A Wavelet Spectrum Technique for Machinery Fault Diagnosis

Derek Kanneg¹, Wilson Wang²

¹eMech Systems Inc., Thunder Bay, Canada; ²Department of Mechanical Engineering, Lakehead University, Thunder Bay, Canada. Email: Derek@emechsys.com, wilson.wang@lakeheadu.ca

Received June 13th, 2011; revised August 3rd, 2011; accepted August 12th, 2011.

ABSTRACT

Rotary machines are widely used in various applications. A reliable machinery fault detection technique is critically needed in industries to prevent the machinery system’s performance degradation, malfunction, or even catastrophic failures. The challenge for reliable fault diagnosis is related to the analysis of non-stationary features. In this paper, a wavelet spectrum (WS) technique is proposed to tackle the challenge of feature extraction from these non-stationary signatures; this work will focus on fault detection in rolling element bearings. The vibration signatures are first analyzed by a wavelet transform to demodulate representative features; the periodic features are then enhanced by cross-correlating the resulting wavelet coefficient functions over several contributive neighboring wavelet bands. The effectiveness of the proposed technique is examined by experimental tests corresponding to different bearing conditions. Test results show that the developed WS technique is an effective signal processing approach for non-stationary feature extraction and analysis, and it can be applied effectively for bearing fault detection.

Keywords: Machinery Condition Monitoring, Rotary Machines, Bearing Fault Detection, Non-Stationary Signal, Wavelet Transform, Resonance Feature

1. Introduction

Rolling element bearings are widely used in rotary machinery. A reliable bearing fault diagnostic technique is critically needed in a wide array of industries to prevent machinery performance degradation, malfunction, or even catastrophic failures [1]. Bearing condition monitoring usually involves two sequential processes: feature extraction and fault diagnosis [2]. Feature extraction is a process in which health condition related features are extracted by appropriate signal processing techniques, whereas fault diagnosis is a decision-making process to estimate bearing health conditions based on the extracted representative features. Therefore, feature extraction plays the key role for bearing health condition monitoring, whereas non-robust features may lead to false alarms (i.e., an alarm is triggered by some noise instead of a real bearing fault) or missed alarms (i.e., the monitoring tool cannot recognize the existence of a bearing defect) in diagnostic operations [3].

Several techniques have been proposed in the literature for bearing fault-related feature extraction, in which the analysis can be performed in the time domain, the frequency domain, or the time-frequency domain [4-6]. In time-domain analysis, for example, a bearing fault is detected by monitoring the variation of some statistical indices such as root-mean-square value, crest factor or kurtosis. A bearing is believed to be damaged if the monitoring indices exceed predetermined thresholds; however, it is usually challenging to determine robust thresholds in real-world applications. Frequency-domain analysis is based on the transformed signal in the frequency domain. The advantage of frequency-domain analysis over time-domain analysis is its capability to easily identify and isolate certain spectral components of interest [7]. Bearing health conditions are assessed by examining the fault related characteristic frequency components in a spectrum or in some extended spectral expressions such as bispectrum or cepstrum maps [8,9]. Frequency-based techniques are usually supplemented with certain signal analysis methods to enhance representative spectral components, which include frequency filters, envelope analysis, and modulation sidebands analysis [10]. Frequency-domain techniques, however, are not suitable for the analysis of non-stationary signatures that are generally related to machinery defects. Non-stationary or transient signatures can be analyzed by applying time-frequency do-
main techniques such as the short-time Fourier transform (FT) [11], Wigner-Ville distribution [12], spectral kurtosis [13], cyclostationary analysis [14], or wavelet transform (WT) [15]. In bearing fault diagnosis, the WT is a favorite technique, because it does not contain such cross terms as those in the Wigner-Ville transform, while it can provide a more flexible multi-resolution solution than the short-time FT. According to signal decomposition paradigms, the WT can be classified as the continuous WT, discrete WT, wavelet packet analysis, and those WT with post-processing schemes [16-19].

