Non-stationary signal processing for bearing health monitoring

R.X. Gao and R. Yan

Department of Mechanical and Industrial Engineering,
University of Massachusetts,
Amherst, MA 01003, USA
Email: gao@ecs.umass.edu

Abstract: Signals generated by transient vibrations in rolling bearings due to structural defects are non-stationary in nature, and reflect upon the operation condition of the bearing. Consequently, effective processing of non-stationary signals is critical to bearing health monitoring. This paper presents a comparative study of four representative time-frequency analysis techniques commonly employed for non-stationary signal processing. The analytical framework of the short-time Fourier transform, wavelet transform, wavelet packet transform, and Hilbert-Huang transform are first presented. The effectiveness of each technique in detecting transient features from a time-varying signal is then examined, using an analytically formulated test signal. Subsequently, the performance of each technique is experimentally evaluated, using realistic vibration signals measured from a bearing test system. The results demonstrate that selecting appropriate signal processing technique can significantly affect defect identification and consequently, improve the reliability of bearing health monitoring.

Keywords: bearing health monitoring; non-stationary signals; time-frequency analysis.

Reference to this paper should be made as follows: Gao, R.X. and Yan, R. (2006) 'Non-stationary signal processing for bearing health monitoring', *Int. J. Manufacturing Research*, Vol. 1, No. 1, pp.18–40.

Biographical Notes: Robert X. Gao received his PhD degree from the Technical University Berlin, Germany, in 1991, and is presently a Professor at the University of Massachusetts Amherst (UMass). His research interests include sensing methodology, energy-efficient sensor networks and advanced signal processing for system health monitoring, diagnosis and prognosis. Dr. Gao received the *CAREER Award* from the National Science Foundation in 1996 and the *Outstanding Engineering Junior Faculty Award* from the UMass in 1999. He has published over 140 refereed journals and conference papers, and serves on the editorial boards for three international journals.

Ruqiang Yan received his BS and MS degrees in Mechanical Engineering from the University of Science and Technology of China, in 1997 and 2002, respectively. He is currently a PhD candidate and Research Assistant at the Department of Mechanical and Industrial Engineering, University of Massachusetts Amherst. His research interests include non-linear time series analysis, multi-domain signal processing, machine condition monitoring and health diagnosis, sensors and sensor networks.

1 Introduction

Growing demand for high-quality and low-cost production has increased the need for intelligent manufacturing systems with online monitoring and control capabilities (Tönshoff et al., 1988; Byrne et al., 1996; Ganesan, Das and Venkataraman, 2004; Liang, Hecker and Landers, 2004). Consequently, sensor-based machine health diagnosis has gained increasing attention from the research community worldwide (Teti, 1996; DimlaSnr, 2000; Tseng and Chou, 2002; Frankowiak, Grosvenor and Prickett, 2005). The goal of these efforts is to obtain information in real-time about the operation status of the machines and use the information to:

1 Identify machine faults at the incipient stage such that proper corrective measures can be taken before the faults have progressed to cause significant structural damage and costly downtime, thus enabling *adaptive* maintenance and production scheduling.

2 More accurately control the quality of products being manufactured, which is directly related to the working conditions of the machine.

In addition to monitoring individual machines, data gathered from the sensors provide insight into the manufacturing process itself, and can be used to assist in high-level decision-making for production optimisation.

Over the past decades, various sensing and signal processing techniques have been developed to improve the effectiveness and efficiency in diagnosing machine conditions. A common approach is to measure machine vibrations, as they are directly related to the structural dynamics and subsequently, the working conditions of the machine (El-Wardany, Gao and Elbestawi, 1996; Li, Tso and Wang, 2000; Goumas, Zervakis and Stavrakakis, 2002). Also, acoustic emission has been investigated for condition assessment of rolling bearings (Morhain and Mba, 2003), tool failure detection (Lee and Tang, 1999), tool-life monitoring (Kim et al., 1999), and flank wear estimation (Kamarthi, Kumara and Cohen, 2000). In general, signals encountered in manufacturing machines can be classified as either *stationary* or *non-stationary*. Stationary signals are characterised by time-invariant statistical properties, such as the mean value or auto-correlation function (Julius and Allan, 1986). A typical example is the periodic vibrations caused by structural imbalance in a rolling bearing, when the bearings' gravitational centre does not coincide with the rotational centre. Analysis of stationary signals has largely been based on well-known spectral techniques such as the Fourier transform, which identifies the constituent frequency components within the signal. In contrast, the statistical properties of a non-stationary signal change over time, making the time-averaging approach adopted in the Fourier transform ineffective. Typically, non-stationary signals encountered in manufacturing machines consist of three major components:

1 A periodic component resulting from the cyclic interactions between the interfacing elements of the machine, such as vibrations caused by the interaction between the rolling elements and the raceway.

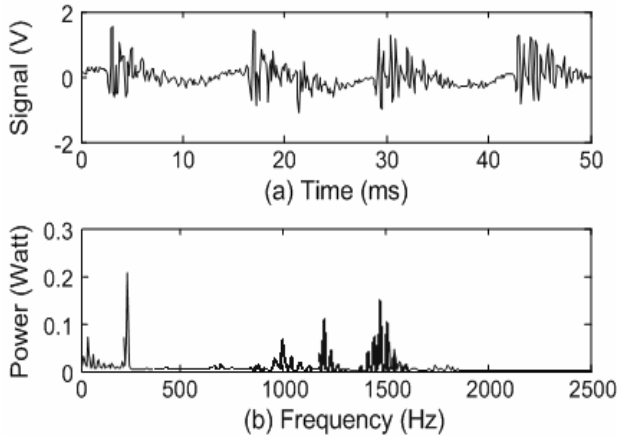
2 A transient component caused by 'one-time' events, such as the sudden breakage of a drilling bit or the inception of a crack inside a workpiece.

3 Broad-band background noise.

Effective and timely detection of transient components hidden in a non-stationary signal is of significant importance to reliable machine condition monitoring, since they represent the precursor of potential machine failure. However, due to their 'one-shot' nature and the inherently short duration and weak magnitude, reliable identification of transient signals has remained a challenging task (Gu, Ni and Yuan, 2002; Padovese, 2004; Shi, Tsung and Unsworth, 2004).

Time-frequency techniques decompose one-dimensional time-series signals into a two-dimensional plane by exposing the time-dependent variations of characteristic frequencies within the signal, thus presenting a valid and effective tool for non-stationary signal analysis than the Fourier transform. As an example, Figure 1a illustrates the waveform of vibration signals measured from a ball bearing that contained a localised defect on its outer raceway. While the spectrum in Figure 1b indicates five major frequency components at 40, 250, 1,000, 1,200 and 1,500 Hz, respectively, it does not provide information on the temporal variations of these frequencies during the course of the bearing test. When the same vibration signals were analysed using the Short-Time Fourier Transform (STFT) with a 12.8 ms window, the contour plot of the signal's time-frequency distribution in Figure 2 indicates that one of the frequency components has shifted from 1,200 to 1,000 Hz at the time point of 20 ms. Examination of the bearing assembly revealed that the defect has grown from the initial localised configuration into a larger area. Such a result indicates that frequency change along the time line can be used as an indicator of the bearing defect progression.

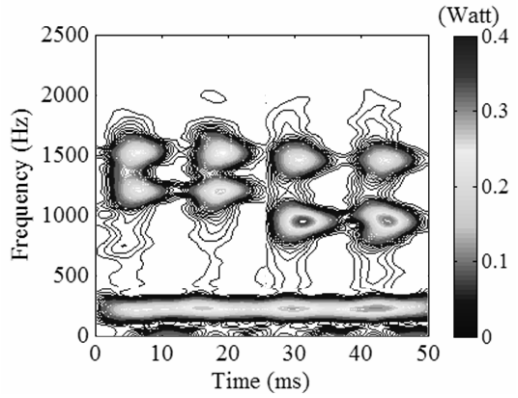
Figure 1 Bearing vibration signal and its spectrum



In addition to the short-time Fourier transform, other time-frequency analysis techniques, such as wavelet transform, wavelet packet transform, and Hilbert-Huang Transform (HHT), have been investigated for the condition monitoring of machine tools (Lee and Tang, 1999; Li, Tso and Wang, 2000; Wu, Escande and Du, 2001), rolling bearings (Mori et al., 1996; Wang and Gao, 2003; Lou and Loparo, 2004; Yan and Gao, 2005; Yu, Cheng and Yang, 2005), gearboxes (Wang and McFadden, 1995; Zheng, Li and Chen, 2002; Chen and Wang, 2002), washing machines (Goumas, Zervakis and Stavrakakis, 2002), power transformers (Satish, 1998) and automotive brakes (Charley, Bodovillé and Degallaix, 2001). Since each of these techniques has shown certain effectiveness in

extracting features hidden within the signals, a comparative evaluation contrasting the relative strength and weakness of the various techniques would provide useful guidance to selecting better-suited techniques for specific applications. Driven by such a motivation, this paper investigates four representative time-frequency techniques – short-time Fourier transform, wavelet transform, wavelet packet transform, and Hilbert-Huang Transform – with respect to their effectiveness in decomposing non-stationary signals. After a quantitative evaluation of the four techniques using a test signal, the four techniques were applied to analysing realistic vibration signals measured on a custom-designed bearing test system. The experiments confirmed the conclusions drawn from the test signal analysis, and established a common ground for evaluating the various time-frequency techniques.

