the rolling elements and unbalance of the motor pump. Gaussian noise, representing random vibrations from other sources of the motor pump is added to the synthetic signal granting a more realistic character to the data.

The resulting signal is composed of two sources: a low frequency vibration, emulating the unbalance of the rotating parts of the motor pump and a damped harmonic oscillator, emulating the mechanic shock between the dynamic and static parts inside the bearing, for instance, caused by the passage of a ball over a fissure in a raceway.

A. Damped Oscillations with one Degree of Freedom

If the source of a vibration is detectable by the accelerometers, we are interested in the displacement $x(t)$, caused by the beatings of a ball in an irregularity inside the bearing. Consider an isolated system. Adding to the balance of force (Hooke's law) $F = m\ddot{x} = -kx$ of a simple harmonic oscillation, a damping proportional to the velocity, we get

$$m\ddot{x} = -kx - c\dot{x} \tag{13}$$

where m is the dislocated mass, k is the spring constant and c is the damping constant. With the initial conditions $x(t = 0) = x_0, \dot{x}(t = 0) = v_0$ and supposing a underdamped system, $c^2 - 4mk < 0$, the solution of the second-order ordinary differential equation (13) gives us the damped vibration.

$$x(t) = Ae^{-\lambda\omega_0 t}\cos(\omega t - \phi_0) \tag{14}$$

where A is the maximum amplitude of the oscillation, $\lambda = \frac{c}{2m}$ the *damping coefficient*, $\omega_0 = \sqrt{\frac{k}{m}}$ the natural frequency of the oscillator, $\omega = \sqrt{\omega_0^2 - \lambda^2}$ the frequency of the damped system and ϕ_0 the phase of the oscillation.

V. RESULTS

To prove the previously presented fault detection method, the results of two tests are shown: one with artificial data from the simulator and another with real data from a submersible motor pump. We will show that the use of pattern recognition techniques avoids heuristics for filtering the relevant information out of the spectrum of the signal envelope. That means that there is no necessity to explicitly define a frequency band where we expect the faults to manifest themselves. Information filtering methods are used, especially feature selection.

A. Synthetic Data

With the simulation being executed with parameters from a real bearing, it was possible to generate a set of signals for the corroboration of the proposed methods. The simulator was configured to rotate at 1800 RPM (F_S = 30Hz), containing 12 balls, with a diameter of 38.1 mm each, in a cage of 165 mm of diameter and considering the contact angle equal to $37°$. The resonance frequency of the rolling elements was adjusted to 4 KHz and 1024 points were sampled at a sampling frequency of 21 KHz.

Fig. 5 illustrates the signal generated by the simulator according to the aforementioned configuration. For better visualization no noise was included or any other fault source was activated, like unbalance.
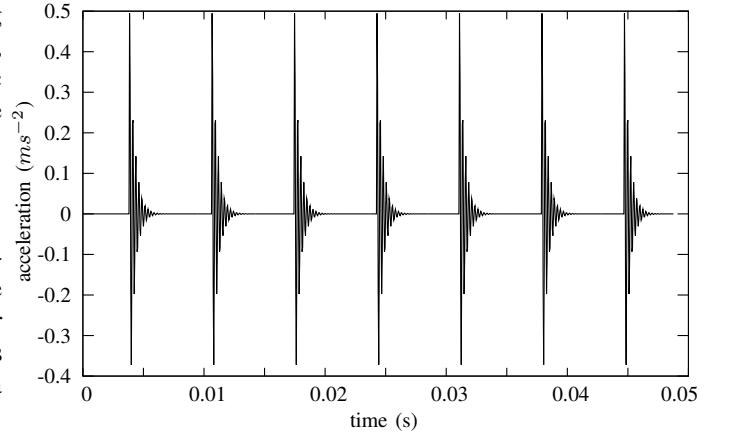

Fig. 5. Simulated original signal in the time domain. No noise or other faults are active, only resonance.

With the data utilized to adjust the simulator, the values of the characteristic fault frequencies of the equations (1) to (4) are: F_C = 12.24 Hz, F_B = 62.93 Hz, F_{BPO} = 146.89 Hz e F_{BPI} = 213.11 Hz. As these fault frequencies appear when the bearing is defective, it is possible to identify the source of the bearing problem by observing the spectrum of the envelope signal.

