



Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi



Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction



Huan Huang*, Natalie Baddour, Ming Liang

Department of Mechanical Engineering, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

ARTICLE INFO

Article history:

Received 24 April 2017

Received in revised form 31 October 2017

Accepted 3 November 2017

Available online 13 November 2017

Keywords:

Bearing fault diagnosis

Unknown time-varying rotational speed

Multiple time-frequency curve extraction

Instantaneous fault characteristic frequency

Fault characteristic coefficient

ABSTRACT

Under normal operating conditions, bearings often run under time-varying rotational speed conditions. Under such circumstances, the bearing vibrational signal is non-stationary, which renders ineffective the techniques used for bearing fault diagnosis under constant running conditions. One of the conventional methods of bearing fault diagnosis under time-varying speed conditions is resampling the non-stationary signal to a stationary signal via order tracking with the measured variable speed. With the resampled signal, the methods available for constant condition cases are thus applicable. However, the accuracy of the order tracking is often inadequate and the time-varying speed is sometimes not measurable. Thus, resampling-free methods are of interest for bearing fault diagnosis under time-varying rotational speed for use without tachometers. With the development of time-frequency analysis, the time-varying fault character manifests as curves in the time-frequency domain. By extracting the Instantaneous Fault Characteristic Frequency (IFCF) from the Time-Frequency Representation (TFR) and converting the IFCF, its harmonics, and the Instantaneous Shaft Rotational Frequency (ISRF) into straight lines, the bearing fault can be detected and diagnosed without resampling. However, so far, the extraction of the IFCF for bearing fault diagnosis is mostly based on the assumption that at each moment the IFCF has the highest amplitude in the TFR, which is not always true. Hence, a more reliable T-F curve extraction approach should be investigated. Moreover, if the T-F curves including the IFCF, its harmonic, and the ISRF can be all extracted from the TFR directly, no extra processing is needed for fault diagnosis. Therefore, this paper proposes an algorithm for multiple T-F curve extraction from the TFR based on a fast path optimization which is more reliable for T-F curve extraction. Then, a new procedure for bearing fault diagnosis under unknown time-varying speed conditions is developed based on the proposed algorithm and a new fault diagnosis strategy. The average curve-to-curve ratios are utilized to describe the relationship of the extracted curves and fault diagnosis can then be achieved by comparing the ratios to the fault characteristic coefficients. The effectiveness of the proposed method is validated by simulated and experimental signals.

© 2017 Elsevier Ltd. All rights reserved.

* Corresponding author.

E-mail address: hhuan061@uottawa.ca (H. Huang).

1. Introduction

Vibration signal analysis is one of the most widely used techniques for bearing fault diagnosis [1]. Once a local bearing defect is developed, it induces an impulse excitation to the vibration system and generates impulsive responses in the vibration signal which amplitude-modulate the carrier frequencies, i.e., the high resonance frequencies [2]. In addition, the impulsive responses appear at a certain frequency called the Fault Characteristic Frequency (FCF). For each type of fault, its FCF is given as the product of the shaft rotational frequency and a specific coefficient related to the dimensions of the bearing [2]. Therefore, by observing the FCF and its harmonics in the frequency spectrum of the demodulated vibration signal, bearing faults can be detected and diagnosed. However, in real applications, bearings often operate under time-varying rotational speed conditions. In such cases, the signal is non-stationary and the FCF is no longer constant. Thus, bearing faults cannot be diagnosed in the frequency spectrum directly.

One of the conventional methods for use with vibration signals for time-varying speed cases is signal resampling via order tracking. The essence of order tracking is the resampling of the vibration signal at a constant angle interval rather than constant time interval [3]. After resampling, the time-varying FCF present in the signal sampled at constant time intervals is converted into a constant fault characteristic order (multiple of the reference shaft speed) in the resampled signal [4]. Subsequently, bearing fault diagnosis can be conducted in the order domain. In addition, tachometer-free order-tracking methods have been proposed for bearing fault diagnosis under time-varying speed conditions [5,6], since the shaft rotational speed is not always measurable due to the lack of measuring devices and/or the infeasibility of installation of such devices. However, the accuracy of order tracking is impacted by factors such as the precision of the obtained rotational speed and the method of interpolation [3]. Moreover, the order tracking process is affected by the transmission path phase [7]. Thus, resampling free methods are of particular interests for bearing fault diagnosis under time-varying speed conditions.

