

Feature Extraction for Rolling Element Bearing Faults Using Resonance Sparse Signal Decomposition

W. Huang¹ · H. Sun¹ · Y. Liu² · W. Wang¹

Received: 9 October 2016 / Accepted: 13 February 2017 / Published online: 27 February 2017
© The Society for Experimental Mechanics, Inc 2017

Abstract Rolling element bearings are widely used in a variety of rotating machineries. If the rolling bearing elements are damaged, a cyclical impact transient signal and the vibration signal modulation phenomenon appears when the fault surface contacts other components of the rolling element bearing. To demodulate the cyclical impact signal and extract the bearing fault information, this paper proposes a new method based on resonance-based sparse signal decomposition (RSSD). First, the bearing vibration signal is decomposed into three components via RSSD. The high-resonance component contains a sustained oscillation cycle signal, the low-resonance component contains the impact transient signal, and the final component is the residual. The sub-bands near the natural bands are extracted for demodulation into two components. Two main sub-bands are obtained by summing these sub-bands. Next, these two main sub-bands are summed to obtain the original signal's main sub-band. Finally, the auto power spectrum is extracted using envelope signal autocorrelation processing, and it reflects the degree and location of the fault in the rolling bearing. To verify its effectiveness in extracting fault information, the proposed method is applied to two practical application examples with an inner race fault and an outer race fault in a rolling bearing, respectively. Compared with envelope analysis and wavelet analysis, the results indicate that the spectra obtained with this method exhibit less burrs and a higher signal-noise ratio, and outperforms the other

spectra in terms of revealing the amplitude modulation frequency of the fault impact response.

Keywords Rolling element bearing · Fault diagnosis · Feature extraction · Resonance sparse signal decomposition · Main sub-bands

Introduction

Condition monitoring and fault diagnosis of machines is gaining importance because of the need to improve product reliability and decrease the possible loss of production due to machine breakdown [1]. Due to the importance of rolling bearings as one of the most widely used industrial machinery elements, the development of a proper monitoring and fault diagnosis procedure is necessary to prevent the malfunctioning and failure of rolling bearings [2]. Rolling bearing faults can occur for many reasons, e.g., improper design, improper mounting, acid corrosion, poor lubrication, and plastic deformation. The most common defect is spalling due to material fatigue after a period of operation, which is initiated by micro cracking under the surface of bearing elements. These cracks propagate toward the surface under the effect of periodical loads and cause cavities or crushing on the surface. Statistical data [1] indicate that 90 % of faults occurring in rolling bearings are due to cracks in the inner and outer race, and the remainder is due to cracks in the balls or cage.

If the inner race, outer race, or rolling element of rolling bearings is damaged, a cyclical impact transient signal is generated with the rotation of the bearing when the fault surface contacts other components of the rolling element bearing. At the same time, a vibration signal modulation phenomenon appears with a characteristic structure and moving

✉ W. Huang
hwt@hit.edu.cn

¹ School of Mechatronics Engineering, Harbin Institute of Technology, 92 West Dazhi St, Harbin, China

² China Aviation Power Machinery Research Institute, Zhuzhou, China

relationship. Therefore, extraction and demodulation of cyclical impact signal is the key in bearing fault diagnosis [3].

The vibration signals of defective bearings are typically complicated and are mixed with a variety of signals [4]. For local faults, especially early faults, the fault features are not sufficiently strong to be extracted from vibration signals because of the varying degrees of impact from the transmission path, the normal vibration of other parts, and the presence of noise [5]. Therefore, the study of an appropriate signal processing method that can separate the rolling bearing vibration signal components effectively has caught the attention of many scholars.

In 2009, Li et al. [6] proposed a Hilbert-Huang transform and marginal spectrum for the detection and diagnosis of local faults in rolling bearings. In 2013, Osman et al. [7] developed an enhanced Hilbert-Huang transform technique to diagnose rolling bearing faults and noise. In 2010, Zvokelj et al. [8] reported a new multivariate and multi-scale method that combined ensemble empirical mode decomposition (EEMD) and principal component analysis for large-size and low-speed bearing fault monitoring. In 2015, Cai et al. [9] presented a novel and adaptive procedure based on EEMD and Hilbert marginal spectrum. The effectiveness was proven by the successful diagnosis of an axle bearing with a single fault or multiple composite faults. In 2012, Guo et al. [10] researched methods for extracting the bearing fault signal from strong noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. The Mahalanobis-Taguchi system based on EMD-SVD was applied to fault diagnosis and health assessment of bearings by Wang et al. [11]. In 2013, Zheng et al. [12] introduced an adaptive data-driven analysis method, known as the generalized empirical mode decomposition (GEMD), which was applied to rolling bearing fault diagnosis. In 2011, Li et al. [13] applied a weighted multi-scale morphological gradient filter to detect rolling element bearing faults. In 2012, Li et al. [14] proposed a continuous-scale mathematical morphology based on the optimal-scale band demodulation of impulsive features for bearing fault diagnosis. In 2013, Raj et al. [15] focused on the early classification of bearing faults using morphological operators and fuzzy inference. Wavelet analysis, a superior multi-scale time-frequency analysis method, was successfully applied to signal de-noising [16–18] and fault feature extraction [19–21] for rolling bearing fault signals.

The exploration and research of these scholars advanced the progress of rolling bearing fault diagnosis. However, the commonly used methods include wavelet analysis and empirical mode decomposition (EMD). These two signal processing methods are both based on frequency band division and cannot effectively extract the impact component if the center frequency of the impact signal and the center frequency of other signals overlap. The wavelet basis function cannot be changed once it is chosen, and the rolling bearing vibration signal is

complex; thus, it is not suitable for decomposition using a single wavelet basis function. These obstacles present a challenge to rolling bearing fault diagnosis.

In 2011, Selesnick [22] proposed the use of the resonance-based sparse signal decomposition method (RSSD). In contrast to the traditional signal decomposition method based on the frequency band division, a complex signal is decomposed by RSSD into high- and low-resonance components according to the quality factor (defined as the ratio of the center frequency and frequency bandwidth, denoted as Q). The high-resonance component is composed of sustained oscillation cycle signals. Here, an oscillation cycle signal can be thought as the signal that oscillates about a cycle of a sine in the time domain. Thus, the high-resonance component usually exhibits sustained oscillatory behavior. Unlike the high-resonance component, low-resonance component is composed of non-oscillatory transient impulses that do not exhibit sustained oscillatory behavior. The transient impulse signal is a wideband signal that possesses a low Q -factor, whereas the sustained oscillation periodic signal is a narrowband signal with a high Q -factor. Therefore, effective separation of transient impact signal and sustained oscillation cycle signals can be achieved based on the different Q -factors. In 2013, Chen [3] and Mo [2] assumed that the impacts of rubbing faults and rolling bearing faults are transient, and therefore, during RSSD of the vibration signal, the low-resonance component was used to represent these fault impacts. In 2015, a nonlinear demodulation analysis method based on resonance was introduced [23], and it was applied successfully to the fault diagnosis of rolling bearings. In 2014, Wang [24] combined EEMD with tunable Q -factor wavelet transform (TQWT) to obtain satisfactory extraction result. An ensemble super-wavelet transform based on the combination of TQWT and Hilbert transform was put forward for investigating vibration features of motor bearing faults [25]. In 2015, Li et al. [26] published a new method based on the resonance-based sparse signal decomposition with the optimal Q -factor and extracted impact fault feature from low-resonance component by Hilbert envelope demodulation method. In 2016, Tang [27] decomposed the original signal by TQWT, and adopted kurtosis criterion and correlation coefficient criterion to guide the analysis of the remained signal components. However, among the most methods, only the low-resonance component was used to extract the fault information. In fact, when RSSD is applied to rolling bearing fault signal, because the signal is complex and the quality factor wavelet is selectable, both the high- and low-resonance components contain a wealth of fault information.

This work proposes a new method that combines the main sub-bands of high- and low-resonance components based on RSSD and is applied to rolling bearing fault diagnosis. First, the bearing fault signal is decomposed into three components using RSSD. The high-resonance component contains a sustained oscillation cycle signal, the low-resonance component contains the rolling fault impact transient signal, and the

final component is the residual. Next, according to the proposed concept of the main sub-band, envelope demodulation analysis is applied to main sub-bands of both high- and low-resonance components. The bearing fault features are extracted according to the envelope demodulation spectrum. The analysis example illustrates that this method is able to extract the impact elements in bearing fault diagnosis and highlights the fault features effectively. The method proposed in this paper provides a new solution for feature extraction of rolling bearing fault signals.

Basic Theory of RSSD

In contrast to the traditional signal decomposition method, RSSD is a new signal processing method based on morphological component analysis. The signal is decomposed by RSSD based on the characteristics of the signal waveform (vibration time, not vibration speed). The RSSD considers resonance to be a property of the signal. The high-resonance component is defined as a signal consisting of multiple simultaneous sustained oscillations, and the low-resonance component is defined as a signal consisting of non-oscillatory transients of unspecified shape and duration. Therefore, a complex signal can be decomposed by RSSD into high- and low-resonance components. The concept of the Q -factor is used to quantify the degree of the resonance. As shown in Fig. 1, a signal with a higher resonance degree exhibits a higher degree of frequency aggregation, a higher Q -factor, and more visible oscillations in the time-domain waveform at the same time (pulses 2 and 4 in Fig. 1), and vice versa (pulses 1 and 3 in Fig. 1). The high-resonance component is the sum of a sparse series of wavelets with the same high Q -factor Q_1 , and the low-resonance component is the sum of a sparse series of wavelets with the same low Q -factor Q_2 .

