Wavelet Analysis and Envelope Detection For Rolling Element Bearing Fault Diagnosis—Their Effectiveness and Flexibilities

Peter W. Tse
e-mail: meptse@cityu.edu.hk
Mem. ASME

Y. H. Peng
Mem. ASME

Richard Yam

Smart Asset Management Laboratory,
Department of Manufacturing Engineering & Engineering Management,
City University of Hong Kong,
Tat Chee Ave., Hong Kong

The components which often fail in a rolling element bearing are the outer-race, the inner-race, the rollers, and the cage. Such failures generate a series of impact vibrations in short time intervals, which occur at Bearing Characteristic Frequencies (BCF). Since BCF contain very little energy, and are usually overwhelmed by noise and higher levels of macro-structural vibrations, they are difficult to find in their frequency spectra when using the common technique of Fast Fourier Transforms (FFT). Therefore, Envelope Detection (ED) is always used with FFT to identify faults occurring at the BCF. However, the computation of ED is complicated, and requires expensive equipment and experienced operators to process. This, coupled with the incapacity of FFT to detect nonstationary signals, makes wavelet analysis a popular alternative for machine fault diagnosis. Wavelet analysis provides multi-resolution in time-frequency distribution for easier detection of abnormal vibration signals. From the results of extensive experiments performed in a series of motor-pump driven systems, the methods of wavelet analysis and FFT with ED are proven to be efficient in detecting some types of bearing faults. Since wavelet analysis can detect both periodic and nonperiodic signals, it allows the machine operator to more easily detect the remaining types of bearing faults which are impossible by the method of FFT with ED. Hence, wavelet analysis is a better fault diagnostic tool for the practice in maintenance. [DOI: 10.1115/1.1379745]

1 Introduction

By reducing costs and shortening the time of repair, condition-based predictive maintenance becomes an efficient strategy for modern industry [1]. Since rolling element bearings are widely used in rotary machines, faults occurring in bearings must be detected as early as possible to avoid fatal breakdowns of machines that may lead to loss of production and human casualties. Therefore, the process of monitoring the operating conditions of machines for advanced warning of defects receives considerable attention in industrial maintenance [2]. Faults which typically occur in rolling element bearings are usually caused by localized defects in the outer-race, the inner-race, the rollers, or the cage [3]. Such defects generate a series of impact vibrations every time a running roller passes over the surfaces of the defects. These vibrations occur at Bearing Characteristic Frequencies (BCF), which are estimated based on the running speed of shaft, the geometry of the bearing, and the location of the defect [2,4]. Theoretically, using the conventional analysis technique, the Fast Fourier Transforms (FFT) spectra of BCF can be generated. Hence, by observing the change of vibration amplitude in these spectra, the machine operator should be able to detect the existence of faults, and diagnose their causes. Because the impact vibration generated by a bearing fault has relatively low energy, it is often overwhelmed by noise with higher energy and vibrations generated from other macrostructural components. Therefore, it is difficult to identify the bearing fault in the spectra of BCF using the conventional FFT method.

To allow for the easier detection of such faults, the Envelope Detection (ED) technique has been used together with FFT [5]. Although various procedures have been proposed for the use of ED in bearing fault diagnosis, the most common practice is to apply ED at the bearing resonance frequency located in the high frequency range [6]. The modulated amplitudes of repetitive impacts are often excited at the bearing structural resonance frequency. Hence, the amplitude demodulation provided by ED allows the detection of localized defects without the interference of vibrations generated by other macro-structural sources [7]. Because ED requires intensive computation, it is preferable to use ED hardware devices to make the process faster. Moreover, since the range of bearing resonance has to be known before the ED method can be applied, several runs of impact tests are needed to find the location of the bearing resonance frequency. Therefore, expensive hardware and an experienced operator are required. This, coupled with the incapacity of FFT to detect faults which exhibit nonstationary impact signals, makes the industry seeking for another alternative which is simpler to use without scarifying the effectiveness in fault diagnosis provided by FFT with ED.

To overcome the limitation of FFT in analyzing nonstationary signals, three types of analyses for nonstationary signals—the Short-time Fourier Transform (STFT) [8], the Wigner-Ville Distribution (WVD) [9], and wavelet analysis [10]—have recently been introduced. These analyses offer simultaneous representations in both time and frequency domains, which are essential for the analysis of nonstationary signals. Moreover, the ability to display multi-resolution in a time-frequency distribution diagram is important in vibration-based machine fault diagnosis. In the low frequency range, a fine resolution in frequency helps the operator to identify the macro-structural vibrations caused by misalignment and unbalance of shafts and couplers. Unfortunately, the micro-structural components of bearings that cause the impact vibrations are difficult to be distinguished in the low frequency range, due to the mashing of vibrations generated by other components. The usual practice is to view these micro-structural vibrations at the bearing resonance frequency located in the high frequency range. It is because the signal to noise ratio is much higher and the interference from other vibration sources can be avoided [6].

Contributed by the Technical Committee on Vibration and Sound for publication in the JOURNAL OF VIBRATION AND ACOUSTICS. Manuscript received April 2000; revised Mar. 2001. Associate Editor: G. T. Flowers.

Since different types of defective components will generate impact vibrations at different time intervals, a fine resolution in time in the high frequency range is essential to identify the particular type of defective component that is causing the impact vibration. Therefore, a fine frequency resolution in the low frequency range and a fine time resolution in the high frequency range are desirable for vibration-based machine fault diagnosis.

In STFT, the resolutions in the time and frequency domains are limited by the width of the analyzing window. Since the width of the window is the same through the entire STFT analysis, the resolutions in time and frequency are always constant. Hence, STFT is deemed unsuitable for fault diagnosis. WVD is a combination of FFT and auto-correlation calculation. It provides an energy distribution of the signal in both the time and frequency domains. Unfortunately, WVD may lead to the emergence of negative energy levels and spectrum aliasing [10]. Therefore, the results obtained by WVD are sometimes difficult to interpret. Wavelet analysis has been developed as an alternative time-frequency analysis for both stationary and nonstationary signals. Since wavelet analysis can provide multi-resolution in both time and frequency, it is uniquely suitable to detect bearing faults. The existence of bearing faults can be revealed by an increase of vibration energy in the low frequency range using a fine resolution in frequency, and the appearance of impacts in the high frequency range. Moreover, the causes of bearing faults can be identified in this high frequency range with fine resolution in time.

