Bearing defect detection and diagnosis using a time encoded signal processing and pattern recognition method

S Abdusslam, P Raharjo, F Gu and A Ball

University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

E-mail: Shukri Abdusslam@hud.ac.uk

Abstract. Many new bearing monitoring and diagnosis methods have been explored in the last two decades to provide a technique that is capable of picking up an incipient bearing fault. Vibration analysis is a commonly used condition monitoring technique in world industry and has proved an effective method for rolling bearing monitoring systems. The focus of this paper is to combine two conventional methods: wavelet transform and envelope analysis with the Time Encoded Signal Processing and Recognition (TESPAR) to develop a better technique for detection of small bearing faults. Results show that TESPAR with these two combinations provides good fault discrimination in terms of location and severity for different bearing conditions.

Keywords: Vibration, condition monitoring, bearing fault, TESPAR, envelope, wavelet.

1. Introduction

The potential for machine Condition Monitoring (CM) to improve system performance and prevent harmful failures has been rising considerably. Various novel CM techniques have been suggested in recent years in order to produce powerful analysis techniques capable of precisely distinguishing bearing conditions.

In bearing CM, vibration based envelope spectrum and wavelet transforms are now widespread for detecting bearing defects. Nevertheless, these techniques are complex analytically and require a large memory space to accommodate the sampled data, and a high power computer necessity to perform the analyses. These demands make it costly for this technique to be realised in an intelligent sensor node which today typically has very limited computational capability, storage and power resources [1].

TESPAR describes signal waveforms in regard to its shapes by coding the signals and putting them in different kinds of representation such as S and A matrices [2].

The paper compares two different techniques resulting from combining TESPAR with two conventional methods, envelope spectrum and the wavelet analysis, and reports an experimental study of bearing faults detection using TESPAR in combination with wavelet signals and envelope signals to assess the capability of two new combination methods in detecting fault locations.

The paper also evaluates the performance of the two methods for roller element bearings with three different kinds of faults. It is revealed that, using bearing vibration signals, both the combinations of TESPAR with envelope signal and the TESPAR with wavelet signal are proven sufficient for bearing conditions discrimination. TESPAR both combinations can classify results accurately and differentiate bearing defects.
2. Envelope analysis

Envelope analysis is a common technique for bearing faults detection [3]. Suppose the resonance frequency is a continuous pure sine and the impulses have the form of a sine wave of much lower frequency, the signal detected will be a modulated signal which has the form of the individual signals multiplied together, as shown in figure 1.

![Figure 1. Illustrative example of amplitude modulation [3].](image)

Basically, envelope analysis extracts the time history of the signals modulating the amplitude of the measured vibration signal. This is very useful for the class of problems where faults have an amplitude modulating effect on the characteristic frequencies of the bearing. Envelope analysis requires taking the FFT of the modulating signal. Different bearing defects produce different patterns of the envelope spectrum and so this is a suitable method for successfully detecting rolling element faults [4].

Figure 2 shows a schematic illustration of the application of envelope spectrum. First the measured signal, which contains low frequency noise, is passed through a band pass filter to attenuate both the low frequency noise and other unwanted noise occurring outside the frequency band of interest. Next, the band pass signal is “rectified” to produce the “absolute waveform”. This is then processed using a low-pass filter to extract the low frequency envelope waveform. Finally, the FFT is applied to the envelope waveform to obtain its spectrum.
In reality, the measured signal may be so noisy that even after passing through the band pass filter the signal may still contain undesirable elements which increase the difficulty of identifying information in the envelope spectrum. Also, the results of the envelope analysis depend on the frequency band chosen. In envelope analysis for bearing CM, the high resonant frequency technique (HRFT) uses band pass filters selected to include the frequency band where the defect signal is relatively strong [5, 6, 7]. However, this band may not be known initially and also may change as the bearing operating condition change. Envelope analysis is likely to be more successful if the operator has prior knowledge of the carrier frequencies before selecting the band pass filters. Additional problems can arise when a number of spectra are summed; and components at the same frequencies may interfere giving either cancellation or reinforcement, depending on their phase differences.

3. Wavelet Transform

The Fourier transform of a signal contains no information on when a particular event occurred. This is not a problem for stationary signals which repeat themselves, but can be important when investigating the development of faults where transitory phenomena often constitute the most significant part of the signal.

In an effort to correct this deficiency, Gabor [8] adapted the Fourier transform to analyze only a small section of the signal at a time - a technique called windowing the signal. Essentially Gabor divided the time-domain signal into a series of contiguous sections and the signal contained in each window is then transformed into the frequency domain. Mathematically the process is not so simple, as the window is a multiplying function which must be tailored to avoid introducing spurious results. Today a large number of such windows are available and most have the property that the initial and final values are close to zero so that the transform does not see the time domain signal as a series of step functions. This process is known as the Short-Time Fourier Transform (STFT) and produces a series of spectra, one for each window. As the duration of each window is known the spectra can be layered into a three dimensional plot (frequency, time and amplitude).