The basic theory of RSSD is a sparse representation of a complex signal using two wavelet basis functions with different Q -factors. The higher-oscillation component of the signals can be represented with high Q -factor wavelets, and the lower-oscillation component can be represented with low Q -factor wavelets. From this perspective, the theoretical foundation of RSSD contains wavelets with tunable Q -factors and sparse signal decomposition.

Wavelet with a Tunable Q -Factor

In RSSD, the Q -factor is defined as the ratio of the center frequency and frequency bandwidth, which is widely used in filter design, control, and the physics dynamical systems. The quantity reflects the frequency aggregation degree of the signal. A higher Q -factor indicates a higher degree of frequency aggregation and more oscillations in the time-domain waveform; this is the primary problem in

how to obtain proper wavelet function according to the Q -factor in RSSD. The binary wavelet transform is a common method with a constant Q -factor, but it is not suitable for high-frequency resolution situations due to its small Q -factor. The rational-dilation wavelet transform (RADWT) was proposed by Bayram [28]. Selesnick [22] proposed a tunable Q -factor wavelet transform (TQWT) based on RADWT. The wavelets can be designed using the Q -factor and redundancy factor r in TQWT, which further increases the flexibility in selecting the Q -factor and makes it more convenient to obtain the wavelets. The low-pass scaling factor α and high-pass scaling factor β [22] can be obtained according to formula (1):

$$\beta = \frac{2}{Q+1}, \alpha = 1 - \frac{\beta}{r} \quad (1)$$

In TQWT, all of the wavelet basis function libraries generated by the determined Q and r have the same Q -factor. As shown in Fig. 2, the center frequency (f_c) and bandwidth (BW) are different in different layers in the same wavelet basis function library. The component represented by the j -th level wavelet in a signal is known as the j -th sub-band. It assumes that the sampling frequency of the input signal is f_s [22]. The center frequency f_c and BW of each sub-band can be obtained according to the following two formulas [22]:

$$f_c \approx \alpha^j \frac{2^{-\beta}}{4\alpha} f_s \quad (2)$$

$$\text{BW} = \frac{1}{2} \beta \alpha^{j-1} \pi \quad (3)$$

Sparse Signal Decomposition

Different signal components based on the oscillation characteristics are nonlinearly decomposed by RSSD using morphological component analysis [16]. The method establishes the optimal sparseness of the high- and low-resonance components. The specific steps are described as follows:

Step 1 Choose the appropriate Q_1 and Q_2 according to the waveform characteristics of the actual signal and obtain the corresponding wavelet basis function libraries S_1 and S_2 via a wavelet transform with a tunable Q -factor.

For many vibration signals (i.e., the rolling bearing vibration signal), a sustained oscillation component (i.e., fault impact response) and abrupt peaks (i.e., random noise peaks) will exist. These two components are not easy to distinguish via oscillation frequency but can be distinguished by the quality factor. After obtaining the wavelet basis functions S_1 , S_2 with Q_1 and Q_2 ($Q_1 > Q_2$) via TQWT, signal X can be expressed as

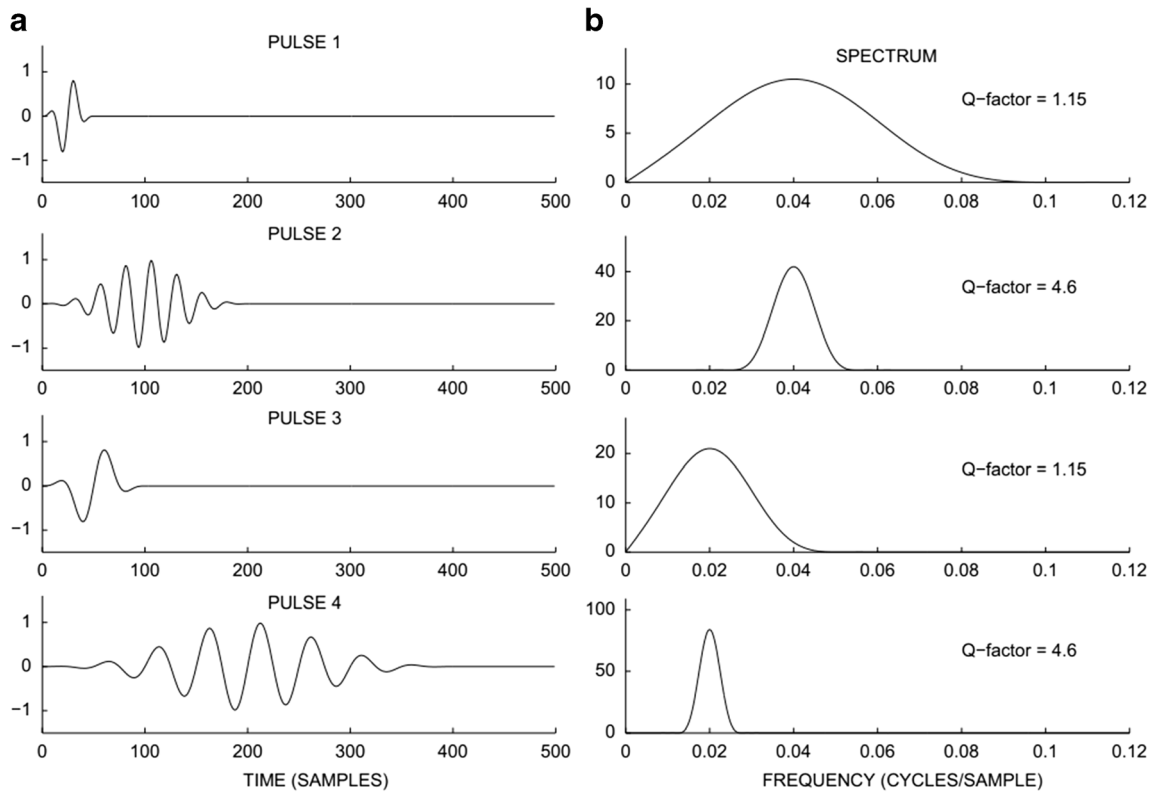


Fig. 1 Signal resonance: (a) Signals. (b) Spectra

$$X = W_1 S_1 + W_2 S_2 + n \tag{4}$$

In equation (4), W_1 and W_2 are the wavelet coefficients of S_1 and S_2 , respectively, $W_1 S_1$ and $W_2 S_2$ are the high- and low-resonance components of the signal X , respectively, and n is the component that cannot be represented by $W_1 S_1$ and $W_2 S_2$ in signal X , which is known as the named residual component or error term. If many minimal values in coefficient matrixes W_1 , W_2 exist and n is notably small at the same time, equation (4) represents the sparse expression of signal X .

Step 2 Define the objective function: Obtain the optimal transformation coefficient matrix under the minimal objective function.

The objective of RSSD is to separate the different resonance components from one signal and minimize the coupling between these two parts. The objective function is defined as

$$J(W_1, W_2) = \|X - S_1 W_1 - S_2 W_2\|_2^2 + \sum_{j=1}^{J_1+1} \lambda_{1,j} \|W_{1,j}\|_1 + \sum_{j=1}^{J_2+1} \lambda_{2,j} \|W_{2,j}\|_1 \tag{5}$$

where J_1, J_2 are the decomposition levels of the high- and low-resonance components, respectively, $W_{1,j}, W_{2,j}$ are the coefficient matrices of the high- and low-resonance components of each level wavelet, respectively, and $\lambda_{1,j}, \lambda_{2,j}$ denote

the weight coefficients of the high- and low-resonance components, respectively. The value of $\lambda_{1,j}$ is correlated with the energy (2 norm of $S_{i,j}$) of $S_{i,j}$ ($i = 1, 2, j = 1, 2 \dots J_i + 1$). Generally, the value of $\lambda_{i,j}$ is proportional to the energy of $S_{i,j}$, denoted as

$$\lambda_{1,j} = L_{1i} \|S_{1,j}\|_2 \quad (i = 1, 2, \dots, J_1 + 1) \tag{6}$$

$$\lambda_{2,j} = L_{2i} \|S_{2,j}\|_2 \quad (i = 1, 2, \dots, J_2 + 1) \tag{7}$$

the value ranges of $\lambda_{1,j}$ ($j = 1, 2 \dots J_1 + 1$) and $\lambda_{2,j}$ ($j = 1, 2 \dots J_2 + 1$) are (0,1,1).

The split augmentation Lagrange search algorithm (SALSA) is applied to the iteration update, and the optimal wavelet coefficients W_1^* and W_2^* are obtained to minimize the objective function value.

Step 3 Reconstruct the original signal, and obtain the high-resonance component and low-resonance component.