Recently, extensive research has applied wavelet analysis to the analysis of mechanical vibration signals. Dowling described the use of wavelet analysis and WVD in processing nonstationary signals for fault diagnosis of industrial machines [11]. He has briefly stated the inadequacies of FFT and suggested that wavelet analysis, due to its capacity for the analysis of nonstationary signals, is ideal in the detection of incipient signal changes that usually indicate failures in mechanical devices. However, no result was given on the detection of localized defects occurring in rotary machines. Newland proposed a method to measure the changing spectral composition of nonstationary signals using wavelet maps [12]. He has also developed the harmonic wavelet that provides box-like spectra to eliminate frequency aliasing and phase-locking function to make the identification of fault-related transient signals easier [13]. Debabrata used wavelet analysis in monitoring mechanical processes [14]. He suggested that the defects of a process could be identified according to the changes of various forms of wavelet coefficients. Several fault detection indices, such as the energy of the spectrum and the extreme autocorrelation, could be obtained using wavelet analysis. A case study has been presented on the detection of bearing failures through the significant changes of total energy and band energy using wavelet analysis. However, the method for identifying the cause of failure was not given. Mori and Kasashima et al. defined several failure indices using the discrete type of wavelet analysis for the predication of spalling in ball bearings from vibration signals [15]. Li and Ma used continuous wavelet analysis to detect localized bearing defects based on vibration signals [16]. They used an index to estimate the interval of bearing impacts, and identified the type of bearing faults based on calculation of the autocorrelation of wavelet coefficients. However, this method is too complicated for use in industry. Wang and McFadden [17] used wavelets to detect gearbox faults which exhibit modulated signals similar to those of running bearings. None of these studies have compared the efficiency of conventional FFT with ED to wavelet analysis. Without a real comparison, it is difficult to convince the industry that wavelet analysis is a more suitable tool for machine fault diagnosis. The intent of this paper is to perform such a comparative study and propose a more convenient method for the machine operator to use in bearing fault diagnosis.

The general procedures of using conventional FFT and ED in bearing fault diagnosis are described in Section 2, as are the experimental platform used for testing their capacities and limitations. Section 3 introduces wavelet analysis and the type of wavelet analysis that we propose to use for bearing fault diagnosis. In Section 4, a series of motor-pump driven systems are used to test the effectiveness of both FFT with ED and wavelet analysis. This is followed by the presentations of results and discussions. Finally, concluding remarks are given, and the benefits of using wavelet analysis for machine-condition based maintenance are described in Section 5.

2 Faults of Rolling Element Bearings

2.1 Fault Related Bearing Characteristic Frequencies. The types of defective components that typically cause rolling element bearing failure are the outer-race, the inner-race, the rollers, and the cage. In this study, we use the first three types of bearing defects for testing purposes. As mentioned in the previous section, when a bearing rotates, each type of defect will generate a particular frequency of impact vibration, which is a kind of BCF. By identifying the type of BCF that occurs, the cause of the defect can be determined [18,19]. The BCF of the first three types of bearing defects can be estimated by using the following three equations.

Outer-race defects are revealed at the Ball-Passing Frequency Outer-race (BPFO):

$$\text{BPFO (in Hz)} = \frac{n}{2} f_r \left(1 + \frac{d}{D} \cos \beta \right) \tag{1}$$

Inner-race defects are revealed at the Ball-Passing Frequency Inner-race (BPFI):

$$\text{BPFI (in Hz)} = \frac{n}{2} f_r \left(1 - \frac{d}{D} \cos \beta \right) \tag{2}$$

Rolling element defects are revealed at the Ball-Passing Frequency Roller (BPFR) or the Ball Spin Frequency (BSF):

$$\text{BPFR or BSF (in Hz)} = \frac{D}{d} f_r \left[1 - \left(\frac{d}{D} \cos \beta \right)^2 \right] \tag{3}$$

where f_r is the relative rotating frequency between inner and outer races, n is the number of rollers, β is the contact angle between the race and the roller, d is the roller diameter and D is the bearing's pitch diameter.

These equations are based on the assumption of a pure rolling motion. However, in reality, some sliding motion may occur, which causes slight deviation of the BCF locations. Therefore, these equations should be regarded as approximations only.

2.2 Envelope Detection. Normally, BCF appear in the low frequency range, and are revealed in a narrow band spectrum generated when using FFT. In ideal circumstances, by extracting BCF in a narrower band through the use of zooming or band-pass filtering, the repetitive impact vibrations will be revealed. However, in practice, the isolation of BCF may not be possible, as the energy involved in periodic impacts is very low and usually buried in much higher level of vibration generated by macro-structural components. In Fig. 1(*a*) a bearing has a defective outer-race, but the related BPFO and its harmonics are difficult to detect in the corresponding spectrum, where all the types of BCF should be located. Moreover, due to small changes in rotational speed, the heavily zoomed spectrum will be smeared to such a degree that the line spectrum from the original pulses will disappear. Therefore, the ED technique should be used together with FFT to ensure that the impulsive frequency of BPFO can be more easily found.

When there are localized defects occurring on the races or rollers, the vibrations caused become amplitude modulated due to periodic changes in the forces. Since ED is a signal processing technique for amplitude demodulation, the repetitive impacts, which have been modulated, will be demodulated and appear in the envelope spectrum. Fig. 1(*b*) shows the use of the envelope

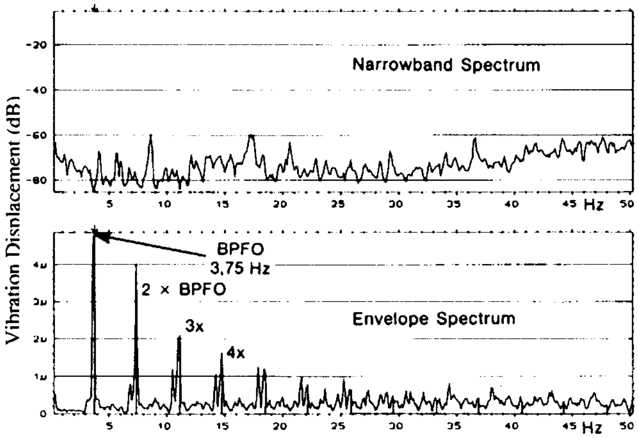

Fig. 1 A narrow band spectrum in (*a*) shows no sign of bearing fault, but in (*b*) an envelope detection shows outer-race defect [20]

