

Extraction of the largest amplitude impact transients for diagnosing rolling element defects in bearings

Lei Hu ^{a,b}, Lun Zhang ^b, Fengshou Gu ^{c,*}, Niaoqing Hu ^b, Andrew Ball ^c

^a College of Traffic Engineering, Hunan University of Technology, Zhuzhou, 412007, PR China

^b Laboratory of Science and Technology on Integrated Logistics Support, College of Mechatronics and Automation, National University of Defense Technology, Changsha, 410073, PR China

^c School of Computing and Engineering, University of Huddersfield, Huddersfield, HD1 3HD, United Kingdom

Abstract

This paper presents a method based on the extraction of the largest amplitude impact transients (ELAIT) for diagnosing the rolling element defect in bearings. As a defected rolling element causes two largest amplitude impact transients (LAITs) during a spin period when the element passes the load zone centre, LAITs are separated for each rolling element according to the kinematics of the bearing operation. By applying band-pass filtering, demodulation, low-pass filtering, and ensemble averaging to these LAITs, an enhanced signature named envelope ensemble average (EEA) is obtained for each rolling element, which allows a reliable indication of the defected elements. The robustness of the method is evaluated by investigating the localised fault model of rolling bearings with the inclusion of phase errors caused by rotational speed oscillation and rolling element slippage along with additive white noises. Evaluation results show that EEA signatures are very sensitive to element defects and give an accurate indication of the most probably defected element, and the ELAIT method is robust to rotational speed oscillation and slippage. The same performance is also achieved when the method was validated with experimental signals from a test rig of machinery fault simulation, showing effectiveness and robustness in detecting rolling element defects in an operated bearing. Besides, the proposed method can be easily implemented online as it does not need a tachometer and is implemented at low computation cost.

Keywords: Fault diagnosis; rolling bearings; extraction of the largest amplitude impact transients; noise reduction

1. Introduction

Rolling bearing faults can cause the failure of rotating machinery if maintenance action is not taken in time. Fault diagnosis of rolling bearings supports the maintenance decision and then helps to prevent the rotating machinery from unexpected failures. Vibration based fault diagnosis approaches are based on the assumption that impacts of rolling elements cause vibration transients (or transient responses) when passing through the localised defect. In general, these successive transients exhibit largely periodical or repetitive responses. The periodicity is usually expressed as characteristic fault frequencies [1][2]. For no slippage between the rolling elements and races, fault frequencies can be calculated according to [3][4]:

$$\text{Ball pass frequency of outer race: } f_{\text{BPO}} = \frac{1}{2}Z \left(1 - \frac{d}{D} \cos \alpha \right) f_r \quad (1)$$

$$\text{Ball pass frequency of inner race: } f_{\text{BPI}} = \frac{1}{2}Z \left(1 + \frac{d}{D} \cos \alpha \right) f_r \quad (2)$$

$$\text{Ball spin frequency: } f_{\text{BS}} = \frac{D}{2d} \left(1 - \left(\frac{d}{D} \right)^2 \cos^2 \alpha \right) f_r \quad (3)$$

where Z is the number of rolling elements, d is the element diameter, D is the pitch diameter, α is the contact angle, and f_r is the shaft rotational speed. A localised defect of a rolling element impacts the outer race and the inner race once a spin and two transients are generated. Thus $2f_{\text{BS}}$ is normally used as the fault frequency to indicate element defects.

The rolling element defect and the inner race defect are more difficult to detect than the outer race defect, especially in cases when the defects are incipient and the components of fault frequencies are submerged in noise. Fig. 1 shows

* Corresponding author. Tel: +44-1484-473548; Fax: +44-1484- 473075.

E-mail address: f.gu@hud.ac.uk (Fengshou Gu); hulei0819@foxmail.com (Lei Hu)

the load zone in a bearing for the case of a vertical external loading. As the outer ring is fixed, the outer race defect normally appears in the load zone and the transients caused by the contacts between defect and elements are not modulated. On the contrary, the inner ring and the elements are moving parts and their position to the load zone is always changing during rotation. Thus amplitude modulation is caused by the defect moving in and out of the load zone. The inner ring rotates at shaft speed and the modulation frequency is f_r for inner race defect case. The rolling elements rotate at cage speed and the modulation frequency is $f_c = f_{BP0}/Z$ for element defect case. Transients of different kinds of bearing defects are shown in Fig. 1. For equal sized defects, the transient intensity of the outer race defect would be larger than those of the inner race defect and the rolling element defect [5].

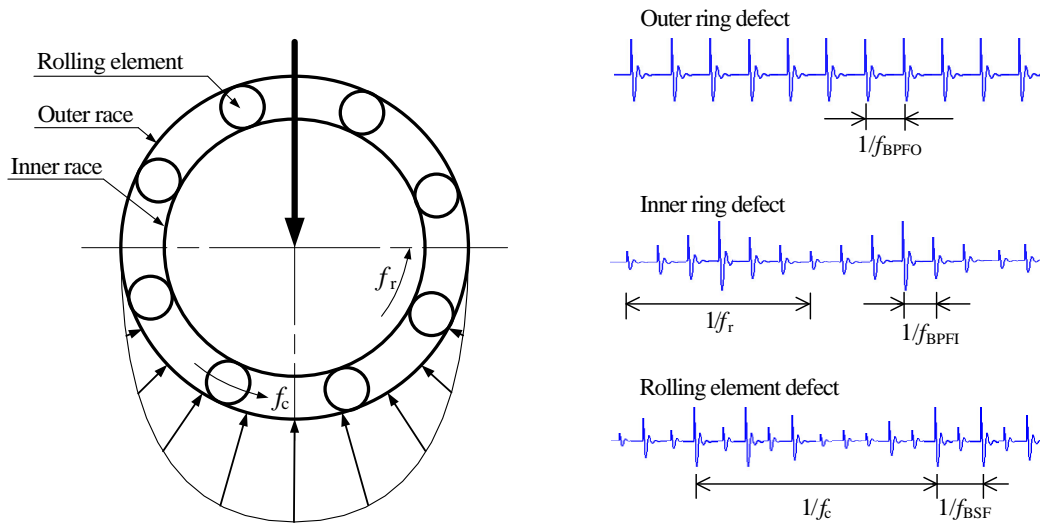


Fig. 1. The load distribution in a bearing and time domain response of bearing defects.

Many vibration signal processing methods have been introduced to characterize the bearing vibration signals for fault diagnosis. These methods, such as fast Fourier transform (FFT), short time Fourier transform (STFT), empirical mode decomposition (EMD) [6][7], Kurtogram method [8][9][10], modulation signal bispectrum (MSB) [11][12], etc., are used to extract effective diagnostic features from vibration signals, and then the extracted features can be used as inputs of a pattern recognition model for intelligent fault diagnosis. In order to improve the separation capability of the extracted features, preprocessing methods are used to improve the signal noise ratio (SNR) of vibration signals.

Time synchronous averaging (TSA) is a well-known signal preprocessing method for fixed-axis gears [13]. TSA divides a long gear vibration signals into successive short segments, the length of which is the interesting period T that corresponds to the shaft rotational frequency, $f_r = 1/T$. Then a synchronous signal is yielded by averaging these segments,

$$x_{\text{TSA}}(t) = \frac{1}{N} \sum_{n=1}^N x(t + nT). \quad (4)$$

The amplitude frequency characteristic of Eq. (4) is presented as a comb filter, which consists of a series of uniformly distributed band-pass filters and their sidebands. The central frequencies of these band-pass filters are the shaft rotational frequency and its harmonics kf_r , $k = 1, 2, \dots$, thus noise as well as components whose frequencies are not integer multiples of the shaft rotational frequency are suppressed significantly.

Rigorous synchronization of divided signal segments is critical for TSA, since small phase error can cause great information loss. Thus vibration was traditionally resampled in the angle domain with a key-phase signal captured using a photoelectric encoder or a tachometer, and then by synchronising with a trigger signal with the shaft rotational frequency, components at integer multiples of shaft rotational frequency will pass the comb filter. As fault frequency of fixed-axis gears is typically Z multiples of the shaft rotational frequency where Z is the number of gear teeth, TSA is successfully used for fault diagnosis of fixed-axis gears.

However, the fault frequencies of rolling bearings are usually not harmonics of the shaft speed, thus shaft speed

cannot be used as triggers for TSA of bearing vibrations. Thus, for enhanced fault diagnosis of rolling bearings, Hecke et al. [14] proposed a spectral average-based method to reduce noise by performing averaging in frequency domain instead of time domain. McFadden and Toozhy [15] applied TSA to monitor inner race defect of rolling bearings. As the ball pass frequency of inner race f_{BPI} , the cage speed f_c , and the shaft speed f_r satisfy the equation $f_{BPI} = Z(f_r - f_c)$, by synchronising with a trigger signal with the shaft rotational frequency relative to the cage frequency, $(f_r - f_c)$, components at integer multiples of $(f_r - f_c)$, such as f_{BPI} will be passed. The above synchronous averaging method is only used for enhanced diagnosis of inner race defect, and the requirement of cage speed is a great drawback as mounting of embedded optical sensors into a bearing to measure cage speed is not allowed in most cases. Siegel et al. [16] generated a trigger signal from the outer race fault frequency and proposed a tachometer-less synchronously averaged envelope (TLSAE) method to detect different levels of damage on the bearing outer race.

The kinematic motion of a rolling bearing is similar to a planetary gearbox. The inner ring of the bearing corresponds to the sun gear, the rolling elements correspond to the planet gears, and the outer ring corresponds to the ring of a planetary gearbox. For planetary gearbox, McFadden proposed an improved TSA method which can be used to achieve signal averages of the individual planets and sun. This improved TSA is named vibration separation [17][18]. The assumption that underlies the vibration separation method is that when a given planet gear is near an accelerometer, the vibrations measured by the accelerometer are dominated by the meshing of the specific planet gear with the sun and ring gears. A sequence of vibration segments, which are associated with a single rotation of the planet or sun gear, can be separated based on tooth meshing sequence [19]. Then the separated vibration segments can be used to perform TSA and result in improved fault detection. Thus, it is natural to introduce the vibration separation method into the bearing case. However, as there is no tooth meshing in bearing, the vibration segments cannot be located according to the meshing sequence. Especially, rolling element slippage and rotational speed oscillation can happen inevitably, which cause phase errors and degrade the performance of vibration separation for bearing cases.

Therefore, this paper improves the vibration separation method so as to be suitable for the bearing case and thereby proposes a method based on the extraction of the largest amplitude impact transients (ELAIT) to achieve more accurate diagnosis of element defects. A sequence of vibration segments that correspond to the spin periods during which a particular rolling element passes the load zone centre and causes largest amplitude impact transients (LAIT) is separated according to the kinematics of the bearing. Then these LAIT segments are enhanced with band-pass filtering, demodulation, and low-pass filtering. Envelope ensemble average (EEA) is yielded at last by averaging the processed LAITs for each rolling element to enhance the envelope waveform of the transients. The method is firstly validated with simulation signals which have phase errors due to rotational speed oscillation and rolling element slippage. Then, its effectiveness is verified by measured vibration signals.

2. Methodology

2.1. LAIT spin separation for defected rolling elements

The schematic of a rolling bearing is shown in Fig. 2. There are Z rolling elements that spin around their own centres at the speed of f_{BS} while they all revolve around the shaft centre at the cage speed of f_c . The centre point of load zone is at P . The first element that is about to pass the point P is named as RE 1 for short. The following $Z-1$ elements that will pass P successively are respectively named as RE 2, RE 3, ..., and RE Z .

For the RE i , there is just one spin occurring across P for each revolution of the cage. It is this spin that is most interesting. Obviously, the element is most heavily loaded in the spin and a defect on the element can induce largest amplitude impact transients (LAITs). For convenience, this spin is denoted as the LAIT spin. Taking the RE 1 as an example, shown in Fig. 2, Spin 1 is the first LAIT spin corresponding to the first revolution and Spin 7 is the successive LAIT spin in the following revolution and so on.

When just these LAITs which have most significant amplitudes are used for fault diagnosis, a more reliable and accurate diagnostic results can be obtained. Comparatively, conventional methods such as envelope analysis can perform less effectively as it also involves other impulse responses which usually are smaller in amplitudes and can be interfered more with noise.