If a bearing is damaged, the generated vibration signatures could be either stationary or non-stationary. It is relatively easier to analyze the stationary signatures using some classical fault detection techniques [20]. However, it still remains a challenging task to extract robust representative features from the non-stationary vibration signals (e.g., those generated from a fault on bearing rotating components), particularly in real-world industrial applications. This is because: 1) a bearing is a system instead of a simple mechanical component, which consists of inner/outer rings as well as a number of rolling elements; 2) slippage often occurs between the rolling elements and rings in operations; and 3) the machinery operation conditions are usually noisy. Correspondingly, the objective of this paper is to develop a wavelet spectrum (WS) technique to tackle this challenge in which the representative periodic features will be enhanced by an integration process over several contributive wavelet bands.

2. The Wavelet Spectrum (WS) Technique

Whenever a fault occurs on a bearing component, impacts are generated in operation, which in turn excite the bearing and its support structures. The resulting resonance signatures are usually amplitude modulated by the bearing defect [2]; therefore, the analysis of these resonance signatures plays a key role in vibration-based bearing fault detection. Figure 1(a) shows part of a typical acceleration signal, measured from the housing of a tested bearing with an inner-race defect when the shaft speed \( f_t = 35 \text{ Hz} \). When a defect occurs on a bearing rotating component, the modes and magnitudes of the resulting resonances often vary over time due to the variation in angular position of the impacts [20]; this non-stationary characteristic of condition-related signatures makes bearing fault detection still remain a very challenging task in both research and industrial applications. In this work, a WS technique is proposed to investigate the characteristics of these non-stationary resonance signatures for the purpose of bearing fault detection. The WS technique involves five steps for signal processing, as discussed as follows.

The first step is to apply the WT to demodulate the resonance vibration signatures, both stationary and non-stationary, over a series of wavelet bands. Given a continuous signal \( x(t) \), the wavelet coefficients are determined by

\[
W(t,s) = \int_{-\infty}^{\infty} x(\tau) \sqrt{s} w^*(s(\tau - t)) \, d\tau
\]

(1)

Where \( w^*(t) \) denotes the complex conjugation of mother wavelet function \( w(t) \); \( s \) and \( t \) are the scale and time variables, respectively, which produce dilation and translation [20]. The choice of an appropriate mother wavelet depends on the signal properties and the purpose of the analysis. By testing and comparison, Morlet wavelet is selected as the mother wavelet for the signal analysis in this work, which is a modulated Gaussian function:

\[
w(t) = \exp\left(-\frac{t^2}{2b_0^2}\right) \exp\left(j2\pi f_0 t\right)
\]

(2)

where \( b_0 \) is the spread of the Gaussian function and \( f_0 \) is the center frequency of the pass-band of the mother wavelet. As \( b_0 f_0 \) increases, the duration of the wavelet expands, and the time resolution will decrease correspondingly. As a result, the obtained mother wavelet \( w(t) \) may not be suitable to analyze fast-decaying transient signatures. To solve this problem, the product of the spread and the scaled center frequency is kept as a constant in this work, i.e.,

\[
b_i f_i = \frac{b_i}{s_i} = b_0 f_0 = \frac{1}{\sqrt{2 \ln 2}}
\]

(3)

where \( 2\pi b_i f_0 = \pi \sqrt{2/\ln 2} \) was given in [20]; \( s_i \) represents the \( i \)-th selected scale; \( b_i \) and \( f_i \) are the corresponding \( i \)-th spread and center frequency, respectively. Based on the relation between \( b_i \) and \( f_i \) as in Equation (3), the mean of the obtained mother wavelet \( w(t) \) will be kept less than \( 10^{-12} \) in this case, and the effective support will vary with the scaled center frequency to accommodate the variation of the signatures of interest. At each wavelet scale \( s_i \), the magnitude of wavelet coefficient function \( |W(t,s_i)| \) that represents the demodulated envelope signal is normalized by its standard deviation, that is

\[
\bar{W}(t,s_i) = \frac{|W(t,s_i)|}{\sqrt{\sum_{l=1}^{L} \left| W(t_l,s_i) \right|^2 - \frac{1}{L} \sum_{l=1}^{L} |W(t_l,s_i)|^2}} / (L-1)
\]

