

Feasibility study on diagnostic methods for detection of bearing faults at an early stage

A. FERNÁNDEZ, * J. BILBAO, I. BEDIAGA, A. GASTÓN, J. HERNÁNDEZ

Control Engineering Department

Ideko Research Centre

Pol. Ind. Arriaga, 2 20870 Elgoibar

*Department of Applied Mathematics

University of the Basque Country

School of Engineering Alda. Urkijo, s/n 48013 - Bilbao
SPAIN

<http://www.ideko.es>

Abstract: - This article presents a comparison study on the performance of methods for fault detection in ball bearings. Different kinds of actual faults have been acquired in real time from a test platform and a signal of a fault in an early stage has been created. The effectiveness of prediction of several methods is tested when faults are in different stages of development. Furthermore, the implementation of these techniques for actual systems and real-time application is discussed.

Key-Words: - fault detection, bearings, incipient fault

1 Introduction

Automatic system diagnosis has always attracted considerable interest in control engineering due to its applications in machinery condition monitoring and maintenance, and its effects increasing safety and product quality. The greatest challenge in this field of knowledge is to achieve the prognosis of faults, that is, not just detect, isolate and assess the severity of a fault but also predict when the fault will be unbearable for the operation requirements.

In the case of mechanical transmissions, vibration signal analysis has been proven to be the one of the most effective techniques for the detection and diagnosis [1]. The process followed in these systems consists in a digital processing of the vibration signals and a posterior analysis to determine which element of the specific mechanic component is at fault. Particularly, rolling element bearings and gears are the most critical components in a mechanic transmission and their faults can provoke a great damage to the transmission and the whole machine. In the area of rolling element bearings, McFadden and Smith published several articles for modelling the bearing vibrations caused by single and multiple defects [2], [3]. Since then, the research about fault detection has led to the development and application of different signal processing techniques: starting from the traditional ones more used in industrial applications, [4]; [5]; [6] as FFT, cepstrum, amplitude or frequency demodulation, Hilbert transforms [7] to more recent methods as cyclostationary analysis [8]; [9]; [10], neural networks [11] and wavelet transforms [12]; [13]. Signal models have also been used: MA (moving average) models, AR

(autoregressive) models for prediction [14]. The development of fault models for simulation and diagnostics searches the improvement of the diagnosis [15]. In [16] Randall carried out a comprehensive review of all these methods. More specifically, Baillie discussed the adequacy of autoregressive modelling for fault detection in rolling element bearings [17]. Despite the extensive research carried out into different methods for fault detection and diagnosis, most of them are based on simulated signals or signals from a fault in an advanced stage of development. This study pretends to focus on the comparison of the suitability of all these techniques for the detection of faults in an early stage. In general, faults in ball bearings develop gradually, therefore their detection at an early stage can enable the development of procedures and techniques for machine condition prognosis. Currently, prognosis is evaluated by processing trends or by fault signal models. Hence, after a reviewing current diagnostic methods and the processing trends, the final aim of this research is the automation of the detection, diagnosis and prognosis of ball bearings in real-time. The first step in this process has been the creation of a DSP based device, which allows the integration of monitoring and diagnostic strategies in real-time.

2 Signal acquisition

The system used for obtaining the signals is a test bench of a mechanical transmission, which is composed of several rotating shafts, an engine, a pair of gears, a belt and several ball bearings. Fig. 1 shows the test bench. The monitoring and diagnostic

platform implementation has been divided into three stages. The first stage is corresponding to the signal acquisition. The acquisition system performs the sampling of the signal coming from a piezoelectric accelerometer (4371 Brüel&Kjaer) after being filtered and amplified by a signal conditioning system (2635 Brüel&Kjaer). The sampling frequency is 20000 Hz, which will allow the analysis of signal frequencies up to 10000 Hz. This value is high enough to detect faults in ball bearings since the fault frequency and resonance frequencies do not exceed 3000 Hz for the test conditions. The number of samples stored is 4096, corresponding to a spectrum resolution of 5 Hz, is enough to detect the fault. Finally, the a/d conversion is achieved by means of the ADC TLV1570 (Texas Instruments) which combines a high acquisition speed of 1.25 Msps with a 10-bit resolution.

The second stage consists in the signal processing. The vibration acquired is processed with different diagnostic algorithms, which have been implemented in the DSP. The processing board used is based on the DSP TMS320C6711 of Texas Instruments, which operates in floating point at 150 MHz.