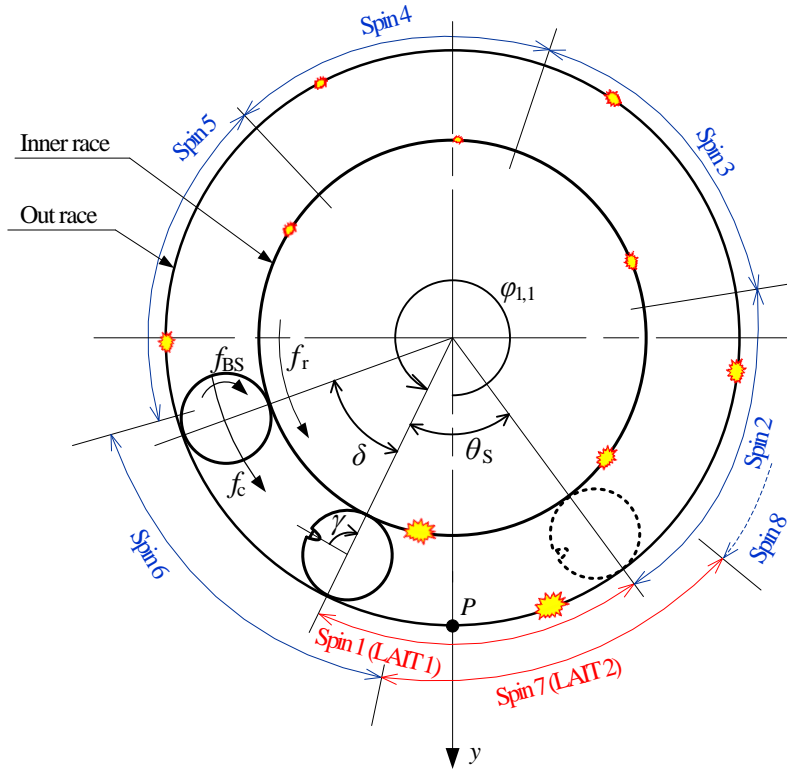


Fig. 2. The kinematics of a rolling bearing.

To separate the vibration signal segments associated with the LAIT spin periods, the phase relationship between LAITs needs to be developed. The distance that an element moves along the outer race is πd during one spin, where d is the diameter of the element. The subtended angle θ_s corresponding to the distance is given by

$$\theta_s = 2\pi d / (d_o \cos(\alpha)) \quad (5)$$

where d_o is the outer race diameter, and α is the contact angle. For brevity, the angle θ_s is also denoted as roller rotation subtended angle (RRSA). It is the RRSA that a rolling element revolves around the shaft centre during each spin period.

As shown in Fig. 2, supposing the initial phase of Spin 1 for RE 1 is $\varphi_{1,1}$ with respect to y-axis, the phase of Spin j for RE 1 can be expressed as

$$\varphi_{1,j} = \varphi_{1,1} + (j - 1)\theta_s. \quad (6)$$

The first 8 spins of RE 1 are illustrated in Fig. 2 anticlockwise. In the same way, the progressive phase $\varphi_{i,j}$ for Spin j of RE i can be calculated by

$$\varphi_{i,j} = \varphi_{1,j} - (i - 1)\delta. \quad (7)$$

where δ is the separation subtended angle between two successive elements, and $\delta = 2\pi/Z$.

According to the motion relationship as shown in Fig. 2, it is obvious that the phase $\varphi_{i,j}$ associated with a LAIT spin must meet the angular condition:

$$2\pi - \theta_s < \text{mod}(\varphi_{i,j}, 2\pi) \leq 2\pi, \quad (8)$$

i.e. the starting phase of each LAIT can be found from the wrapped reference phase of $\varphi_{i,j}$.

For the time sequence of vibration signal, RRSA θ_s corresponds to a spin period T_s , which can be calculated by

$$T_s = \frac{\theta_s}{2\pi f_c}. \quad (9)$$

The spin period also can be expressed as $T_s = 1/f_{BS}$ and calculated from Eq. (3). As two transients are generated during a spin, the spin period is twice as much as the fault period T_B of element defect. Similarly the time increment

sequence $t_{i,j}$ associates with the progressive phase $\varphi_{i,j}$ is

$$t_{i,j} = \frac{\varphi_{i,j}}{2\pi f_c} \quad (10)$$

Therefore, the condition in the time domain can be expressed as

$$T_R - T_S < \text{mod}(t_{i,j}, T_R) \leq T_R, \quad (11)$$

where $T_R = 1/f_c$ is the revolution period of the cage. Based on this condition, it is straightforward to find the starting time instant for each LAIT and thereby a LAIT segment with duration of T_S can be separated.

For a rolling bearing, RRSA is usually not an integer factor of 2π , which is designed for more uniform wear over the race ways and elements. This means that the relative phases $\text{mod}(\varphi_{i,j}, 2\pi)$ vary between LAIT spins, as illustrated by the two LAITs in Fig. 2, which leads to differences in the number of non-LAIT spins between two successive LAIT spins. Consequently the separated LAITs are not uniformly distributed in the time domain even if the rotational speed is constant.

It should be noted that the separated LAITs are not rigorously synchronous due to speed oscillation and slippage. Even an extremely small phase fluctuation will cause great loss of information about the transient waveform after an ensemble averaging. To overcome this deficiency, band-pass filtering, demodulation, and low-pass filtering are applied to the separated LAIT segments of vibration signals before an ensemble averaging. The processing results in a smoother envelope that is much less sensitive to the phase errors, and thereby an ELAIT based method for diagnosis of rolling element defect can be developed. The flowchart of the method will be introduced in the next section.

2.2. ELAIT based diagnosis method

The proposed fault diagnosis method based on ELAIT consists of three main steps: Step 1 is LAIT spin separation; Step 2 is LAIT processing; and Step 3 is LAIT ensemble averaging and normalisation. The schematic diagram of the method is shown in Fig. 3 which is depicted below.

2.2.1. LAIT spin separation

LAIT spin separation is the primary step in which a sequence of indices of LAIT spins for each rolling element is calculated according to the bearing kinematic motions as explained in Section 2.1. Then these indices are referenced to segregate short segments as raw LAIT signatures.

Based on the length of signal $X = [x_1, x_2, \dots, x_N]$ sampled with the sampling frequency of f_s , an angular increment sequence with J spins can be constructed to be $\Psi_i = [\varphi_{i,1}, \varphi_{i,1} + \theta_S, \varphi_{i,1} + 2\theta_S, \dots, \varphi_{i,1} + (J-1)\theta_S] \in \mathbb{R}^J$, for RE i , where $\varphi_{i,1} = \varphi_{1,1} - (i-1)\delta$, and $J = \lfloor Nf_{BS}/f_s \rfloor$. After wrapping the values of the sequence between $[0, 2\pi]$, the indices of LAIT spins can be then identified by applying the condition (8), and denoted as $D_i = [d_{i,1}, d_{i,2}, \dots, d_{i,k}, \dots, d_{i,K}] \in \mathbb{N}^K$, where $d_{i,k}$ is the signal index from which Spin k of RE i starts, $K = \lfloor Nf_c/f_s \rfloor$ is the number of LAIT spins. Finally a LAIT matrix is constructed for RE i as

$$\mathbf{M}_i = \begin{bmatrix} M_{i,1} \\ M_{i,2} \\ \vdots \\ M_{i,K} \end{bmatrix} = \begin{bmatrix} x(d_{i,1}) & x(d_{i,1}+1) & x(d_{i,1}+2) & \cdots & x(d_{i,1}+(L_2-1)) \\ x(d_{i,2}) & x(d_{i,2}+1) & x(d_{i,2}+2) & \cdots & x(d_{i,2}+(L_2-1)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(d_{i,K}) & x(d_{i,K}+1) & x(d_{i,K}+1) & \cdots & x(d_{i,K}+(L_2-1)) \end{bmatrix} \quad (12)$$

where $M_{i,k}$ is the k th LAIT segment of RE i , L_2 is the rounded length of spins, $L_2 = \lceil (\theta_S/2\pi)f_s/f_c \rceil$.

2.2.2. LAIT processing

LAIT processing is the second step in which all separated LAITs are processed with band-pass filtering, demodulation, and low-pass filtering before performing ensemble averaging in the next step. The information in a band with sidebands of fault frequency extending out either side of the natural frequency is enhanced by filtering the LAITs with a band-pass filter. A fast Kurtogram algorithm [9], which has been proved to be a very powerful and practical tool in bearing fault diagnosis, can be employed to search for the optimal frequency band for band-pass filtering.

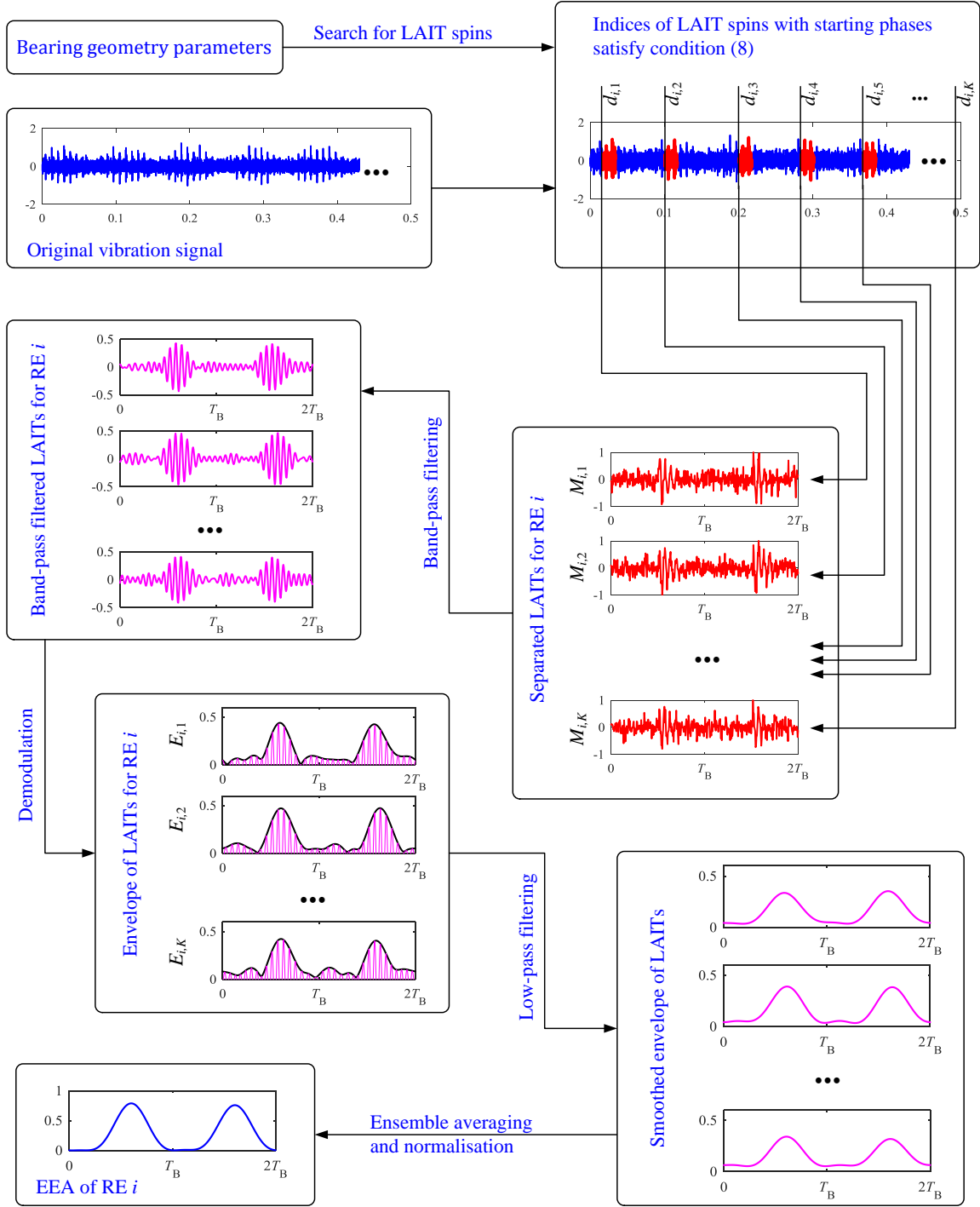


Fig. 3. Schematic diagram of ELAIT based diagnosis method.