(4)

where \( l = 1, 2, \cdots, L \), and \( L \) is the total number of samples; \( i = 1, 2, \cdots, I \), and \( I \) is the number of wavelet scales; \( |W(t_l,s_i)| \) is the \( l \)-th sample of \( W(t,s_i) \). To reduce the
interference effects from the low-frequency noisy components, in this work, the overall frequency band of interest is chosen as \([Zf_r, f_s/2.56]\), where \(f_s\) denotes the shaft rotation speed, \(Z\) is the order of shaft harmonics, \(Zf_r\) represents the lower bound frequency for feature extraction (\(Z = 35\) is used in this case); \(f_s\) is the sampling frequency, and the constant 2.56 is selected to avoid aliasing effects. The centre frequencies of the wavelet should be deployed properly to implement the WT over this designated frequency band \([Zf_r, f_s/2.56]\), without the overlapping between the wavelet frequency bands. Based on the FT of the dilated wavelet \(w(st)\), the 3-dB bandwidth \(BW_i\) for the \(i\)th centre frequency \(f_i\) is derived as follows: \(BW_i = \left[1 - \lambda, 1 + \lambda\right] f_s\), where \(\lambda = \ln2/\sqrt{2\pi}\) is a constant. Beginning with the lower bound frequency \(Nf_i\), the centre frequencies \(f_i\) can be recursively calculated and positioned as:

\[
f_i = \left(\frac{1+\lambda}{1-\lambda}\right)^{-1} Zf_r, \quad i = 1, 2, \cdots, I - 1 \tag{5}
\]

\[
f_i = \frac{1}{2} \left[ f_i \frac{f_s}{2.56} + f_{i-1} (1+\lambda) \right], \quad i = I \tag{6}
\]

where \(I\) is the number of the wavelet scales (\(I = 7\) in this case). Figures 1(b)-(h) show the respective normalized wavelet coefficients \(\tilde{W}(t,s_i)\) over seven wavelet bands, which are determined based on the vibration signal as shown in Figure 1(a). It can be seen that the resonance signatures in Figure 1(a) are usually demonstrated in several consecutive wavelet bands (Figures 1(b)-(h)) due to the variation of the transient modes.

The second step to implement the proposed WS technique is to cross-correlate the wavelet coefficients from the neighboring wavelet bands to enhance the defect-related periodic features, that is,

\[
X_i(l) = E\left[ \tilde{W}(t,s_i) \tilde{W}^*(t+l,s_{i+1}) \right] \tag{7}
\]

\[
\overline{X}_i(l) = \frac{X_i(l) - \mu_i}{\sigma_i} \tag{8}
\]

where \(E[\cdot]\) denotes the expectation function; \(X_i(l)\) are the cross-correlation sequences that are normalized by their standard deviation \(\sigma_i\) around the mean \(\mu_i\). In estimating the correlation sequence, the method adopted here is slightly different from the commonly used Pearson’s approach. Pearson product-moment correlation estimation limits \(\overline{X}_i(l)\) to the range of \([-1,1]\) whereas the proposed one can prevent such a restriction. It is also noted that the cross-correlations are performed on the neighboring wavelet bands; this is because the demodulated features from the resonance signatures are usually reflected on the adjacent wavelet bands. An example is illustrated in Figure 1 where the extracted features are marked by arrows. From a physical perspective, each time as a bearing incipient fault encounters its mating components, an impact is generated, which in turn induces the resonance of the local structure. Corresponding to each impulse, the resonant response usually occurs over consecutive frequency bands in a random nature.