The third and final stage deals with the communication with the PC. Once the signal has been processed and the diagnostic has been obtained, the DSP device transmits that information to the PC. The implemented communication is performed through the parallel port by means of the DSP kernel tools named RTDX (Real Time Data Exchange).

The experiments carried out for the sensing strategy have demonstrated that the best point for the signal measurements is given for a radial and horizontal location of the sensor on the bearing.

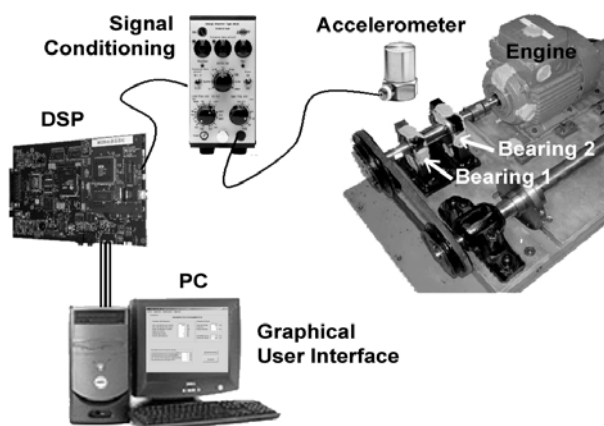


Fig. 1. Data acquisition process for mechanical transmission test bench.

3 Fault description

The most typical faults in ball bearings are produced by a localised wear in the inner race, the outer race or the balls. Localised defects include cracks, pits and

spalls on the rolling surface, although the dominant mode of fault is the spalling of the races. When the ball strikes the defect, a shock is produced, exciting high frequency resonances of the structure. The presence of such defect causes a significant increase in the vibration level. The frequency of the shocks can be calculated by the following formulae (1):

$$\begin{aligned} BPF_O &= \frac{n f_r}{2} \left(1 - \frac{BD}{PD} \cos \alpha \right) \\ BPF_I &= \frac{n f_r}{2} \left(1 + \frac{BD}{PD} \cos \alpha \right) \\ BFF &= \frac{PD}{BD} f_r \left[1 - \left(\frac{BD}{PD} \cos \alpha \right)^2 \right] \end{aligned} \quad (1)$$

where BPF_O is the ballpass frequency for outer race, BPF_I is the ballpass frequency for inner race and BFF is the ball fault frequency. These fault frequencies are dependent on the number of balls (n), shaft speed (f_r), contact angle (α) and ball (BD) and pitch (PD) diameters. The ball bearing characteristics are shown in table I. However, due to small gaps among the balls or a bad lubrication, the fault frequency can suffer slight variations.

The analysis of the fault frequencies from the acquired signals has traditionally been the way to decide whether a fault has produced or not. An increase in the amplitude of the vibration at those frequencies can be interpreted as the result of the periodic shock caused by the fault.

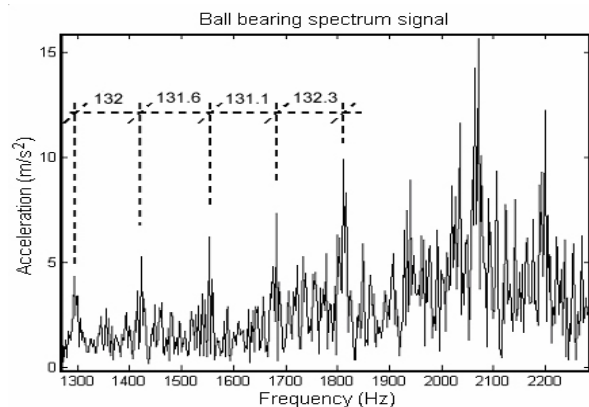


Fig. 3. Spectrum for a bearing with an outer race fault (shaft speed 25 Hz).

In order to check the ability of different diagnostic methods to predict the evolution of damage in ball bearings, a model of a fault at an early stage has been developed. This signal has been created from a non-faulty bearing signal to which a train of impulses has been added. The frequency of this added signal will vary depending on the fault we want to simulate. The faulty signal must have similar characteristics as the non-faulty one, i. e., the fault must not be evident in

the time representation and the resonance frequencies must not be very excited in the frequency representation.