Figure 2 STFT of the bearing vibration signal



2 Formulation of test signal

To enable quantitative evaluation of the four time-frequency analysis techniques, a non-stationary test signal $x(t)$ that mimics the realistic bearing vibration signal shown in Figure 1a is first analytically formulated. Based on the five major frequency components identified in Figure 1b and the time-dependency of the 1,000, 1,200 and 1,500 Hz components, as revealed by the short-time Fourier transform, the test signal was formulated as consisting of three constituent components:

$$x(t) = x_1(t) + x_2(t) + e(t) \tag{1}$$

In Equation (1), the term $x_1(t)$ represents the periodic component in the signal, given by:

$$x_1(t) = \sum_{i=1}^{N_1} A_i \times \sin\left[2\pi \times f_A(i) \times t + \theta_A(i)\right] \tag{2}$$

with N_1 being the number of sinusoidal elements, and A_i, $\theta_A(i)$, and $f_A(i)$ representing the amplitude, initial phase, and frequency of the i^{th} sinusoidal element, respectively. Spectral analysis indicated that the fundamental periodicity of the bearing signal can be reconstructed from the 40 and 250 Hz components. Specific values of other parameters

were determined using a pair of Fourier and inverse Fourier transforms, and are listed in Table 1.

Table 1 Parameters related to $x_1(t)$ in the test signal

	Sinusoidal Element (N_1)	
Parameter	*1*	*2*
A_i *[V]*	0.33	0.12
θ_A *(i) [rad]*	0	0.32π
f_A*(i) [Hz]*	250	40

The term $x_2(t)$ represents the transient component in the signal, which relates to the inception and progression of structural defects within the bearing (e.g. on the bearing raceways or surface of the rolling balls). The transient component was modelled as a series of exponentially attenuated vibrations (Harris, 1991), given by:

$$x_2(t) = \sum_{i=1}^{N_2} B_i \times \Theta(t - t_i) \times e^{-C_i \cdot (t - t_i)} \times \sin[2\pi \times f_B(i) \times (t - t_i) + \theta_B(i)] \tag{3}$$

where N_2 is the number of vibration cycles, and B_i, C_i, t_i, $\theta_B(i)$, and $f_B(i)$ are the amplitude, attenuation factor, time-delay, initial phase, and frequency of the i^{th} vibration cycle, respectively. Based on the result of the Short-Time Fourier Transform (STFT) analysis (Figure 2), a total of eight elements with the highest energy concentration were used to construct the transient component of the test signal. The specific values of the parameters were determined through least square error-based curve fitting method, as listed in Table 2.

Table 2 Parameters of $x_2(t)$ in the test signal

	Transient Component (N_2)							
	Group 1		*Group 2*		*Group 3*		*Group 4*	
Parameter	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
B_i *[V]*	0.6	0.5	0.6	0.5	0.6	0.5	0.6	0.5
C_i	377.3	301.8	377.3	301.8	377.3	251.5	377.3	251.5
t_i *[ms]*	4	4	16.5	16.5	29	29	41.5	41.5
θ_B*(i) [rad]*	0	0.33π	0	0.33π	0	0.33π	0	0.33π
f_B*(i) [Hz]*	1,500	1,200	1,500	1,200	1,500	1,000	1,500	1,000

In Equation (3), the functional item $\Theta(t - t_i)$ is defined as:

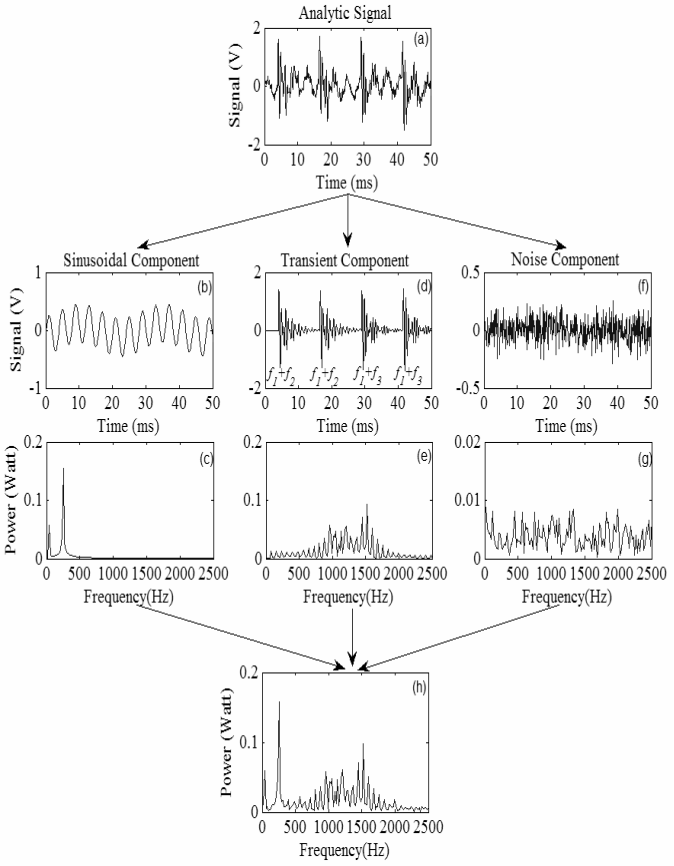
$$\Theta(t - t_i) = \begin{cases} 1 & t - t_i \geq 0 \\ 0 & t - t_i < 0 \end{cases} \tag{4}$$

It is included in Equation (3) to identify the point in time at which a transient element occurs.

The term $e(t)$ in Equation (1) represents a broad-band background noise that commonly exists in realistic vibration signals. In the presented study, it was modelled as a Gaussian white noise. The energy content of the $e(t)$ component was determined by the difference between the energy of the measured bearing vibration signal and the sum of the energy of the periodic and transient components contained in the test signal.

In Figure 3, the waveform and spectrum of the constructed test signal with its three constituent components (i.e. periodic, transient, and noise) are illustrated. The spectrum of each of the three components is shown beneath their respective time-domain representation. To simulate the transient component, a total of four groups of pulse trains were used. The four groups were separated from each other by a 12.5 ms time interval, and each group contains two specific frequency elements (e.g. f_1 and f_2 or f_1 and f_3). Within each pulse group, the two transient frequency elements were time-overlapped. Altogether, three frequencies were chosen for the transient component construction: $f_1 = 1.5$ kHz, $f_2 = 1.2$ kHz, and $f_3 = 1.0$ kHz. The specific configuration as well as the values of the specific parameters were chosen such that the test signal has returned a high correlation with the measured bearing vibration signal, with a high correlation coefficient of 0.97. As a result, the test signal was considered satisfactory for quantitative evaluation of the four time frequency techniques.

Figure 3 Test signal generated based on bearing vibrations. (a) Waveform of the test signal $x(t)$; (b) periodic component $x_1(t)$; (c) Spectrum of $x_1(t)$; (d) Transient component $x_2(t)$; (e) Spectrum of $x_2(t)$; (f) Noise component e(t); (g) Spectrum of e(t); (h) Spectrum of the test signal



Note: f_1= 1.5 kHz, f_2=1.2 kHz, f_3,=1.0 kHz

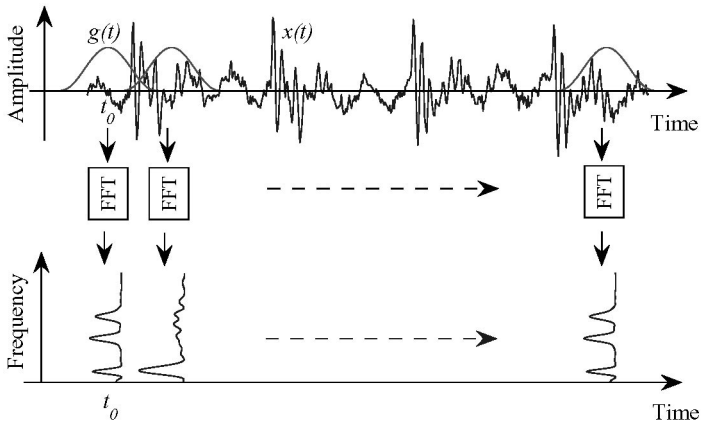
3 Time-frequency analysis of test signal

Using the formulated test signal, the strength and weakness of the four representative time-frequency analysis techniques were comparatively investigated, as described below.