Time-Frequency analysis enables resampling-free methods for machinery fault diagnosis under time-varying speed conditions, since the time-varying fault characters appear as “ridges” in the well processed Time-Frequency Representation (TFR) [8]. Various time-frequency analysis methods have been developed to obtain the TFR, among which the Short-Time Fourier Transform (STFT) and wavelet transform are the most popular [8]. Furthermore, to improve the readability of the TFR, synchrosqueezed wavelet transforms have been developed by Daubechies et al. [9]. However, the synchrosqueezing method could be handicapped by time-dimension diffusions of the wavelet coefficients. To deal with diffusions in both time and frequency dimensions, a generalized synchrosqueezing transform approach was proposed by Li et al. [10].

Methods based on Generalized Demodulation (GD) have been investigated to utilize the TFR for gearbox fault diagnosis under time-varying speed conditions by Feng et al. [11–13]. Generalized demodulation can be utilized to convert the arbitrary instantaneous frequency trajectories into constant frequency lines in the time-frequency plane. By iteratively applying generalized demodulation, multiple instantaneous frequency trajectories can be converted into constant frequency lines. A similar strategy has been investigated and applied to bearing fault diagnosis under time-varying speed conditions by Shi et al. [14]. However, to apply generalized demodulation, the instantaneous frequency of the target signal component, i.e. the Instantaneous Fault Characteristic Frequency (IFCF), has to be estimated. Theoretically, the IFCF and its harmonics manifest themselves as ridges in the TFR and ideally the IFCF shows a more significant ridge than its harmonics. Thus, in Ref. [14], the IFCF is estimated as the frequency with the maximum amplitude at each moment in the TFR obtained via the STFT. However, the peak of the FCF is not always the highest peak in the frequency spectrum. Thus, a multiple amplitude superposition algorithm was proposed to extract the IFCF by Wang et al. [5]. However, extracting the Time-Frequency (T-F) curve by relying only on the amplitude of the peaks may result in the extracted curve not being consistent with the curve itself, for example the extracted curves contains unexpected frequency jumps. The results can be worse if the signal is noisy. Therefore, more reliable methods for T-F curve extraction from the TFR need to be investigated. Moreover, if multiple T-F curves can be directly extracted from the TFR, then it is not necessary to apply generalized demodulation to convert the instantaneous frequency trajectories into constant frequency lines. By doing so, the process of bearing fault diagnosis can be simplified and the computational efficiency can be improved accordingly.

T-F curve extraction (also called Instantaneous Frequency (IF) extraction or ridge extraction) from the TFR has been investigated by many researchers [15–18]. The TFR can be regarded as an image, thus, image processing technique can be applied to locate the ridges in the TFR. An algorithm composed of steps of image binary, dilatation, and skeleton operation was proposed by Borda et al. to estimate the IF [15]. The IF is estimated as the result of the skeleton obtained. Another ridge extraction method based on the image processing technique of active contours was proposed by Terrin et al. for the analysis of uterine electromyogram [17]. However, the connectivity of the extracted IF or ridge is not ensured via these image processing techniques since the result is a skeleton or contour composed of pixels without direction. A ridge extraction method by use of a dynamic programming algorithm was proposed by Liebling et al. for spectral interferometry imaging [18]. The extracted ridge is a path composed of points with directions via this method. Following a similar idea, a T-F curve extraction approach, called fast path optimization, was proposed by Iatsenko et al. [16]. The T-F curve is extracted by solving a chosen optimal function composed of three terms, including one related to the amplitude of the peaks in the TFR and the other two related to the frequency and the changes in the frequency of the peaks, respectively. The latter two terms can effectively prevent the unexpected frequency jumps which lead to the extracted T-F curve more accurately represents the ridge in the TFR. Therefore, with this approach, the T-F curve is extracted from the TFR more reliably. However, fast path optimization has not been used for T-F curve extraction from the TFR of bearing vibration signals. Moreover, bearing fault diagnosis requires multiple T-F

curves to be extracted, including those corresponding to the IFCF and its harmonics, and the T-F curve associated with the Instantaneous Shaft Rotational Frequency (ISRF).