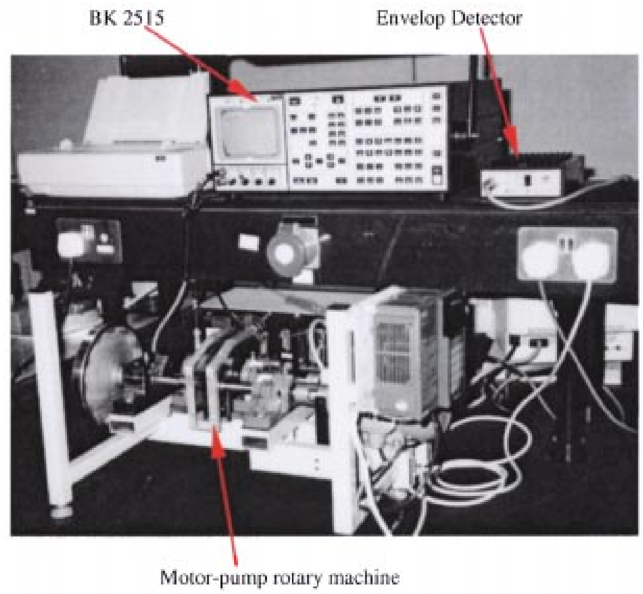

Fig. 3 Additional hardware for FFT and envelope detection

spectrum to identify a bearing outer race defect. The BPFO and its related harmonics are clearly shown, and there is no doubt that an outer-race defect is occurring.

Usually, four steps are involved in the process of ED [6]. First, the high frequency structural excitation range of a bearing is determined by performing impact tests or the like. Since the signal to noise ratio is higher in this range, the amplitude modulated signal should be easily found in this range which is highlighted by dashed line in Fig. 2, step A'. Second, the temporal vibration signal then passes through a band-pass filter centered around the bearing excitation frequency. The low frequency rotational speed of the rotor and the very high frequency random noise are rejected by the band-pass filter to leave bursts of the narrow band bearing signal as shown in Fig. 2, step B [18]. Third, the signals are rectified (carrier) to keep only the envelope. At this stage, a series of enveloped time signal should be revealed as shown in Fig. 2, step C. Finally, the enveloped temporal signal is analyzed in a low frequency range of a so-called ''envelope spectrum'' to identify the de-modulated signal as shown in Fig. 2, step D. Depending on the type of defect, the corresponding BCF should be clearly revealed in this envelope spectrum.

However, in practical application, the ED technique has the following drawbacks:

1 ED hardware device is required as the process is computational intensive. Figure 3 shows the FFT based spectrum analyzer (BK 2515) and the ED hardware device necessary to perform the analysis of FFT with ED.

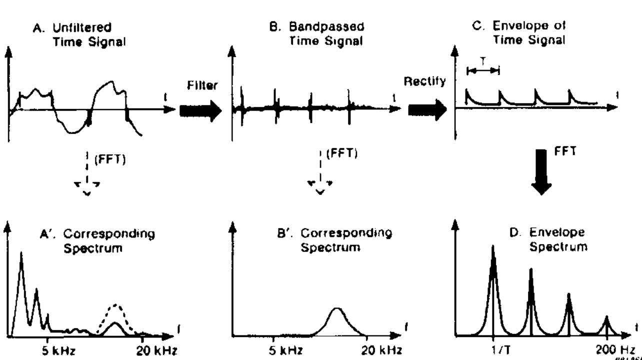

Fig. 2 Process of envelope detection to reveal periodic impulses [21]

2 Impact tests are needed to determine the excitation frequency of a particular bearing system. The additional hardware required are an impact hammer or a vibration exciter and its controller.

3 A skillful operator is needed to perform the impact tests for finding the excitation range of the bearing system, the process of ED, and the selection of an appropriate band-pass filter.

These requirements may not be available to some companies due to limited capital and insufficient human resources.

3 Wavelet Analysis (WA)

The conventional FFT provides a representation of a time signal $x(t)$ in its frequency form $Y_x(\omega)$. By using FFT, a time signal $x(t)$ can be converted to its frequency representation $Y_x(f)$ with $f=\omega/2\pi$ as [22]:

$$Y_x(\omega)=\int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt. \tag{4}$$

FFT does not give a simultaneous representation in the time and frequency domains. Moreover, FFT cannot capture nonstationary signals. Thus, it is handicapped in detecting impulses generated by the random contacts of defective surfaces.

The major advantage of WA is the presentation of signals in time-frequency distribution diagrams with multi-resolution in time and frequency. As already mentioned, this property is essential in the detection of bearing faults using vibration analysis. WA is a scale-time decomposition of a temporal signal $x(t)$ into components $\psi(t-b/a)$ which are localized in time, and is defined by [23]:

$$W_x(a,b)=\int_{-\infty}^{+\infty} x(t)\psi^*_{a,b}(t)dt, \tag{5}$$

where $\psi^*_{a,b}(t)$ is the complex conjugate of $\psi_{a,b}(t)$ which is the scaled and shifted versions of a so-called ''mother wavelet'' function defined as

$$\psi_{a,b}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{6}$$

Here $a>0$ is the scaling parameter and b is the translation or time shifting parameter. The scaling operation performs the stretching and compressing operations on the mother wavelet function, which in turn can be used to represent the signal to be analyzed in a different frequency range. On the other hand, the time shifting operation shifts the mother wavelet along the time axis. The shifted version is used to catch the time information of the signal to be analyzed. Therefore, a family of scaled and shifted wavelets can be created by varying the scaling a and the shifting b parameters, which serve as the base to form a family of wavelets to analyze the captured signals. The factor, $1/\sqrt{a}$, is used to ensure that the energy of the scaled and shifted versions are the same as the mother wavelet.

For practical purposes, regarding the possibility of time-frequency localization [23], the mother wavelet $\psi(t)$ must be a window function, meaning that $\int_{-\infty}^{+\infty}|\psi(t)|dt<\infty$, and must satisfy the admissibility condition as:

$$C_g=2\pi\int\frac{|Y_\psi(\omega)|^2}{\omega}d\omega<\infty, \qquad (7)$$

where $Y_\psi(\omega)$ is the FFT of $\psi(t)$. Following from Eq. (7) then $Y_\psi(\omega)=0$ for $\omega=0$, that is the wavelet function $\psi(t)$ has a zero average value by having $\int_{-\infty}^{+\infty}\psi(t)dt=0$. This is the reason why $\psi(t)$ is a ''small wave''. Equation (7) is needed to obtain the inverse of the WA as

$$x(t)=\frac{1}{C_g}\int_{-\infty}^{\infty}\int_{0}^{\infty}W_x(a,b)\psi_{a,b}(t)\frac{1}{a^2}da\,db. \qquad (8)$$

The scaling parameter, a, which represents a frequency band, has a center and a width of frequency band as ω_c/a and $2\Delta/a$ respectively. Here ω_c and Δ are the center frequency and bandwidth of the mother wavelet function $\psi(t)$ respectively. Therefore, the wavelet analysis characterizes the temporal signal $x(t)$ by decomposing it into a series of components with different frequency bandwidths. In the representation of time-frequency distribution, the band energy is used. Note that the scaling parameter, a, corresponds to a frequency band, $\Delta_a=2\Delta/a$. The band energy is calculated by

$$E_a=\int|W_a(a,b)|db. \qquad (9)$$

This represents the signal energy contents contained in the frequency bandwidth of Δ_a.