3.1 Short-time fourier transform

The STFT employs a sliding window function $g(t)$ centered at time τ to perform 'time-localised' Fourier transform of the signal $x(t)$ consecutively, and the result reveals variation of the signal's frequency content as time evolves, as illustrated in Figure 4.

Figure 4 Illustration of short-time Fourier transform on the test signal $x(t)$



Note: FFT-Fast Fourier Transform

Mathematically, the STFT is defined as:

$$STFT(\tau, f) = \int x(t)g(t-\tau)e^{-j2\pi ft}dt = \int x(t)g^*_{\tau,f}(t)dt \tag{5}$$

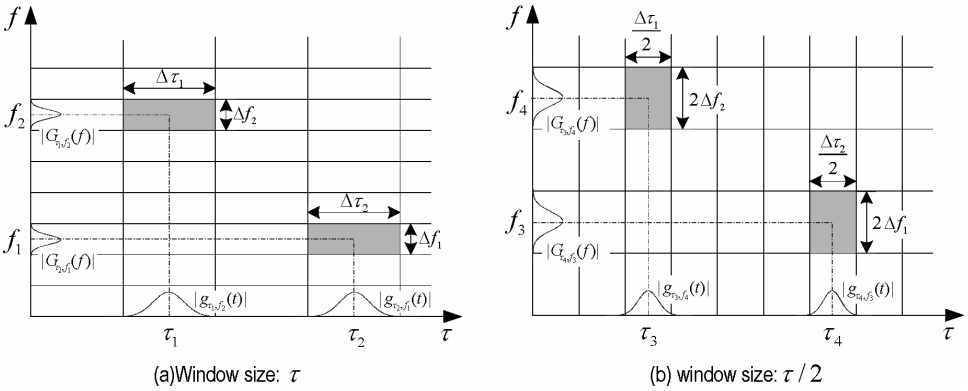
where (•)* denotes the complex conjugate of the function (•), The window function directly affects the time and frequency resolutions of the signal decomposition. While higher resolution provides better separation of the constituent components within a signal, the time and frequency resolutions for the STFT cannot be chosen arbitrarily at the same time, according to the uncertainty principle (Cohen, 1989). Specifically, the product of the time and frequency resolutions is lower bounded by:

$$\Delta\tau \cdot \Delta f \geq (4\pi)^{-1} \tag{6}$$

where $\Delta\tau$, the time resolution, is measured by the root-mean-square time width, and Δf, the frequency resolution, is measured by the root-mean-square bandwidth (Rioul and Vetterli, 1991). As an example, the Gaussian window function $g(t) = \exp(-\alpha t^2\tau^2)$ (with α being a constant and τ controlling the window width) has the time and frequency resolutions of $\Delta\tau = \tau/(2\sqrt{\alpha})$ and $\Delta f = \sqrt{\alpha}/(\tau \times 2\pi)$, respectively, and thus possesses the optimal time-frequency resolution, since the product of $\Delta\tau$ and Δf is exactly

$(4\pi)^{-1}$. Since the time and frequency resolutions of a window function are dependent on the parameter τ only, once the window function is chosen, the time and frequency resolutions over the entire time-frequency plane is fixed. Illustrated in Figure 5 are two scenarios where the product of the time and frequency resolutions of the window function (i.e. the area defined by $\Delta\tau \times \Delta f$) will remain the same, regardless of the actual window size (τ or $\tau/2$) employed.

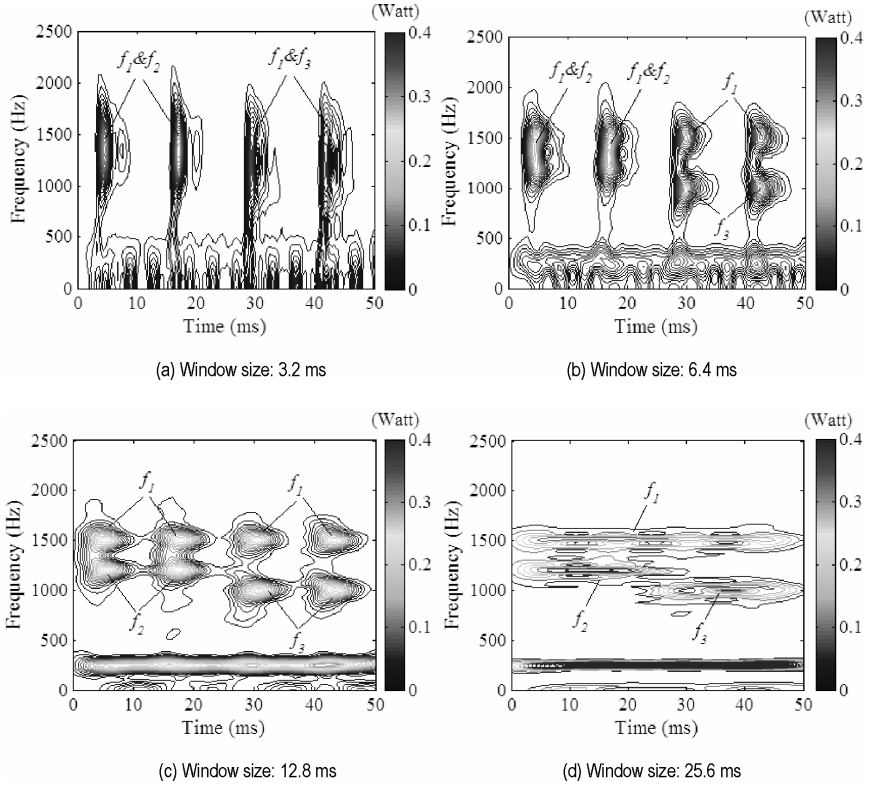
Figure 5 Time-frequency resolutions associated with the STFT technique



The effect of the window size τ on the time and frequency resolutions is illustrated in Figure 6, where STFT with the Gaussian window was performed on the test signal shown in Figure 3, using four different window sizes: 3.2, 6.4, 12.8 and 25.6 ms.

While the smallest window of 3.2 ms has provided good time resolution in separating the four pulse trains contained in the signal, as illustrated in Figure 6a, its frequency resolution was too low to differentiate the two time-overlapped transient elements within each group. As a result, the frequency elements f_1 = 1.5 kHz and f_2 = 1.2 kHz are displayed as one lumped group on the time-frequency plane, and so does the group containing f_1 = 1.5 kHz and f_3 = 1.0 kHz. In contrast, the largest window of 25.6 ms provided a good frequency resolution to illustrate the three frequency elements of 1.0, 1.2 and 1.5 kHz in Figure 6d, but the time resolution was insufficient to differentiate the four pulse trains that are timely separated with a 12.5 ms interval. The best overall performance is given by the window of 12.8 ms, shown in Figure 6c, which allowed for all of the transients to be adequately differentiated on the time-frequency plane. Given that the specific frequency content of a signal encountered during machine monitoring operations are generally not known *a priori*, selection of a suitable window size for effective signal decomposition using the STFT technique is not guaranteed. In addition to the transient elements, the two fundamental frequencies of 40 and 250 Hz were also represented in Figure 6, where the largest window size of 25.6 ms has returned the best frequency resolution.

Figure 6 Results of the STFT of the test signal using four different window sizes



(a) Window size: 3.2 ms (b) Window size: 6.4 ms

(c) Window size: 12.8 ms (d) Window size: 25.6 ms

3.2 Wavelet transform

In contrast to STFT where the window size is *fixed*, the wavelet transform enables *variable* window sizes in analysing different frequency components within a signal: small window for high frequencies and large windows for low frequencies (Daubechies, 1990; Rioul and Vetterli, 1991; Newland, 1994). The result is a *constant relative-bandwidth* frequency analysis, which is well-suited for capturing transient features embedded in time-varying signals. Unlike the Fourier transform, which expresses a signal as the sum of a series of single-frequency sine and cosine functions, the wavelet transform decomposes a signal into a set of functions obtained from the *scaling* (i.e. dilation and contraction) and *shift* (i.e. time translation) of a *mother* wavelet. Mathematically, the wavelet transform of a signal $x(t)$ is defined as (Daubechies, 1990):

$$WT(s,\tau) = \frac{1}{\sqrt{s}} \int x(t) \psi^* \left(\frac{t-\tau}{s} \right) dt \tag{7}$$

where $\psi^*(\bullet)$ denotes the complex conjugation of the mother wavelet $\psi(t)$, τ is the shifting parameter, and $s > 0$ represents the scaling parameter. The parameter s determines the time and frequency resolutions of the scaled mother wavelet $\psi((t-\tau)/s)$.