The optimal transformation coefficient matrix is obtained by reconstructing the observed signal as follows:

$$X = W_1^* S_1 + W_2^* S_2 + n \tag{8}$$



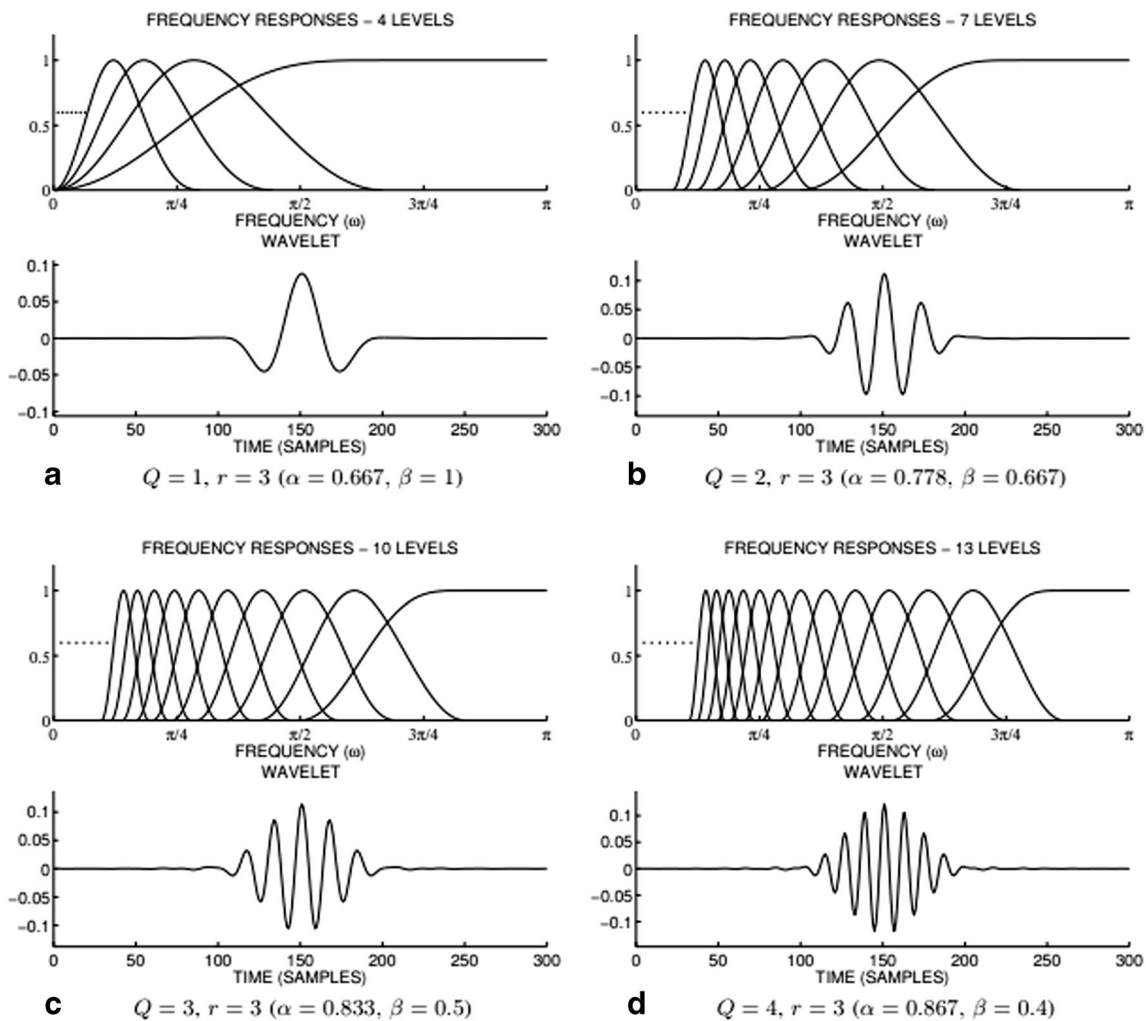


Fig. 2 Wavelet shape and frequency features in TQWT

In equation (8), $\mathbf{W}_1^* \mathbf{S}_1$ is the high-resonance component and $\mathbf{W}_2^* \mathbf{S}_2$ is the low-resonance component.

When the object function $J(\mathbf{W}_1, \mathbf{W}_1)$ reaches a minimal value, the coupling degree between the high- and low-resonance components and the residual component reaches a minimum. Greater and lesser oscillations in the signals are maximally decomposed into high- and low-resonance components, respectively. The high- and low-resonance components are effectively separated using the above steps.

Experimental Setup

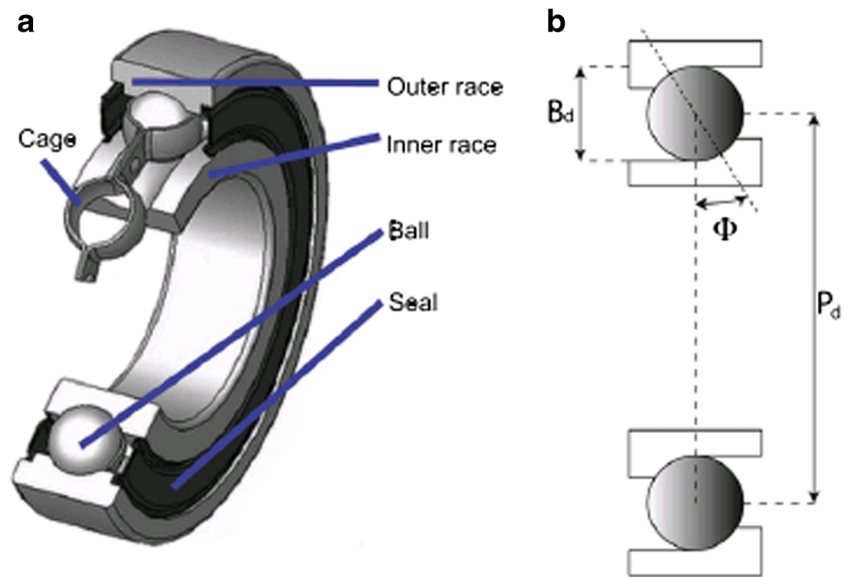
Two types of vibration arise when a rolling element passes local defects (such as spalling) at a high speed. One type is characterized by the relationship between the structure and vibration movement, which produces

the impact of periodic vibration when the rolling element contacts with the raceway at the local defect. Rolling fault elements can be diagnosed by the vibration frequency, which is caused by the impact frequency between the defects and rolling elements. This frequency is known as the characteristic defect frequency. The other type is the natural frequency of the bearing system due to the effect of impact and is characterized by the attenuation oscillation of each pulse. Based on a single-row angular contact ball bearing, as shown in Fig. 3 [1], the formula for the various characteristic frequencies is used [1, 29]. It is assumed that the motion between the rolling elements and race is characterized by pure rolling and that the inner race rotates with the shaft.

The ball pass frequency of the outer race is

$$F_{BPO} = F_s \left(\frac{N_b}{2} \right) \left(1 - \frac{B_d}{P_d} \cos \phi \right) \tag{9}$$

Fig. 3 Rolling bearing components



The ball pass frequency of the inner race is

$$F_{BPI} = F_S \left(\frac{N_b}{2} \right) \left(1 + \frac{B_d}{P_d} \cos \phi \right) \quad (10)$$

The ball (roller) spin frequency is

$$F_B = F_S \left(\frac{P_d}{2B_d} \right) \left(1 - \frac{B_d^2}{P_d^2} \cos^2 \phi \right) \quad (11)$$

The fundamental train frequency (cage speed) is

$$F_C = F_S \left(\frac{1}{2} \right) \left(1 - \frac{B_d}{P_d} \cos \phi \right) \quad (12)$$

where F_s is the shaft rotational frequency (Hz), N_b is the number of rolling elements, B_d is the diameter of the rolling element (mm), P_d is the pitch diameter of the rolling bearing (mm), and ϕ is the bearing contact angle ($^\circ$).

In this paper, all experiments were conducted on the machinery fault simulator (MFS) test bench manufactured by SpectraQuest Inc, as shown in Fig. 4. The experimental setup is able to simulate a fault in the bearing, gear, or belt as well as indicate mispositioning and off-center faults. The driver is a three-phase motor (1 horsepower). To reduce deviation caused by motor vibration, a short shaft with a 3/4-inch diameter is attached to the motor via a flexible coupling. The shaft is supported by two rolling element bearings on the two sides. The bearing located closer to the motor is the bearing under test that simulates all bearing faults and the bearing at the farther end is a good bearing. To ensure effective coupling of the signal, a piezoelectric accelerometer of type SQI608A11-3 F/8 is fixed on the bearing base via standard stud mounting. The bearing vibration signals are input to a computer through a data acquisition system with an 8-

channel data acquisition card for analysis and extraction of the signal characteristics.

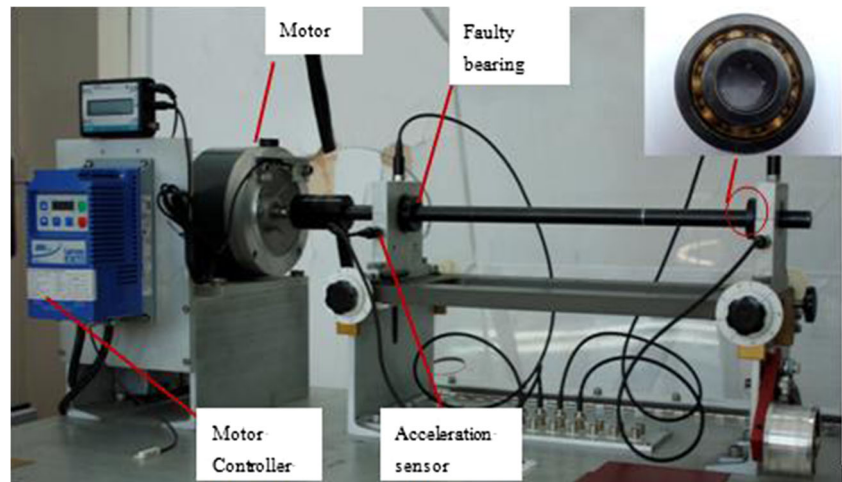
All test objects in this research represent the most commonly used industrial single-row deep-groove ball bearings of type ER-12 T, as shown in Fig. 4. The sampling frequency, sampling length, and rotation rate are 51.2 kHz, 65,536 points, and 30 Hz, respectively. The basic parameters of the bearings and the fault characteristic frequency in the experiments are shown in Table 1.

Engineering Validation

Experiment with an Inner Race Fault

A single point fault is introduced to the inner race using an electrode via charge machining with fault diameters of 0.4 mm and a depth of 0.2 mm. To simulate an early weak fault signal, the load on the shaft supported by the rolling bearings is removed, and the test bench is tapped rapidly with a hammer (approximately 5 times per second). The collected fault vibration signal and the spectrogram are shown in Fig. 5. According to the spectrogram, it indicates that the harmonics of the shaft rotation frequency are primarily concentrated below 2 kHz, and there are two natural frequency bands (also known as formants) in the bearing system at approximately 3.1 and 10 kHz. Because the formant peak at approximately 3.1 kHz is larger and narrower, this natural frequency band will be chosen as the demodulation band.