The first step of the investigation is spectral filtering. The filtered frequency band was [2800 Hz, 5100 Hz], because this region contains the resonance of the material.

The step following the filtering is envelope detection, allowing the identification of the origin of the fault. In Fig. 6 the peaks in the frequency 144.1 Hz and its harmonics are made evident, enabling the detection of a fault in the outer raceway.

Next we postulate that the use of pattern recognition techniques [19], [4], [3], especially *feature selection*, allows an automatic discrimination of the bearing condition classes, i.e. the normal state and the various fault types. In general, measures of specific bands and its derived measures are defined to compose the set of features that are the base of the classification [7], [5], [22].

On the contrary to this explicit definition of the useful features, we apply automatic information filtering techniques to discover the most useful of the features, in this case of
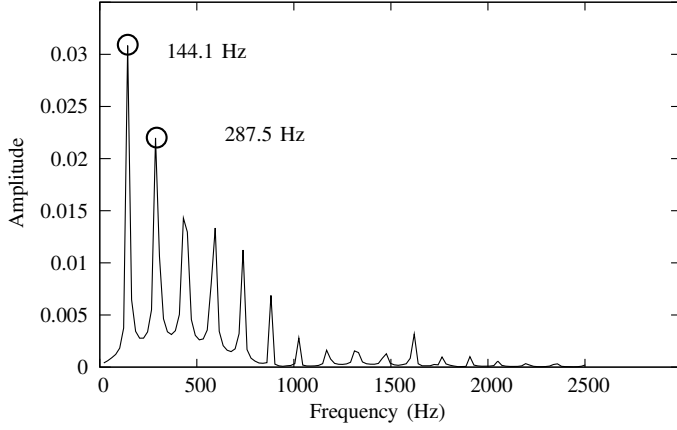
Fig. 6. Envelope spectrum emphasizing the fault in the outer raceway of the bearing (F_{BPO} = 146.89 Hz)

the signal envelope frequencies. We consider therefore the complete envelope spectrum, eq. (12), as the initial feature vector, in this case of dimension $d = 512$. For each of the 5 classes (without fault and the 4 faults of the equations (1) to (4)) 50 samples were generated. Gaussian noise with variance $\sigma^2 = 0.04$ was added to an original synthetic signal in the time domain (1024 samples). For example, for the "inner raceway" fault class, the maximum absolute value is 0.62167, the absolute average value is 0.18079 and the median of the absolute value is 0.15269. Thus, the signal-to-noise ratio is quite low and makes it difficult to filter out the discriminative information.

The technique of *feature selection* [19] [3] has two main advantages. First, it reduces the dimension of the feature vector, facilitating the subsequent classification. For example, training a multi-layer perceptron with 512 entries is much more complex than using only the 20 best features.

In this work we use the *Sequential Forward Selection* strategy [19] because it is a good trade-off between computational cost and search space complexity. Only 20 of the original 512 features were selected, representing a complexity reduction of 96%. As the selection criterion we used inter-class Euclidean distance. The sequence of selected frequencies was: 0 Hz, 123.047 Hz, 287.109 Hz, 840.82 Hz, 2317.38 Hz, 1066.41 Hz, 758.789 Hz, 143.555 Hz, 307.617 Hz, 451.172 Hz, 471.68 Hz, 41.0156 Hz, 225.586 Hz, 594.727 Hz, 922.852 Hz, 1004.88 Hz, 205.078 Hz, 328.125 Hz, 574.219 Hz, 1025.39 Hz. In this sequence one can encounter some of the fault frequencies and their harmonics which evidences the discriminative importance of the characteristic fault frequencies.