In this paper, an algorithm for multiple time-frequency curve extraction for the purpose of bearing fault diagnostics is proposed based on the fast path optimization approach. The rest of the paper is structured as follows. In section 2, the fast path optimization approach is introduced and a multiple T-F curve extraction algorithm is proposed based on the fast path optimization. In section 3, a procedure of bearing fault diagnosis under time-varying speed conditions is developed based on the multiple T-F curve extraction and a fault diagnosis strategy is designed accordingly. Then, in section 4, simulations are conducted to illustrate the effectiveness of the proposed method. Signals of linearly varying and non-linearly varying rotational frequencies are used. Additionally, the signal of linearly varying rotating speed is categorized as IFCF dominant or not dominant in the TFR to cover more types of situations. Following the simulations, in section 5, experimental data are used to further test the effectiveness of the proposed method. Vibration signals collected from a bearing with an outer race defect, a bearing with an inner race defect, and a healthy bearing are utilized for the validation. Finally, conclusions are drawn based on the results of simulations and the experimental validation.

2. Multiple time-frequency curve extraction

Extracting the time-frequency curves from the TFR can be accomplished by extracting the ridges in the TFR and finding the corresponding time-frequency curves where the ridges locate. One of the simplest approaches for ridge extraction from the TFR is extracting the ridge point-by-point which is called a “one-step” approach [16]. The procedure of the “one-step” is locating an initial ridge point first and then searching for the ridge points forwards and backwards following specific optimal functions. This approach is simple and straightforward. However, it is not reliable since one wrongly picked point may totally change the result. Thus, a more reliable approach is proposed based on fast path optimization for ridge extraction [16]. Additionally, for multiple ridge curve extraction, a common way is to find the curve associated with the dominant component, which can then be reconstructed and subtracted from the signal. Subsequently, the other curves can then be extracted by repeating the procedure. In this paper, an algorithm for multiple time-frequency curve extraction is proposed based on the fast path optimization approach without signal reconstruction.

2.1. Time-frequency curve extraction via fast path optimization

Time-frequency curve extraction from the TFR can be performed by finding the ridges in the TFR and taking the corresponding frequencies along the time span as the time-frequency curve. Thus, the ridge extraction is critical for time-frequency extraction. In this paper, the STFT is applied to obtain the TFR, the TFR of signal $x(t)$ is

$$X(\tau, f) = \int_{-\infty}^{+\infty} x(t)w(t - \tau)e^{-j2\pi ft} dt \quad (1)$$

where $w()$ is the window function, τ is the time variable, and f is the frequency variable. For simplicity, τ and f are regarded as continuous variables in Eq. (1). However, in practice, both τ and f are discretized. In the following, we assume that the length of the TFR in time span is N and the length of the TFR in frequency span is N_f . Generally, ridge extraction is conducted by choosing a start point and locating the connected ridges forwards or backwards point-by-point along the time span following a certain optimization function. The main drawback of this approach as mentioned earlier is that a wrongly selected point may completely change the extracted ridge. Thus, a more accurate approach to extract the curve is via optimizing the whole profile of the curve, which is called path optimization. However, the computational cost will be significantly increased if the optimization is performed over all possible peaks directly. To take advantage of the path optimization approach and to reduce the high computational cost, a fast path optimization is proposed in Ref. [16].

The fast path optimization is dependent on only a finite number of the past points, rather than all peaks [16]. To demonstrate the approach, the number of peaks at time τ_n ($n = 1, 2, \dots, N$) in the TFR is denoted as $N_p(\tau_n)$, the amplitude of m th peak at τ_n is denoted as $Q_m(\tau_n)$, and the corresponding frequency of the m th peak at τ_n is denoted as $\nu_m(\tau_n)$. All peaks at τ_n can be determined via

$$\nu_m(\tau_n) = f \quad s.t. \begin{cases} \frac{d[X(\tau_n, f)]}{df} = 0 \\ \frac{d^2[X(\tau_n, f)]}{df^2} < 0 \end{cases}, m = 1, 2, \dots, N_p \quad (2)$$

All the peaks together generate a peak map, Fig. 1(a). The circles denote the peaks in the TFR. To extract a time-frequency curve from the TFR, one needs to determine which peak should be extracted as the ridge at each moment. Thus, for a TFR that has a time span $[\tau_1, \dots, \tau_N]$, the path optimization can be described as [16]