There are many types of wavelet functions available to use for different purposes, such as the Harr, Dabechies, Gaussian, Meyer, Mexican Hat, and Morlet functions. Generally, continuous WA is preferable for vibration-based machine fault diagnosis, as the resolution is higher compared to the dyadic type of WA [24]. Moreover, continuous WA is not as orthogonal for the different scale, a, as the discrete WA [25]. Although one can make the wavelet orthogonal by selecting a discrete set of a, this will mean that the inspected signals have a scale different from the selected set. Hence, the continuous WA is easier to ''match'' with the inspected signals. In this study, the continuous type of Gaussian wavelet is used.

4 Comparison of Using Wavelet Analysis and FFT With ED in the Detection of Bearing Faults

4.1 Experimental Platform and Testing Machine. For test purposes, a series of motor-pump rotary machines were used to generate vibration patterns caused by a variety of bearing faults. As shown in Fig. 4, each machine was composed of a variable speed drive, an AC motor driving a hydraulic pump, flexible couplers, shafts, a number of ball bearings, journal bearings, a gear coupler, and a flywheel for balancing. The load of the pump could be adjusted via a variable displacement valve.

The instruments used for the experiments include a Bruel & Kjaer FFT based frequency spectrum analyzer (BK2515), charge amplifiers, an envelop detector (ED), a number of piezoelectric accelerometers with tri-axial adapters, a tachometer for speed compensation, a digital data recorder for storing the vibration signals, and a PC for data transfer and manipulation. The vibration signals were measured by a number of piezoelectric accelerometers with tri-axial adapters mounted on each of the bearing housings. Therefore, the vibrations occurring at the vertical, the horizontal and the axial directions could be measured.

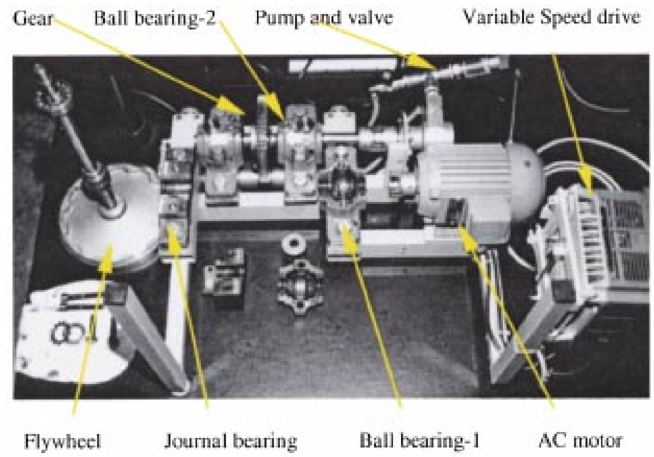

Fig. 4 The components of a motor-pump rotary machine

4.2 Experimental Results. To compare the effectiveness of WA and FFT with ED in bearing fault diagnosis, a variety of artificial faults were introduced into the SKF1205E ball bearings. The types of faults included a defective outer-race, a defective inner-race, and a defective roller. Although the motor was set to rotate at 20 Hz, the actual rotating speed monitored by the tachometer was found to be 23.2 Hz. Using Eqs. (1) to (3), and given that the geometric parameters of the bearings $-d=7.13$ mm, $D=32.02$ mm, $\beta=5$ deg$\sim$12 deg, and $n=24$—the calculated BCF for each type of fault are presented in Table 1. The vibration signals were measured from the accelerometers at a sampling rate of 65.4 kHz.

4.2.1 Results of Fault Diagnosis Using FFT With ED. The envelope spectra of various types of bearing faults obtained by FFT with ED are shown in Fig. 5. Since the result of impact test shows that the range of bearing excitation frequencies were from 8.6 kHz to around 12.8 kHz, the center of the band-pass filter of ED was set at 10 kHz [26]. Using the selected filter of ED, an envelope spectrum was created, as shown in Fig. 5(a). The impact repetition frequency, f_0, at 222 Hz and its harmonics ($2xf_0\approx444$ Hz) can be clearly recognized. This frequency f_0 is very close to the calculated BPFO at 220.14 Hz as listed in Table 1, hence it was identified as the outer-race defect.

The envelope spectra caused by the inner-race and roller defects are shown in Figs. 5(b) and 5(c) respectively. In Fig. 5(b) the

Table 1 Calculated BCF for different types of faults

Fault Types	Outer-race (BPFO)	Inner-race (BPFI)	Roller (BPFR)
Calculated BCF	216.64~ 220.14Hz	336.65~ 340.20Hz	99.06~ 99.63Hz
Actual BCF	222Hz	304Hz	114Hz
Interval	4.5ms	3.3ms	8.8ms