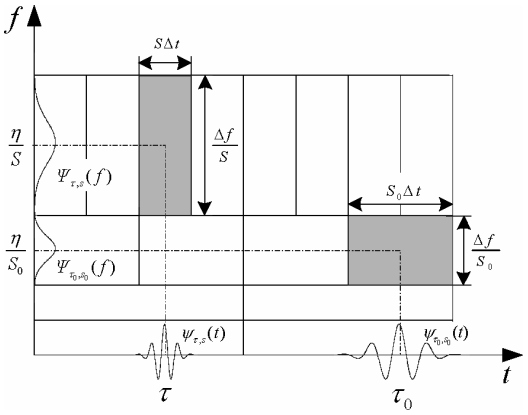
As an example, if the Morlet wavelet $\psi(t) = \exp(i2\pi f_0 t)\exp(-\alpha t^2 / \beta^2)$ is chosen as the mother wavelet, its scaled version will be expressed as:

$$\psi\left[(t-\tau)/s\right] = \exp\left[i2\pi f_0 (t-\tau)/s\right]\exp\left[-\alpha(t-\tau)^2 / s^2\beta^2\right] \tag{8}$$

with the parameters f_0, α, and β being constants. The corresponding time and frequency resolutions of the Morlet wavelet are calculated as $\Delta t = s\beta / 2\sqrt{\alpha}$ and $\Delta f = \sqrt{\alpha} / (s \cdot 2\pi\beta)$, respectively. The expressions indicate that the time and frequency resolutions are directly and inversely proportional to the scaling parameter *s*. In Figure 7, variation of the time and frequency resolutions of the Morlet wavelet at two locations $(\tau, \eta/s)$ and $(\tau_0, \eta/s_0)$ are illustrated, Changing the scale from *s* at the location $(\tau, \eta/s)$ to s_0 = 2*s* at $(\tau_0, \eta/s_0)$ decreases the time resolution by half while doubling the frequency resolution.

The transform operation, in essence, is a *matching* operation that looks for the 'similarity' between the signal being analysed and the mother wavelet that serves as a template. Through variations of the scales and time shifts of the mother wavelet function, the wavelet transform is capable of extracting signal features over the entire signal spectrum, without expecting the signal to have a dominant frequency band as the Fourier transform does.

Figure 7 Time frequency resolutions of the wavelet transform ($s_0 = 2s$)



Signal decomposition through Discrete Wavelet Transform (DWT) can be efficiently realised by means of a pair of low-pass and high-pass wavelet filters (Mallat, 1989), denoted as $h(k)$ and $g(k) = (-1)^k h(1-k)$, respectively. These filters, also known as Quadrature Mirror Filters (QMF), are constructed from the selected mother wavelet $\psi(t)$ and its corresponding scaling function $\phi(t)$, expressed as:

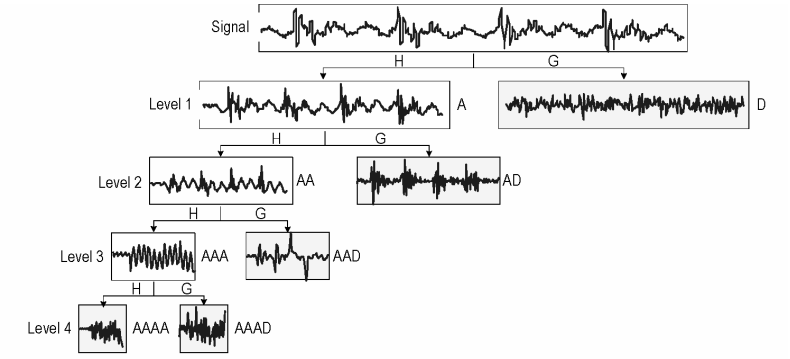
$$\begin{cases} \phi(t) = \sqrt{2}\sum_k h(k)\phi(2t-k) \\ \psi(t) = \sqrt{2}\sum_k g(k)\phi(2t-k) \end{cases} \tag{9}$$

Using the wavelet filters, the signal being analysed is decomposed into a set of low and high-frequency components (Mallat, 1999):

$$\begin{cases} a_{j,k} = \sum_m h(2k-m)a_{m,j-1} \\ d_{j,k} = \sum_m g(2k-m)a_{m,j-1} \end{cases} \tag{10}$$

In Equation (10), $a_{j,k}$ is the *approximation* coefficient, which represents the low-frequency component of the signal, and $d_{j,k}$ is the *detail* coefficient, which corresponds to the high-frequency component. The approximate and detail coefficients at wavelet scale 2^j (with j denoting the level) are obtained by convolving the approximate coefficients at the previous level (j–1) with the low-pass and high-pass filter coefficients, respectively. Such a process is schematically shown in Figure 8, which illustrates how the test signal was decomposed by a four-level discrete wavelet transform. After passing through the high-pass and low-pass filters on the first level, the output of the low-pass filter, denoted as the approximation coefficients, is filtered again by the second-level filter banks. The process repeats itself, and at the end of the fourth level decomposition, the test signal is decomposed into five feature groups: one group containing the lowest frequency components, denoted as the *approximate* information and labeled as *AAAA* in Figure 8, and four groups containing progressively higher frequency components, which are called the *detailed* information and labeled as *AAAD*, *AAD*, *AD*, and *D*. The levels 1–4 correspond to the wavelet scales $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, and $2^4 = 16$, respectively.

Figure 8 Procedure of a four-level signal decomposition using discrete wavelet transform

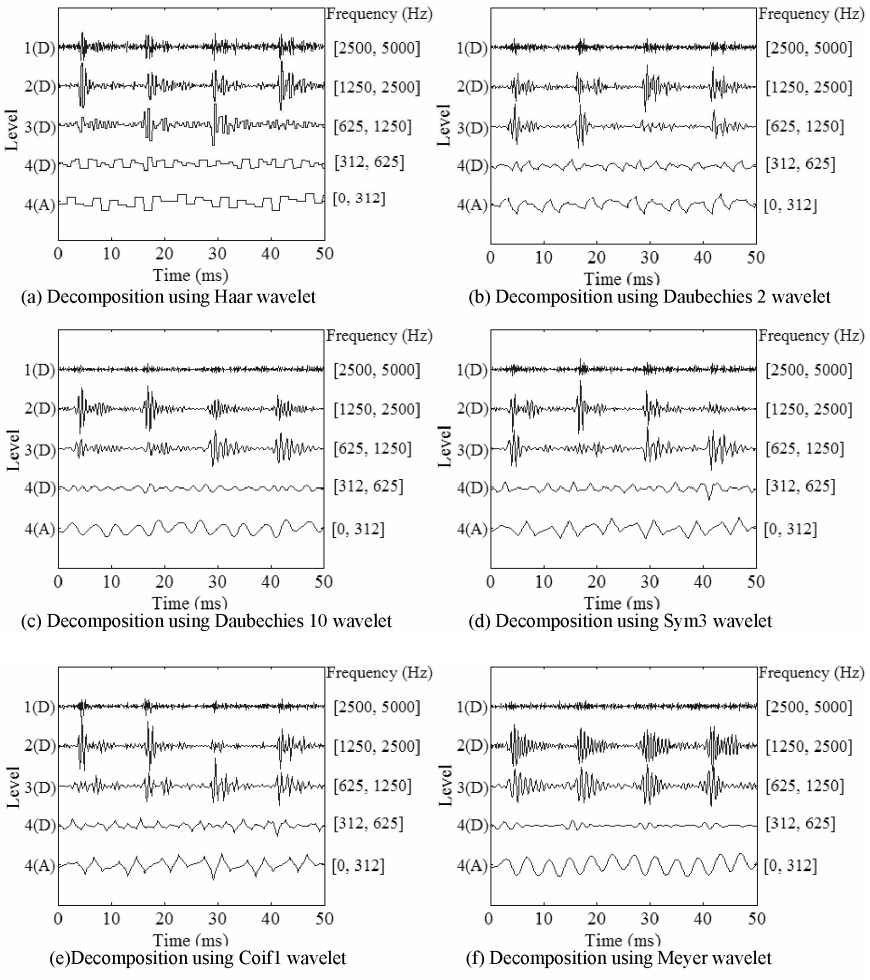


Note: H - Low pass filter; G - High pass filter; A - Approximate information; D - Detailed information

The results of a four-level decomposition of the test signal is shown in Figure 9, where six mother wavelets were employed: Haar, Daubechies 2 and Daubechies 10, Symlet 3, Coiflet 1, and Meyer wavelet (Daubechies, 1992; Chui, 1997; Yan and Gao, 2003). The results indicate that the type of mother wavelet employed directly affects the effectiveness in transient elements identification. For the presented study, transient elements detection was realised by setting up a threshold $t_{hr} = m + 2s_t$, with m and s_t being the mean value and standard deviation of the wavelet coefficients at each decomposition level, respectively (Nikolaou and Antoniadis, 2002). For example, the

Haar wavelet and Daubechies 10 wavelet have shown, in Figures 9a and c, to be able to identify two transient elements at the decomposition level 3 (which corresponds to the 625–1,250 Hz frequency range) and four transients at the decomposition level 2 (corresponding to 1,250–2,500 Hz), out of a total of eight transient elements contained in the test signal. The Symlet 3 wavelet, in Figure 9d, and Coiflet 1 wavelet, in Figure 9e, identified three transient components at decomposition level 2 and 3, respectively, and the Daubechies 2 wavelet, in Figure 9b, identified three transient components at level 3 and four transients at level 2. In comparison, all of the eight transients were successfully identified by the Meyer wavelet in Figure 9f. However, the transient elements f_2 and f_3, with the centre frequencies of 1.2 and 1.0 kHz, respectively, were found at the same decomposition level 3, corresponding to frequency range from 625 to 1,250 Hz. As a result, they could not be differentiated.