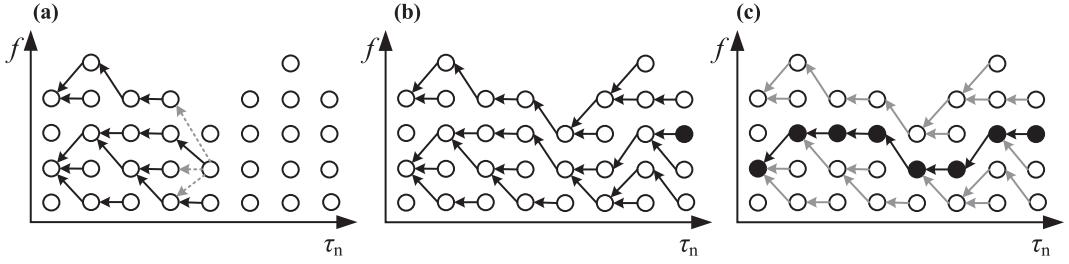


Fig. 1. Scheme of the T-F curve extraction (cited from Ref. [18]). (a) Step 1; (b) step 2; (c) step 3.

$$\{m_c(\tau_1), \dots, m_c(\tau_N)\} = \underset{\{m_1, \dots, m_N\}}{\operatorname{argmax}} \sum_{n=1}^N F[\tau_n, Q_{m_n}(\tau_n), v_{m_n}(\tau_n), \{v_{m_1}(\tau_1), \dots, v_{m_N}(\tau_N)\}] \quad (3)$$

where $m_c(\tau_n)$ determines the peak to be extracted as the ridge at τ_n , $F[\cdot]$ is the chosen support function for the optimization and $\{m_1, \dots, m_N\}$ refers to a sequence of the peak numbers along the time span. The path optimization is performed over all peaks at the whole time span. The computational complexity is on the order of $O(N_p(\tau_1) \cdot N_p(\tau_2) \cdots N_p(\tau_N))$ which can be up to $O(M_p^N)$ with $M_p = \text{Max}(N_p(\tau_n))$. Additionally, all peaks in the TFR have to be determined prior to the extraction. Therefore, the computational complexity is on the order of $O(N_f N + M_p^N)$ in total. It can be seen that the computational cost is extremely high if the optimization is directly applied to all possible peaks. Thus, a fast path optimization is proposed to be only dependent on a finite number of the previous points with the support function [16]

$$F[] = \begin{cases} \log Q_m(\tau_n) & n = 1 \\ \log Q_m(\tau_n) + w_2(v_m(\tau_n), m[f_d], \text{IQR}[f_d], \lambda_2) & n = 2 \\ \log Q_m(\tau_n) + w_2(v_m(\tau_n), m[f_d], \text{IQR}[f_d], \lambda_2) \\ + w_1(v_m(\tau_n) - f_d(\tau_{n-1}), m[\Delta f_d], \text{IQR}[\Delta f_d], \lambda_1) & n \geq 3 \end{cases} \quad (4)$$

where

$$w_1(v_m(\tau_n) - f_d(\tau_{n-1}), m[\Delta f_d], \text{IQR}[\Delta f_d], \lambda_1) = -\lambda_1 \left| \frac{v_{m_n}(\tau_n) - f_d(\tau_{n-1}) - m[\Delta f_d]}{\text{IQR}[\Delta f_d]} \right| \quad (5)$$

$$w_1(v_m(\tau_n) - f_d(\tau_{n-1}), m[\Delta f_d], \text{IQR}[\Delta f_d], \lambda_1) = -\lambda_1 \left| \frac{v_{m_n}(\tau_n) - f_d(\tau_{n-1}) - m[\Delta f_d]}{\text{IQR}[\Delta f_d]} \right| \quad (6)$$

$$m[] \equiv \text{perc}_{0.5}[], \quad \text{IQR}[] \equiv \text{perc}_{0.75}[] - \text{perc}_{0.25}[] \quad (7)$$

where $f_d(\tau_{n-1})$ is the frequency of the candidate ridge point at τ_{n-1} , f_d are the frequencies of a series of candidate ridge points in history $[\tau_1, \dots, \tau_{n-1}]$, Δf_d is the derivative of f_d , $m[f(t)]$ is the median of $f(t)$, IQR is the interquartile range, defined in Eq. (7), where $\text{perc}_p[f(t)]$ denotes the p th quantile of $f(t)$, and λ_1 and λ_2 are penalty factors which can be taken as 1 according to the analysis in Ref. [16]. Weight functions $w_1()$ and $w_2()$ are employed to suppress the atypical variations of the ridge frequency's value and derivative, respectively. With the chosen function given in Eq. (4), the frequency jumps can be prevented in the extracted T-F curve.