Then a matrix of LAIT envelopes can be expressed as

$$\mathbf{E}_i = \begin{bmatrix} E_{i,1} \\ E_{i,2} \\ \vdots \\ E_{i,K} \end{bmatrix} = \begin{bmatrix} e(d_{i,1}) & e(d_{i,1}+1) & e(d_{i,1}+2) & \cdots & e(d_{i,1}+(L_2-1)) \\ e(d_{i,2}) & e(d_{i,2}+1) & e(d_{i,2}+2) & \cdots & e(d_{i,2}+(L_2-1)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e(d_{i,K}) & e(d_{i,K}+1) & e(d_{i,K}+1) & \cdots & e(d_{i,K}+(L_2-1)) \end{bmatrix} \quad (13)$$

where $E_{i,k}$ is the magnitude of the analytic signal of $M'_{i,k}$, $E_{i,k} = |M'_{i,k} + i\mathcal{H}(M'_{i,k})|$, $M'_{i,k}$ is the filtered signal segment of $M_{i,k}$, and $\mathcal{H}(M'_{i,k})$ is the Hilbert transform of $M'_{i,k}$.

At last the envelope waveform of impulses is made clearer by smoothing with a low-pass filter.

2.2.3. LAIT ensemble averaging and normalisation

LAIT ensemble averaging is the last step which performs the noise reduction philosophy of TSA. For RE i , the ensemble averaging result of the envelopes can be expressed as

$$A_i(t) = \frac{1}{K} \sum_{k=1}^K e(d_{i,k} + t - 1), t = 1, 2, \dots, L_2 \quad (14)$$

where $e(d_{i,k} + t - 1), t = 1, 2, \dots, L_2$ is the envelope of the k th LAIT for RE i as shown in Eq. (13). In order to obtain reliable and comparable results, the signatures of $A_i(t)$ are normalised with respect to the local minimal and global maximal by

$$A'_i(t) = \frac{\left(A_i(t) - \min_t A_i(t)\right)}{\max_{i,t} \left(A_i(t) - \min_t A_i(t)\right)}. \quad (15)$$

In this way, the normalised $A'_i(t)$ has a local minimum 0 and a global maximum 1. For brevity, this normalised ensemble averaging result $A'_i(t)$ is named as envelope ensemble average (EEA). Since the transient intensity is higher during LAIT spins of the defected element, peaks for the defected elements can be close to 1, whereas they can become ambiguous for elements away from the defected one.

2.3. Discussion

(1) The ELAIT based diagnosis method defines the element that is about to pass the load zone centre P first as RE 1. As the initial phase of rolling elements can differ before each test, different elements can be defined as RE 1 during different tests, and different elements can be located as the defected one. However, the purpose of the proposed method is not to locate the defected element in particular, but to extract enhanced features to indicate rolling element defects.

(2) When ignoring the angle γ (as shown in Fig. 2) between the defect point and the line connecting the shaft centre and the rolling element centre, the initial phase $\varphi_{1,1}$ between RE 1 and y-axis is the only unknown variable in the separation method. Once $\varphi_{1,1}$ is set, the phase $\varphi_{i,j}$ of Spin j for RE i can be set according to Eq. (6) ~ (7). The initial phase $\varphi_{1,1}$ can be set according to the cage speed signal measured with optical sensors, but mounting an embedded optical sensor into bearing to measure the cage speed signal is not allowed in most cases. As the initial phase $\varphi_{1,1} \in (2\pi - \delta, 2\pi]$, if we set $\varphi_{1,1}$ to be the median of the interval, $\bar{\varphi}_{1,1} = 2\pi - \delta/2$, the error of the phase $e_{ph} = |\varphi_{i,j} - \bar{\varphi}_{i,j}| = |\varphi_{1,1} - \bar{\varphi}_{1,1}| \leq \delta/2, \forall i$. It means that the phase error between a selected LAIT spin and the ideal LAIT spin has a maximum of $\delta/2$. As the primary task for the bearing fault detection is usually to find if any element is defected, the information regarding to which of elements is not necessary. In this sense, the initial phase $\varphi_{1,1}$ can be set to any value.

(3) Two elements may undergo through the load zone centre P during a LAIT spin. If the separation subtended angle δ between two successive elements is smaller than RRSA θ_S , then in cases of

$$2\pi - (\theta_S - \delta) < \text{mod}(\varphi_{i,j}, 2\pi) \leq 2\pi, \quad (16)$$

both the phase $\varphi_{i,j}$ and $\varphi_{i+1,j} = \varphi_{i,j} - \delta$ satisfy inequality (8), which means that the spins whose phases satisfy inequalities (16) are selected as the LAIT spins for both RE i and RE $i+1$.

(4) The characteristic fault frequencies of rolling bearings are based on the assumption that there is pure rolling contact between the elements and raceways. However, pure rolling contact is an ideal assumption and slippage as well as speed oscillation happens inevitably in reality. The contact angle varies with the position of each rolling element in the bearing as the ratio of local radial to axial load changes. Thus, each rolling element has a different effective rolling diameter and is trying to roll at a different speed, but the cage limits the deviation of the rolling elements from their mean position, thus causing some random slippage, which is typically less than 2% of the fault period for steady operation systems [2]. Slippage and speed oscillation can cause phase errors and then non-synchronous LAITs. Typically, the transients have a waveform of damping oscillation, which lasts several natural periods. When performing ensemble averaging of the raw LAITs, the LAITs are required to be synchronous in the time scale of the natural period. As the natural period of the transients is several times smaller than the length of the transients, the envelope is much

less sensitive to the phase errors than the damping oscillation itself. Thus the ELAIT based diagnosis method enhances the envelope waveform of the two transient impulses in a spin period by performing ensemble averaging of the LAIT envelopes instead of the LAITs themselves.

3. Performance Evaluation

To evaluate the performance of the proposed ELAIT method, vibration signals simulated by a model of rolling bearings and acquired from a test rig are based for various cases. Simulation signals illustrate the efficiency of the ELAIT method in cases where there are known phase errors caused by speed oscillation, rolling element slippage and noise. And experimental validation illustrates the efficiency of the ELAIT method for real bearings that can have unexpected influences such as random signals from other components of a machine.

The tested bearings are of MB ER-10K deep groove ball bearings with 8 balls. Other geometric parameters are the rolling element diameter $d = 7.9375$ mm, the pitch diameter $D = 33.5026$ mm, and the design contact angle $\alpha = 0$. Then according to Eq. (1) and (3), the ball-spin frequency $f_{BS} = 1.992f_r$ and the revolution frequency $f_c = f_{BPO}/Z = 0.3815f_r$.

3.1. Simulation validation

3.1.1. Localised fault model of rolling bearings with phase errors

Phase errors of fault transients have received high research attention due to their significant influences on bearing diagnostics. Two different models were successively proposed in literature. The first one in Ref. [20] adds a small random variable $\delta\tau_i$ to the occurrence time of the i^{th} impact:

$$T_i = iT + \delta\tau_i. \quad (17)$$

where T is the impact period associated with the fault frequency. This model gives a cyclostationary signal with the uncertainty of occurrence independent of the number of periods into the future. A more acceptable model was later introduced in Ref. [2] and [21] to define the difference ΔT_i between two successive occurrence times to be a random variable, i.e.

$$\Delta T_i = T_{i+1} - T_i \quad (18)$$

The second model gives a pseudo-cyclostationary signal with the uncertainty of occurrence increasing with the number of periods into the future.

As the occurrence time of the i^{th} impact can be obtained by using both the shaft period T_r and bearing period factor P_f by

$$T_i = iP_f T_r, \quad (19)$$

applying differentiation to both sides of Eq. (19) yields

$$dT_i = iP_f dT_r + iT_r dP_f, \quad (20)$$

which shows that the phase error dT_i of occurrence time can be caused by both speed oscillation dT_r and variance of P_f . For a machine with a speed controller, variation of shaft period is nearly random and independent of the number of periods into the future, thus $iP_f dT_r$ can be regarded as the random error $\delta\tau_i$ between real occurrence time T_i and theoretical occurrence time iT , thus the first model Eq. (17) should be applied. For slippage the system has no memory and dP_f changes in each period, thus $iT_r dP_f$ can be regarded as cumulus $\sum_{j=0}^{i-1} \delta\tau'_j$ where $\delta\tau'_j$ is a random variable of slippage. It means that errors caused by slippage cumulate with the number of periods and the second model Eq. (18) should be chosen.

Then the phase error of Eq. (20) can be expressed as

$$dT_i = \delta\tau_i + \sum_{j=0}^{i-1} \delta\tau'_j. \quad (21)$$

and $T_i = iT + dT_i$.

Hereby a localised fault model of rolling bearings is presented as

$$x(t) = \sum_{i=0}^M (A_0 \cos(2\pi f_m T_i)) \left(e^{-B(t-T_i)} \cos(2\pi f_n(t-T_i)) \right) + n(t), \quad (22)$$

where A_0 is the amplitude of the impulse produced by defect, f_m is the modulation frequency, B is the attenuation coefficient of transient impulse, f_n is the bearing natural frequency, $n(t)$ is external random noise, and T_i is the occurrence time of impact with phase error $\delta\tau_i$ caused by speed oscillation and phase error $\sum_{j=0}^{i-1} \delta\tau'_j$ caused by slippage.

Here definitions are given as follows. The speed oscillation degree (*SOD*) is defined as the ratio of mean of absolute speed variation to speed. As fault period is proportional to speed, the speed oscillation degree can be given as

$$SOD = \frac{\text{mean}_i |\delta\tau_i|}{T} \times 100\%, \quad (23)$$

The slippage degree (*SD*) is defined as the ratio of the mean of absolute slippage to the theoretical period,

$$SD = \frac{\text{mean}_i |\delta\tau'_i|}{T} \times 100\%, \quad (24)$$

The SNR of the simulation signal is defined as

$$SNR = 10 \log_{10} \frac{P_s}{P_N}, \quad (25)$$

where P_s is the power of the noise-free signal and P_N is the power of noise.

In next 2 sections simulation signals with phase errors caused by speed oscillation and element slippage are generated respectively to validate the ELAIT method. In all cases the sampling frequency $f_s = 25.6$ kHz, the shaft speed is at $f_r = 30$ Hz and the natural frequency is set as $f_n = 2000$ Hz. Then the ball-spin frequency $f_{BS} = 1.992 f_r = 59.7582$ Hz and the modulation frequency $f_c = 0.3815 f_r = 11.4462$ Hz. The fault frequency $2f_{BS} = 119.516$ Hz. A spin lasts a duration of $T_S = 1/f_{BS} = 0.0167$ s, which is double of the fault period $T_B = 0.0084$ s. The simulation signal lasts $L = 10$ s, during which there are $[L \cdot f_{BS}] = 597$ whole spins and 1194 transient impulses. The cage revolves $[L \cdot f_c] = 114$ whole cycles, thus 114 LAIT spins are separated for each element.

3.1.2. Validation for speed oscillation cases

A sequence of random variables with uniform distribution is generated as speed oscillations (as shown in Fig. 4; there is no proof that speed oscillation or slippage has uniform distribution or Gaussian distribution). The speed oscillation degree is yielded according to Eq. (23), $SOD = 5\%$.

Fig. 5 shows the first case (named as Case 1) of simulation signals with the above speed oscillations. The signal noise ratio is -18.05 dB and the localised defect is on RE 4. Changes in the occurrence times of impacts can be observed from the partial enlarged detail of the noise free signals as shown in Fig. 5(b). The natural frequency and sidebands on both sides cannot be seen in the spectrum. A bandwidth of 800 Hz is selected for the band-pass filter as shown by two red lines in the spectrum subfigure of Fig. 5(c), thus at least 3 sidebands are included in each side of the natural frequency. The envelope spectrum of the band-pass filtered signals shows clearly the harmonics of fault frequency with sidebands of cage frequency, typifying the features of defect on a rolling element.