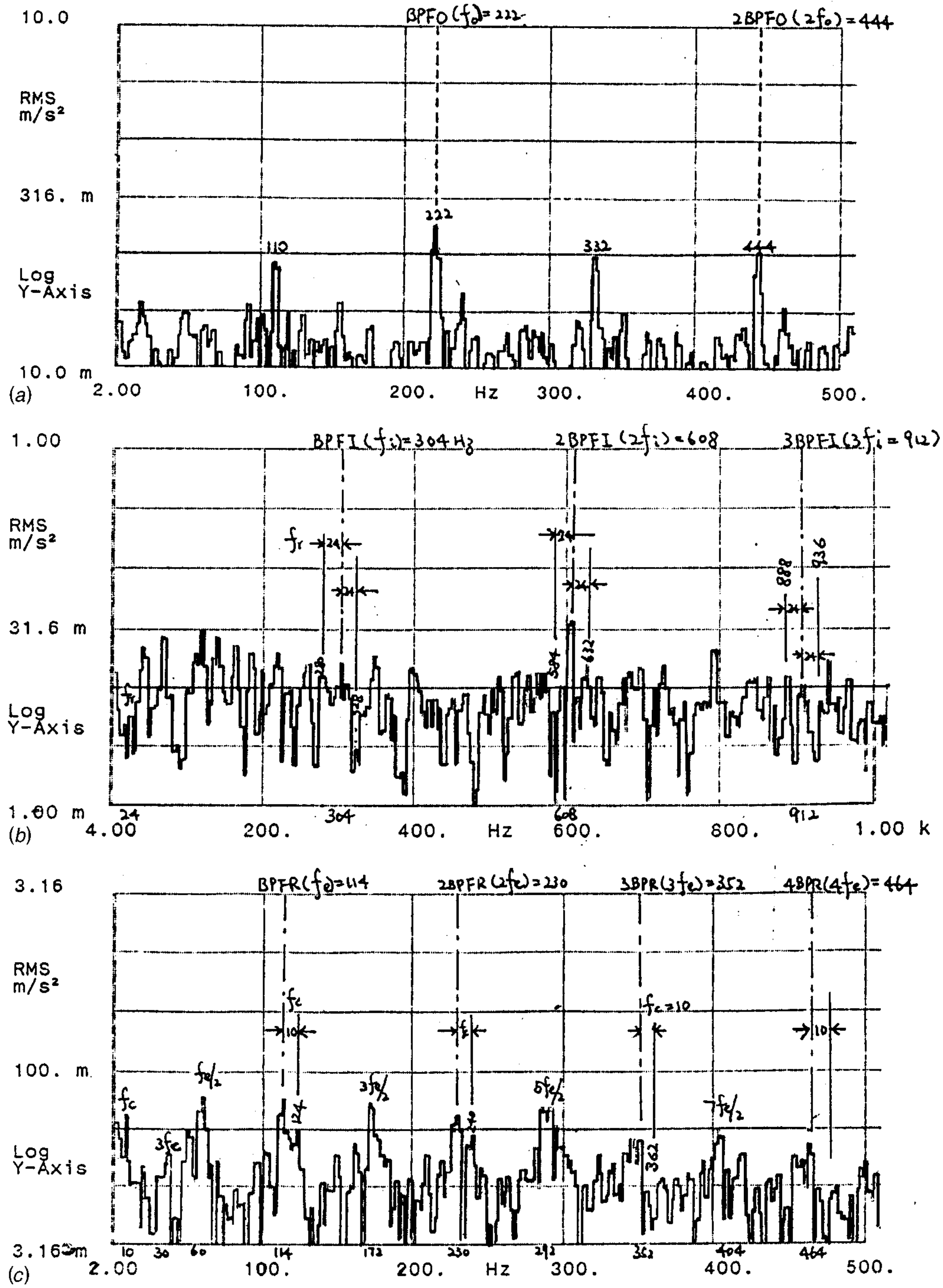

Fig. 5 (*a*) Envelope spectrum used to identify the outer-race defect (*b*) Envelope spectrum used to identify the inner-race defect (*c*) Envelope spectrum used to identify the roller defect

appearance of the inner-race defect frequency, f_i or BPFI (around 304 Hz), its harmonics (2×BPFI≈608 Hz, 3×BPFI≈912 Hz), as well as its sidebands (modulation sidebands at spacing about 24 Hz, roughly equal to the shaft rotational frequency, f_r at 23.2 Hz) are hardly identified. In order to make the readers easier to understand the theory of ED, we have artificially inserted remarks at the frequencies which are suspected to be the BPFI, its harmonics and related sidebands as shown in Fig. 5(*b*).

In Fig. 5(*c*), the roller defect frequency, f_e or BPFR (around 114 Hz), its harmonics (2×BPFR≈228 Hz, 3×BPFR≈344 Hz), and its sidebands (modulation sidebands at spacing of around 10 Hz, roughly equal to the cage rotation frequency, f_c, at 9.8 Hz) could be identified only by very careful inspection. The repetition of harmonics and modulation sidebands throughout the envelope spectra of the inner-race and roller defects make the identification of inner-race and roller defects not as easy to detect as the outer-race defect (BPFO) as shown in Fig 5(*a*). Again, to make the readers understand the theory of ED, we have artificially inserted remarks at the frequencies which are suspected to be the BPFR, its harmonics and related sidebands as shown in Fig. 5(*c*). The experimental results demonstrated that FFT with ED can be used as a diagnostic tool for bearing faults, especially for outer-race defects. However, the identification of inner-race and roller defects could be very difficult. A skillful operator must present to accu-