The solution of the fast path optimization is given as

$$\begin{aligned} \text{for } n = 1, \dots, N, \quad m = 1, \dots, N_p(\tau_n) \text{ and } k = 1, \dots, N_p(\tau_{n-1}) \\ q(m, \tau_n) = \underset{k}{\operatorname{argmax}} \{F[Q_m(\tau_n), v_m(\tau_n), v_k(\tau_{n-1})] + U(k, \tau_{n-1})\}, \quad n > 1 \\ U(m, \tau_n) = \begin{cases} F[Q_m(\tau_n), v_m(\tau_n), v_m(\tau_n)] & n = 1 \\ F[Q_m(\tau_n), v_m(\tau_n), v_{q(m, \tau_n)}(\tau_{n-1})] + U(q(m, \tau_n), \tau_{n-1}) & n > 1 \end{cases} \end{aligned} \quad (8)$$

where $q(m, \tau_n)$ indicates which peaks at the previous moment τ_{n-1} should be linked to the peaks at the current moment τ_n , and $U(m, \tau_n)$ is an intermediate variable which contributes to the optimization. The computational complexity is reduced to the order of $O(N_f N + M_p^2 N \log N)$, which is significantly lower than $O(N_f N + M_p^N)$. To further illustrate the T-F extraction approach, a scheme is shown in Fig. 1. The spots represent the peaks in the TFR determined via Eq. (2). The extraction is composed of three steps, shown as (a), (b) and (c) in Fig. 1, respectively. First, by applying the solution given in Eq. (8), $q(m, \tau_n)$ is calculated for all peaks from τ_2 to τ_N ; then, the peak with the maximum value of $U(m, \tau_n)$ at τ_N is selected as the last point of

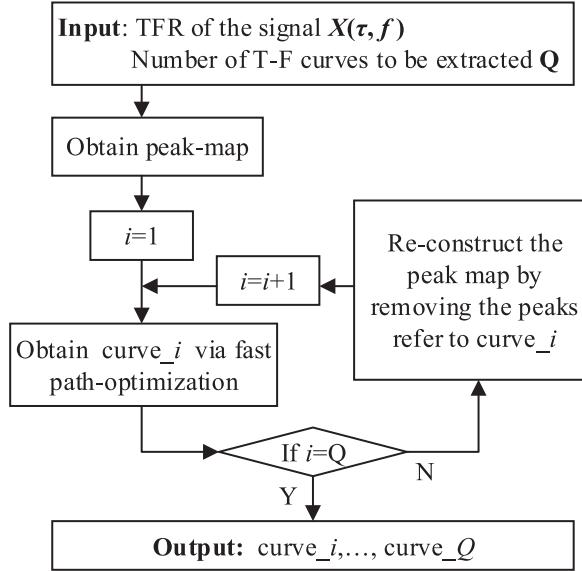


Fig. 2. Flowchart of multiple T-F curve extraction.

the T-F curve, shown as the solid spot in Fig. 1(b); finally, from the selected point, the whole curve can be determined following the track. The extracted curve is noted as $f_p(\tau_n)$.

2.2. Multiple time-frequency curve extraction based on fast path optimization

Based on the fast path optimization approach for T-F curve extraction as introduced in the previous section, a multiple T-F curve extraction algorithm is proposed. The flowchart of the algorithm is shown in Fig. 2. Fast path optimization is iteratively applied to the TFR to extract multiple T-F curves. After one T-F curve is extracted, the peaks from the extracted T-F curve are removed from the current peak map to generate a new peak map. Then, with the new peak map, the next T-F curve can be extracted by applying fast path optimization again.

The number of curves to be extracted should be selected as sufficient to show the IFCF and at least one of its harmonics. Obviously, the more curves the better. However, more curves lead to higher computational cost. The computational complexity of the multiple T-F curve extraction algorithm is on the order of $O(QN_f N + QM_p^2 N \log N)$. Ideally, the Q value should be changed for different signals according to the composition of the signal and the intensity of the noise. Therefore, a guideline is that the Q should be high enough to reveal the IFCF and its harmonics, while the computational cost remains acceptable. In this paper, $Q = 5$ is selected for all the tests in simulations and experiments. However, different Q values can be selected for different cases, as long as the selection satisfies the guideline.