Figures 2(a)-(f) illustrate the \(\overline{X}_i(l)\) array determined from six pairs of neighboring wavelet bands. It is seen that some periodic features are prominent (e.g., in Figures 2(b), (e), and (f)) whereas others are less pronounced (e.g., in Figures 2(a), (c), and (d)). Correspondingly, another key process in bearing incipient fault detection is how to properly choose the more contributive wavelet bands to integrate cross-correlation coefficient functions to highlight the periodic features.

Each periodic feature with high amplitudes will modify the distribution of correlation sequence and cause the distribution more skewed and/or tailed, which could be detected by the Jarque-Bera (JB) statistic [21,22]. In this work, the correlation coefficient from each pair of neigh-
Securing wavelet bands is treated as a discrete random variable, and its probability distribution is examined. As an example, Figures 2(a’)-(f’) show the probability distributions of the correlation sequences in the corresponding Figures 2(a)-(f). It is seen that the properties of the tails of the distribution function vary with respect to the bandwidth. To characterize this effect, a JB statistic-based performance index \( J_i \) is proposed as:

\[
J_i = \frac{1}{6} \left[ S_i^2 + \frac{(K_i - 3)^2}{4} \right]
\]

(9)

where \( S_i \) and \( K_i \) are, respectively, the skewness and kurtosis that are estimated by using a large number of samples \( (L = 327,680 \text{ in this case}) \). In bearing fault detection, a larger \( J_i \) is expected when the bearing is faulty, since it indicates that the periodic features are highlighted (i.e., with higher magnitudes). Accordingly, the third step in the implementation of the WS technique is to choose more contributive bandwidths in which the correlation sequences could bring about larger \( J_i \).

The fourth step of the proposed WS technique is to integrate the correlation coefficients from the contributive bandwidths to achieve a 1-D feature representation. In this work, a \( J \)-weighted function is suggested for the integration process:

\[
H(l) = \frac{\sum_{m \in C} J_i R_i(l)}{\sum_{m \in C} J_i}
\]

(10)

where \( C \) is a subset of \( \{1,2,\cdots,J-1\} \). The selection of the members of \( C \) depends on applications; in this case, \( C \) takes the top half of the members of \( \{1,2,\cdots,J-1\} \) whose corresponding correlation sequences generate greater \( J_i \). Figure 3(a) shows some examples of the integrated correlation sequence \( H(l) \) derived using Equation (10). It is seen that the periodic features, carried by the vibration signal in Figure 1(a), can be clearly recognized. These periodic features are spaced by an interval of 118 samples, or with the repeat rate of approximately 173 Hz (i.e., the inner-race defect frequency) for a 20480 Hz sampling frequency.

Once the integrated cross correlation sequences are obtained, the fifth and final step is to examine the characteristic defect frequencies (i.e., the inner race defect frequency \( f_{id} \), the outer race defect frequency \( f_{od} \), and the rolling element defect frequency \( f_{od} \) [8]) by constructing the averaged autocorrelation spectrum. This autocorrelation spectrum analysis involves two processes [20]: performing the autocorrelation on \( H(l) \) to further enhance the involved periodic features, and conducting the spectral analysis (FT) for periodic feature extraction. Specifically,

\[
r(\kappa) = E[H(l)H^*(l+\kappa)]
\]

(11)

\[
R(f) = F[r(\kappa)]
\]

(12)

\[
\Phi(f) = \Phi[R(f)]
\]

(13)

where \( F[\cdot] \) denotes the FT, \( \kappa = 0,1,2,\cdots,L-1 \).