Ball bearings		
Model	SKF 1205	
Ball number	13	
Pitch diameter (PD)	39 mm	
Ball diameter (BD)	7.5 mm	
Contact angle (α)	0 rad	
Fault frequencies (Hz)		
Shaft speed	19.3 Hz	25 Hz
Outer race (BPFO)	101.3	131.25
Inner race (BPF1)	149.57	193.75
Ball (BSF)	96.65	125.19

Table 1. Bearing characteristics

3 Comparison of diagnostic methods

3.1 Traditional methods

Traditional methods are thoroughly presented in [5]. In this study we compare the performance of different traditional techniques for the diagnosis of developed faults and faults at an early stage. The techniques involved in this comparison are: Fast Fourier Transform (FFT) analysis, cepstrum, amplitude demodulation analysis, and Hilbert transform technique for the four types of faults previously described: outer race fault, inner race fault, ball fault and fault at an early stage.

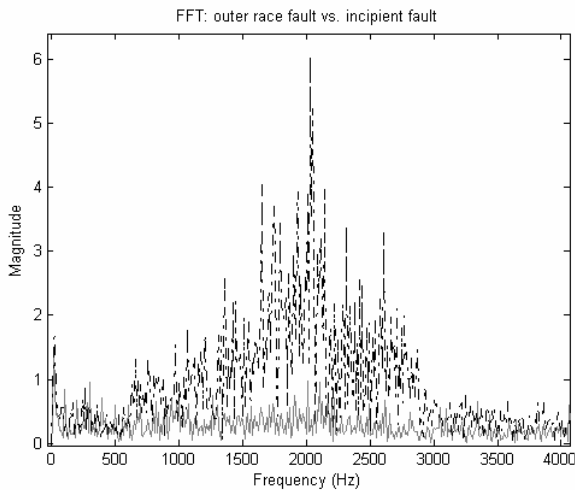


Fig. 4. Comparison spectra for big and incipient fault (shaft speed 19.3 Hz).

3.1.1 Fast Fourier Transform analysis

One of the most current characteristics among vibration signals generated by bearings is that they present different types of series of harmonics. These series are composed of multiples of the fault frequency and lateral bands of the rotation frequency.

For an outer race fault, for instance, there will be a frequency spacing among peaks at the resonance frequencies as it can be seen in fig. 3.

Nonetheless, for undeveloped faults this method does not show peaks of enough magnitude at the resonance frequencies in order to achieve a fault assesment (see fig. 4).

3.1.2 Cepstrum

The cepstral analysis calculates the spectrum of the logarithm of the energy spectrum. This technique discovers periodicities in the frequency spectrum. However, inexactitudes in the spacing of frequencies (as in fig. 4) decrease the effectiveness of this method. Fig. 5 shows that the frequency peak is not strong enough to be distinguished, even for a big fault. Consequently this is not a proper method for discovering incipient faults in bearings.

3.1.3 Amplitude demodulation

Amplitude demodulation is an adequate technique for bearing fault diagnosis since the fault signal is similar to a signal modulated in amplitude. Through a demodulation process, the corresponding signal to the impulses (the envelope) can be separated from that corresponding to the resonances (carrier). In the demodulation process it is essential that the filtering of the signal is correct. The signal must be filtered by a low-pass or band-pass filter in order to eliminate the high frequencies (the carrier). The cut-frequency must be selected above the bearing fault frequency. As a result, the spectrum will show harmonics of the fault frequency (see Fig. 6).

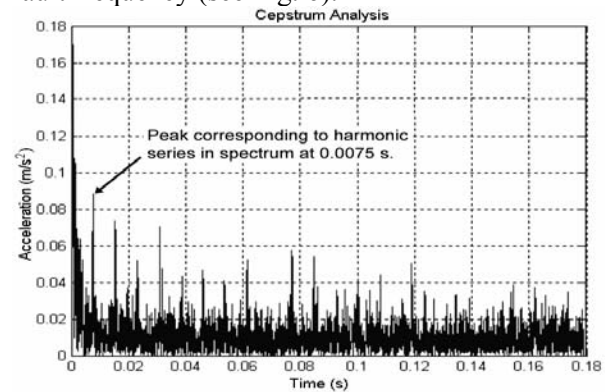


Fig. 5. Cepstral analysis for an outer race fault (shaft speed 25 Hz).

3.1.4 Hilbert transform

The Hilbert transform is the relationship between real and imaginary parts of the Fourier transform. The Hilbert transform is used to demodulate the signal so as to obtain the low frequency variations in a high frequency signal, as it is needed in diagnosing faulty bearings. Fig. 7 shows the Hilbert transform for the

vibration signal of a bearing with a developed fault in the outer race for a shaft speed of 19.3 Hz. The advantage of the Hilbert transform is that this method does not need to filter the high frequencies to obtain a good result. This method is the most sensitive to fault among the classical ones and gives the clearer representation. Therefore it is the most indicated to analyze incipient faults.