Figure 9 Discrete wavelet transform of the test signal using six different mother wavelets



Note: A: Approximate information; D: Detailed information

3.3 Wavelet packet transform

While discrete wavelet transform provides flexible time-frequency resolution as opposed to the STFT, it suffers from a relatively low resolution in the high-frequency region, which leads to difficulty in differentiating high frequency transients, as illustrated in Figure 9. The Wavelet Packet Transform (WPT), in comparison, further decomposes the *detailed* information of the signal in the high frequency region, thus overcomes the drawback. To perform wavelet packet transform of a signal at a certain level (e.g. level 4), the functions in Equation (9) are unified as:

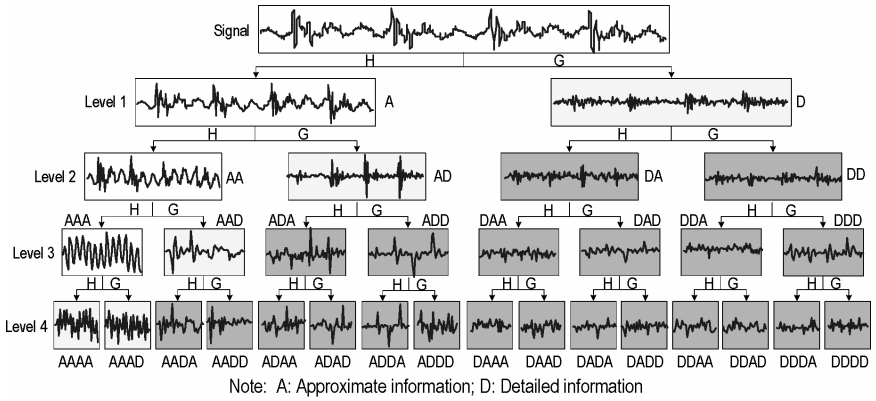
$$\begin{cases} u_{2n}(t) = \sqrt{2}\sum_k h(k)u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2}\sum_k g(k)u_n(2t-k) \end{cases} \tag{11}$$

where $u_0(t) = \phi(t)$, and $u_1(t) = \psi(t)$. Correspondingly, the signal is decomposed as:

$$\begin{cases} d_{j+1,2n} = \sum_m h(m-2k)d_{j,n} \\ d_{j+1,2n+1} = \sum_m g(m-2k)d_{j,n} \end{cases} \tag{12}$$

where $d_{j,n}$ denotes the wavelet coefficients at the j level, n sub-band, $d_{j+1,2n}$ and $d_{j+1,2n+1}$ denotes the wavelet coefficients at the j+1 level, 2n and 2n+1 sub-bands, respectively, and m is the number of the wavelet coefficients. As illustrated in Figure 10, a four-level wavelet packet transform produces a total of sixteen feature groups, with each group covering one sixteenth of the signal frequency spectrum. The enhanced signal decomposition capability makes WPT an attractive tool for detecting and differentiating transient elements with high frequency characteristics.

Figure 10 Procedure for signal decomposition using wavelet packet transform



Another advantage of WPT is that there are multiple ways (> 2^L) to analyse a signal using a L-level decomposition (Mallat, 1999), making it possible to optimise the process and improve the effectiveness. A widely applied criterion for optimal WPT-based signal

representation is the Shannon entropy (Coifman and Wickerhauser, 1992). For wavelet coefficients at the n^{th} sub frequency-band within the level j, $d_{j,n} = \{d_{j,n} : n = 1, 2, ...2^j\}$, the Shannon entropy is defined as:

$$Entropy(d_{j,n}) = -\sum_i p_i \cdot \log(p_i) \tag{13}$$

where p_i is the probability distribution of the energy contained in the wavelet coefficients at the n^{th} sub frequency-band within the level j. The probability distribution function is defined as:

$$p_i = \left|d_{j,n}(i)\right|^2 / \left\|d_{j,n}\right\|^2 \tag{14}$$

with $\sum_{i=1}^{m} p_i = 1$, and $p_i \cdot \log_2 p_i = 0$ if $p_i = 0$. The symbol m represents the number of wavelet coefficients at the n^{th} sub frequency-band within the level j.

Equations (13) and (14) indicate that the entropy of the wavelet coefficients is bounded by:

$$0 \le E_{\text{entropy}}(d_{j,n}) \le \log_2 m \tag{15}$$

Furthermore, Equation (15) indicates that the Shannon entropy will display a large value if the energy content is spread out among the constituent wavelet coefficients within the sub-frequency band, and assume a small value if the energy is concentrated on a few dominant components. Such an inverse relationship between the entropy and energy concentration can be used to differentiate transient elements located within a sub frequency-band. By search for the minimum Shannon entropy, wavelet coefficients with high-energy concentration that correlate with transients of interest can be identified. Mathematically, such a process involves comparing the entropy of the lower level (e.g. in sub-bands *DAAA* and *DAAD*, Figure 10) of the tree structure with the entropy of the higher source level (e.g. sub-band *DAA*), starting from the bottom of the decomposition (e.g. level 4). If the higher level has returned smaller entropy than the sum of the entropies from the lower level, then the higher level sub frequency-band will be retained. Otherwise, it will be replaced by the two sub frequency-bands at the lower level.

Based on the minimum entropy algorithm introduced above, a wavelet packet transform using the Meyer mother wavelet was performed on the test signal. All of the eight transient elements contained in the signal were successfully identified and differentiated, as shown in the time-frequency plane in Figure 11. This indicates that wavelet packet transform provides better performance than discrete wavelet transform for non-stationary signal processing.

3.4 Hilbert-Huang transform

The Hilbert-Huang transform is a time-frequency analysis technique that combines the *Empirical Mode Decomposition* (EMD) with the Hilbert transform (Huang, Shen and Long, 1998). By assuming that the signal being analysed:

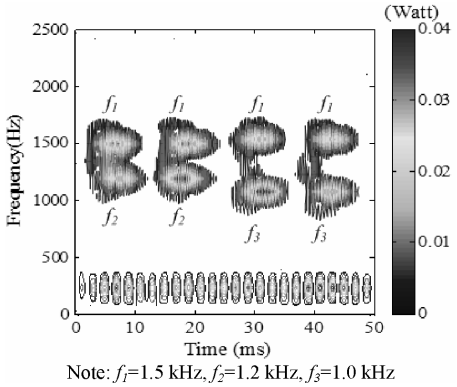
1 has at least two extrema, that is one maximum and one minimum

2 has a characteristic time scale that is defined by the time lapse between successive alternations of the local maxima and minima

3 when having no extrema, still contains inflection points and thus can be differentiated once or more times to reveal the extrema.

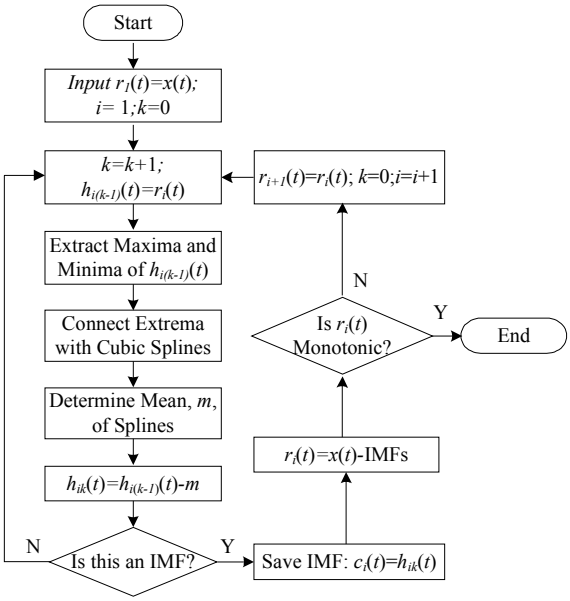
The HHT empirically decomposes a signal into a number of *Intrinsic Mode Functions* (IMFs), with each IMF being a mono-component function, calculated as the difference between the signal and the mean envelope values. Using the cubic Spline function, the upper and lower mean envelopes of the signal are extracted. In Figure 12, the iterative EMD process is shown.