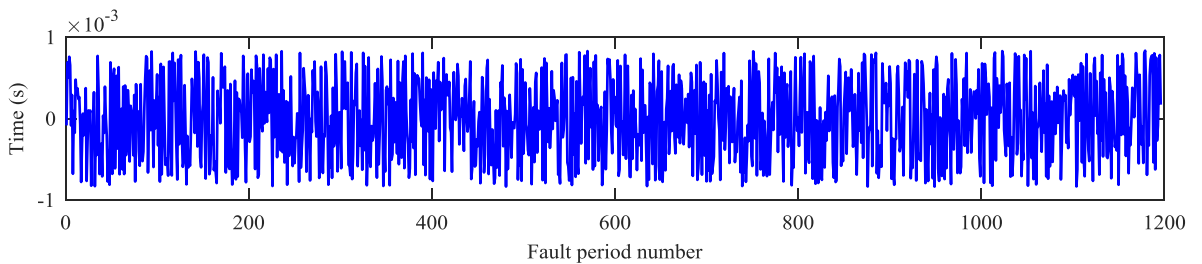


Fig. 4. Phase errors caused by speed oscillation.

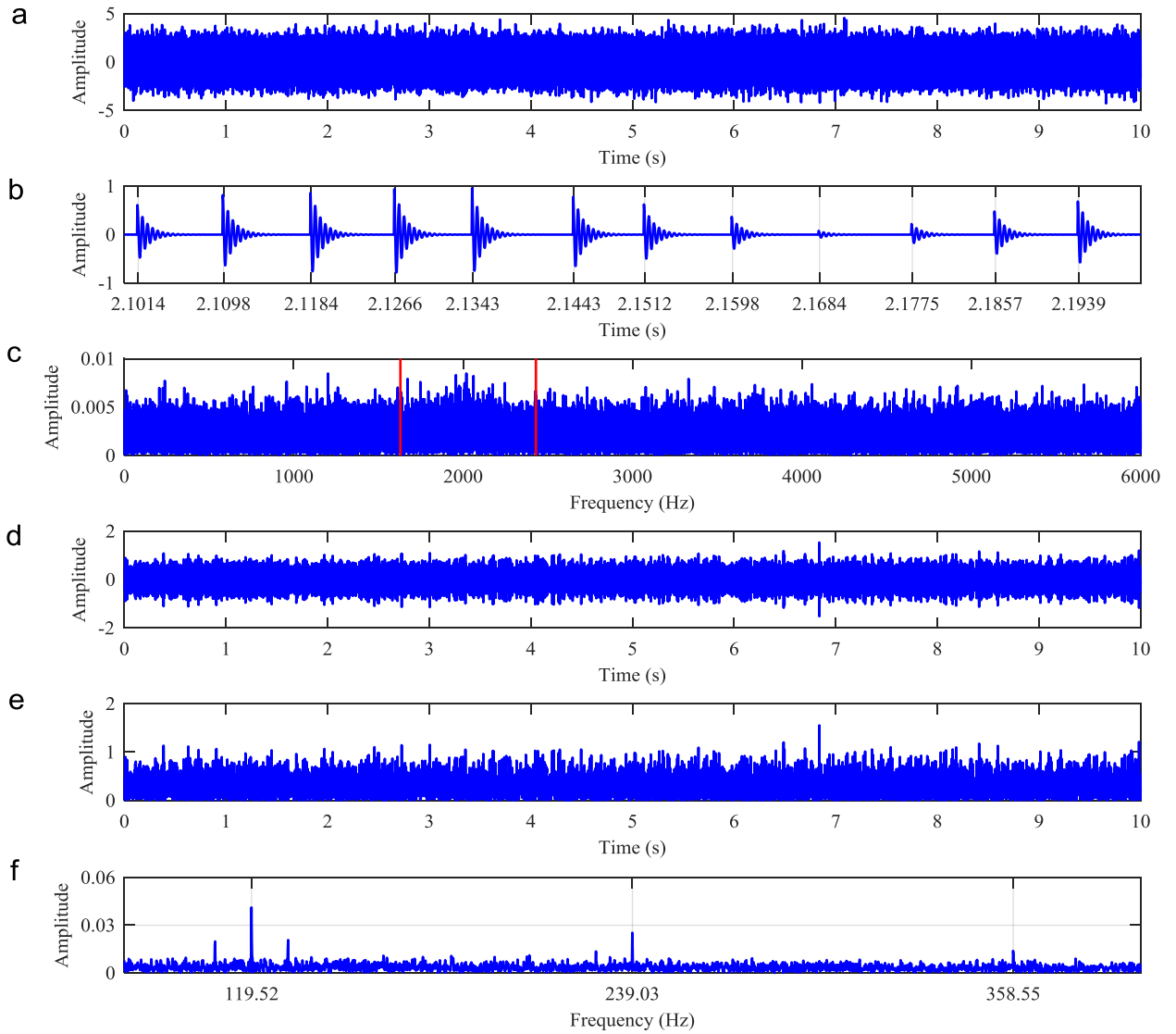


Fig. 5. Simulation signals and their spectra (Case 1): (a) time domain signals; (b) partial enlarged detail of noise free signals; (c) amplitude spectrum; (d) filtered signals; (e) envelope; and (f) envelope spectrum.

Fig. 6 shows the EEAs of the above signals. It can be seen clearly that there are two peaks in each EEA of RE 1 ~ RE 6. These two peaks are the ensemble averages of LAIT envelopes. The two peaks have two significant characters: (1) the two peaks of each EEA are uniformly distributed, and the interval between these two peaks is just one fault period T_B ; (2) these two peaks in the EEAs of RE 1 ~ RE 6 appear at the same time instants. These characters indicate the defect of rolling elements. It also can be seen that the peaks have bigger amplitudes in one or two rolling elements and the amplitudes decrease gradually in both clockwise and anticlockwise directions. Particularly, Fig. 6 shows that peaks have bigger amplitudes in the EEAs of RE 3 and RE 4. Peak amplitudes decrease from RE 4 to RE 6 in the clockwise direction and from RE 3 to RE 1 in the anticlockwise direction. The peaks become ambitious in the EEAs of RE 7 and RE 8. The change features of these peak amplitudes indicate that the defect causes bigger impulses when RE 3 and RE 4 pass the load zone centre. Although it may be hard to determine exactly which of RE 3 or RE 4 is defected, a diagnostic result that at least one of them is defected can be obtained, which is consistent with the simulated case.

On the other hand, when applying the proposed method to the background noise only, the EEAs obtained, shown in Fig. 7, exhibit random oscillations, or have no features of two peaks at the interval of one fault period. This further verifies that the proposed method is only effective when there are fault transients.

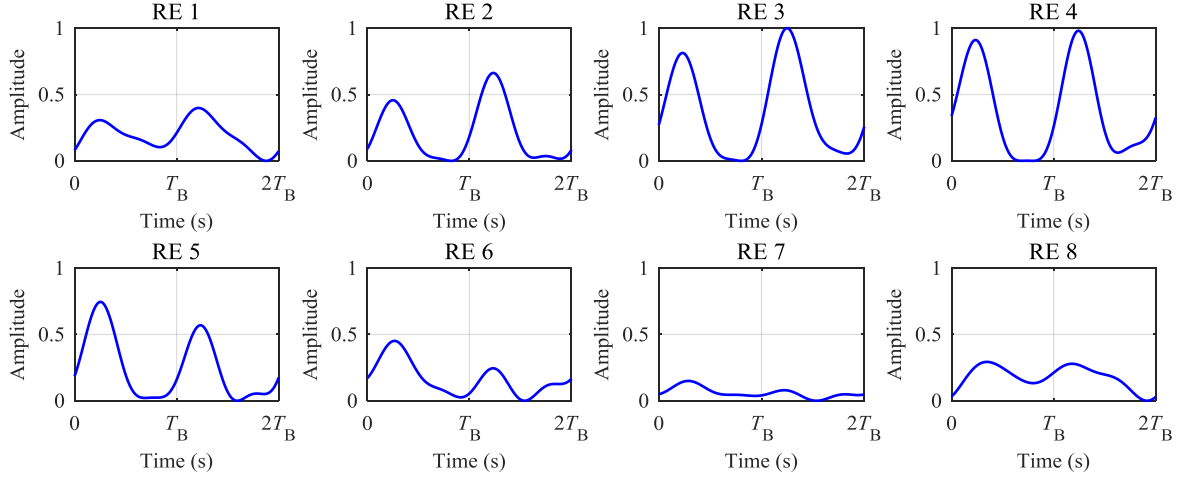


Fig. 6. EEAs of the simulation signals (Case 1).

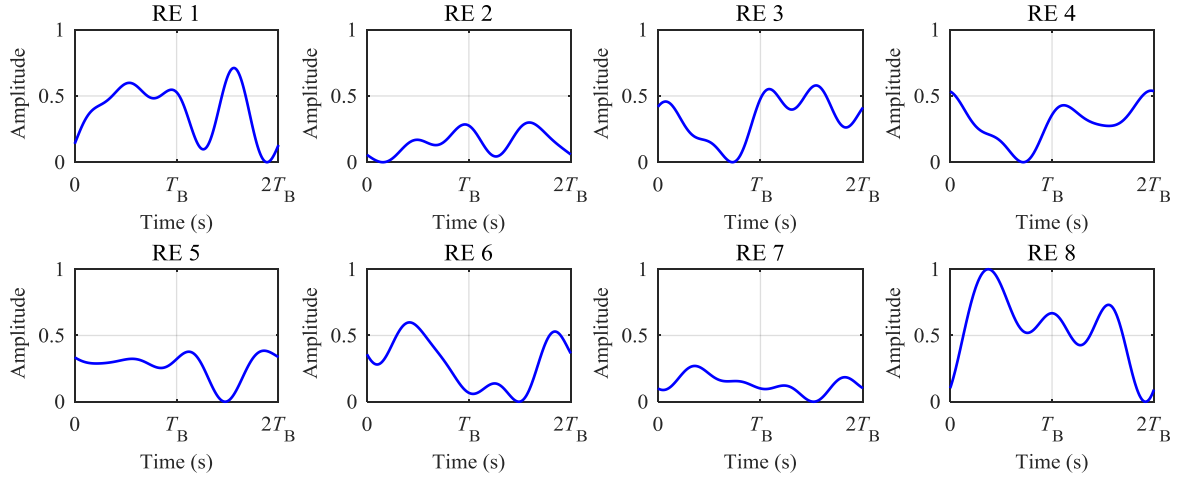


Fig. 7. EEAs from background noise signal (Case 1).

Fig. 8 shows a high noise level case, denoted as Case 2, whose SNR is as low as -24.93 dB. The simulation signals are with the same speed oscillations of Case 1, and the defect is also on RE 4. Changes of impact occurrence times can be observed in the partial enlarged noise free signal. Neither the natural frequency with sidebands on both sides can be seen in the spectrum nor can the harmonics of fault frequency be seen from the envelope spectrum. Harmonics of the fault frequency and sidebands of f_c are submerged in noise.

EEAs obtained from the above signals are shown in Fig. 9, from which two uniformly distributed peaks can be seen in each EEA of RE 2 ~ RE 5. Compared with the peaks in the EEAs yielded in Case 1 (as shown in Fig. 6), the peaks in the second case have flatter top and bigger tailedness. However, the two peaks in each EEA of RE 2 ~ RE 5 exhibit the two particular characters as mentioned in Case 1: the two peaks of each EEA are uniformly distributed, and the interval between these two peaks is one fault period T_B ; and these two peaks in the EEAs of RE 2 ~ RE 5 appear at the same time instants. From Fig. 9 it also can be seen that the peak amplitudes are bigger in the EEAs of RE 3 and RE 4 than those in the EEAs of RE 2 and RE 5. The amplitudes of these peaks indicate that the defect causes bigger impulses when RE 3 and RE 4 pass the load zone centre and at least one of them is defected. The results show the effectiveness of the ELAIT method in case of speed oscillation and very low SNR , demonstrating its high robustness to noise influences.