The two drawbacks with this method are: first, once a particular window size has been selected it cannot then be changed, but many signals can require more precision in determining frequency at some times than at others. Second, the uncertainty principle means that good resolutions in both time and frequency-domains cannot be achieved simultaneously [9]. For example, if the signal to be analyzed is of short duration, obviously a narrow window should be selected, but the narrower the window the wider the associated frequency band and the poorer the frequency resolution.

The logical step was to develop a transform procedure which incorporated variable-sized windowing. Wavelet analysis allows the use of windows of longer duration when more precise frequency information is needed, and shorter duration when more precise time information is required. Because the term frequency is reserved for the Fourier transform we do not speak about time-
frequency domains when discussing wavelets, instead we speak about time-scale representations where scale can be thought of as an inverse function of frequency.

The use of wavelet analysis is attracting considerable attention as a tool for diagnosis of the condition of bearings. Unlike the Fourier transform which expresses the time-domain signal in terms of sine and cosine functions, wavelet analysis expresses a signal over the whole spectrum in terms of wavelet functions of different scales which makes it more useful for the extraction of transient features contained in a signal. Because wavelet analysis uses wavelets of variable scales it can examine the entire spectrum when extracting a defect signal, and so does not require a signal with a dominant frequency band as is needed for frequency-domain filtering.

The windows used with wavelet transforms have the extremely important and useful feature that the width of the window changes (is scaled) as the transform is determined for each spectral component.

4. TESPAR Approach

TESPAR is a digital language that originated as a means of coding signals for speech recognition [10]. TESPAR depicts signal waveforms according to its real and complex zeros based on a mathematical waveforms representation [2] which is different from conventional CM techniques.

TESPAR quantization procedure has been developed to code signals according to the period between two consecutive zero-crossings and the shape of the curve thus contained [11]. This period is named an epoch. Every epoch can be illustrated by two parameters: D (duration - number of samples in the epoch) and S (shape – determined by the number of minima or maxima contained in the epoch) [12]. Figure 3 shows an epoch encoded into its TESPAR parameters where D=17 and S=1.

![Figure 3. TESPAR: single epoch with D = 17, S = 1.](image)

Most signal waveforms can be coded into a limited sequence of numerical descriptors known as the TESPAR symbol stream [13], normally from 1 to 28. In fact 28 symbols have been found to be sufficient to describe most signals adequately [13, 14]. The symbol sequence can be characterized in two ways: a one-dimensional “S-Matrix” vector or two-dimensional vector which is named the “A-Matrix” [14].

The S-matrix can be defined as the TESPAR symbols that record the number of times each TESPAR alphabet symbol occurs in the TESPAR symbol stream, and the A-matrix can be defined as a two-dimensional 28x28 vector matrix that records the number of times each pair of symbols in the alphabet appears in the symbol stream. The A-matrix has the important characteristic of expressing the temporal relationship between pairs of symbols [12] and because parameter n represents the delay between symbols it also provides frequency information. Typically, slowly oscillating patterns have n > 10 while higher frequency patterns have n < 10 [15].
5. Test Rig Facilities

The test rig is displayed in figure 4. The drive power is from a 4.0 kW, 3-phase, 4 pole electric inductions motor. The speed and torque of the motor are controlled by a Siemens Micro Master Controller so the drive shaft can be run at different speeds (to a maximum of 1420 rpm) with different applied torque loads to a maximum of 4.0 kW. Two pairs of matched flexible couplings couple the motor to the brake via three short cylindrical steel shafts. There are two bearing housings to support the shafts, the housing nearer the motor contains a MAC cylindrical roller bearing type N 406 and the other contains an SKF double row self-aligning spherical roller bearing type 22208 EK.

A radial load is applied to the central shaft using a hand pump with a pressure gate to pressurise a hydraulic ram frame, which is connected to a load cell. One Sinocera YD-5 piezoelectric accelerometer with frequency range up to 10 kHz is positioned on the housing of the N 406 bearing to measure the vibration.

Figure 4. Test rig construction.

A data acquisition system was designed to monitor and record vibration, temperature, and other parameters of interest, but it also contains some basic data analysis tools including spectrum analysis for online data examination. A Sinocera type YE6232B, 16 channel, 16 bit data acquisition system was used, see figure 5.