To fit the impulse response curve better, choose the RSSD parameters of $Q_1 = 4$, $Q_2 = 1$, $J_1 = 18$, $J_2 = 20$, and $r_1 = 3.5$, $r_2 = 3.5$, and set the weight coefficients $\lambda_{i,j}$ ($i = 1, 2$; $j = 1, 2 \dots J_i + 1$) at 0.15 times the l_2 -norm of the corresponding

Fig. 4 Fault simulator set up

wavelet $S_{i,j}$ ($i = 1, 2; j = 1, 2 \dots J_i + 1$). Then RSSD is applied to the vibration signal, and the result is shown in Fig. 6.

The following observations can be made from Fig. 6:

- (1) The high-resonance component contains several damped oscillations at each acceleration mutation point (Fig. 6(b)). In the low-resonance component, an oscillation of only a short duration (one to two oscillation cycles) exists at each acceleration mutation point (see Fig. 6(c)). Therefore, although both of the resonance components contain a portion of the fault information, additional fault information is available in the high-resonance component than in the low-resonance component. Careful observation of the high-resonance component waveform indicates that each string of sustained oscillation in the high-resonance component contains two oscillation frequencies corresponding to the two natural frequencies of the bearing system.
- (2) The bearing failure in this experiment is quite weak, and therefore, there are only a few abrupt peaks in the attenuation oscillations provoked by the failure impact, which indicates that the proportion of the low resonance-component is considerably smaller than that of the high-resonance component.
- (3) As shown in Fig. 6, not every impact generated by the contact of rolling elements and fault zone can be manifested in the high-resonance component. One reason for this phenomenon is that the fault is overly weak and the load is removed, and therefore, only the impact

generated by the contact in the load zone may be reflected in the high-resonance component. Another reason is that the rolling elements that experience the fault will not produce an impact in all instances. The rolling element may occasionally directly skim over the peeling area with no impact in terms of the gap between the rolling element and inner and outer race.

It is necessary to analyze the sub-band that highlights the fault impact response of the rolling element bearing. Thus, the energy-dominated sub-band that is highlighted in the analysis of sub-bands near the bearing natural frequency should be analyzed. In this paper, it defines the summation of the sub-bands with center frequencies that are located near the natural bearing frequencies of the main sub-band. The waveforms for 19 sub-bands in the high-resonance component are shown in Fig. 7(a), and 20 sub-band waveforms in the low-resonance component are shown in Fig. 7(b). The sub-band energy distributions of the resonance component are shown in Fig. 7(c) and (d), which illustrate that the energy of the high-resonance component is concentrated in two sub-band groups (5–8 and 15–18) and the energy of the low-resonance component is concentrated in three sub-band groups (1–3, 4–7, and 17–19). Obviously, the energy distribution of the high-resonance component corresponds directly to the two natural frequencies (10 kHz and 3.1 kHz). Therefore, the demodulation band is determined by the sub-band groups of the high-resonance component.

Table 1 Basic parameters of the faulted rolling bearing

Model	Pitch dia. (mm)	R. E. dia. (mm)	Number of R. E.	Contact angle (°)	Characteristic frequency			
					Outer ring	Inner ring	Rolling element	Cage train
ER-12 T	34.170	7.105	9	0	107.06	163.3	69.12	11.90

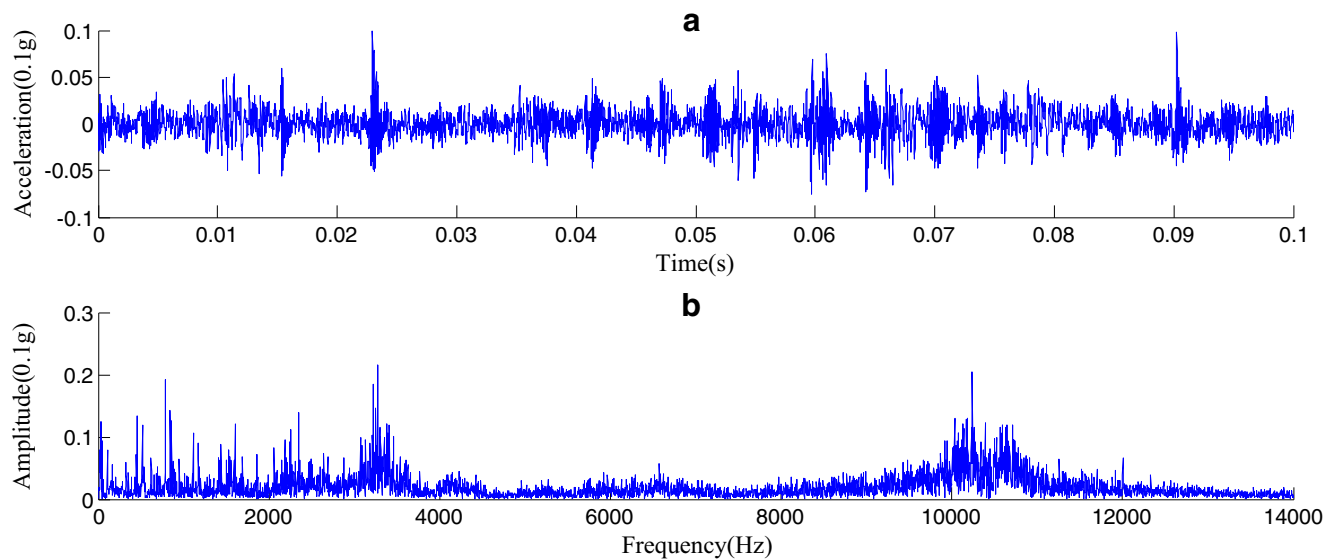


Fig. 5 Bearing vibration acceleration for the inner race fault: (a) Time domain acceleration signal. (b) Spectrum of acceleration signal

When determining the demodulation band, the natural frequency of approximately 3.1 kHz is chosen as the demodulation band, according to equations (2) and (3) or sub-band groups 15–18 in Fig. 7(c) and 4–7 in d. These two sub-band groups are summed up and the main sub-bands of the high- and low-resonance components are obtained, respectively (see Fig. 8(a) and (b)). Both the high- and low-resonance components contain a portion of the fault information, and the fault information is primarily contained in the main sub-bands. Here, we sum up the two main sub-bands and obtain the original signal's main sub-band, which carries most of the fault information (see Fig. 8(c)).

After applying the RSSD to the bearing vibration signal, the main sub-bands of the high- and low-resonance components are selectively analyzed. The envelope demodulation is applied to the main sub-band to obtain the spectrum envelope of the fault frequency. Then the envelope analysis is adopted for the main sub-bands of the high- and low-resonance components and original signal. The results are shown in Fig. 8(d), (e), and (f).

Autocorrelation has a smoothing effect on the main sub-band. As the Fourier transform of the autocorrelation, the auto power spectra of the envelope signal preserve the cycle components so that the fault feature frequencies of the rolling

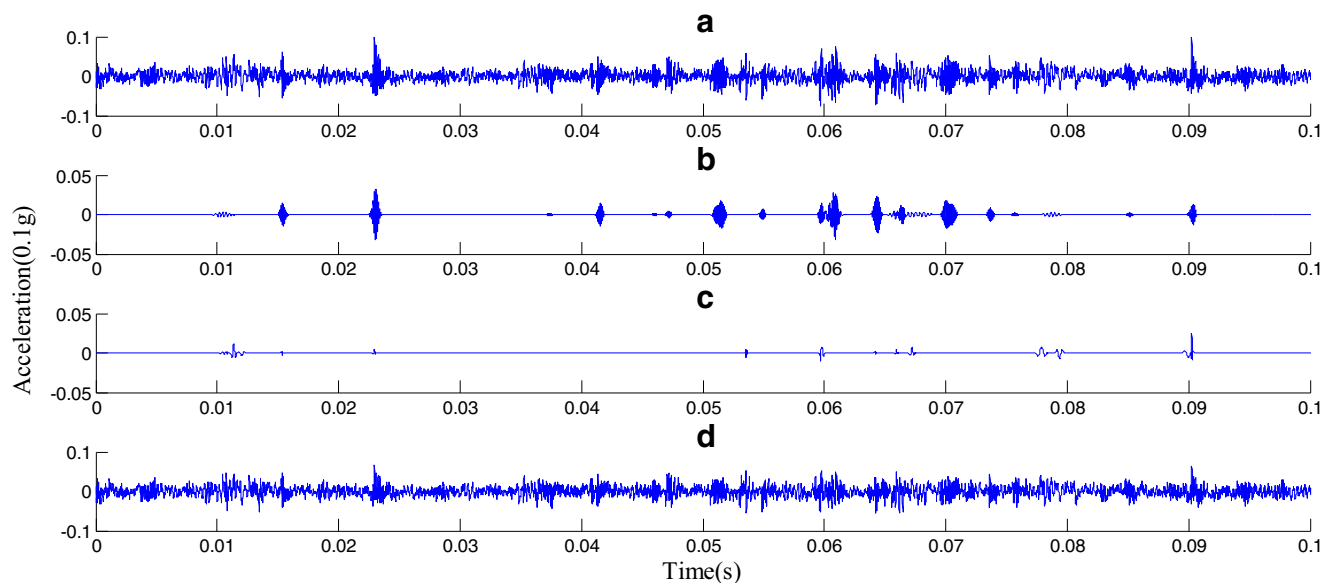


Fig. 6 RSSD result of the inner race fault signal: (a) Time domain waveform. (b) High-resonance component. (c) Low-resonance component. (d) Residual component