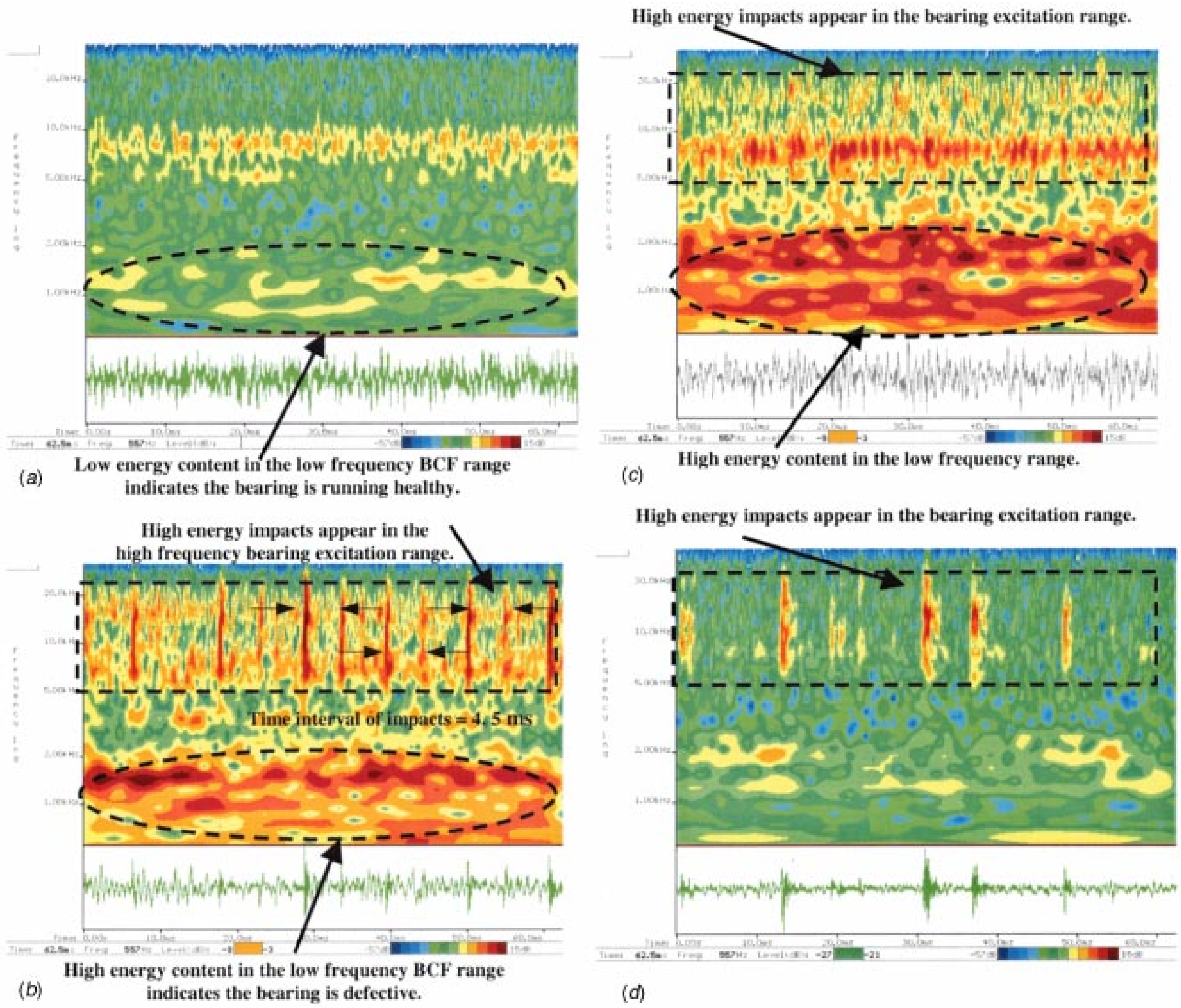

Fig. 6 (*a*) Results of a healthy bearing using wavelet analysis (*b*) Results of outer-race defect using wavelet analysis (*c*) Results of inner-race defect using wavelet analysis (*d*) Results of rolling element defect in a wavelet analysis

rately interpret the results generated from the envelope spectrum, or else false alarms or even ignorance about faults may occur.

4.2.2 Results of Fault Diagnosis Using Wavelet Analysis. As mentioned in Section 4.1, many sets of vibration data were collected at different directions, rotating speeds and bearings of the motor-pump machines. Diagnostic results of all three types of faults have been obtained from their time-frequency distribution diagrams generated by WA. Here, we show some typical results obtained from the vibration signals collected at the radial direction of the bearings running at 23 Hz under full loading condition. Figures 6(*a*) to 6(*d*) show the analyzed results of bearings running under the conditions of normal, outer-race fault, inner-race fault and roller fault respectively.

When the bearing is operating in normal condition, the energy level in the low frequency range, where most of the macro-structural vibrations are located, should be low. As shown in Fig. 6(*a*), the energy levels of vibrations within the range of BCF (under 2 kHz) are low. The dominant energy levels are from -39 dB to -3 dB which are highlighted by contours in light colors. Meanwhile, in Fig. 6(*b*), where an outer-race fault exists in the bearing, the energy levels of vibrations in the low frequency range are high, extending from -9 dB to 15 dB, an increase of around 20 dB. The vibrations are highlighted by dark colored contours in Fig. 6(*b*). The difference of energy levels between a healthy and a defective bearing is obvious in the low frequency range where the BCF are located. Hence, with the help of WA, the operator can detect the existence of faults in a bearing by simply inspecting the energy level at the low frequency range of the time-frequency distribution diagram. If the energy level has changed significantly at the range of BCF, but the operating environment has not changed, then the increase of energy could be caused by the occurrence of a fault.

Further evidence to prove the occurrence of a fault could be obtained by inspecting the energy level at the high frequency range, where the bearing excitation range is located. Figures 6(*b*) to 6(*d*) show the high energy impacts generated by the defects in the outer-race, the inner-race and the roller respectively. The ranges of bearing excitation are bounded by rectangles in dashed lines as shown in Figs. 6(*b*) to 6(*d*). If high concentrations of energy caused by impacts, which are highlighted by strips or lagoons in dark colored contours, are existing in the high frequency bearing excitation range, then faults are occurring in the inspected bearing. Such a simple visual inspection does not require the machine operator to have a lot of experience in fault diagnosis. Moreover, the tedious procedures of using the classical method of FFT with ED for detecting faults are not required.

To save cost in maintenance, sometimes it is desirable to know the cause of a fault so that the defective component of a bearing can be identified and repaired without replacing it with a new bearing. As already mentioned, the cause of the fault can be determined by inspecting the time interval of impacts located at the high frequency bearing excitation range. The length of time interval of the impacts is inversely proportional to the BCF. By measuring the time interval of each impact in the time-frequency distribution diagrams, and then comparing the time intervals listed in Table 1, one can determine the type of fault that is occurring. For instance, as shown in Fig. 6(*b*), the time interval of impacts is estimated as 4.5 ms, which is approximately equal to the inverse of the calculated BPFO (222 Hz). Hence, an outer-race fault is occurring in the inspected bearing, which has a small crack intentionally induced onto the outer-race of the bearing prior to the analysis. However, for the faults of inner-race and roller as shown in Figs. 6(*c*) and 6(*d*) respectively, visually inspecting the time intervals of impacts is difficult.

To let the operator performs the visual inspection of the time intervals of impacts easier, the results of WA are displayed in different frequency ranges or scales as shown in Figs. 7(*a*) to 7(*c*) for the normal, inner-race fault and roller defect respectively. The case of outer-race fault is not compared here as the time interval of impacts can be easily determined in Fig. 6(*b*). Since the bearing excitation range is embedded in the high frequency range of 8.17 kHz to 16.3 kHz (the second range or scale counting from the top of the figures), the inspection of time interval is performed in this range. Due to frequency aliasing inherited from WA [24], the features of impacts also appear in the first and third scales (counting from the top). Ideally, such impacts should only appear in the second scale.

Note that when the bearing is running normally, no distinctive impact can be observed in the frequency range of 8.17 kHz to 16.3 kHz as shown in Fig. 7(*a*). However, if the bearing is defective, then the defect related impacts should appear in this frequency range as shown in Figs. 7(*b*) and 7(*c*).