The 16-channel high speed data acquisition system recorded all the measurements at a sampling rate of 96 kHz. The hardware consisted of three parts: the piezoelectric accelerometer model YD-5, see figure 6, which was connected via a charge amplifier to a data acquisition card which was then connected to a PC.
The amplifier is necessary to amplify the vibration signal, which is often very weak, and it also electronically isolates the accelerometer from the processing and display equipment. The software is data acquisition logic and the analysis software (with other utilities that can be used to access and analyse data from the data acquisition memory of the computer). The data acquisition system specification is shown in Table 1.

### Table 1. Data Acquisition System (DAS) Specification.

| DAQ system | Specification                  |
|------------|--------------------------------|
| Manufacturer | Sinocera YE6232B               |
| Number of channels | 16 channels, selectable voltage/IEPE input |
| A/D conversion resolution | 16 bit                          |
| Sampling rate (maximum) | 100 kHz per channel parallel sampling |
| Input range | ±5V                             |
| Gain        | Selectable either 1, 10 or 100 |
| Filter      | Anti-aliasing filter            |
| Interface   | USB 2.0                         |

A data acquisition control programme developed for Lab-Windows was used during tests and it consists of a main data inspection panel and parameters set-up panel. This software was based on a Windows operating system and had the capability to carry out on-line data sampling. Modifications such as the number of channel, sampling frequency, data length and filenames can be chosen, recorded and stored on separate set-up page of the software package.

### 6. Envelope spectrum and wavelet transforms analyses

To benchmark the TESPAR, two conventional techniques, the envelope and the wavelet analysis, have been used for the four bearing cases. Spectra obtained from these analyses provide base lines against which to measure the effectiveness of TESPAR.

In calculating the envelope spectrum, signals were filtered using a frequency band from 8 kHz to 15 kHz, in which a structural resonant was found to enhance the signal greatly.

For wavelet analysis, there are many different kinds of wavelets that can be selected. In this paper the wavelet function of 4th order of Daubechies was chosen because it gives the best match to bearing defects signal. In addition, the wavelet coefficient at scale 7 was used so that it matches the frequency range in envelop analysis. Figures 7 and 8 show the envelope spectrum and the wavelet transform for the four bearings.
Figure 7. Envelope spectrums for three different small bearings faults.

Figure 8. Daubechie 4th wavelet for three different bearings faults.
When a bearing has no fault it still produces small amplitude vibrations but the characteristic frequencies are be seen in envelope spectrum or wavelet transform, see figures 7 and 8 respectively. This can be illustrated by the blue solid line in figure 7(a) or red dashed line in figure 8(a).

Compared with the faulty case in the same plots, the envelope spectrum and the wavelet transform from the healthy bearing are very flat, i.e. no clear spectral lines can be seen. Figures 7(b) and 8(b) are the spectra for a small outer race fault and the characteristic frequency is identified at 83.3 Hz (plus harmonics). Furthermore, in the same figure, patterns (c) and (d), which represent roller and inner race faults respectively, the characteristic roller and inner race faults frequencies can be seen and identified at 48.3 Hz and 134.4 Hz respectively.

These two spectra suggest that the fault can be diagnosed by corresponding characteristic frequencies. Moreover, it confirms that the signals under investigation did contain sufficient information for bearing monitoring and hence were prepared for use with the TESPAR coding.

7. TESPAR with Wavelet Signal
The wavelet signal for each case was encoded into its TESPAR symbol stream, and then S and A-Matrices were constructed for detection and evaluation of fault location.

Figure 9(a) shows the S-Matrices of wavelet signal for the healthy bearing and three different bearing faults seeded into the outer race, roller and inner race respectively. It can be seen that most of signal content appear in the symbol between 2 and 6, especially, the highest amplitude at Symbol 2. This may show that the structure of the signal is relatively simple. To the eye, visually all symbols display similar amplitudes for the four cases, discrimination is then difficult in this case.

![S-Matrices from Wavelet Signal](image)

(a) S-Matrices from Wavelet Signal

![S-Matrices from Wavelet Signal](image)

(b) S-Matrices from Wavelet Signal

Figure 9. Wavelet S-Matrices for a healthy and three different bearing faults.

However, as shown in figure 9(b) comparing the normalised amplitudes (obtained by dividing all values by the healthy value) between four cases at each symbol exhibit better image, only at Symbol 8 a clear difference can be observed and that may be a basis for assessing differences between the four cases.
When the changing percentage of symbols in term of amplitude is considered as shown in figure 9(b), much clearer image is obtained for symbol 8 as a key symbol that makes the difference, whilst the remaining symbols show less difference between the four cases.

The amplitude for the baseline is the minimum and fault detection can be done based on amplitude of this Symbol 8. Moreover, it has found that only one symbol in wavelet S-Matrices makes difference to differentiate the faults as shown already in figure 9(b).

Figure 10 shows the A-Matrices of wavelet signal for the healthy bearing and three different bearing faults. By looking at the four patterns, there are obviously completely different.