Figure 11 Wavelet packet transform of the test signal



Note: f_1=1.5 kHz, f_2=1.2 kHz, f_3=1.0 kHz

Figure 12 Flow chart of the empirical mode decomposition process



At the end of such a empirical mode decomposition process, the signal $x(t)$ is represented as:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{16}$$

where $c_i(t)$ is the i^{th} intrinsic mode function of the signal, and r_n is the residue. By defining the Hilbert transform of the intrinsic mode function $c_i(t)$ as:

$$H[c_i(t)] = \frac{1}{\pi} \int \frac{c_i(\tau)}{t-\tau} d\tau \tag{17}$$

the analytic signal $z_i(t)$ of $c_i(t)$ can be constructed as:

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t)e^{j\Phi_i(t)} \tag{18}$$

where $a_i(t)$ and $\Phi_i(t)$ represent the amplitude and phase angle of the analytic signal, respectively. Taking derivative of the phase angle yields:

$$\omega_i = \frac{d(\Phi_i(t))}{dt} \tag{19}$$

Substituting ω_i into Equation (18) yields:

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t,\omega_i)e^{j\int \omega_i(t)dt} \tag{20}$$

Since $c_i(t)$ is the real part of the analytic signal $z_i(t)$, by neglecting the residue component r_n, Equation (16) can be modified as:

$$x(t) = \Re\left[\sum_{i=1}^{n} a_i(t,\omega_i)e^{j\int \omega_i(t)dt}\right] \tag{21}$$

where the symbol $\Re(\bullet)$ denotes the real part of the signal, and $a_i(t,\omega_i)$ is the amplitude of the i^{th} IMF at time t with frequency ω_i. As a result, the Hilbert-Huang transform of a signal $x(t)$ can be expressed as:

$$HHT(t,\omega) = \sum_{i=1}^{n} a_i(t,\omega_i) \tag{22}$$

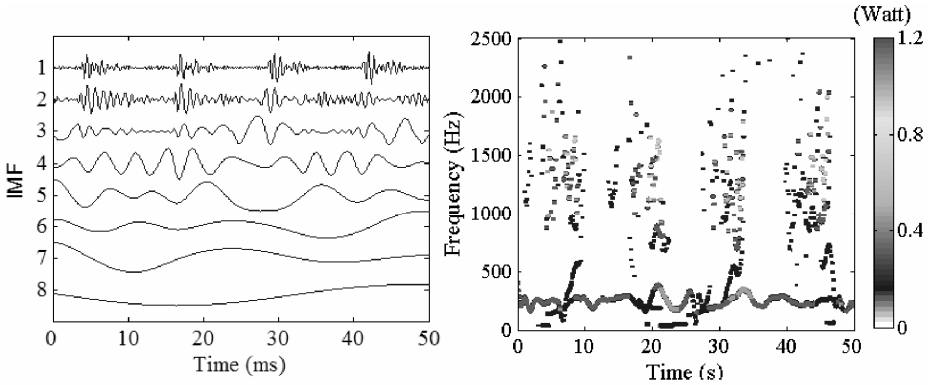
Comparing with the Fourier-based representation of the signal $x(t)$:

$$x(t) = \Re(\sum_{i=1}^{\infty} a_i e^{j\omega_i t}) \tag{23}$$

where a_i and ω_i are constant, it can be seen that the time-varying expression for amplitude and instantaneous frequency, as shown in Equation (21), make the HHT a viable tool for analysing non-stationary signals. Figure 13 illustrates the result of Hilbert-Huang transform of the test signal with the signal's intrinsic mode functions shown in Figure 13a and the Hilbert-Huang spectrum shown in Figure 13b. While the four groups

of transient components contained in the test signal were separated along the time axis, the time-overlapped transients were not properly differentiated along the frequency axis, as the instantaneous frequency components were mixed within each group.

Figure 13 Hilbert-Huang transform of the test signal



(a) IMF extracted from the signal

(b) Hilbert-Huang spectrum of the signal

4 Experimental study

The performance of the each of the four time-frequency analysis techniques was evaluated using vibration signals acquired from a ball bearing undergoing a run-to-failure test. The bearing under test was installed on a bearing test system, as illustrated in Figure 14. Physically, when a localised defect is initiated in a rolling element bearing, for example due to spilling on the surface of the bearing raceway, impacts will be generated every time when the rolling elements roll over the defect. Such impacts subsequently excite the intrinsic modes of the bearing system, giving rise to transient vibrations at the mode-related resonant frequencies. As the defect size increases, different intrinsic modes of the bearing system will be excited, leading to frequency changes of the impact-induced transient vibrations. Therefore, by evaluating the time-frequency distributions of the vibration signal, degradation of the bearing's health condition can be monitored.

To simulate a localised structural defect, a 0.27 mm groove was introduced across the outer raceway of the test bearing (model 1100KR from Timken). The defect size has grown to approximately 5.5 mm, after the bearing has run continually for 2.3 million revolutions. Upon reaching 2.7 million revolutions, the defect had propagated throughout the entire raceway and rendered the bearing practically non-functional. Throughout the test, changes of the amplitude of the bearing vibration were monitored continually using accelerometers. As shown in Figure 15, three defect propagation phases (I–III) could be identified, correlating to the growth of the defect size.

Systematic data analyses were performed on data segments, each of which 100 ms long. To demonstrate the applicability of the time-frequency analysis techniques, two data segments are shown in Figures 16, which have captured the frequency changes along the time and thus can serve as a quantitative measure for bearing defect propagation. The first data segment comprises data points covering the transition from Phase I to Phase II,

and the second segment representing transition from Phase II to Phase III. The power spectrum of the two data segments shown in Figure 16b and 16d only illustrated the average frequency information over the analysed time period, without being able to identify the propagation of the bearing defect.

Figure 14 Schematic view of the experimental bearing test system

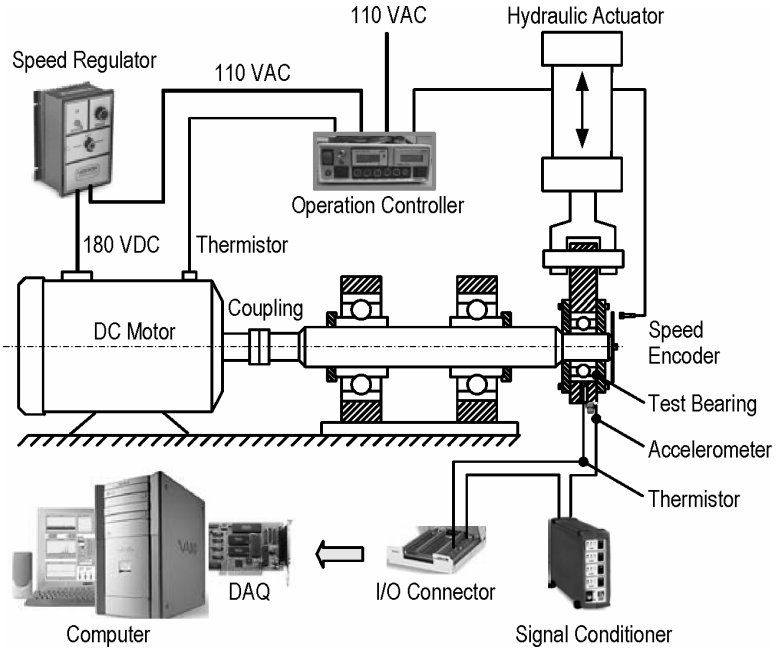


Figure 15 Growth of vibration amplitude as a function of the bearing revolution

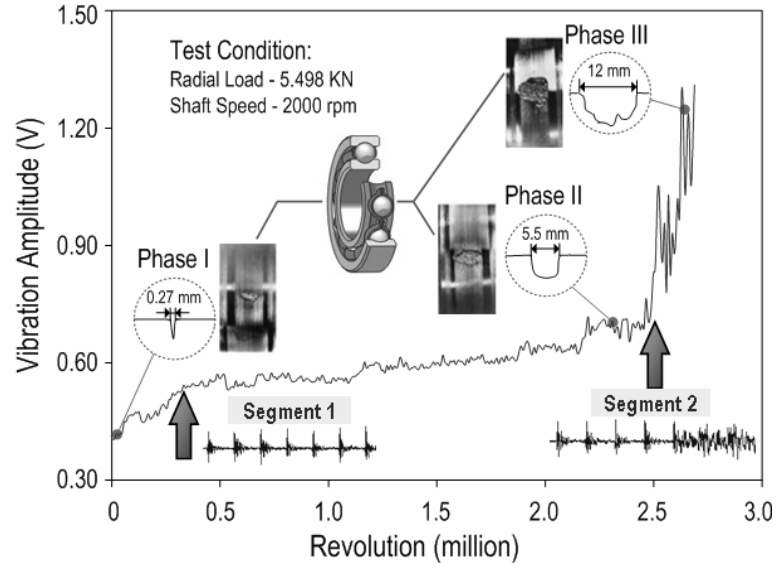
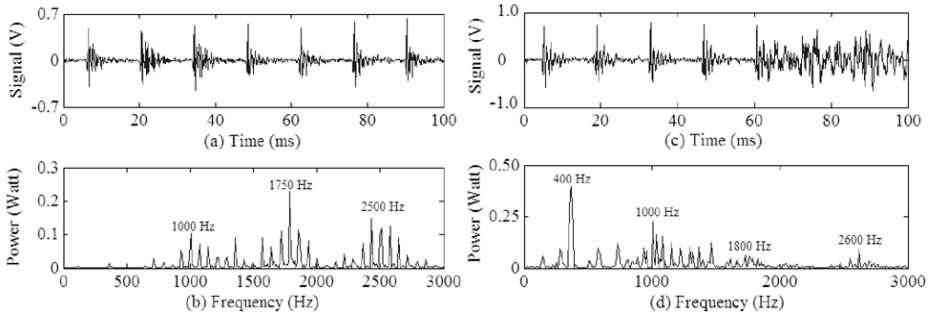