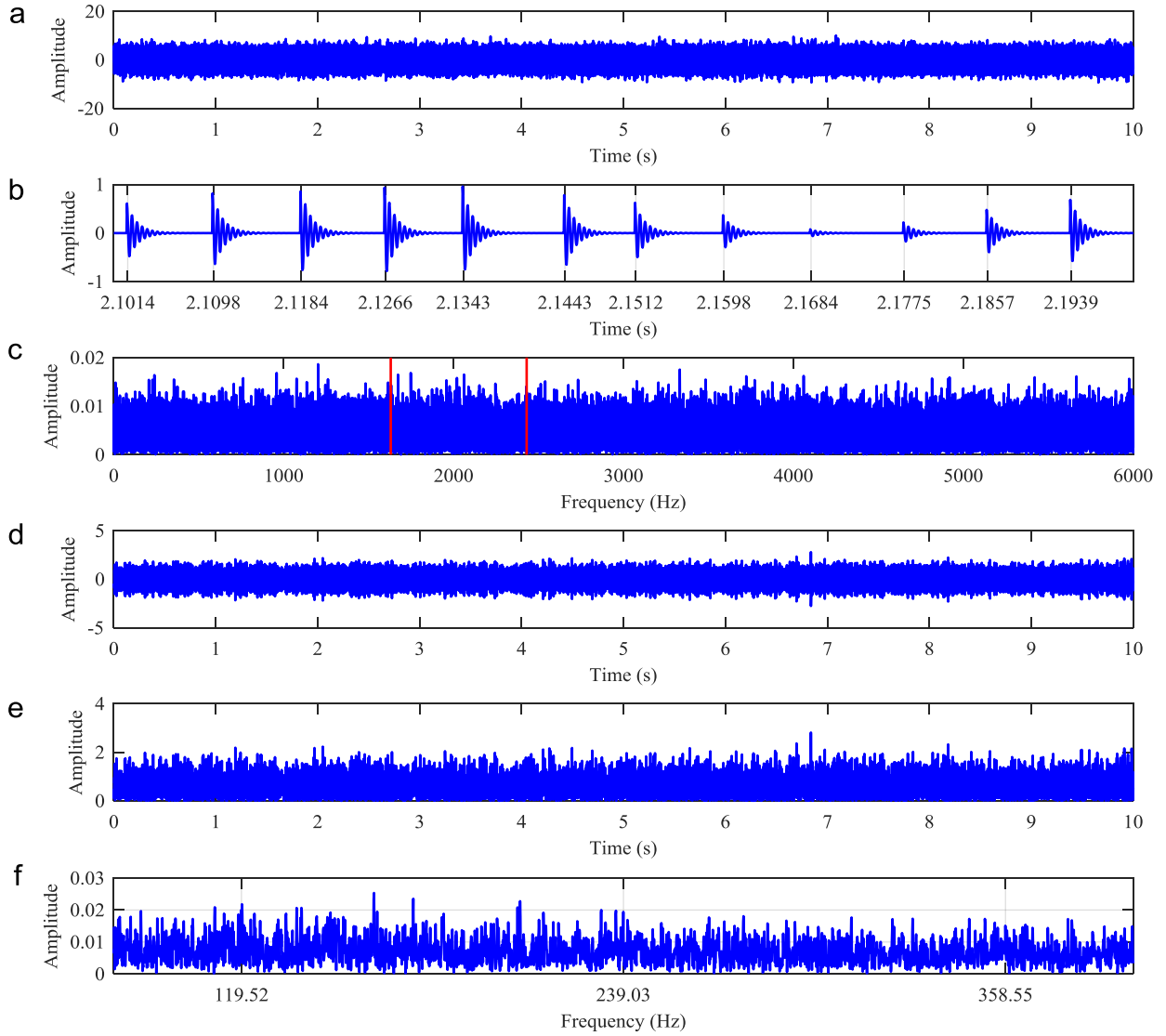


Fig. 8. Simulation signals and their spectra (Case 2): (a) time domain signals; (b) partial enlarged detail of noise free signals; (c) amplitude spectrum; (d) filtered signals; (e) envelope; and (f) envelope spectrum.

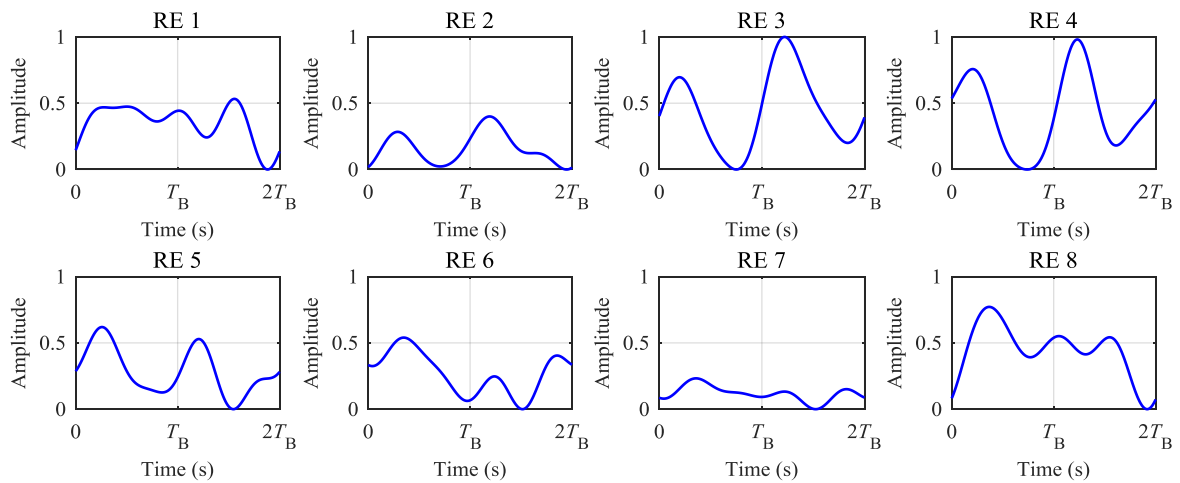


Fig. 9. EEAs of the simulation signals (Case 2).

3.1.3. Validation for slippage cases

A random variation sequence $\delta\tau'_j$ of slippage with uniform distribution is generated 50 times. The slippage sequences have a maximum of $T_B \times 2\%$, and the slippage degrees of these 50 sequences are yielded according to Eq. (24), and $SD \approx 1\%$. The corresponding phase error sequences $\sum_{j=0}^{i-1} \delta\tau'_j$ are shown in Fig. 10. It can be seen that the phase error sequences differ greatly and the variance of phase error increases with the number of fault periods.

Here a phase error sequence is selected as shown in Fig. 11. Fig. 12 shows the first slippage case (named as Case 3) of simulation signals with $SNR = -18.05$ dB and $SD = 1\%$. The localised defect is set on RE 4. Small changes of impact occurrence times can be observed from the partial enlarged noise free signals. The natural frequency and its sidebands of fault frequency on both sides cannot be seen in the spectrum. The first and second order harmonics of fault frequency $2f_{BS}$ and their sidebands of f_c can be seen in the envelope spectrum.

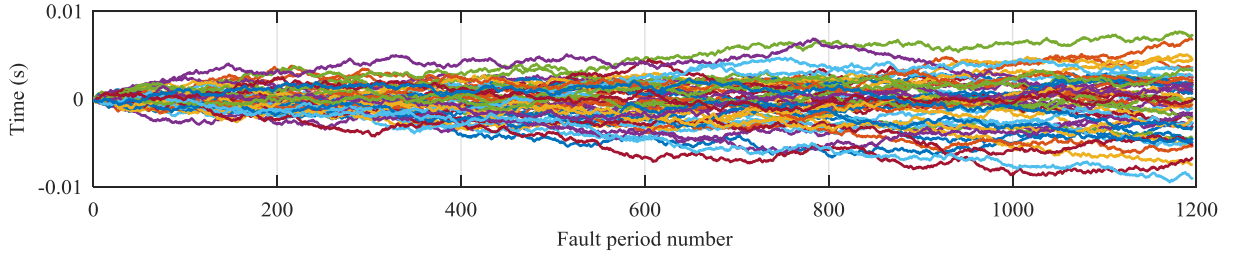


Fig. 10. Phase errors generated for 50 realizations.

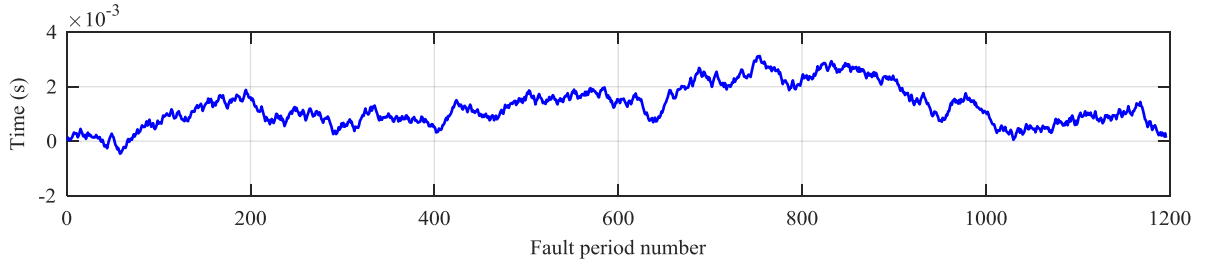


Fig. 11. Phase error caused by slippage.

The ELAIT results for the signals are shown in Fig. 13, from which two uniformly distributed peaks can be seen in the EEAs of RE 1 ~ RE 6. The two peaks in the EEAs exhibit the two particular characters: the interval between the two peaks is just one fault period, and these two peaks appear at the same time instants of the EEAs. Peak amplitudes decrease from RE 3 to RE 1 in the anticlockwise direction and from RE 3 to RE 6 in the clockwise direction. Such peaks cannot be clearly seen in the EEAs of RE 7 and RE 8 which are furthest away from RE 3 and RE 4. The change features of these peaks indicate that the defect causes bigger impulses when RE 3 passes the load zone centre. These peaks illustrate that at least one of RE 1 ~ RE 6 is defected and RE 3 is the most probably defected one. The conclusion is not very agreeable to that RE 4 is the defected element. The reason is that cumulative phase errors caused by slippage and initial phase $\varphi_{1,1}$ shift the locations of the impulses. Nevertheless, a conclusion of rolling element defect can be drawn according to the two peaks shown in the EEAs.

Fig. 14 shows a high noise level case (named as Case 4) whose SNR is as low as -24.5 dB when the simulation signals are with the same $SD = 1\%$, and the defect is also on RE 4. Because of high noise effect, fault features can be found from neither the spectrum nor the envelope spectrum, showing the ineffectiveness of conventional method in detecting the fault. Amazingly, the EEAs, as shown in Fig. 15, present two clear peaks at RE 3 ~ RE 5. The interval between these two peaks is just around one fault period, and these two peaks appear at about the same time instants in the EEAs. These peaks indicate that there are defects on rolling elements even though it may be not so sure which of these three elements are defected.