3. Proposed procedure for bearing fault diagnosis under unknown time-varying speed conditions

With the multiple T-F curve extraction algorithm, a procedure for bearing fault diagnosis under time-varying rotational speed conditions is proposed in the flowchart shown in Fig. 3. This procedure is resampling-free and tachometer-free.

The raw signal is first amplitude demodulated via the Hilbert transform, after which the demodulated signal, also called the envelope signal, is obtained. Then, the STFT is employed to obtain the TFR. Subsequently, the multiple T-F curve extraction algorithm is applied to the TFR to extract a certain number of T-F curves. A fault diagnosis strategy is proposed with the extracted T-F curves.

The flowchart of the fault diagnosis strategy is shown in Fig. 3. The Fault Characteristic Coefficient (FCC), defined as the constant ratio of the FCF to the Shaft Rotating Frequency (SRF) is utilized in the flowchart. The formulae for the FCF of outer race fault (BPFO) and inner race fault (BPFI) are given as [2]

$$\text{BPFO} = \frac{n_b}{2} \left(1 - \frac{d}{D} \cos \phi \right) f_r = \text{FCC}_O f_r \quad (9)$$

$$\text{BPFI} = \frac{n_b}{2} \left(1 + \frac{d}{D} \cos \phi \right) f_r = \text{FCC}_I f_r \quad (10)$$

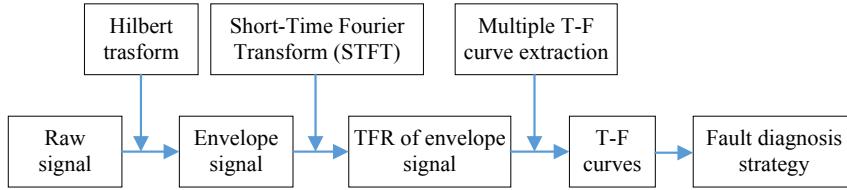


Fig. 3. Procedure for bearing fault diagnosis under unknown time-varying speed conditions.

where n_b is the number of rolling elements, d is the diameter of the rolling element, D is the pitch diameter of the bearing, ϕ is the angle of the load from the radial plane, f_r is the shaft rotational frequency, FCC_O and FCC_I are respectively the outer race and inner race FCCs. It is clear that the ratio of the FCF to the SRF is constant since it is determined by the dimensions of the bearing. Theoretically, for time-varying speed cases, the IFCF should maintain the same ratio with the ISRF. Based on this, if the extracted T-F curves cover the IFCF, its harmonics and the ISRF, then the fault can be detected and diagnosed by matching the average ratios of the curves to the extracted ISRF with FCC, i.e., the ratio of the FCF to f_r . The average ratio between two extracted curves $f_p(\tau_n)_i$ and $f_p(\tau_n)_j$ is defined in this paper as

$$R_a = \frac{1}{N} \sum_{n=1}^N \frac{f_p(\tau_n)_i}{f_p(\tau_n)_j} \quad (11)$$

The bottom of the ratio is taken as the extracted T-F curve which has the lowest frequency since the ISRF is smaller than the IFCF and its harmonic. The ratio then becomes

$$R_a = \frac{1}{N} \sum_{n=1}^N \frac{f_p(\tau_n)_i}{f_p(\tau_n)_l} \quad (12)$$

where $f_p(\tau_n)_l$ denotes the T-F curve with the lowest frequency. It should be pointed out that the calculated average ratio may not completely match the FCC due to the limited resolution of the TFR and the limited precision of the given dimensions of the