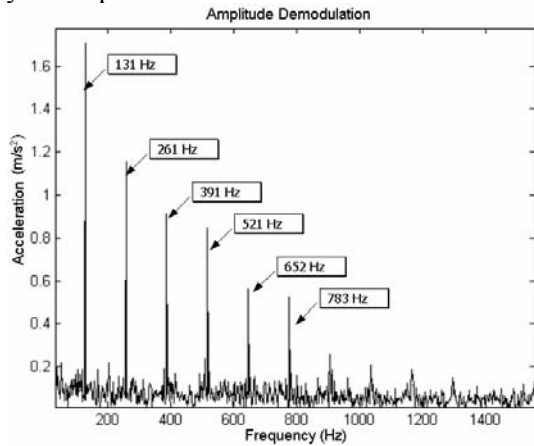


Fig. 6. Amplitude demodulation analysis for an outer race fault (shaft speed 25 Hz).

3.2 Advanced methods

In recent years other approaches to signal processing have been used in fault detection and diagnosis. The most mentioned and published are wavelet transforms and model based diagnosis.

Wavelet transforms can be regarded as a natural evolution of Fourier transforms. Whereas the latter provides a frequential representation of a time series, the former frames that frequential information in time scales. This approach has a great advantage in that it achieves fine time resolution at high frequencies and fine frequency resolution at low frequencies. This characteristic is really helpful when trying to make out faults at an early stage since, as faults in a developed stage, they have to be located both in time and frequency. Separately, the time series and the frequency representation do not make obvious the presence of a fault but their combination may help. Besides, the wavelet transform is a powerful tool to disclose transient information in signals. As a disadvantage for wavelets to be used in real-time processing, we could appoint the fact that they become computationally slow. Analogously with the Fourier domain, where convolutions become computationally fast, we can find a large number of computations that can become computationally efficient thanks to the sparsity of operators when transformed into the wavelet domain [19]. This characteristic makes possible its implementation for real-time processing, as presented in [20] with the continuous wavelet transform. However this

implementation is beyond the aim of this document. Therefore, we will only discuss the possibilities of the technique to detect early faults. In the literature there is a vast research about different wavelet transforms and their applications in fault detection either for gears or rolling element bearings. May [21]; [22] and [23] serve as some recent examples of the interest that this matter raises.

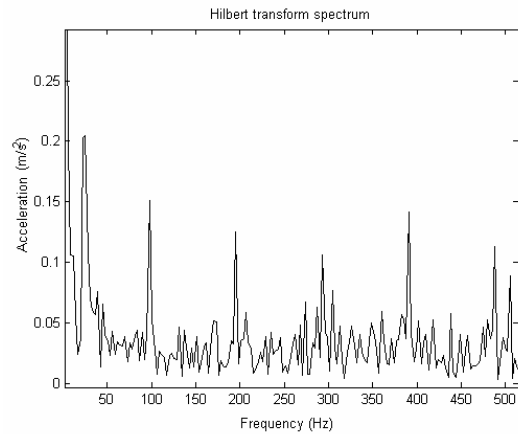


Fig. 7. Hilbert transform for outer race fault (shaft speed 19.3 Hz).

For vibration analysis, several wavelet techniques (DWT, wavelet packets, matching pursuit, CWT, etc.) are capable of enhancing the non-stationary features such as spikes and transients of vibration signals generated in rolling element bearings and gears. In many cases wavelet analysis is more effective than traditional techniques in the fault detection of these mechanical systems. In this case a continuous wavelet transform (CWT) has been used to perform the analysis:

$$\gamma(s, \tau) = \int f(t) \psi_{s,\tau}^*(t) dt \quad (2)$$

where * denotes complex conjugation.

The wavelet transform needs the choice of a mother wavelet, i. e., the basic function from which wavelets will be generated:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (3)$$

where s is the scale factor and τ is the translation factor [24].

We have chosen the Morlet function due to its simple analytic form and because, being a cosine function with exponential decay on both sides, it resembles an impulse (like fault vibrations do) and therefore it is effective for detecting periodic impulse signals:

$$\psi(x) = C e^{-x^2/2} \cos(5x) \quad (4)$$

where the constant C is used for normalization.

The wavelet transform used represented the fault but it does not give a clear diagnostic. The difficulty to

face when working with wavelets is to choose the most appropriate for the specific analysis, what is not always obvious.