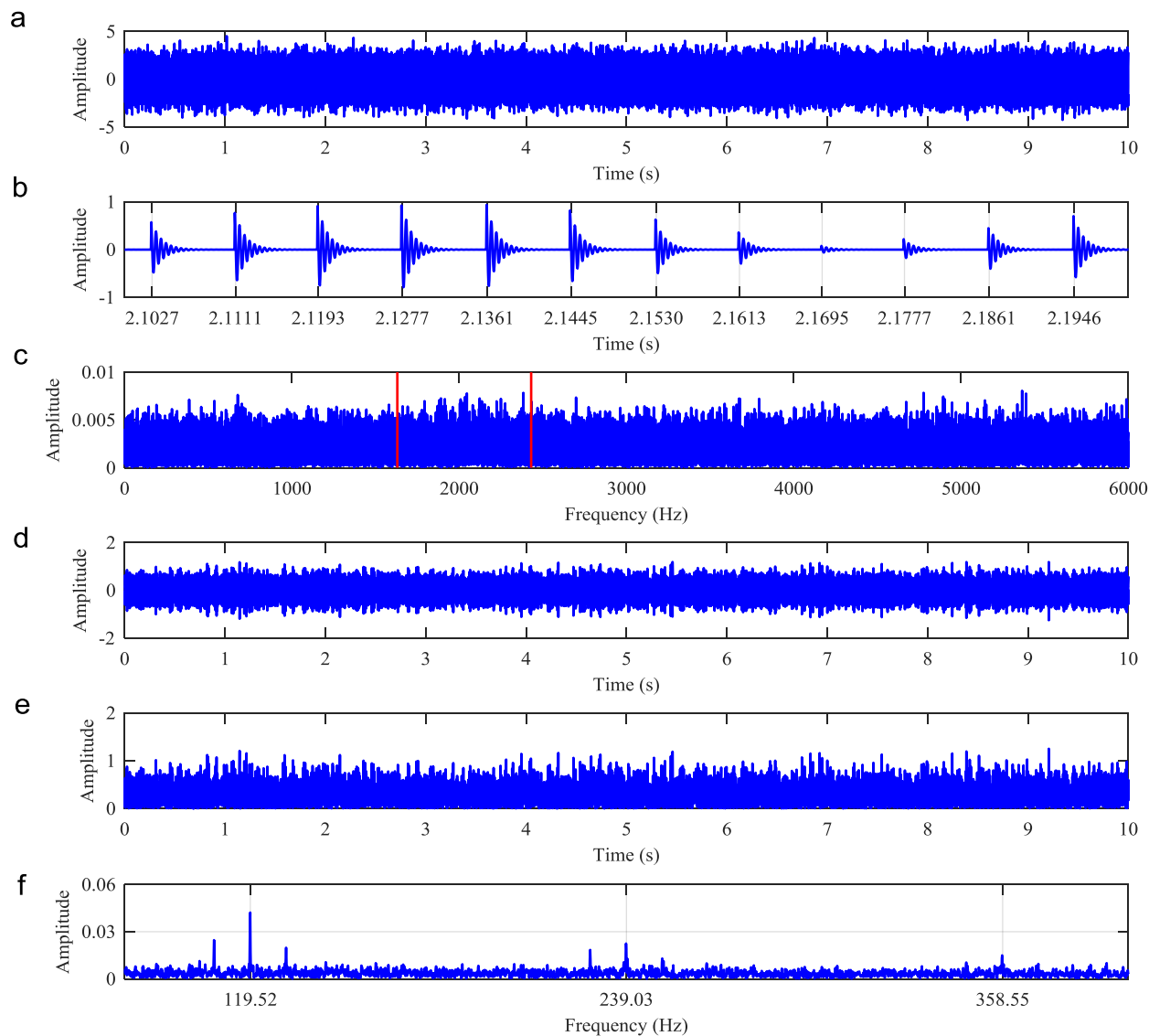


Fig. 12. Simulation signals and their spectra (Case 3): (a) time domain signals; (b) partial enlarged detail of noise free signals; (c) amplitude spectrum; (d) filtered signals; (e) envelope; and (f) envelope spectrum.

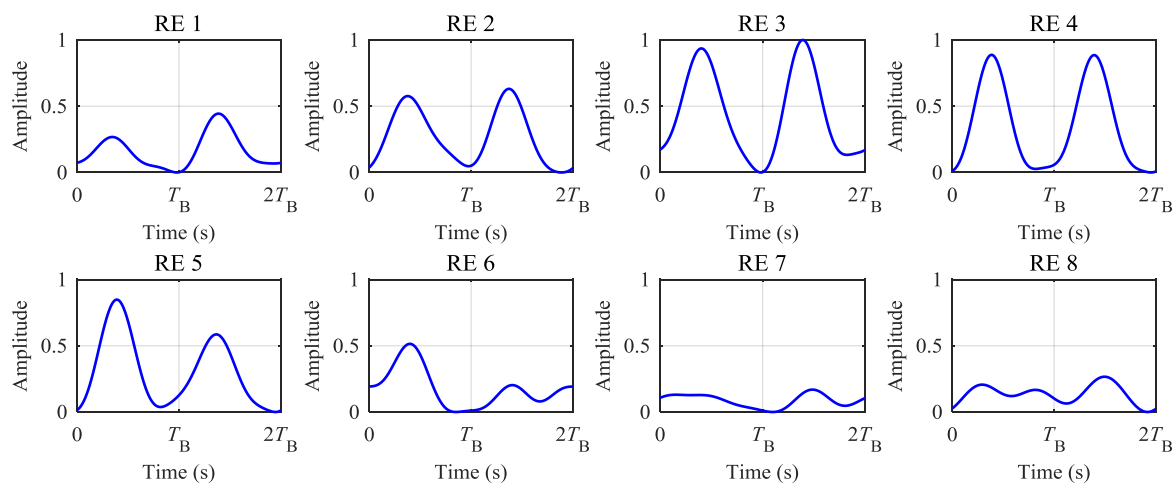


Fig. 13. EEAs of the simulation signals (Case 3).

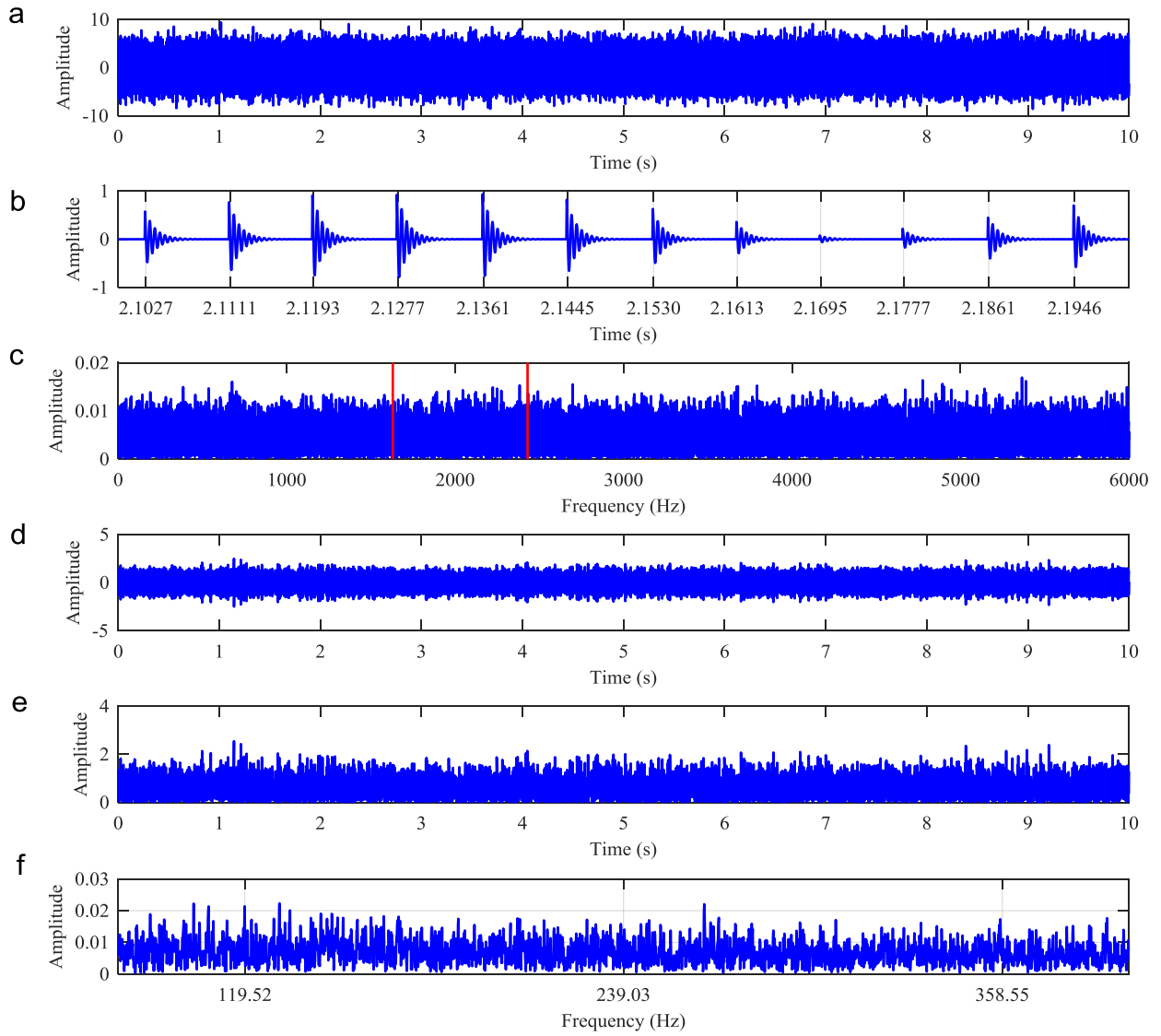


Fig. 14. Simulation signals and their spectra (Case 4): (a) time domain signals; (b) partial enlarged detail of noise free signals; (c) amplitude spectrum; (d) filtered signals; (e) envelope; and (f) envelope spectrum.

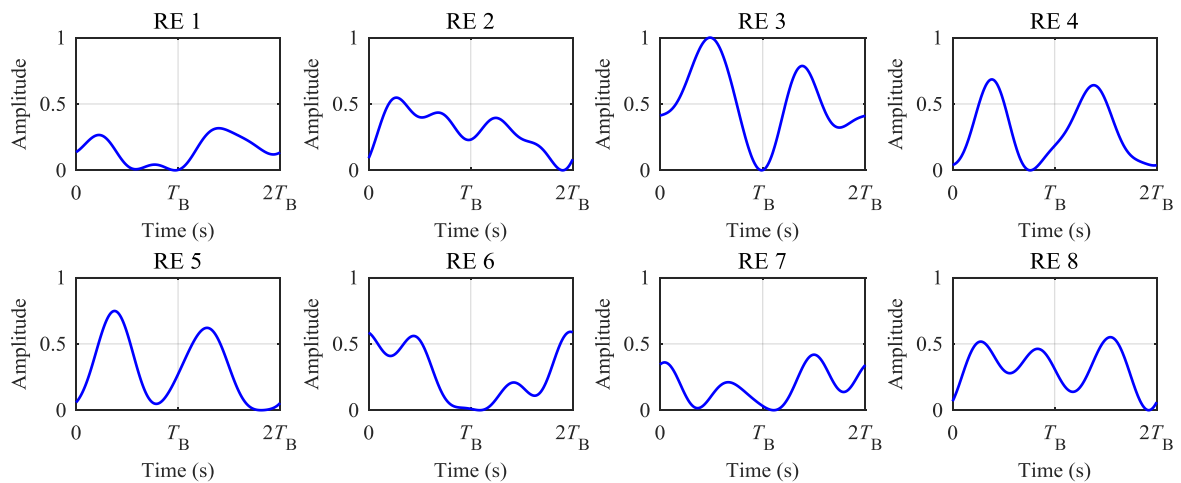


Fig. 15. EEAs of the simulation signals (Case 4).

3.2. Experiment validation

The experiments were carried out based on a machinery fault simulation bench as shown in Fig. 16. The bench mainly consists of a motor, two bearings, an inertia wheel, two double groove V-belt sheaves, a 3-way gearbox, a crank wheel mechanism, and a reciprocating mechanism with a spring. The gravity force ($5 \text{ kg} \times 9.8 \text{ m} \cdot \text{s}^{-2}$) from the inertia wheel is employed to provide the radial load to excite bearing fault. A varying elastic force is generated by the spring during the operation of the reciprocating mechanism. The varying elastic force is employed to provide a varying additional load. The varying load is transmitted to the motor and causes speed oscillation. Meanwhile, the varying load also changes the ratio of local radial to axial load, and causes the contact angle varying with the position of each rolling element in the bearing, and then causes slippage. Two accelerometers with bandwidth from 1 Hz to 9 kHz are mounted on the bearing housing to measure vibrations in both vertical and horizontal radial directions. A tachometer is employed to measure rotational speed of the shaft. The rotational speed is used for validation of the ELAIT method, and performing ELAIT does not need the tachometer signals.

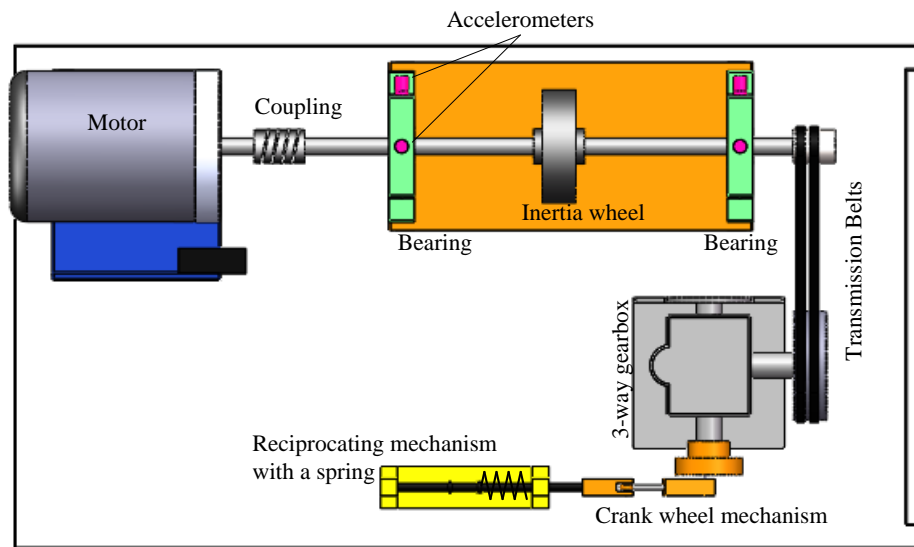


Fig. 16. Machinery fault simulation bench.