Figure 16 Vibration data segments measured from the run-to-failure bearing test



Applying the short-time Fourier transform to the data segments using a 12.8 ms window, which was chosen from an empirical *trial-and-error* search process, two major frequency components at 1,750 and 2,500 Hz and about 14 ms apart from each other were identified. This time interval corresponds to a 71 Hz repetition frequency of the bearing outer raceway based on the geometry of the test bearing and its rotating speed (Harris, 1991). After the time point of 45 ms, another dominant frequency component of 1,000 Hz was identified, as shown in Figure 17a. Such a change of dominant frequency components correlated well with the defect size growth. As the defect further propagates from Phase II to Phase III, another new, dominant frequency component of 400 Hz was identified, as shown in Figure 17b.

Figure 17 Short-time Fourier transform of bearing vibration signals at defect growth phase transitions

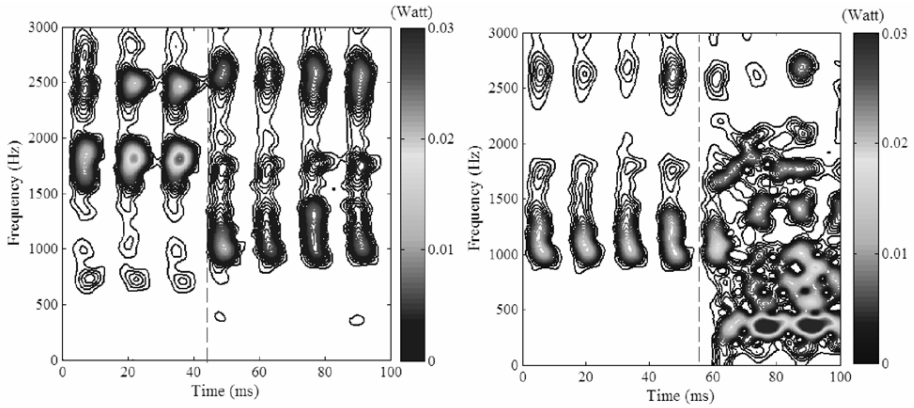
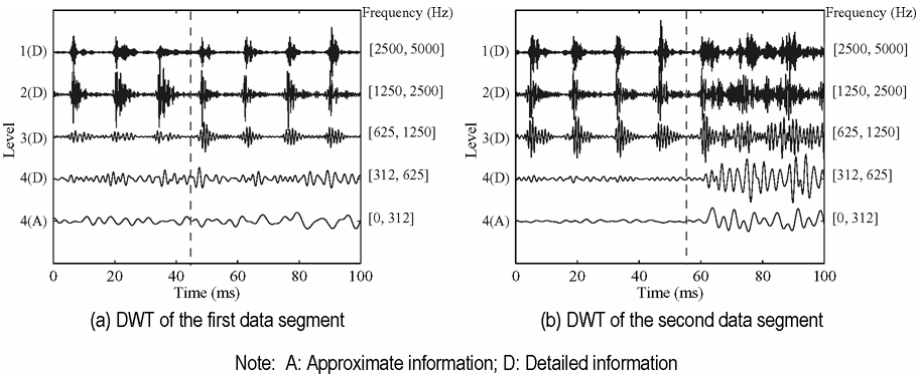


Figure 18 illustrates the result of a four-level discrete wavelet transform of the two data segments, using the Meyer wavelet as the mother wavelet. Major transient elements with a time separation of 14 ms were identified at the first two decomposition levels, for both segments. These two decomposition levels cover the frequency sub-bands of 1,250–2,500 Hz, and 2,500–5,000 Hz, respectively. After the time point of 45 ms, another four transient elements are seen at the third decomposition level (corresponding to the

frequency sub-band of 625–1,250 Hz).. Furthermore, in the data segment corresponding to the Phase II to Phase III transition, significant changes in the wavelet coefficients at each decomposition level can be seen after the time point of 55 ms, as shown in Figure 18b. This indicates that more frequency components were generated during the bearing test, as the bearing health condition continued to deteriorate. Such a conclusion could be confirmed when the test bearing was disassembled for visual inspection.

Figure 18 Discrete wavelet transform of the vibration signal



The result of data analysis using the Hilbert–Huang transform is shown in Figure 19. While the existence of the transient elements was identified in general, the frequency shifts corresponding to the phase transitions could not be clearly seen.

Applying the wavelet packet transform to the two data segments, it is seen in Figure 20 that not only all the major transient elements could be identified, but the corresponding frequency shifts could also be clearly seen. The result also shows the increased number of frequency components after the 45 ms and 55 ms time point for the phase transitions from I to II and from II to III, respectively, reflecting the defect size propagation.

Figure 19 Hilbert-Huang transform of the bearing vibration signal

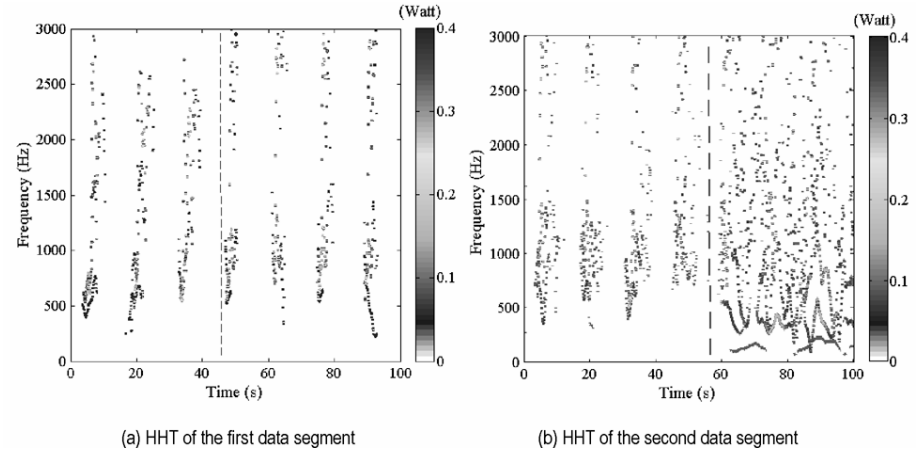
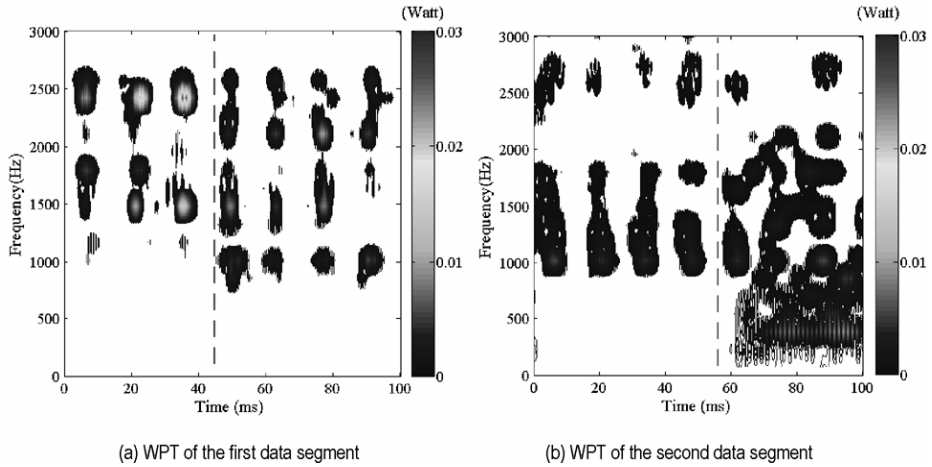


Figure 20 Wavelet packet transform of the bearing vibration signal



5 Conclusions

Choosing a proper signal processing technique is of critical importance to successfully extracting features hidden within vibration signals measured from rolling bearings and other machine components. To overcome the deficiency in conventional spectral analysis technique for non-stationary signal analysis, four representative time-frequency analysis techniques were comparatively studied in this paper. The wavelet packet transform has shown to be able to identify all the transient elements embedded within both the analytically formulated test signal and the realistic bearing vibration signal. In addition, it was able to clearly identify the frequency shifts associated with the phase transitions of defect growth. The short-time Fourier transform, upon selection of appropriate window size, was also able to detect the frequency shifts associated with bearing defect propagation. In comparison, the discrete wavelet transform and Hilbert-Huang transform were not successful in the identification. Research is being continued to further explore the constrains and limitations of time-frequency techniques, with the goal to develop effective and efficient algorithms for machine health monitoring and diagnosis.