Model	Model order
Inner race fault	45
Outer race fault	20
Rolling element fault	20
No fault	10

Table 2. Table borrowed from the paper of [17].

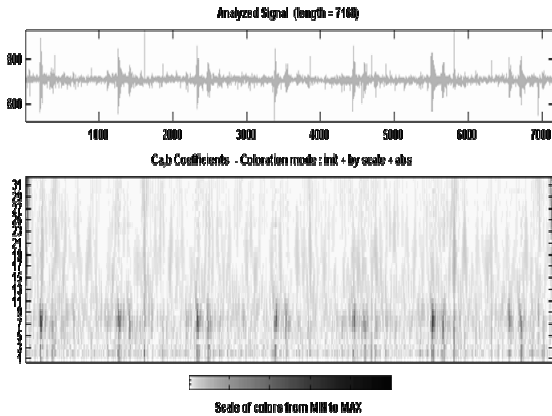


Fig. 8. Wavelet analysis for an inner race fault (shaft speed 19.3 Hz).

The second approach to be considered is model-based diagnosis. In particular, autoregressive (AR) modelling has been proved as an adequate technique for obtaining a prediction in rolling element bearings vibration signals [25]. Additionally, linear regression models allow fast calculation, which is convenient for real-time applications. Thus, a linear autoregressive process has been selected to build the model:

$$y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = v(t) \quad (5)$$

$$\theta^T = (a_1 \dots a_n)$$

where n is the model order, i. e., the number of past inputs required, and $v(t)$ is considered to be white noise. The model order can easily be determined experimentally. An important feature of this type of model is that the prediction is linear in the parameter vector θ , which makes its estimation simple [26].

We can express the linear regression model as a forward linear predictor, where \hat{y} denotes the estimated value for y :

$$\hat{y}(t) = -a_1 y(t-1) - \dots - a_n y(t-n) + v(t) \quad (6)$$

such that the sum of the squares of the errors is minimized:

$$e(n) = \hat{y}(t) - y(t) \quad (7)$$

If the estimation of the signal for the next step is correct, the prediction error will correspond to the value of the white noise. The calculation of the estimates of the model parameters is commonly made through the Yule-Walker equations, which can be solved by the Levinson-Durbin recursion algorithm.

The computation of the parameter estimated requires proportional to $2n^2$ operations [27]. As for the model order, the author in [17] defined the optimal linear autoregressive model orders as shown in Table II.

The process of diagnosing a fault consists in estimating the future state of the signal and comparing it to the original one. When a fault occurs the value will not be well predicted and the autocorrelation sequence of the prediction error will show a deviation.

4 Future work

There are several points to consider in a further research: to develop an actual fault at an early stage to compare the methods and check their performance in R-T and to extend the comparison to other methods. Finally, the implementation of these methods in a real system (such as a mechanical transmission in a machine) instead of a test bench will be the next stage in our research. Real systems' diagnosis and prognosis must be made in operation, which implies several R-T requirements, sensor limitations and noise dealing as the one caused by gears and other mechanical components on the signals measured for a determined component.

5 Conclusion

In this paper we have presented a comparison among different diagnostic methods from the point of view of detecting faults at an early stage and their implementation in real-time. According to the results presented, the Hilbert transform appears to be a technique with possibilities to be successfully implemented to detect early faults with an acceptable computation cost.

6 Acknowledgment

The authors would like to thank to the University of Mondragon, (Mondragon, Spain) for the use of their equipment. This work was partly supported by the Fundaci3n de Centros Tecnol3gicos Iñaki Goenaga under a Grant for postgraduate research.

References:

- [1] J. T. Renwick, "Vibration analysis- a proven technique as a predictive maintenance tool," IEEE Transaction on Industry Applications, vol. 21, pp. 324-332, 1985.
- [2] P. D: McFadden, and J. D. Smith, "Model for the Vibration Produced by a Single Point Defect in a