The time intervals of impacts in Figs. 7(*b*) and 7(*c*) are estimated as 3.3 ms and 8.8 ms respectively. These intervals are approximately equal to the inverse of the calculated BPFI (304 Hz) and BPFR (114 Hz), corresponding to the defects of inner-race and roller respectively. Although the spacing of impacts caused by inner-race fault is not obvious, a series of impacts with short duration does show on the second scale of Fig. 7(*b*). It is a distinguishable feature for inner-race fault as compared to other types of faults. The display of vibration signals at different scales or frequency ranges does provide additional evidence to prove the existence of faults and help the operator to identify their causes. Therefore, by inspecting the time intervals of the impacts, the operator may be able to determine the causes of faults and repair the corresponding defective components of the bearing.

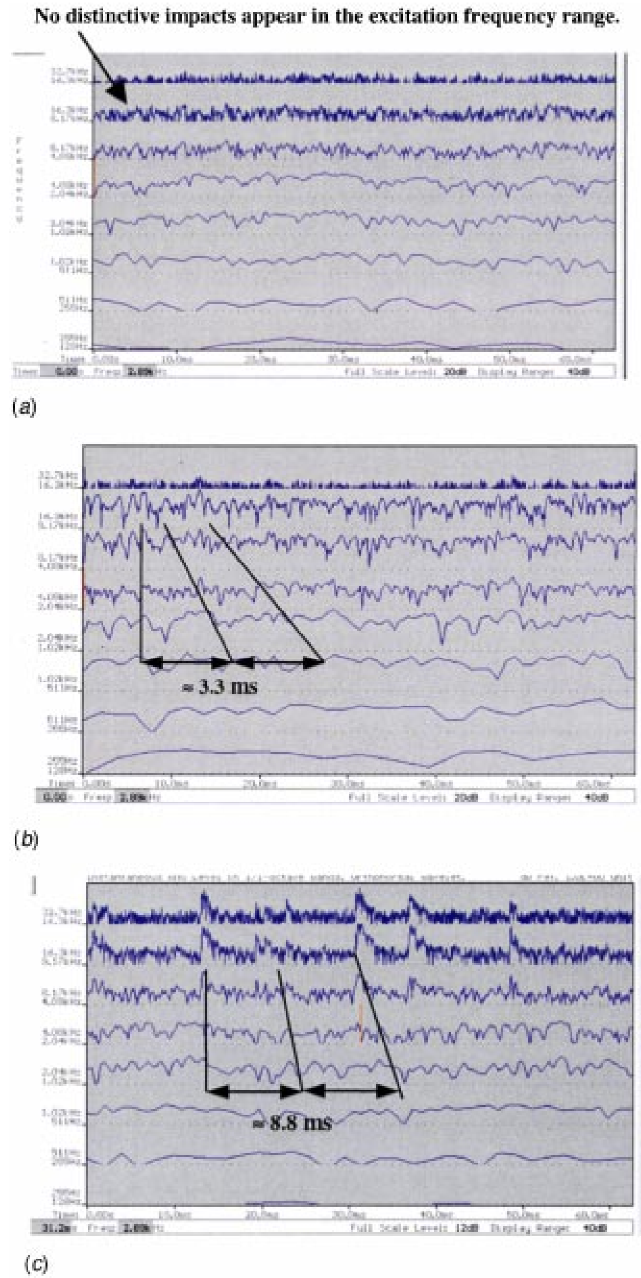

Fig. 7 (***a***) **Normal bearing running condition** (***b***) **Inner-race defect generates impacts at around 3.3 ms** (***c***) **Rolling element defect generates impacts at around 8.8 ms**

4.3 Comparison of Wavelet Analysis and FFT With ED. Both the methods of wavelet analysis and FFT with ED are effective in diagnosing the outer-race defect [Figs. 5(*a*) and 6(*b*)]. However, when using FFT with ED, the faults caused by defective inner-race and roller are difficult to identify using visual inspection. The corresponding BPFI and BPFR are hardly revealed, even when the envelope spectra is used [Figs. 5(*b*) and 5(*c*)]. On the other hand, when using the time-frequency distribution diagrams provided by WA [Figs. 6(*c*) 6(*d*) 7(*b*) and 7(*c*)], the high energy impacts caused by inner-race and roller defects can be identified in the high frequency bearing excitation ranges. For the bearing with roller defect [Fig. 6(*d*)], the high energy content in the low frequency BCF range is not obvious. Similar to Fig. 1(*a*). Figure 6(*d*) provides another piece of evidence that the bearing faults may not be revealed by using the narrowband spectrum only. Unfortunately, such a practice is still commonly used in the industry. However, as shown in Fig. 6(*d*), the impacts do appear in the high frequency bearing excitation range. This, coupled with the existence of impacts shown at the same range [second scale from the top of Fig. 7(*c*)], allows the operator to make a preliminary assumption that the cause of fault may be due to a defective roller. Such a practice is impossible using the FFT with ED method as illustrated in Fig. 5(*c*).

Note that the impacts caused by the faults of inner-race and roller are not repetitive in time as expected [Figs. 6(*c*) and 6(*d*)]. This phenomenon even makes the FFT based method difficult to detect the impacts. With the help from the time-frequency distribution diagrams provided by WA, these occasionally repeating impacts do appear on the diagrams.

Similar results have been obtained by using other motor-pump driven rotary machines running in the same operating conditions. The authors have also found that WA is more superior than FFT in the fault diagnoses of gas engines [27] and gearboxes [24].

5 Conclusions

In summary, both the WA and the FFT with ED methods are effective in finding the bearing outer-race fault. However, to diagnose the faults of inner-race and roller, the tool of WA is found to be easier for the machine operator to interpret the analyzed results. The operator does not require much skill to use WA in the process of fault diagnosis. The hardware for impact tests and ED, which are necessary for the method of FFT with ED, are not required for WA. As shown in Figs. 6 and 7, WA does provide good resolution in frequency at the low frequency range, and fine resolution in time at the high frequency range. Such a multi-resolution capability is essential for vibration-based machine fault diagnosis.

The critical point is the selection of an appropriate set of parameters for the WA. The selected parameters must enable the WA to generate a family of wavelet functions which can adaptively ''match'' with the inspected signals. Some simple and effective methods have been suggested by the authors in another paper [27]. Although the continuous Gaussian WA used here does not have any optimization of parameters, the results are satisfactory in the fault diagnosis of a series of motor-pump machines. In industry, a simple solution is always preferable. In-depth analyses that require adaptive WA may be not welcomed by machine operators. The aim is to fix the problem as soon as possible with minimum interruption to production. Therefore, a general type of continuous Gaussian WA is suitable as long as it can provide sufficient resolutions in time and frequency for fault diagnosis.