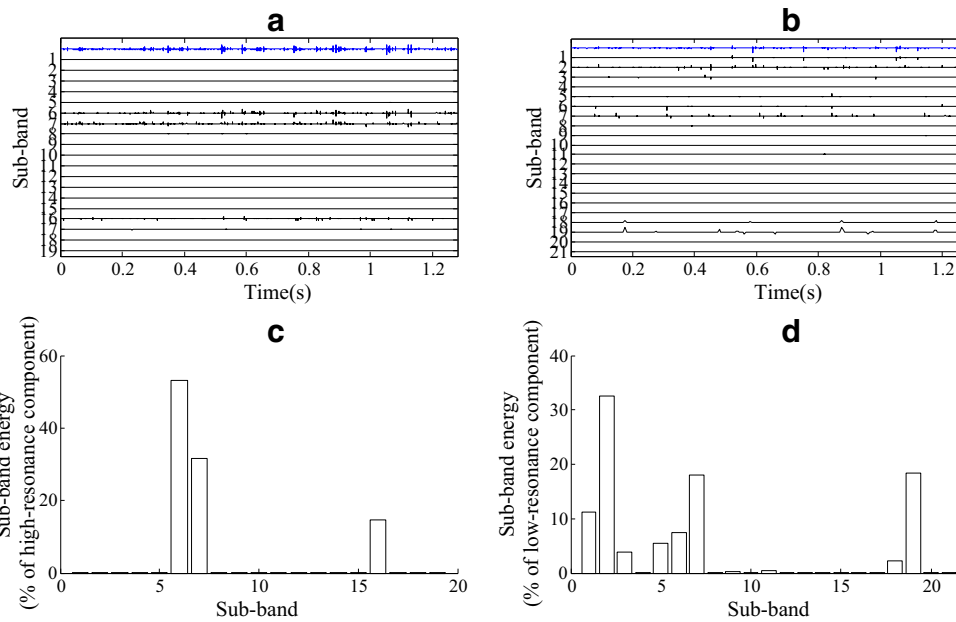


Fig. 7 Sub-bands and their energy distribution in the two resonance components: (a) Sub-bands in the high-resonance component. (b) Sub-bands in the low-resonance component. (c) Energy distribution of sub-

bands in the high-resonance component. (d) Energy distribution of sub-bands in the low-resonance component

bearings can be extracted from them. The auto power spectra of each main sub-band’s envelope signal are shown in Fig. 9.

A clear peak occurs at the fault characteristic frequency (163.5 Hz) in each auto power spectrum shown in Fig. 9. This frequency (163.5 Hz) corresponds to the feature frequency of the bearing inner race from equation (10). In addition, the side bands (133.5 and 193.5 Hz) on both sides of the fault characteristic frequency (163.5 Hz) are quite small or may not exist. An early fault indicates that this rolling element bearing inner race fault signal is weak. Among all of the auto power

spectrums, the peak value at the fault characteristic frequency in the auto power spectrum of the main sub-band envelope signal of the original signal is the largest and most obvious; the peak value at the fault feature frequency in the auto power spectrum of the main sub-band envelope signal of the low-resonance component is the smallest, and the peak is not dominant in the auto power spectrum of low-resonance component main sub-band envelope signal. The failure information contained in the low-resonance component is smaller and more easily interfered with by noise.

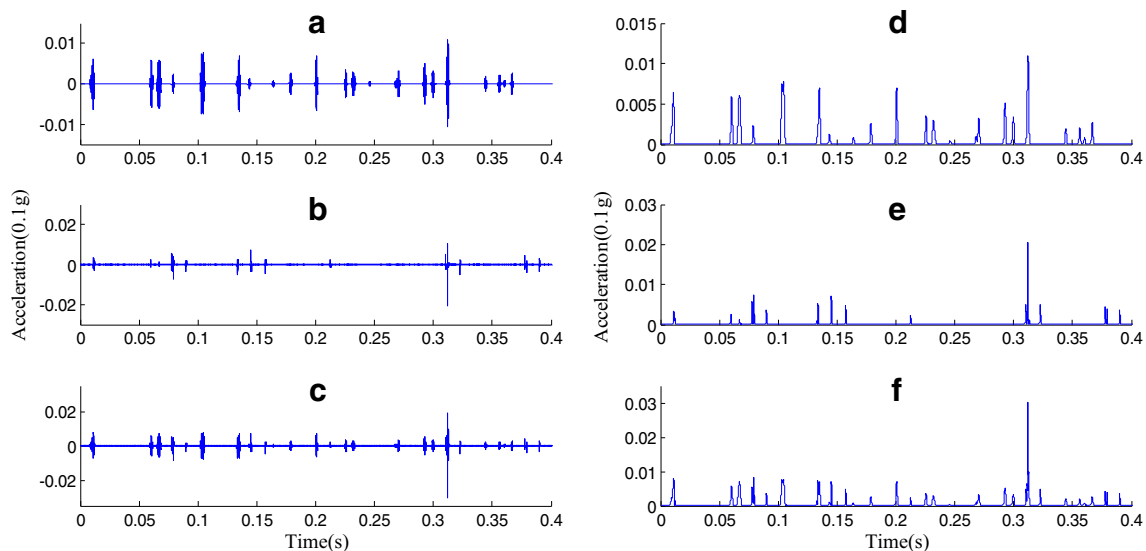


Fig. 8 Waveform of each main sub-band and its envelope signal for the inner race fault: (a) Main sub-band of the high-resonance component. (b) Main sub-band of the low-resonance component. (c) Main sub-band of the original signal. (d) (e) (f) Envelope signal of (a) (b) (c), respectively

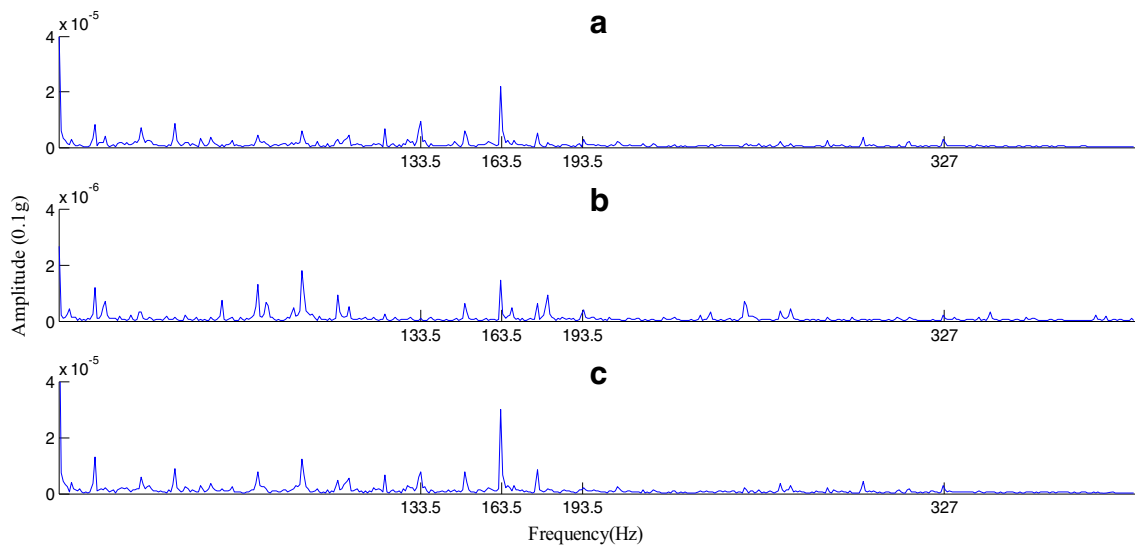


Fig. 9 Auto power spectrum of the main sub-band envelope signal for the inner race fault: (a) Main sub-band of the high-resonance component. (b) Main sub-band of the low-resonance component. (c) Main sub-band of the original signal

Experiment with an Outer Race Fault

In this subsection, the method is validated by another experiment with a hole of diameter 0.3 mm at the outer race. The rotating frequency of the shaft is 30 Hz and the load on the shaft is set to 5 kg. To enhance the noise, the hammer is used to tap the test bench at approximately 5 times per second. The gathered acceleration signal and its frequency spectrum is shown in Fig. 10. In time domain, there is hardly any obvious acceleration mutation caused by fault impacts, which indicates the fault signal is extremely weak. For contrast, these two natural frequency bands are both chosen as demodulation bands.

Because the fault feature is weaker, in order to make more fault information decomposed into high- and low-resonance components, smaller $\lambda_{i,j}$ is chosen at 0.1 times the l_2 -norm of

the wavelet $S_{i,j}$ ($i = 1, 2; j = 1, 2 \dots J_i + 1$). Other parameters are the same as Section 4.1. The RSSD result is shown in Fig. 11.

It can be seen that about six damped oscillations exist in the high-resonance component. Moreover, there is some overlap between two adjacent damped oscillations. This is because two fault impacts are generated when the rolling element enters and leaves the spalling area each time.