In implementation, the spectra obtained by Equation (13) from \( P \) segments of measured signals (\( P = 5 \) in this case) should be normalized and then averaged to reduce the effects of random noise,

\[
\Phi_f(f) = \frac{1}{P} \sum_{p=1}^{P} \left\{ \Phi_p(f) \right\}
\]

(14)

where \( f_u \) is the observation upper-bound frequency that should be larger than the maximum bearing characteristic frequency, and \( f_u = 300 \text{ Hz in this case} \). Bearing health conditions are estimated by analyzing the related characteristic frequency components (i.e., \( f_{id} \), \( f_{od} \), and \( f_{od} \)) in the resulting spectra. Figure 3(b) shows the resulting spectra determined by applying the proposed WS technique on the vibration signal shown in Figure 1(a). It is seen that the defect frequency (approximately 173.17 Hz)
can be clearly detected; in this case, the defect occurs on the bearing’s inner race, that is, \( f_{id} = 173.17 \) Hz when the shaft speed \( f_s = 35 \) Hz.

3. Performance Validation

A number of tests have been conducted to verify the effectiveness of the proposed WS technique in bearing fault detection; the experimental setup that is employed for these tests is shown in Figure 4. The shaft is driven by a 3-hp induction motor. The motor speed ranges from 20 rpm to 4200 rpm, which is manipulated by a speed controller. An optical sensor is used to provide a one pulse per revolution signal for rotation speed detection and advanced signal processing applications. A flexible coupling is employed to damp out high-frequency vibrations generated by the motor. The rolling element bearing under examination is press-fitted into the left bearing housing, and the vibration signals are measured by two accelerometers (ICP-IMI, SN98697) installed on the housing along both the horizontal and vertical directions. Radial loads are applied by two pairs of disks. A data acquisition board (NI PCI-4472) is employed for signal collection.

In the tests, four bearing health conditions are examined: healthy bearings, bearings with outer race defects, bearings with inner race defects, and bearings with rolling element faults. Each bearing is tested under seven shaft speeds (900, 1200, 1500, 1800, 1920, 2100, and 2400 rpm) and two load levels, respectively. The sampling frequency \( f_s \) is set at 20480 Hz.

The performance of the proposed WS technique will be compared with two related classical methods, the one-scale WT [19] and the frequency-domain analysis using envelope demodulation and FT as the supplementary signal processing tools [9], to verify its effectiveness in non-stationary feature extraction and bearing incipient fault detection. In this comparison study, the classical methods are applied on both structural resonance frequency bands and optimally-selected frequency bands. In resonance frequency band investigation, the one-scale WT is employed with the wavelet center frequency at 2000 Hz; the frequency-domain analysis is conducted with the signal that is band-pass filtered around the resonant frequency [1500 2500] Hz of the bearing and housing [20]. The analysis of these two classical methods is also based on the corresponding averaged autocorrelation spectra.

In frequency-based analysis, the bearing fault is detected by checking if there exists a pronounced spectral component in the resulting spectra that corresponds to one of the bearing characteristic defect frequencies. If the frequency-based fault detection technique cannot enhance the bearing health condition-related spectral components (i.e., making them pronounced or dominant in the spectral maps), other supplementary methods, based on either time-domain or time-frequency-domain analysis, should be properly employed to improve the diagnostic accuracy [2,16]. In our investigation, it is found that when the bearing is in its normal condition, the shaft speed dominates the resulting spectra due to unavoidable imperfections (e.g., system unbalance). When an incipient bearing fault (i.e., inner race defect or outer race defect) occurs, the bearing characteristic defect frequency will become pronounced if the proposed WS technique is employed. The results from these examinations are summarized in Table 1, in which the numbers represent the percentages of successful bearing health condition estimation. From Table 1, it is seen that: 1) in general,
Table 1. Comparison of diagnostic results using different methods.