The outer race pattern 10(b) exhibits few peaks in the positive side of the scale with a prominent peak rising to 0.02; however, the roller fault pattern 10(c) reveals more peaks in the positive side while the biggest peak is exhibited in the negative side. The inner race fault case as shown in 10(d), also displays a different pattern, more and larger peaks in the positive side and one large peak on the negative side. These three-dimensional patterns can be used to completely discriminate the faults.

8. TESPAR with Envelope Signal

Figure 11 shows the S-Matrices of the envelope signal and the change in normalised values, expressed as a percentage of the envelope signal symbols for the healthy bearing and for three bearing faults seeded into the outer race, roller element and inner race of three similar bearings.
Figure 11. Envelope signal S-Matrices for a healthy and three different bearing faults.

As shown in figure 11(a) the healthy bearing is defined by its difference from the faulty conditions for symbols 2, 6 and 8, symbol 2 shows the most difference between the healthy and faulty bearings while the symbols for the faulty bearings have close amplitudes perhaps except symbol 3.

Figure 11(b) reveals the symbols’ normalization to the healthy case, symbol 8 exhibits the most difference between the four cases, and in addition, the faulty bearings in this symbol show the biggest difference comparing with the healthy case.

The A-Matrices shown in figure 10 demonstrate that the separation between the healthy condition and the three faults is clear. The changes in peak positions and amplitudes with the introduction of the faults can clearly discriminate all conditions. With the faulty cases there are statistically significant changes peak amplitudes indicating that a fault is present. The differences in the 3-D A- matrix patterns are terms of peak positions and magnitudes for the bearing faults investigated here offer an excellent method for fault location.

The TESPAR combination with envelope signal is a new practice that offers a real chance to discriminate between healthy and faulty bearings with identification of fault locations.
9. Conclusion
In this paper TESPAR was combined with two established signal processing methods: wavelet transform and envelope analysis in an effort to produce a powerful combination that in common can distinguish faults in bearing components earlier than separately. To assess the capability and competence of these novels combination methods, their results were compared with those attained using the envelope spectrum and the wavelet analysis alone.

When TESPAR was applied to encoding wavelet filtered data the resultant A-Matrices allow complete differentiation between the healthy and faulty bearings and between different fault cases. The S-Matrices are capable of discrimination between bearing conditions by using just one symbol.

However, the TESPAR and envelope signal combination also proved to be sensitive to bearing faults. The A-Matrices also exhibit good diagnosis performance due to its clear different features from the healthy bearing and between different bearing faults. The S-Matrices have three symbols, each being able to separate the faulty bearings.

References
[1] George, MH. 2007 TESPAR Paves the Way to Smart Sensor Review MCB University Press 17
[2] King, R. and Gosling, W. 1978 Electronic Letters, 14 456–457
[3] Sawahli, N. and Randall, R. 2007 Simulation of vibrations produced by localized faults in rolling elements of bearings in gearboxes 5th Australasian Congress on Applied Mechanics ACAM 10-12 December Brisbane Australia
[4] Konstantin-Hansen, H. 2003 Envelope analysis for diagnostics of local faults in rolling element bearings B&Q Application Note BO 0501 02/03
[5] Burgess, P. 1988 Anti-friction bearing fault detection using envelope detection Trans. Professional Engineers Institution Vol 15 No 2 pp77-82
[6] Friedman, A. 1997 Demodulation an essential tool for vibration analysis Shock and Vibration
[7] Toersen, H. 1998 Application of an envelope technique in the detection of ball bearing defects in a laboratory experiment TriboTest Vol 4 No 3 pp 297-308
[8] Gabor, D. 1946 Theory of communication Proc IEE 93 pp 429-457
[9] Lee, J. and Kim, H. 2001 Development of enhanced Wigner-Ville distribution function Mechanical Systems and Signal Processing 15 2 pp 367-398
[10] Rodwell, G. and King, R. 1996 TESPAR/FANN Architectures for low-power low cost Condition Monitoring Applications Proceedings of COMADEM 96 Sheffield
[11] King, R. and Phipps, T. 1999 TESPAR and approximation strategies ICSPAT 98 Vol 18 pp 445-453 Great Britain
[12] Licklider, J. Pollack, I. 1948 Effects of differentiation Integration and infinite peak clipping upon the intelligibility of speech Journal of the Acoustical Society of America Vol 20 No 1 pp 42-51
[13] Hagan, M. Demuth, H.and Beale, M. 1995 Neural network design. International Thomson Publishing
[14] Yang, S. Er, J. and Gao, Y. 2001 A high performance neural network based speech recognition system Proceeding of International Joint Conference on Neural Networks Vol 2 pp 1527
[15] Rabiner, L. and Juang, B. 1993 Fundamentals of speech recognition. Prentice Hall