The ball bearing seeded with a localised defect was mounted in the left bearing housing. Fig. 17(a) shows the vibration signal in the horizontal direction of interest. The signal length is $L = 14 \text{ s}$. The sampling frequency is $f_s = 25.6 \text{ kHz}$. In addition to the effect of load zone, the defected ball can also rotate the axial direction. It means that the defect will not always interact with the inner and outer races, and therefore impulses are much less significant, compared with that when it always interacts with races. Consequently the signal shows highly nonstationary in that some of transients exhibit higher amplitudes than others, as shown in Fig. 17(a).

Fig. 17(b) shows the rotational speed calculated with zero cross detection method from the tachometer signals. It can be seen that the speed oscillates between 18.9 Hz-19.2 Hz around the mean value of 19.06 Hz. The speed oscillation degree is $SOD = 1.14\%$, which demonstrates inevitable speed variation existing in such a test rig. **It is worth noting that** the mean speed can also be estimated from the amplitude spectrum of the vibration signals using the higher order harmonics of the shaft frequency. With the estimated speed, fault frequency is calculated to be $2f_{BS} = 75.93 \text{ Hz}$ and the fault period is 0.0132 s. The modulation frequency of the cage speed is $f_c = 7.27 \text{ Hz}$. Thus the signal covers $[L \cdot f_{BS}] = 531$ whole spins and 1062 transient impulses. And the cage revolves $[L \cdot f_c] = 101$ cycles, allowing 101 LAIT spins to be separated for each rolling element.

The Kurtogram analysis result of the signal is shown in Fig. 17(c). It yields an ‘optimal’ band of 11,733-12,800 Hz, denoted as Band 1, in which the kurtosis value is 226.8. However, the signal in Band 1 may be affected by aliasing as it is close to $f_s/2$ and there is inevitable transient of the built-in antialiasing filter. Considering the avoidance of this potential effect, four possible bands are selected, in which high amplitudes relating resonances are in Band 2 of 1500-3600 Hz, Band 3 of 4100-4500 Hz, Band 4 of 4800-5620 Hz, and Band 5 of 9000-11,200 Hz. These bands are highlighted by the two vertical lines in Fig. 17(d).

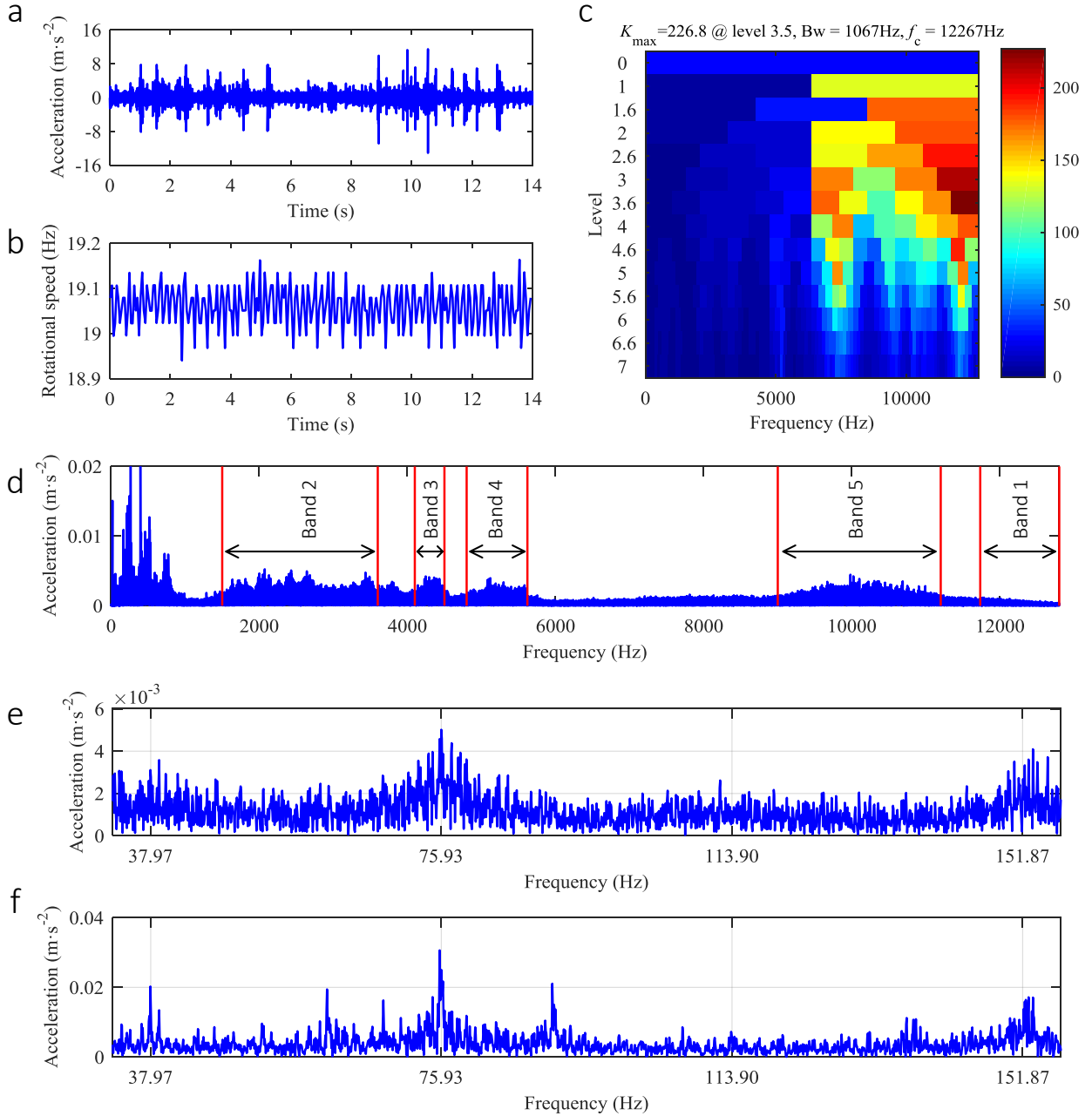


Fig. 17. Test signals and spectra of a faulty bearing: (a) raw vibration signals; (b) rotational speed; (c) Kurtogram results of the vibration signals; (d) amplitude spectrum; (e) envelope spectrum in Band 1; and (f) envelope spectrum in Band 2.

The envelope spectrum from the signal of Band 1 is shown in Fig. 17(e). It can be seen that the peaks at the frequencies of $2f_{BS}$ and $4f_{BS}$ are badly smeared, causing high ambiguity in detecting the fault. Especially, there are high background noise components, making more difficulty in determine the existence of the fault. Meanwhile, envelope spectra from other four bands, especially Band 2, can exhibit a more significant peak at the fault frequency. As shown in Fig. 17(f), characteristic peaks at f_{BS} , $2f_{BS}$, and $4f_{BS}$ are presented better in Band 2. This may demonstrate that Kurtogram can produce a biased band due to the high nonstationarity of the signal consisting high local amplitudes that can cause more aliasing effect too. Nevertheless, the spectrum from Band 2 still shows high smearing effects and noise that influence diagnostic results.

Performing ELAIT resulted in the EEAs for the signals in Band 1 and Band 2, as presented in Fig. 18 and Fig. 19 respectively. Due to the effect of noise suppression, both of these two figures show two clear peaks in the EEAs of RE 1 ~ RE 7. These two peaks have two particular characters: (1) the interval between these two peaks is about one fault period; (2) these two peaks appear at about the same time instants in the EEAs. In addition, it also can be seen that the EEAs have higher peak amplitudes in RE 2 ~ RE 6 and smaller peak amplitudes in RE 1 and RE 7, showing

asymmetric effect of impulse responses. However, the differences between the EEAs for this test case are not very apparent because of the small size of the bearing unit. Yet, EEA signatures with clear two peaks in RE 1 ~ RE 7 can be based on to indicate that at least one of the rolling elements is defective. In addition, comparing EEAs between these two bands, they both show clear features for diagnosing the faults. Therefore, ELAIT method is not very sensitive to band selection; likely resonance bands can be effective.

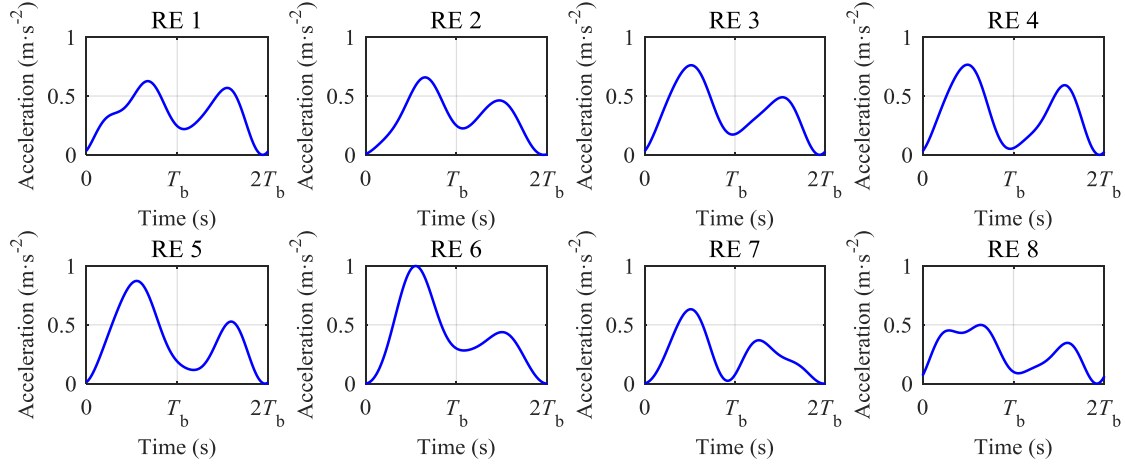


Fig. 18. EEAs of the faulty bearing using ELAIT with Band 1 from Kurtogram.

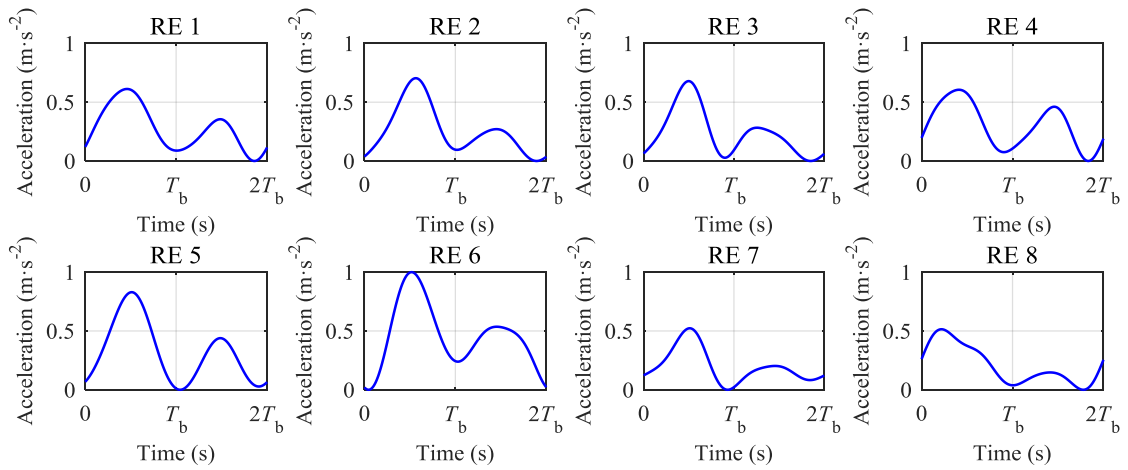


Fig. 19. EEAs of the faulty bearing yielded by performing ELAIT with Band 2.

To verify further the robustness of the ELAIT method, vibration data collected from a healthy bearing is examined to show the EEA characteristics that do not have the two character peaks. Fig. 20(a) shows the vibration signal. The speed as shown in Fig. 20(b) oscillates around the mean value of 18.87 Hz. The speed oscillation degree is $SOD = 1.54\%$. The corresponding fault frequency is $2f_{BS} = 75.18$ Hz and the fault period is 0.0133 s. The modulation frequency of the cage speed is $f_c = 7.20$ Hz. There are $[L \cdot f_{BS}] = 526$ whole spins and 1052 transient impulses during the test. The cage revolves $[L \cdot f_c] = 100$ whole cycles, thus 100 LAIT spins are separated for each rolling element.