In terms of different demodulation bands, the auto power spectra of each main sub-band's envelope signal is shown in Fig. 12. Obviously, peaks exist at the fault characteristic frequency (107.06 Hz) and its harmonics in each auto power spectrum. Comparing the peak values in different components, it is easy to find that the peak values in main sub-band envelope signal of the original signal are the largest. In addition, the peaks at 3.1 kHz demodulation band are much

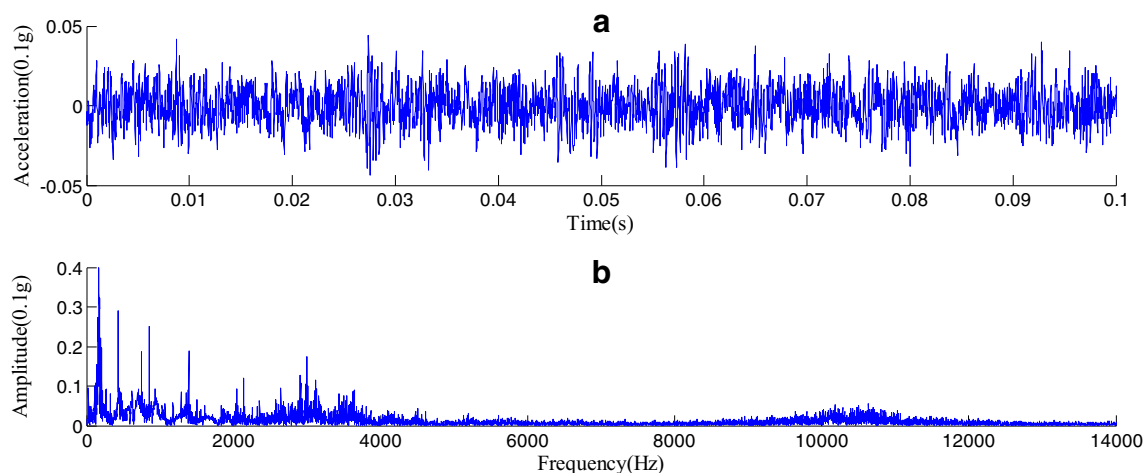


Fig. 10 Bearing vibration acceleration for the outer race fault: (a) Time domain acceleration signal. (b) Spectrum of acceleration signal

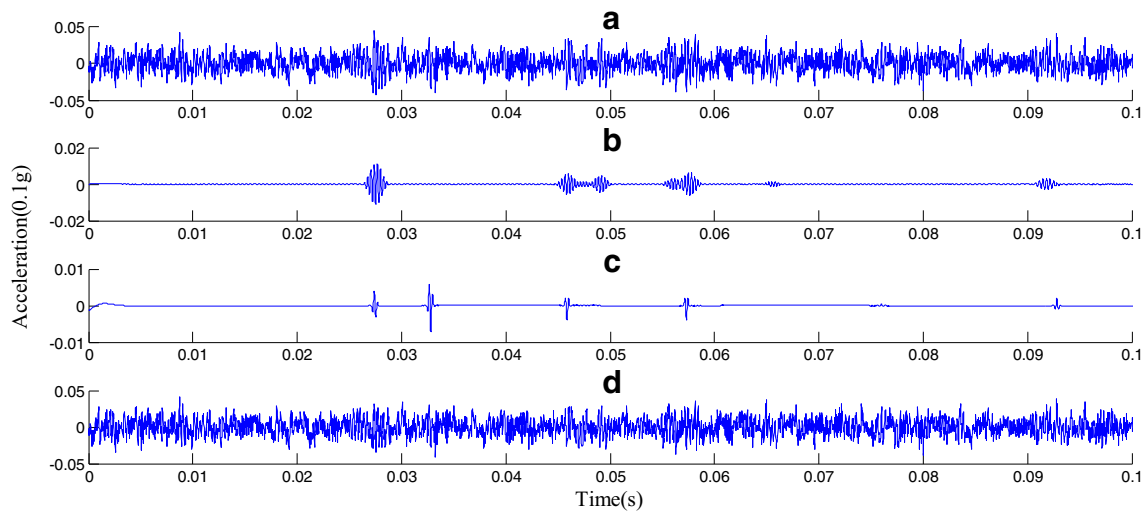


Fig. 11 RSSD result of the outer race fault signal: (a) Time domain waveform. (b) High-resonance component. (c) Low-resonance component. (d) Residual component

clearer than those at 10 kHz. And the peak values at the fault feature frequency in Fig. 12(a), (b), (c) are much larger than those in d,e,f. It illustrates that more obvious fault information can be extracted when choosing the frequency band at approximately 3.1 kHz as the demodulation band. Therefore, when there are several natural frequency bands, a larger and narrower natural frequency band is preferable.

Comparative Analysis

To verify the effectiveness of the RSSD in bearing fault diagnosis, this paper compares the RSSD results with the results

from the envelope analysis and wavelet analysis for the same signals.

Comparison with the envelope analysis

For the inner and outer race faults, in terms of the larger and narrower formant peak at approximately 3.1 kHz, the envelope decomposition frequency bands are defined at 2,700–3,700 Hz and 2,500–3,300 Hz, respectively. The envelope analysis results are shown in Figs. 13 and 14. Figures 13 and 14(c) present the frequency spectra after the envelope

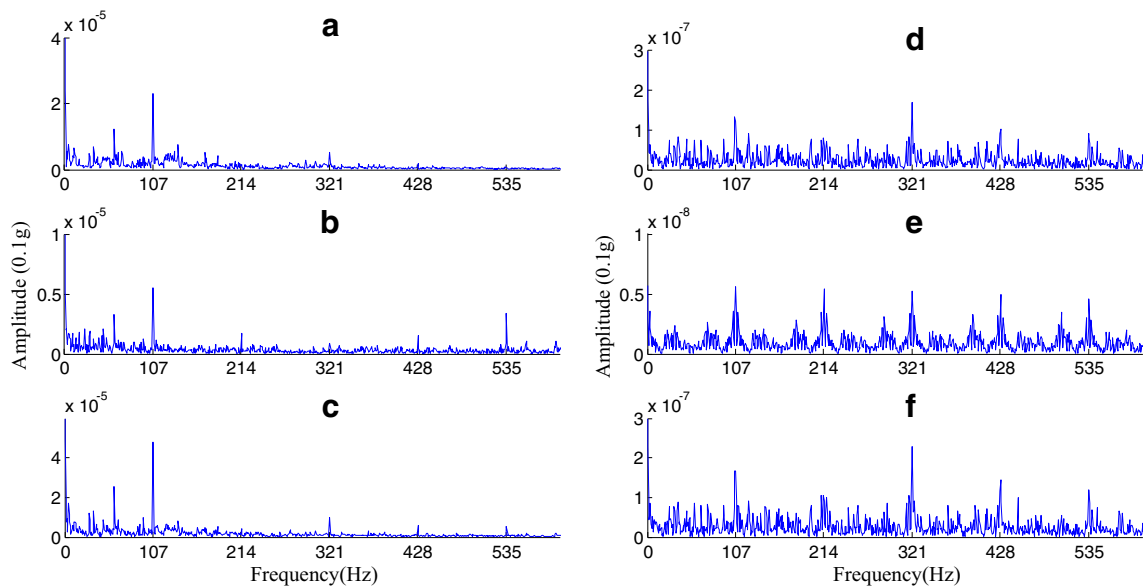


Fig. 12 Auto power spectrum of the main sub-band envelope signal for the inner race fault: (a) Main sub-band of the high-resonance component at 3.1 kHz. (b) Main sub-band of the low-resonance component at 3.1 kHz. (c) Main sub-band of the original signal at 3.1 kHz. (d) Main

sub-band of the high-resonance component at 10 kHz. (e) Main sub-band of the low-resonance component at 10 kHz. (f) Main sub-band of the original signal at 10 kHz

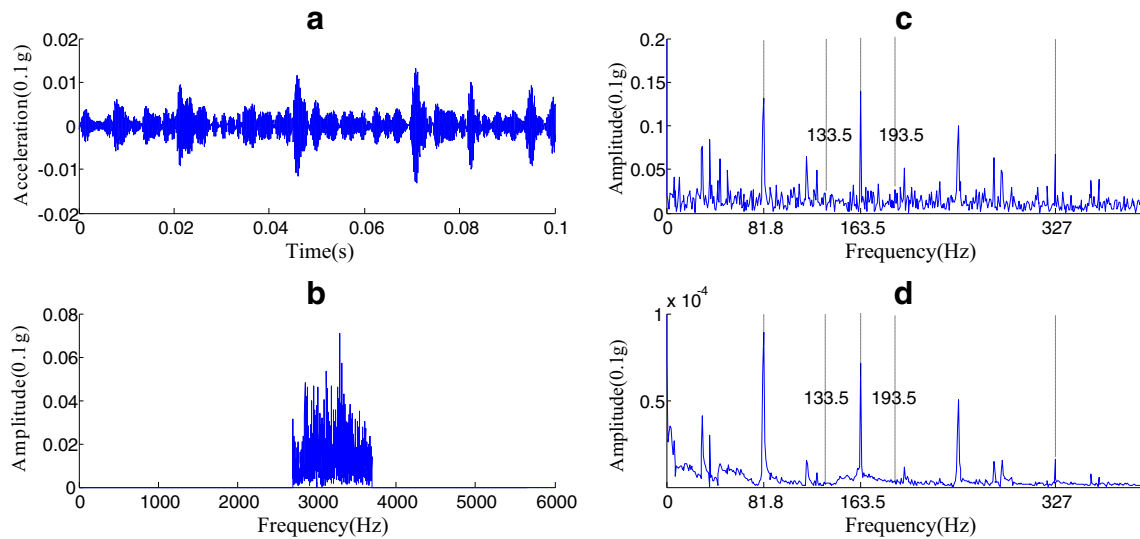


Fig. 13 Envelope analysis result of the inner race fault signal: (a) 2,700–3,700 Hz waveform in the time domain after band-pass filtering. (b) Spectrum after filtering. (c) Envelope spectrum. (d) Auto power spectrum of envelope signal

analysis, and Figs. 13 and 14(d) present the auto power spectra of the envelope signals.

For the inner race fault, Fig. 13(c) and (d) illustrate that the main spectral peaks occur at half, one time, and twice the inner race fault feature frequency (163.5 Hz), but there are no side peaks on both sides of 30 Hz from the main peak. An obvious peak occurs at the fault characteristic frequency (163.5 Hz) in Fig. 9(c), and the side bands on both sides of the fault characteristic frequency are also rather obvious. For the outer race fault, Fig. 14(c) and (d) show that the peaks only occur at one time, and triple the outer race fault feature frequency (107 Hz). In contrast, the peaks in Fig. 12(c) are more dominant due to fewer noise burrs and the high signal-noise ratio.