An extremely useful tool for high-dimensional data visualizing is the Sammon Plot [15], that in two or three dimensions reproduces the mutual Euclidean distance between each example (here 250) in the original dimension (the 20 best features determined by the feature selection). Over the usually employed two-dimensional visualization by the first two principal components of the Principal Component Analysis (PCA) [19], the Sammon plot has the advantage that it preserves the

Classifier	Estimated error rate
Linear Machine	1.45%
Quadratic Gaussian	3.50%
1-Nearest-Neighbor	2.50%
Multi-Layer Perceptron	1.00%

nonlinear relationships among the data points and that it does not cut off non-principal components. Fig. 7 clearly illustrates the separability of the classes, i.e, the feature model potential to diagnose the health of the bearing.
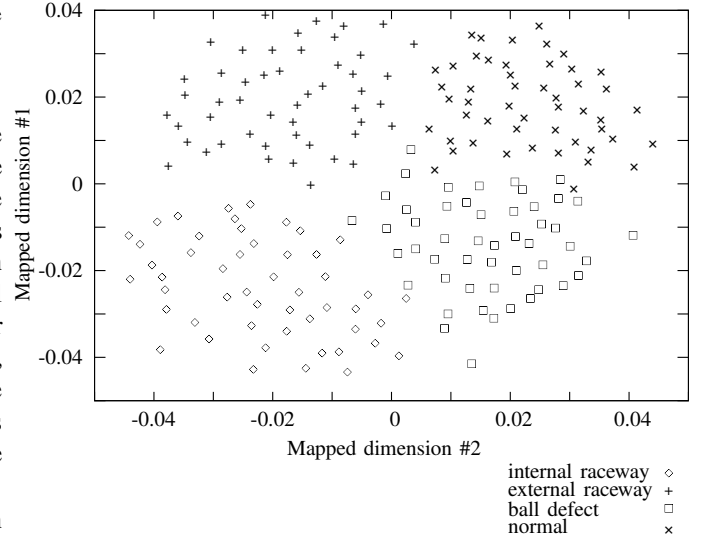

Fig. 7. Sammon Plot for the signal of the envelope of original dimension 20, i.e. after feature selection, mapped to 2-D

An empirical comparison with various classifier models [19] was made to confirm the separability of the data. To estimate the error rate, the "Leave-One-Out" method was used. Table I shows the result of the performance estimates experiences of various classifiers: Linear Machine, Bayes with multivariate Gaussian distribution, 1-Nearest-Neighbor and Single-Hidden-Layer Perceptron.

B. Real Data

The second test was conducted with real data from a submersible centrifugal motor pump, produced by Landustrie, The Netherlands [21]. The rolling-element bearing used in the test was the SKF6305, with fundamental shaft frequency, F_{S} equal to 25 Hz. The values of the frequencies originated from potential damages are: $F_{\mathrm{C}} = 9.2$ Hz, $F_{\mathrm{B}} = 44$ Hz, $F_{\mathrm{BPO}} = 64.4$ Hz and $F_{\mathrm{BPI}} = 111$ Hz. A total of 16384 points was measured with a sampling frequency of 51.2 KHz. Fig. 8 illustrates the signal in the time domain.

Fig. 9 shows the envelope spectrum of the signal filtered from 4 KHz to 8 KHz. The clearly visible peaks at the 106 Hz frequency and its harmonics evidences a fault at the inner raceway.
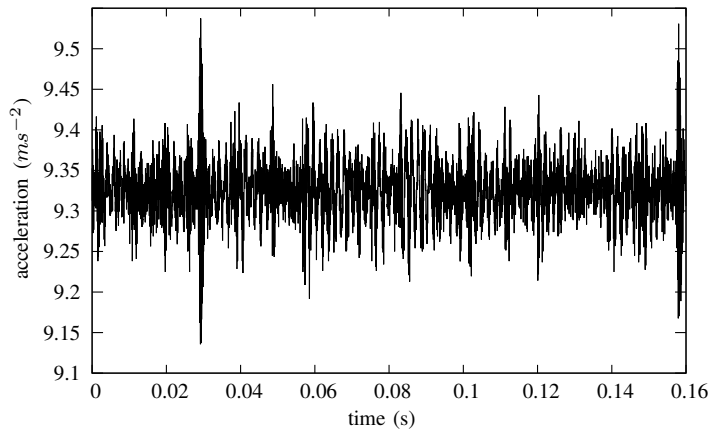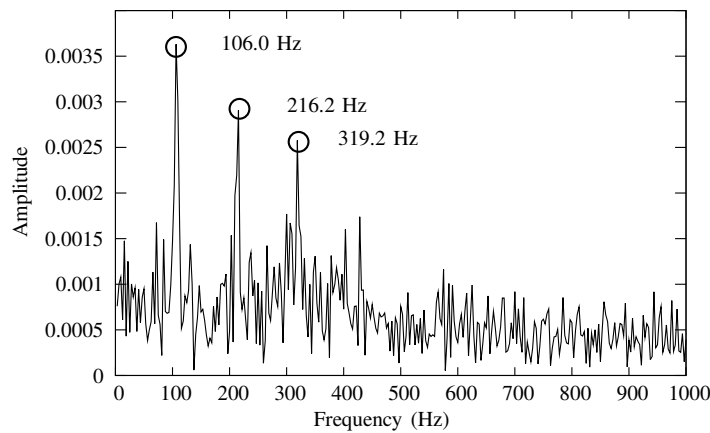
Fig. 8.   Original vibration signal in the time domain