The Kurtogram analysis result of the vibration signal is shown in Fig. 20(c). It can be seen that the optimal band from Kurtogram is from 8200Hz to 8400Hz, denoted as Band 1, in which the kurtosis is 25.2. In addition, two more potential bands, in which there are high amplitudes relating resonances, are also chosen, and denoted as Band 2 of 1400-2400 Hz and Band 3 of 9600-10,800 Hz respectively. These three bands are highlighted by the two vertical lines in Fig. 20(d).

The EEAs of the healthy bearing for the signals in Band 1 and Band 2 are shown in Fig. 21 and Fig. 22 respectively. The two peaks in EEA signatures as mentioned above cannot be found, meaning the bearing is in health condition. This then further verifies that the proposed method is reliable in indicating the fault of rolling element defect.

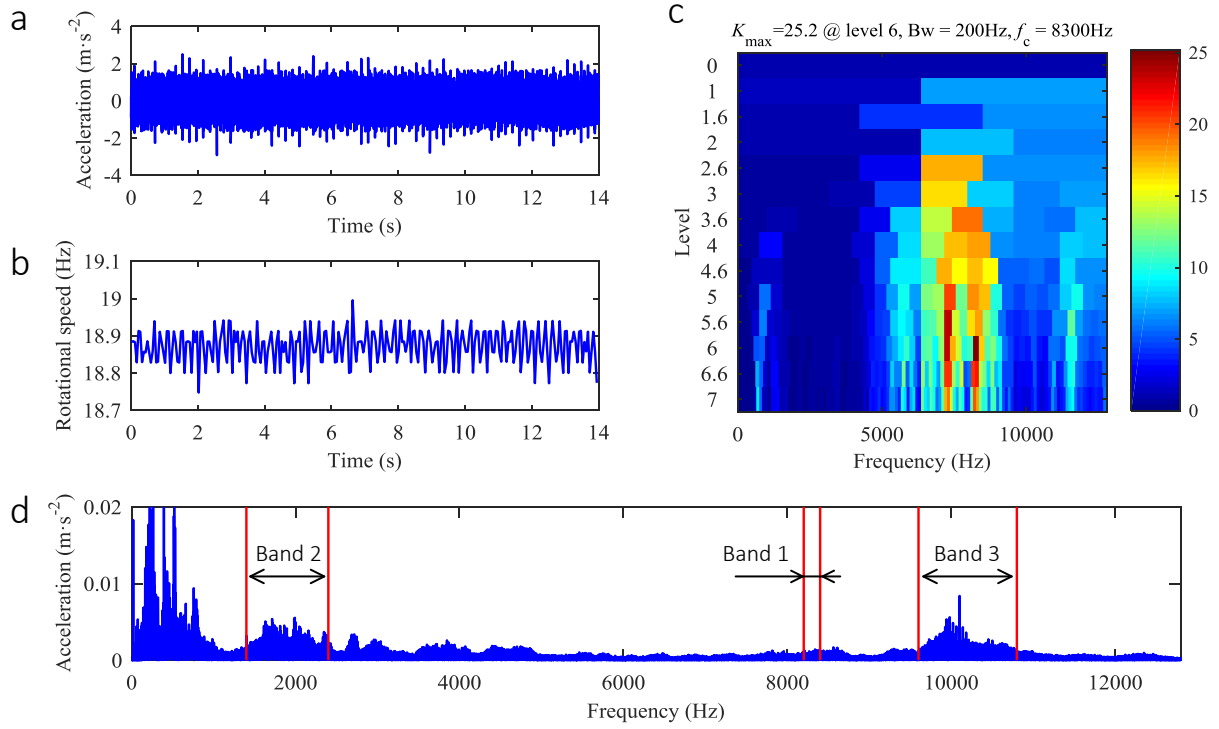


Fig. 20. Test signals and spectra of a healthy bearing: (a) raw vibration signals; (b) rotational speed; (c) Kurtogram results of the vibration signals; and (d) amplitude spectrum.

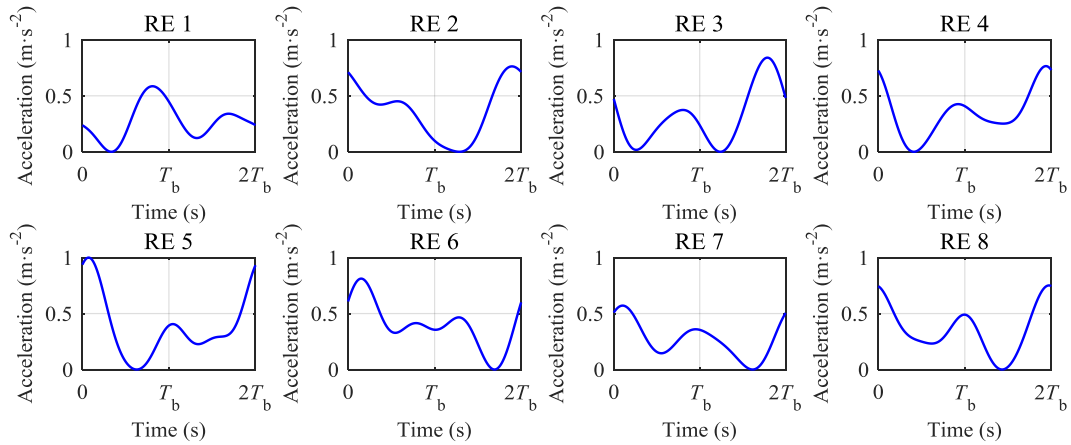


Fig. 21. EEAs of the healthy bearing yielded by performing ELAIT with Band 1.

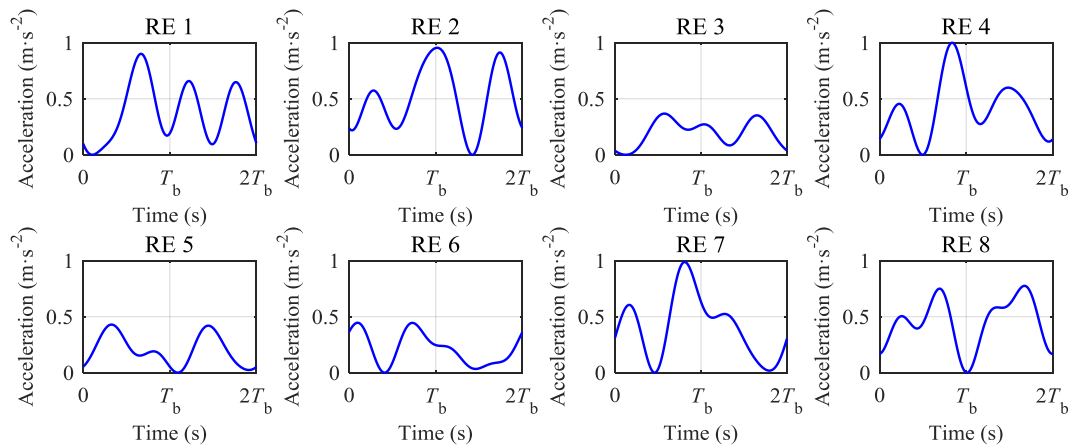


Fig. 22. EEAs of the healthy bearing yielded by performing ELAIT with Band 2.

4. Conclusion

An ELAIT based method is proposed for achieving enhanced diagnosis of rolling element defects in bearings. According to the characteristic that a defected rolling element causes two LAITs when passing the load zone centre, this method separates LAITs corresponding to the LAIT spins during which a particular rolling element passes the load zone centre. The envelope waveform of the two LAITs in a LAIT spin is enhanced with band-pass filtering, demodulation, low-pass filtering, and ensemble averaging. In this way, a clean EEA is yielded for each rolling element. The EEAs of different rolling elements show robustness when used for indications of rolling element defects. This method can be very effective and convenient for real-time and online applications because it has a number of key merits: (1) it is sensitive to the fault transients due to a series of noise reduction techniques including signal separation, band-pass filtering, demodulation, low-pass filtering, and ensemble averaging are used; (2) It is robust to speed oscillation and slippage since the information used for diagnosis is the envelope waveform of LAITs, and the envelope waveform of LAITs is not sensitive to phase error; (3) Tachometer or photoelectric encoder is not needed; (4) Since signal processing techniques in the schematic diagram are applied to the short LAITs separated from a long data stream, the method has low computation cost.

Acknowledgment

This work was supported by the National Natural Science Foundation of China [grant numbers 51575518, 51475463].

References

- [1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, D. Rémond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, *Mechanical Systems and Signal Processing* 24 (7) (2010) 1998-2011.
- [2] R. Randall, J. Antoni, Rolling element bearing diagnostics – A tutorial, *Mechanical Systems and Signal Processing* 25 (2) (2011) 485–520.
- [3] P. D. McFadden, J. D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, *Journal of Sound and Vibration* 96 (1) (1984) 69-82.
- [4] R. Randall, Modern envelope analysis for bearing diagnostics, In: 28th International Congress of Condition Monitoring and Diagnostic Engineering Management and the 10th Regional Non-Destructive & Structural Testing Congress, Buenos Aires, Argentina, 2015.
- [5] N. Tandon, A. Choudhury, An analytical model for the prediction of the vibration response of rolling element bearings duo to a localized defect, *Journal of Sound and Vibration* 205 (3) (1997) 275-292.
- [6] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, *Mechanical Systems and Signal Processing* 19 (2) (2005) 259–270.
- [7] J. Cheng, D. Yu, Y. Yang, A fault diagnosis approach for roller bearings based on EMD method and AR model, *Mechanical Systems and Signal Processing* 20 (2) (2006) 350–362.
- [8] J. Antoni, R. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, *Mechanical Systems and Signal Processing* 20 (2) (2006) 308–331.
- [9] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, *Mechanical Systems & Signal Processing*, 21 (2007) 108-124.
- [10] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, *Mechanical Systems and Signal Processing* 25 (5) (2011) 1738–1749.
- [11] J. R. Stack, R. G. Hartley, T. G. Habetler, An amplitude modulation detector for fault diagnosis in rolling element bearings, *IEEE Transactions on Industrial Electronics* 51 (5) (2004) 1097–1102.
- [12] F. Gu, Y. Shao, N. Hu, A. Naid, A.D. Ball, Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment, *Mechanical Systems and Signal Processing*, 25 (1) (2011) 360–372.
- [13] S. Braun, The extraction of periodic waveforms by time domain averaging. *Acta Acustica united with Acustica* 32 (2) (1975) 69-77(9).
- [14] B. V. Hecke, Y. Qu, D. He, E. Bechhoefer, A new spectral average-based bearing fault diagnostic approach, *Journal of Failure Analysis and Prevention*, 14 (3) (2014) 354–362.
- [15] P. D. McFadden, M. Toozhy, Application of synchronous averaging to vibration monitoring of rolling element bearing, *Mechanical systems and signals processing*, 14 (6) (2000) 891-906.

- [16] D. Siegel, H. Al-Atat, V. Shauche, L. Liao, J. Snyder, J. Lee, Novel method for rolling element bearing health assessment—A tachometer-less synchronously averaged envelope feature extraction technique, *Mechanical Systems and Signal Processing* 29 (1) (2012) 362–376.
- [17] P. D. McFadden, A technique for calculating the time domain averages of the vibration of the individual planet gears and sun gear in an epicyclic gearbox, *Journal of Sound and Vibration* 144 (1) (1991) 163-172.
- [18] P. D. Samuel, D. J. Pines, Vibration separation and diagnostics of planetary gears, in: *Proceedings of the American Helicopter Society 56th Annual Forum*, Virginia Beach, Va, 2000.
- [19] P. D. Samuel, J. K. Conroy, D.J. Pines, *Planetary transmission diagnostics*, NASA/CR 2004.
- [20] R. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelop analysis in the diagnostics of bearing faults and other cyclostationary machine signals, *Mechanical Systems and Signal Processing* 15 (5) (2001) 945–962.
- [21] J. Antoni, R. Randall, Differential diagnosis of gear and bearing faults, *ASME Journal of Vibration and Acoustics* 124 (2002) 165–171.