From above analysis, it demonstrates that RSSD is superior to the envelope analysis and auto power analysis of the envelope signal in fault feature extraction for rolling element bearing.

Comparison with the wavelet analysis

In the subsection, the db10 wavelet is chosen as a basis function for the five-layer wavelet decomposition. The results are shown in Figs. 15 and 16. Because the central frequency of the third-layer wavelet is close to 3.1 kHz, the envelope analysis is performed on d3.

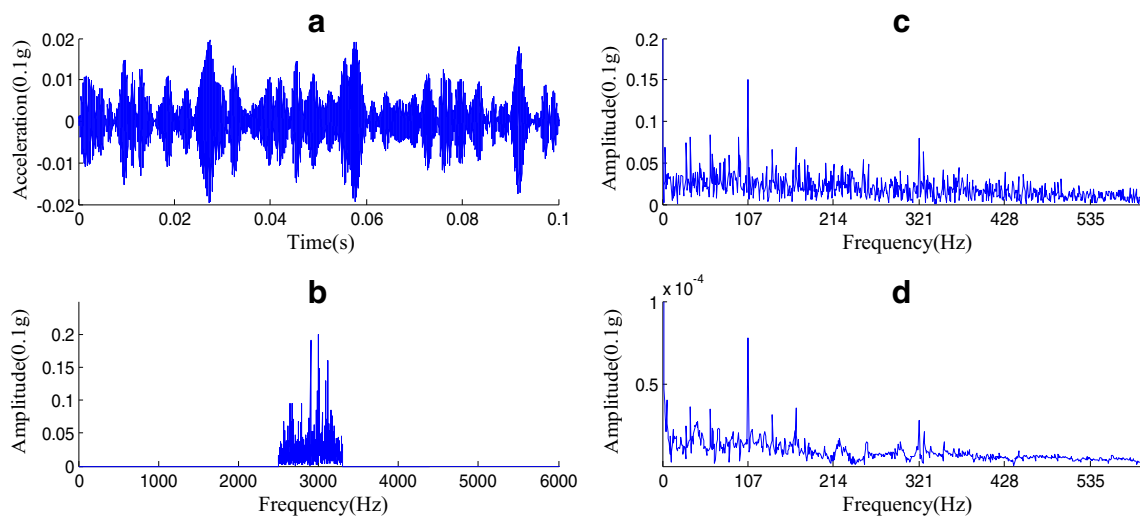


Fig. 14 Envelope analysis result of the outer race fault signal: (a) 2,500–3,300 Hz waveform in the time domain after band-pass filtering. (b) Spectrum after filtering. (c) Envelope spectrum. (d) Auto power spectrum of envelope signal

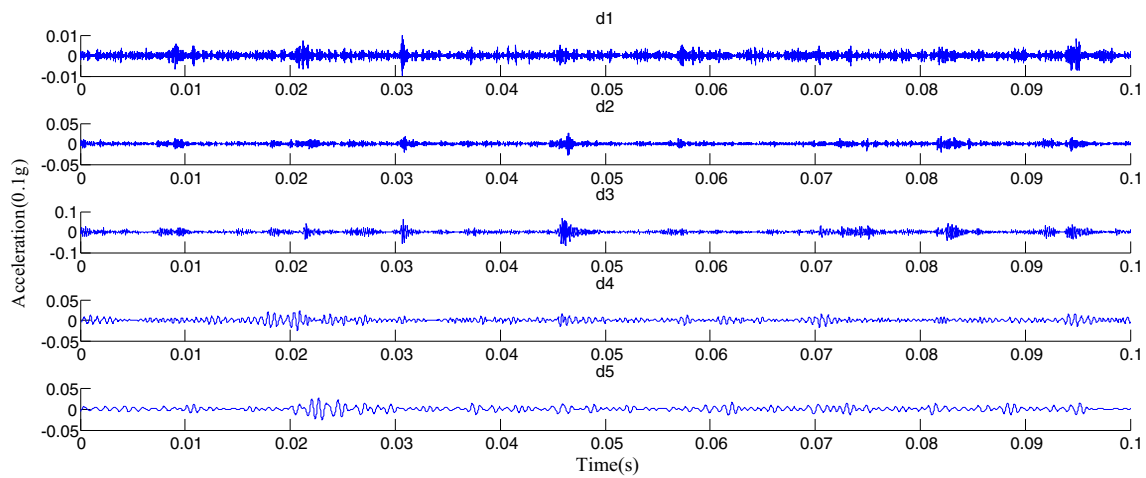


Fig. 15 Wavelet decomposition result of the inner race fault signal

Figures 17 and 18 present the spectra and auto power spectra of the d3 envelope signals. In Fig. 17, peaks occur at half of and twice the inner race fault feature frequency of 163.5 Hz, and another peak is visible at 160 Hz. Compared with the spectrum from RSSD in Fig. 9(c), these two spectra display additional burrs, the main peak in the spectrum is not dominant, and the side bands on both sides of main peak are not obvious. One reason for this observation is that because of the small frequency resolution of the traditional wavelet, it is difficult to accurately extract the natural frequency oscillation component from the fault signal, and the error of fitting complex signals with a single wavelet is more significant. In Fig. 18, peaks exist at one time, and triple the outer race fault feature frequency (107 Hz), especially at one time. Compared with Fig. 12(c), there are more burrs in Fig. 18, although the feature frequency can be also observed.

To sum up, compared with the wavelet analysis, RSSD has been proven more advantageous in fault information extraction for rolling element bearing.

Discussion

- (1) Rolling fault transient shock and vibration signals were generated by rolling bearing fault coupling spread through multiple interfaces such that the impulse response of the signal detected by sensors is an oscillating exponential decay. Therefore, it is improper to denote fault impact composition by the low-resonance component in RSSD. In fact, due to the complexity of the signal and selection of TQWT, both the high- and low-resonance components contain a wealth of fault information. This paper proposes the use of RSSD, which combines high- and low-resonance components in the rolling bearing fault signal.
- (2) For the existing fault diagnosis method with RSSD, the low-resonance components are envelope decomposed directly after RSSD without an analysis of the different sub-bands. This method can decompose fault information only in the case of small noise signals. Additional

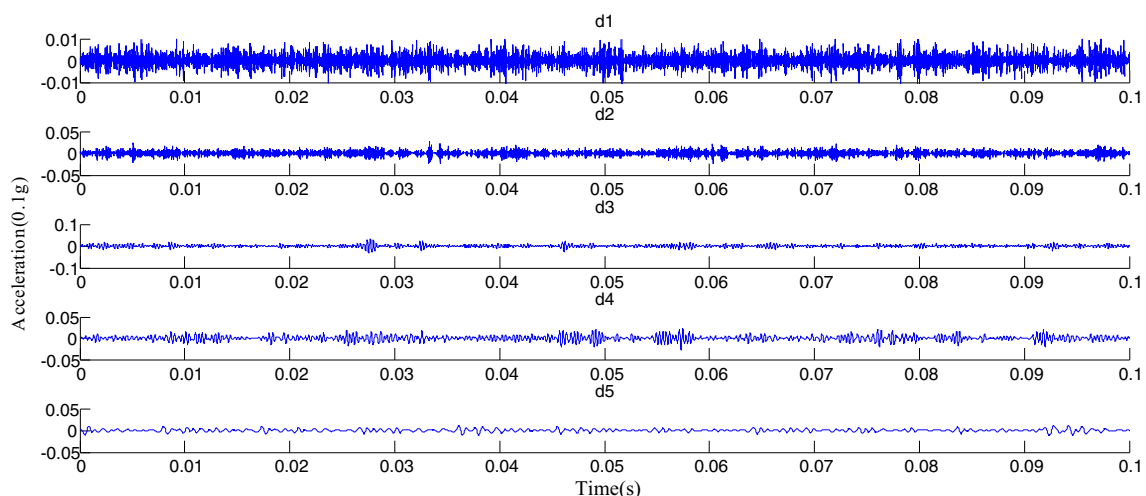


Fig. 16 Wavelet decomposition result of the outer race fault signal

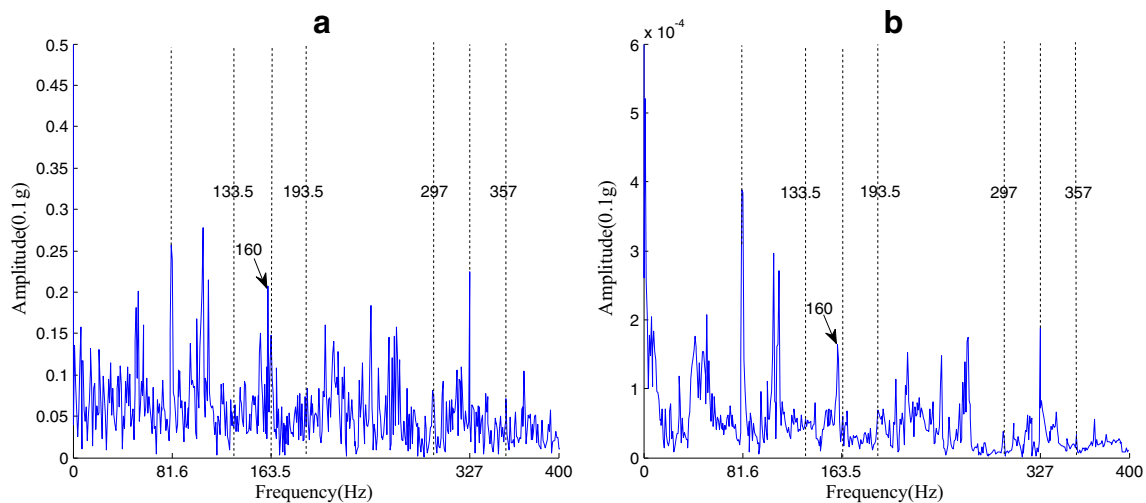


Fig. 17 Envelope spectrum and envelope auto power spectrum of the layer d3 wavelet for the inner race fault: (a) Spectrum of the envelope signal. (b) Auto power spectrum of the envelope signal

noise will exacerbate the decomposition effect due to the low-resonance component, which contains the fault impact signals but also contains the noise impact signals. In this paper, the summation of the sub-bands with center frequencies that are near the natural frequencies of the bearing is defined as the main sub-band, decompose the main sub-band of the high- and low-resonance components, and increase the signal-to-noise ratio of the fault vibration effectively.