<table>
<thead>
<tr>
<th></th>
<th>One-scale WT</th>
<th>Frequency analysis</th>
<th>WS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy bearing</td>
<td>64.3%</td>
<td>71.4%</td>
<td>100%</td>
</tr>
<tr>
<td>Bearing with outer race defect</td>
<td>85.7%</td>
<td>85.7%</td>
<td>100%</td>
</tr>
<tr>
<td>Bearing with inner race defect</td>
<td>50.0%</td>
<td>57.1%</td>
<td>92.9%</td>
</tr>
</tbody>
</table>

the classical methods with entropy-based frequency band selection can be more reliable in detecting a bearing fault than those methods focusing only on structural resonance frequency band; 2) the proposed WS outperforms these two classical methods in terms of bearing fault diagnostic accuracy, no matter which frequency band is examined.

In the following context, the processing results from one testing case will be used, as an example, to compare the performance of different bearing fault detection techniques.

**Healthy Bearing:** As mentioned earlier, when the bearing is in its normal condition, the shaft speed dominates the resulting spectra due to some unavoidable shaft imperfections (e.g., unbalance) and the varying compliance. For example, Figures 5(a)-7(a) show the respective processing results from these three methods for a healthy bearing ($f_s = 35$ Hz). It is seen that the shaft speed can be clearly recognized by using the WS technique (Figure 5(a)). By contrast, the shaft speed information cannot be clearly identified by using the one-scale WT and the frequency-domain analysis (Figures 6(a) and 7(a)); instead, the third harmonic of the shaft speed dominates the resulting spectra (Figures 6(a) and 7(a)), which is in fact very close to the outer race defect frequency ($f_{od} = 106.83$ Hz) in this case.

### 3.1. Outer Race Fault Detection

Outer race fault detection is a relatively easy task because the ring is fixed and the defect-related resonance modes do not change dramatically. Table 1 illustrates that the proposed WS technique can detect the outer race faults in all test cases in which the outer race defect frequency dominates the resulting spectra; one example is shown in Figure 5(b) when $f_{od} = 106.83$ Hz. It is noted that the one-scale WT and the employed frequency-domain technique can also detect the outer race defects although in some test cases the harmonics of the outer race defect frequency dominate the resulting spectra, as seen in Figures 6(b) and 7(b).

### 3.2. Inner Race Fault Detection

The detection of a fault on an inner race is more challenging than on a fixed outer ring because the modes of the generated resonance signatures vary over time. Test results in Table 1 demonstrate that the WS technique is more reliable (e.g., Figure 3(b)) than the related classical methods (e.g., Figures 6(c) and 7(c)) in detecting bearing faults on rotating rings and in suppressing the noisy spectral components. This is because the WS technique is capable of integrating the periodic features from several contributive wavelet bands.
3.3. Rolling Element Fault Detection

The detection of a rolling element fault for ball bearings is one of the most challenging tasks in bearing health condition monitoring, especially when the fault is at its initial stage. This is because: a) the resonance signatures generated by a ball defect are non-stationary; and b) the impacts are random since the defect may not always strike the races. In our tests, it is found that the rolling element defect frequency can be detected as long as the shaft speed is sufficiently high (e.g., over 30 Hz in this test), although the related defect spectral component is not the dominant one in the resulting spectra. It is also seen that the defect frequency processed by using the WS technique (e.g., Figure 5(c)) is more prominent than those from the two classical methods (e.g., Figures 6(a)-7(d)).

4. Conclusions

A wavelet spectrum (WS) technique is proposed in this paper for representative feature extraction and bearing incipient fault detection. The WS technique performs feature extraction by demodulating the non-stationary resonance signatures generated by bearing incipient defects and then correlating the periodic patterns over more contributive wavelet bands. A Jarque-Bera statistic-based performance indicator is suggested to guide the wavelet band selection. The effectiveness of the proposed WS technique is verified by a series of experiments corresponding to different bearing conditions. Test results show that the WS technique is an effective approach for non-stationary feature extraction and bearing fault detection. It outperforms the related classical methods such as one-scale wavelet transform and the employed frequency-domain technique.

REFERENCES

A Wavelet Spectrum Technique for Machinery Fault Diagnosis

538-552. doi:10.1016/j.jsv.2007.01.035