Acknowledgements

The authors gratefully acknowledge funding provided to this research by the National Science Foundation under award # DMI-0218161. Experimental support from the SKF and Timken companies are appreciated.

References

Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., König, W. and Teti, R. (1996) 'Tool condition monitoring (TCM) – the status of research and industrial application', *Annals of the CIRP*, Vol. 44, pp.541–567.

Charley, J., Bodovillé, G. and Degallaix, G. (2001) 'Analysis of braking noise and vibration measurements by time-frequency approaches', *Proceedings of the IMechE, Part C: Journal of Mechanical Engineering Science,* Vol. 215, pp.1381–1400.

Chen, D. and Wang, W.J. (2002) 'Classification of wavelet map patterns using multi-layer neural networks for gear fault detection', *Mechanical Systems and Signal Processing*, Vol. 16, pp.695–704.

Chui, C.K. (1997) *Wavelet: a Mathematical Tool for Signal Analysis.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Cohen, L. (1989) 'Time-frequency distributions – a review', *Proceedings of the IEEE*, Vol. 77, pp.941–981.

Coifman, R.R. and Wickerhauser, M.V. (1992) 'Entropy based algorithms for best basis selection', *IEEE Transactions on Information Theory,* Vol. 38, pp.713–718.

Daubechies, I. (1990) 'The wavelet transform, time-frequency localization and signal analysis', *IEEE Transactions on Information Theory*, Vol. 36, pp.961–1005.

Daubechies, I. (1992) *Ten Lectures on Wavelets.* Philadelphia, PA: SIAM.

DimlaSnr, D.E. (2000) 'Sensor signals for tool-wear monitoring in metal cutting operations –a review of methods', *Int. J. of Machine Tools and Manufacture,* Vol. 40, pp.1073–1098.

El-Wardany, T.I., Gao, D. and Elbestawi, M.A. (1996) 'Tool condition monitoring in drilling using vibration signature analysis', *Int. J. of Machine Tools and Manufacture,* Vol. 36, pp.687–711.

Frankowiak, M., Grosvenor, R. and Prickett, P. (2005) 'A review of the evolution of microcontroller-based machine and process monitoring', *Int. J. of Machine Tools and Manufacture*, Vol. 45, pp.573–582.

Ganesan, R., Das, T.K. and Venkataraman, V. (2004) 'Wavelet-based multiscale statistical process monitoring: a literature review', *IIE Transactions*, Vol. 36, pp.787–806.

Goumas, S., Zervakis, M. and Stavrakakis, G. (2002) 'Classification of washing machine vibration signals using discrete wavelet analysis for feature extraction', *IEEE Transactions on Instrumentation and Measurement*, Vol.51, pp.497–508.

Gu, S., Ni, J. and Yuan, J. (2002) 'Non-stationary signal analysis and transient machining process condition monitoring', *Int. J. of Machine Tools & Manufacture,* Vol. 42, pp.41–51.

Harris, T.A. (1991) *Rolling Bearing Analysis* (3rd ed.), New York: Wiley.

Huang, N.E., Shen, Z. and Long, S.R. (1998) 'The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis', *Proceedings of the Royal Society of London, A*, Vol. 454, pp.903–995.

Julius, S.B. and Allan, G.P. (1986) *Random Data Analysis and Measurement Procedures*, (2nd ed.) New York: John Wiley & Sons.

Kamarthi, S., Kumara, S. and Cohen P. (2000) 'Flank wear estimation in turning through wavelet representation of acoustic emission signals', *ASME J. of Manufacturing Science and Engineering*, Vol. 122, pp.12–19.

Kim, J., Kang, M., Ryu, B. and Ji, Y. (1999) 'Development of an on-line tool-life monitoring system using acoustic emission signals in gear shaping', *Int. J. of Machine Tools and Manufacture,* Vol. 39, pp.1761–1777.

Lee, B. and Tang, Y. (1999) 'Application of the discrete wavelet transform to the monitoring of tool failure in end milling using the spindle motor current', *Int. J. of Advanced Manufacturing Technology*, Vol. 15, pp.238–243.

Li, X., Tso, S. and Wang, J. (2000) 'Real-time tool condition monitoring using wavelet transforms and fuzzy techniques', *IEEE Transactions on Systems, Man and Cybernetics – Part C: Application and Reviews*, Vol. 30, pp.352–357.

Liang, S., Hecker, R. and Landers R. (2004) 'Machining process monitoring and control: the state-of-the-art', *ASME J of Manufacturing Science and Engineering*, Vol. 126, pp.297–310.

Lou, X. and Loparo, K.A. (2004) 'Bearing fault diagnosis based on wavelet transform and fuzzy inference', *Mechanical Systems and Signal Processing*, Vol. 18, pp.1077–1095.

Mallat, S. (1989) 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, pp.674–693.

Mallat, S. (1999) *A Wavelet Tour of Signal Processing*, San Diego, CA: Academic Press.

Morhain, A. and Mba, D. (2003) 'Bearing defect diagnosis and acoustic emission', *Proceedings of IMechE, Part J: Journal of Engineering Tribology*, Vol. 217, pp.257–271.

Mori, K., Kasashima, N., Yoshioka, T. and Ueno, Y. (1996) 'Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals', *Wear*, Vol. 195, pp.162–168.

Newland, D.E. (1994) 'Wavelet analysis of vibration – Part I: Theory; Part II: Wavelet maps', *ASME J. of Vibration and Acoustics,* Vol. 116, pp.409–425.

Nikolaou, N.G. and Antoniadis, I.A. (2002) 'Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted Morlet wavelets', *Mechanical Systems and Signal Processing*, Vol. 16, pp.677–694.

Padovese, L.R. (2004) 'Hybrid time–frequency methods for non-stationary mechanical signal analysis', *Mechanical Systems and Signal Processing,* Vol.18, pp.1047–1064.

Rioul, O. and Vetterli, M. (1991) 'Wavelets and signal processing', *IEEE Signal Processing Magazine*, pp.14–38.

Satish, L. (1998) 'Short-time Fourier and wavelet transform for fault detection in power transformers during impulse tests', *IEE Proceedings – Science, Measurement and Technology*, Vol. 145, pp.77–84.

Shi, D.F., Tsung, F. and Unsworth, P.J. (2004) 'Adaptive time–frequency decomposition for transient vibration monitoring of rotating machinery', *Mechanical Systems and Signal Processing,* Vol. 18, pp.127–141.

Teti, R. (1996) 'A review of tool condition monitoring literature data base', *Annals of the CIRP*, Vol. 44 , pp.659–666.

Tönshoff, H., Wulfsberg, J., Kals, H. and König, W. (1988) 'Developments and trends in monitoring and control of machining processes', *Annals of the CIRP*, Vol. 37, pp.611–622.

Tseng, P.C. and Chou, A. (2002) 'The intelligent on-line monitoring of end milling', *Int. J. of Machine Tools and Manufacture,* Vol. 42, pp.89–97.

Wang, C. and Gao, R. (2003) 'Wavelet transform with spectral post-processing for enhanced feature extraction', *IEEE Transactions on Instrumentation and Measurement*, Vol. 52, pp.1296–1301.

Wang, W.J. and McFadden, P.D. (1995) 'Application of orthogonal wavelets to early gear damage detection', *Mechanical Systems and Signal Processing*, Vol. 9, pp.497–507.

Wu, Y., Escande, P. and Du, R. (2001) 'A new method for real-time tool condition monitoring in transfer machining stations', *ASME J. of Manufacturing Science and Engineering*, Vol. 123, pp.339–347.

Yan, R. and Gao, R. (2003) 'A hybrid signal processing approach to sensor data analysis', *ASME International Mechanical Engineering Congress and Exposition (IMECE), Symposium on Intelligent Sensors and Sensor Networks, paper # IMECE2003-42586, Washington, D.C.*, pp.1159–1166.

Yan, R. and Gao, R. (2005) 'An efficient approach to machine health diagnosis based on harmonic wavelet packet transform', *Robotics and Computer-Integrated Manufacturing*, Vol. 21, pp.291–301.

Yu, D., Cheng, J. and Yang, Y. (2005) 'Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings', *Mechanical Systems and Signal Processing*, Vol. 19, pp.259–270.

Zheng, H., Li, Z. and Chen, X. (2002) 'Gear fault diagnosis based on continuous wavelet transform', *Mechanical Systems and Signal Processing*, Vol. 16, pp.447–457.