- (3) According to the comparison of the auto power spectra of the envelope signals of the rolling bearing high- and low-resonance components, it indicates that the peak value at the fault frequency in the auto power spectrum of the low-resonance component's main sub-band envelope signal is the smallest, and the peak is not the most dominant in the auto power spectrum of the low-resonance

component main sub-band envelope signal. The failure information contained in low-resonance component is smaller and more easily interfered with by noise. In contrast, the high-resonance component contains additional failure information, the main sub-bands can be more precisely positioned near the natural frequency, and the high-resonance component is less easily interfered with by noise.

- (4) In contrast to the envelope analysis and wavelet analysis, RSSD expresses a signal using two wavelet basis functions. These two different types of wavelet basis functions can express different types of rolling bearing vibration signal components. According to specific objects, the *Q*-factor of the two wavelet basis functions can be altered to adjust the waveform, which greatly increases the adaptability and flexibility of RSSD.

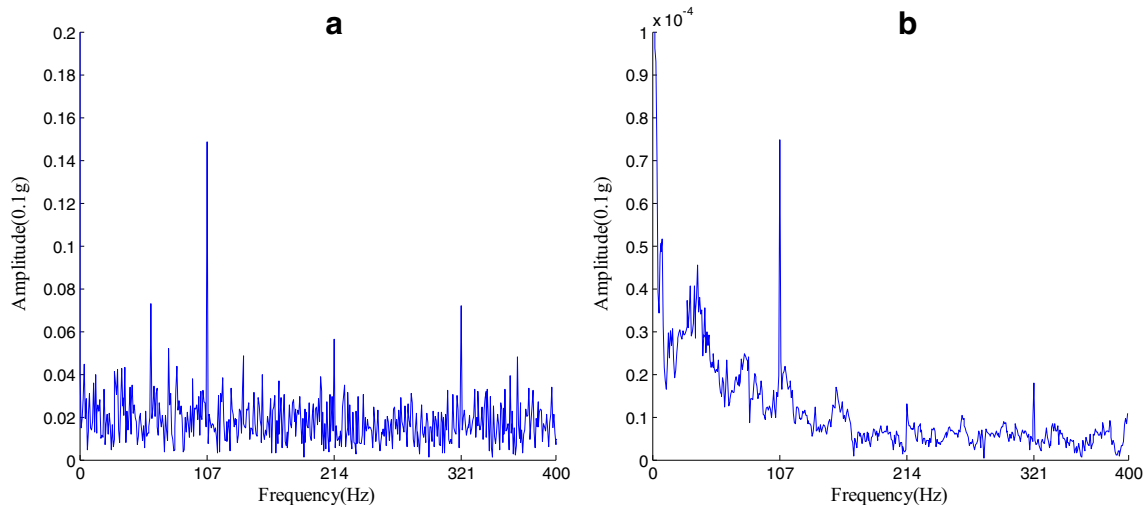


Fig. 18 Envelope spectrum and envelope auto power spectrum of the layer d3 wavelet for the outer race fault: (a) Spectrum of the envelope signal. (b) Auto power spectrum of the envelope signal

Conclusions

To address the disadvantage of using envelope decomposition of the low-resonance component in fault diagnosis with RSSD, this paper proposed a new method based on RSSD that combines the high- and low-resonance components and applied this method to rolling bearing fault diagnosis. The bearing fault signal is decomposed into three components by RSSD: the high-resonance component containing the sustained oscillation cycle signal, the low-resonance component containing the impact transient signal, and the residual component. The concept of resonance of the main sub-band was introduced. Envelope decomposition was applied to the main sub-band of the high and low resonances. The analyzed examples demonstrated that this method was able to extract the impact element for bearing fault diagnosis and highlight the fault features effectively. The proposed method in this paper provides a new solution for feature extraction of rolling bearing fault signals.

Acknowledgements This work is supported by the National Natural Science Foundation of China (51175102) and the Fundamental Research Funds for the Central Universities (HIT.NSRIF.201638). The authors thank Prof. Selesnick of the Polytechnic Institute of New York University for providing the programs used to implement the resonance-based sparse decomposition.

References

- Abbasion S, Rafsanjani A, Farshidianfar A, Irani N (2007) Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech Syst Signal Process* 21(7):2933–2945
- Daiyi MO (2013) Sparse signal decomposition method based on the dual Q-factor and its application to rolling bearing early fault diagnosis. *J Mech Eng* 49(9):533–540
- Chen X, Yu D, Luo J (2013) Early rub-impact diagnosis of rotors by using resonance-based sparse signal decomposition. *Zhongguo Jixie Gongcheng/China Mech Eng* 24(1):35–41
- Slavic J, Brkovic A, Boltezar M (2013) Typical bearing-fault rating using force measurements application to real data. *Hum Mutat* 9(1): 91–94
- Duan C, He Z, Jiang H (2004) New method for weak fault feature extraction based on second generation wavelet transform and its application. *Chin J Mech Eng* 17(4):543–547
- Li H, Zhang Y, Zheng H (2009) Hilbert-Huang transform and marginal spectrum for detection and diagnosis of localized defects in roller bearings. *J Mech Sci Technol* 23(2):291–301
- Osman S, Wang W (2013) An enhanced Hilbert-Huang transform technique for bearing condition monitoring. *Meas Sci Technol* 24(24):50–61
- Žvokelj M, Zupan S, Prebil I (2010) Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis. *Mech Syst Signal Process* 24(4):1049–1067
- Cai Y, Lin J, Zhang W, Ding J (2015) Faults diagnostics of railway axle bearings based on IMF's confidence index algorithm for ensemble EMD. *Sensors* 15(5):10991–11011
- Guo W, Tse PW, Djordjevich A (2012) Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. *Measurement* 45(5):1308–1322
- Wang Z, Lu C, Wang Z, Liu H, Fan H (2013) Fault diagnosis and health assessment for bearings using the mahalanobis–taguchi system based on EMD-SVD. *Trans Inst Meas Control* 35(6):798–807
- Zheng J, Cheng J, Yang Y (2013) Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis. *Mech Syst Signal Process* 40(1):136–153
- Li B, Zhang PL, Wang ZJ, Mi SS, Liu DS (2011) A weighted multi-scale morphological gradient filter for rolling element bearing fault detection. *ISA Trans* 50(4):599–608
- Li C, Liang M (2012) Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis. *J Sound Vib* 331(26):5864–5879
- Raj AS, Murali N (2013) Early classification of bearing faults using morphological operators and fuzzy inference. *IEEE Trans Ind Electron* 60(2):567–574
- Al-Raheem KF, Roy A, Ramachandran KP, Harrison DK, Grainger S (2008) Rolling element bearing faults diagnosis based on autocorrelation of optimized: wavelet de-noising technique. *Int J Adv Manuf Technol* 40(3):393–402
- Wang X, Zi Y, He Z (2011) Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis. *Mech Syst Signal Process* 25(1):285–304
- Su W, Wang F, Zhu H, Zhang Z, Guo Z (2010) Rolling element bearing faults diagnosis based on optimal morlet wavelet filter and autocorrelation enhancement. *Mech Syst Signal Process* 24(5): 1458–1472
- Kankar PK, Sharma SC, Harsha SP (2013) Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform. *Neurocomputing* 110(8):9–17
- Liu J, Wang W, Ma F, Liu J, Ma F (2012) Bearing system health condition monitoring using a wavelet cross-spectrum analysis technique. *J Vib Control* 18(7):953–963
- Rafiee J, Rafiee MA, Tse PW (2010) Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Syst Appl* 37(6):4568–4579
- Selesnick IW (2011) Resonance-based signal decomposition: a new sparsity-enabled signal analysis method ☆. *Signal Process* 91(12): 2793–2809
- Cui L, Mo D, Wang H, Chen P (2015) Resonance-based nonlinear demodulation analysis method of rolling bearing fault. *Adv Mech Eng* 5(5):420694–420694
- Wang H, Chen J, Dong G (2014) Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform. *Mech Syst Signal Process* 48(1–2):103–119
- He W, Zi Y, Chen B, Wu F, He Z (2014) Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. *Mech Syst Signal Process* 54: 457–480
- Li X, Yu D-J, Zhang D-C (2015) Fault diagnosis of rolling bearings based on the resonance-based sparse signal decomposition with optimal Q-factor. *Zhendong Gongcheng Xuebao/J Vib Eng* 28(6): 998–1005
- Tang G, Wang X (2016) Application of tunable Q-factor wavelet transform to feature extraction of weak fault for rolling bearing. *Proceedings of the CSEE* 36(3):746–754
- Bayram I, Selesnick IW (2009) Frequency-domain design of over complete rational-dilation wavelet transforms. *IEEE Trans Signal Process* 57(8):2957–2972
- Liu WY, Han JG, Jiang JL (2013) A novel ball bearing fault diagnosis approach based on auto term window method. *Measurement* 46(10):4032–4037