The procedures of using WA in the fault diagnosis of rolling element bearings can be divided into two stages. The first stage is to detect the existence of any bearing fault in a fast process. If high energy impacts are observed in the high frequency range of the time-frequency distribution diagram, then faults are occurring in the bearing. Additional evidence to prove the existence of faults may be obtained by inspecting whether a substantial increase of vibration energy is found in the low frequency range of the time-frequency distribution diagram. The exceptional case is the roller defect. It shows impacts in the high frequency range only.

If the cause of fault must be identified, then the second stage of inspecting the time interval of impacts in the high frequency bearing excitation range should be performed. Therefore, depending on the need and the time constraint of the company, the machine operator can use the quick first stage for simple fault detection, and replace the defective bearing with a new one. He may also perform the second stage to find the cause of the fault so that he can repair or replace the defective part of the bearing. Hence, the purchase of a new bearing is not necessary. Such a simple and flexible inspection method provided by WA for bearing fault diagnosis is definitely welcomed by the industry.

Acknowledgment

The work described in this paper was supported by Competitive Earmarked Research Grant No. 9040223, which was obtained from the Hong Kong Research Committee Council, and is gratefully acknowledged.

References

[1] Tse, P., 1998, ''Neural Networks Based Robust Machine Fault Diagnostic & Life Span Predicting System,'' Ph. D. Thesis, The University of Sussex, United Kingdom.

[2] Mitchell, J. C., 1993, *Introduction to Machinery Analysis and Monitoring*, PennWell, Tulsa, Okla.

[3] Li, C. J., and Wu, S. M., 1989, ''On-line Detection of Localized Defects in Bearings by Pattern Recognition Analysis,'' ASME J. Eng. Ind., **111**, pp. 331–336.

[4] Sandy, J., 1988, ''Monitoring and Diagnostics for Rolling Element Bearings,'' Sound Vib., **22**, No. 6, pp. 16–20.

[5] Brown, D. N., 1989, ''Envelop Analysis Detects Bearing Faults Before Major Damage Occurs,'' Pulp Pap., **63**, pp. 113–117.

[6] Courrech, J., 2000, ''Envelop Analysis for Effective Rolling—Element Fault Detection—Facts or Fiction?'' Up Time Magazine, **8**, No. 1, pp. 14–17.

[7] McFadden, P. D., and Smith, J. D., 1984, ''Vibration Monitoring of Rolling Element Bearings by the High Frequency Resonance Technique: A Review,'' Tribol. Int., **17**, pp. 3–10.

[8] Gade, S., and Herlufsen, H., 1994, ''Digital Filter techniques vs. FFT Techniques for Damping Measurements,'' Bruel & Kjaer Technical Review, No. 1.

[9] Ville, J., 1984, ''Theory et Application de la Notion de Signal Analytique,'' Cables et Transmissions, **20A**.

[10] Gade, S., and Gram-Hansen, K., 1996, ''Non-stationary Signal Analysis Using Wavelet Transform, Short-time Fourier Transform and Wigner-Ville Distribution,'' Bruel & Kjaer Technical Review, No. 2.

[11] Dowling, H. M., 1993, ''Application of Non-stationary Analysis to Machinery Monitoring,'' *Proceedings of IEEE Int Conf. On Acoustics, Speech and Signal Processing*, pp. 59–62.

[12] Newland, D. E., 1994, ''Wavelet Analysis of Vibration, Part 1: Theory, Part 2: Wavelet Maps,'' ASME J. Vibr. Acoust., **116**, pp. 409–425.

[13] Newland, D. E., 1999, ''Ridge and Phase Identification in the Frequency Analysis of Transient Signals by Harmonic Wavelets,'' ASME J. Vibr. Acoust., **121**, pp. 149–155.

[14] Debabrata, P., 1994, ''Detection of Change in Process Using Wavelets,'' *Proc. of the IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis*, pp. 174–177.

[15] Mori, K., Kasashima, N., Yoshika, T., and Ueno, Y., 1996, ''Prediction of Spalling on Ball Bearing by Applying the Discrete Wavelet Transform to Vibration Signals,'' Wear, **195**, pp. 162–168.

[16] Li, C. J., and Ma, J., 1997, ''Wavelet Decomposition of Vibrations for Detection of Bearing-Localized Defects,'' NDT & E Int., **30**, pp. 143–149.

[17] Wang, W. J., and McFadded, P. D., 1996, ''Application of Wavelets to Gearbox Vibration Signals for Fault Detection,'' J. Sound Vib., **192**, No. 5, pp. 927–939.

[18] Ratcliffe, G. A., 1990, ''Condition Monitoring of Rolling Element Bearings Using the Enveloping Technique,'' *Machine Condition Monitoring*, Mechanical Engineering Publications Ltd., London, pp. 55–67.

[19] Tse, P., and Atherton, D., 1999, ''Prediction of Machine Deterioration Using Vibration Based Fault Trends and Recurrent Neural Networks,'' ASME J. Vibr. Acoust., **121**, No. 3, pp. 355–362.

[20] Bruel & Kjaer, *Peak and Envelop Analysis for Bearing Fault Detection, Application Note: BO 0286-11*, Bruel & Kjaer Ltd., Denmark.

[21] Bruel and Kjaer, *Envelop Detector*, Bruel and Kjaer System Development, WB1048.

[22] Dimaragonas, A., 1992, *Vibration for Engineers*, Prentice-Hall, Inc., New Jersey, pp. 33–34.

[23] Chui, ChK., 1992, *An Introduction to Wavelets*, Vol. 1, Academic Press, Boston.

[24] Tse, P., and He, L. S., 2001, ''Can Wavelet Transforms Used for Data Compression Equally Suitable for the Use of Machine Fault Diagnosis,'' *The 18th Biennial Conference on Mechanical Vibration and Noise*, Pittsburgh, USA.

[25] Huang, N. E., et al., 1998, ''The Empirical Mode Decomposition and the Hibert Spectrum for Nonlinear and Non-stationary Time Series Analysis,'' Proc. R. Soc. London, Ser. A, **454**, pp. 903–995.

[26] Lai, H. W., and Tse, P., 1995, *Detection of Bearing Faults in Rotating Motor Using Vibration Analysis*, Technical Report of MEEM Department, City University of Hong Kong, Hong Kong.

[27] Tse, P., Xu, G., Qu, L., and Kumara, S., 2000, ''An Effective and Portable Electronic Ear for Fault Diagnosis Using Machine Operating Sound Directly,'' International Journal of Acoustics and Vibration, an affiliated Journal of the Institute of Acoustics and Vibration (JIAV), **6**, No. 1, pp. 23–31.