- Rolling Element Bearing," *Journal of Sound and Vibration*, vol. 96, no. 1, pp. 69-82, 1984.
- [3] P. D: McFadden, and J. D. Smith, "The vibration produced by multiple point defects in a rolling element bearing," *Journal of Sound and Vibration*, vol. 98, no. 2, pp. 69-82, 1985.
- [4] G. Lipovszky, K. Sólyomvári, and G. Varga, *Vibration testing of machines and their maintenance. Studies in mechanical engineering* 10, Elsevier, 1990.
- [5] R. B. Randall, *Frequency Analysis*. Brüel & Kjaer, 1987.
- [6] T. A. Harris, *Rolling bearing analysis*. 3rd. Edition. Wiley-interscience, 1991.
- [7] D. Ho, and R. B. Randall, "Optimization of bearing diagnostics techniques using simulated and actual bearing fault signals," *Mechanical systems and signal processing*, vol. 14, no. 5, pp. 763-788, 2000.
- [8] J. Antoni, and R. B. Randall, "Differential diagnosis of gear and bearing faults," *Journal of Vibration and Acoustics*, vol. 124, no. 2, pp. 165-171, 2002.
- [9] R. B. Randall, J. Antoni, and S. Chobsaard, "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals," *Mechanical Systems and Signal Processing*, vol. 15, no. 5, pp. 945-962, 2001.
- [10] A. C. McCormick, and A. K. Nandi, "Cyclostationarity in rotating machine vibrations," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 225-242, 1998.
- [11] B. Li, M. Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-Network-Based Motor Rolling Bearing Fault Diagnosis," *IEEE Transactions on industrial electronics*, vol. 47, no. 5, 2000.
- [12] A. Ypma, R. Ligteringen, E. E. E. Frietman, and R. P. W. Duin, "Recognition of bearing failures using wavelets and neural networks," in *Proceedings of 2nd UK Symposium on Applications of Time-frequency and time-scale methods TFTS97*, Coventry (UK), August 27 - 29, 1997.
- [13] D. F. Shi, W. J. Wang, and L. S: Qu, "Defect detection for bearings using envelope spectra of wavelet transform," *Journal of Vibration and Acoustics*, vol. 126, no. 4, pp. 567-573, 2004.
- [14] W. Wang, "An Evaluation of Some Emerging Techniques for Gear Fault Detection," *Structural Health Monitoring*, vol. 2, no. 3, pp 225-242, 2003.
- [15] J. Antoni, and R. B. Randall, "A stochastic model for simulation and diagnostics of rolling element bearings with localized faults," *Journal of vibration and acoustics*, vol. 125, July 2003.
- [16] R. B. Randall, "State of the Art in monitoring rotating machinery," in *Proceedings of ISMA*, vol. 4, pp. 1457-1477, 2002.
- [17] D. C: Baillie, and J. Mathew, "A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 10, no. 1, pp. 1-17, 1996.
- [18] C. Begg,, T. Merdes, C. S. Byington, and K. P: Maynard, "Dynamics modeling for mechanical fault diagnosis and prognostics," presented at *Maintenance and Reliability Conference (MARCON 99)*, Gatlinburg, Tennessee, May 10-12, 1999.
- [19] W. H: Press, S. A. Teuklosky, W. T. Vetterling, and B. P: Flannery, *Numerical recipes in C: The art of scientific computing*, 2nd edition, Cambridge University Press, 1992.
- [20] G. Y. Luo, D. Osypiw, and M. Irle, "On-line vibration analysis with fast wavelet algorithm for condition monitoring of bearing," *Journal of vibration and control*, vol. 9, no. 8, pp. 931-947, 2003.
- [21] D. F. Shi, W. J. Wang, and L. S. Qu, "Defect detection for bearings using envelope spectra of wavelet transform," *Journal of vibration and acoustics*, vol. 126, no. 4, pp. 567-573, 1994.
- [22] P. Tse, Y. H: Peng, and R. Yam, "Wavelet analysis and envelop detection for rolling element bearing fault diagnosis- Their effectiveness and flexibility," in *Transactions of the ASME: Journal of Vibration and Acoustics*, vol. 123, no. 3, pp. 303-310, 2001.
- [23] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjöberg, and J.-O. Stromberg, "Towards automatic detection of local bearing defects in rotating machines," *Mechanical Systems and Signal Processing*, vol. 19, no. 3, pp. 509-535, 2005.
- [24] C. Valens. (1999). A really friendly guide to wavelets [Online]. Available: <http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html>
- [25] C. K. Mechefske, and J. Mathew, "Fault detection and diagnosis in low speed rolling element bearings. Part I: The use of parametric spectra," *Mechanical systems and signal processing*, vol. 6, pp. 297-307, 1992.
- [26] L. Ljung L., and T: Söderström, *Theory and practice of recursive identification*, The MIT Press, Cambridge, Massachusetts, 1987.
- [27] L. Ljung, *System identification: Theory for the user*, Prentice-Hall, 1987.