Fig. 9.   Envelope spectrum emphasizing the bearing defect

VI. Conclusion and Future Works

In this work we employed signal processing and pattern recognition techniques to classify faults in bearings. The envelope analysis provides the feature vector used in the subsequent classification steps. On the contrary to the majority of the works that focus on the fault detection problem, we explore pattern recognition methods to automate the analysis of the obtained features.

In the near future we will be able to acquire real data from an experimental workbench (SpectraQuest MFS2004-PK7) allowing the refinement of the developed techniques. A study of distinct bearing models will be realized. It is also projected to implement the fault diagnosis system in an motor pump environment in the oil extraction industry.

Acknowledgments

References

[1] R. N. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, 2nd edition, 1986.
[2] R. A. Collacott. *Vibration Monitoring and Diagnosis: Techniques for cost-effective plant maintenance*. Wiley, New York, 1979.
[3] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice/ Hall Int., London, 1982.
[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, New York, 2nd edition, 2001.
[5] X. Fan and M. J. Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20:966–982, 2006.
[6] T. A. Harris and A. G. Piersol. *Harris's Shock and Vibration Handbook*. McGraw-Hill, 5th edition, 2002.
[7] S. Ahmed Saleh Al Kazzaz and G. K Singh. Experimental investigation on induction machine condition monitoring and fault diagnosis using digital signal processing. *Eletric Power System Research*, 65:197–221, 2003.
[8] S. M. Kuo and B. H. Lee. *Real Time Signal Processing*. Wiley, 2001.
[9] S. A. McInerny and Y. Dai. Basic vibration signal processing for bearing fault detection. *IEEE Transactions on Education*, 46(1):149–156, 2003.
[10] R. Keith. Mobley. *Root Cause Failure Analysis (Plant Engineering Maintenance Series)*. Butterworth-Heinemann, 1999.
[11] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, 2nd edition, 1998.
[12] Soo-Chang Pei and Min-Hung Yeh. Discrete fractional hilbert transform. *IEEE International Symposium*, IV:506–509, 1998.
[13] K. Ragulskis and A. Yurkauskas. *Vibration of Bearing*. Hemisphere, 1989.
[14] N. F. Rieger and J. F. Crofoot. *Vibration of Rotationary Machinary*. Vibration Institute, 1977.
[15] J. W. Sammon Jr. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, 1969.
[16] C. Scheffer and P. Girdhar. *Pratical Machinery Vibration Analysis and Predictive Maintenance*. Elsevier, 1st edition, 2004.
[17] B. A. Shenoi. *Introduction to Digital Signal Processing and Filter Design*. McGraw-Hill, 2006.
[18] P. J. Tavner and J. Penman. *Conditiong Monitoring of Electrical Machines*. Wiley, 1987.
[19] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Third Edition*. Academic Press, Inc., Orlando, FL, USA, 2006.
[20] V. Čížek. Discrete Hilbert transform. *IEEE Transactions on Audio and Electroacustics*, AU-18(4):340–343, 1970.
[21] A. Ypma, R. Ligteringen, E. Frietman, and R. Duin. Recognition of bearing failures using wavelets and neural networks, 1997.
[22] D. Yu, J. Cheng, and Y. Yang. Application of emd method and hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing*, 19:259–270, 2005.