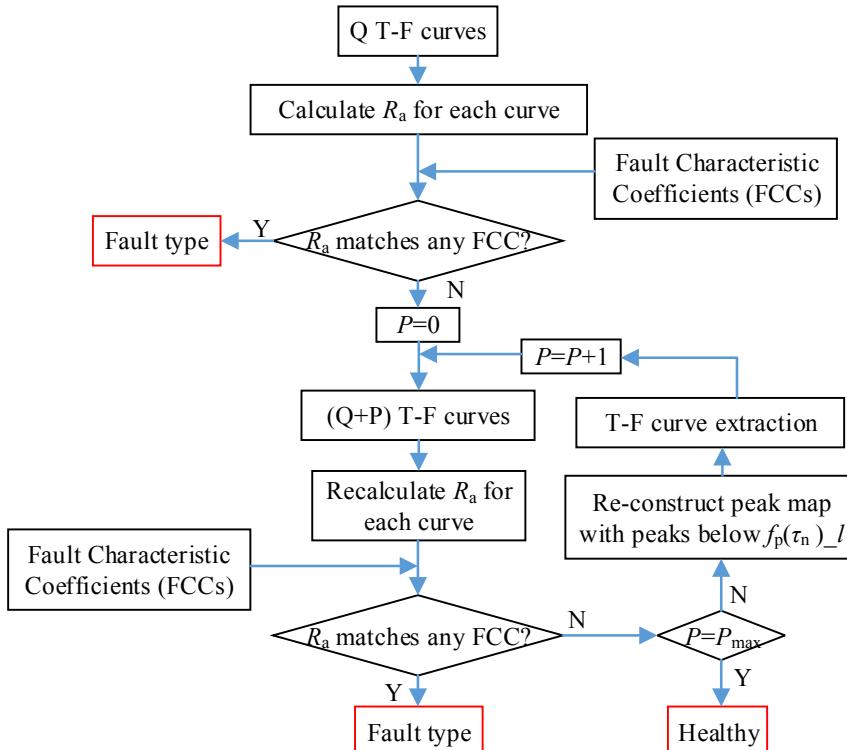


Fig. 4. Fault diagnosis strategy.

bearing. Thus, it is proposed to use a 5% relative error of the calculated average ratio to the FCC for tolerance of the matching, i.e. $|R_a - FCC| \cdot 100\% / FCC \leq 5\%$. If the calculated average ratios of the T-F curves to the curve of the lowest frequency match the FCC and its harmonics (n^*FCC), then the corresponding fault type is determined. Additionally, the sidebands with the frequency $n^*FCF \pm f_r$ may be also observed in the TFR. The average ratios of sidebands to f_r are $n^*FCC \pm 1$ since $FCF = FCC \cdot f_r$.

However, it is possible that the ISRF may not be covered in the extracted Q curves. The ISRF shown in the TFR can be the frequency of the modulating signal of the bearing fault-induced signal [2] or the fundamental frequency of the signal caused by misalignment, eccentric or imbalanced mass [6]. If the bearing fault-induced signal is not amplitude modulated by the modulating signal with the ISRF as the fundamental frequency and the machine operates under a properly installed condition (e.g., with minimal misalignment, eccentric and imbalanced mass), the ISRF shown in the TFR can be relatively weak. Thus, an iterative process is included in the fault diagnosis strategy. The flowchart of the fault diagnosis strategy is shown in Fig. 4. If the average ratios of the extracted Q curves do not match the FCCs, which implies that the ISRF was not included in the Q extracted curves, then extra T-F curves need to be extracted. Each new curve is extracted from the portion below the previously extracted curve of the lowest frequency in the TFR. With the newly extracted T-F curve, the average ratios can be recalculated. If the ratios are still not matched with the FCCs, then the new T-F curve is extracted iteratively until the repeat times reaches a chosen limit P_{max} . Similar to the setting of Q , P_{max} has to be sufficient enough to ensure that the ISRF is extracted if the bearing is faulty. However, it cannot be too large because of the increased computational cost. In this paper, $P_{max} = 3$ was selected in the simulations and experiments. The bearing is considered healthy if no ratio matches one of the FCCs after the complete process.

4. Simulations

Simulations were performed in order to illustrate the application of the proposed procedure for bearing fault diagnosis under unknown time-varying speed conditions and also to test the performance of the procedure.

The vibration signal induced by a bearing defect can be simulated as impulse responses of a 1 degree-of-freedom mass-spring-damper system, expressed as [19]

$$x(t) = \sum_{m=1}^M L_m e^{-\beta(t-mT_p-\sum_{i=1}^m \varepsilon_i)} \sin[\omega_r(t-mT_p-\sum_{i=1}^m \varepsilon_i)] u(t-mT_p-\sum_{i=1}^m \varepsilon_i) \quad (13)$$

where M is the number of the impulse responses, L_m is the amplitude coefficient, T_p is the period of the impulses, the reciprocal of the FCF, β represents the damping characteristic determined by the structure of the system, ω_r is the damped natural frequency of the system, ε_i is the slipping coefficient, which is randomly simulated as $0.01T_p - 0.02T_p$, and $u(t)$ is the unit step function. This model can be considered as an amplitude modulated signal with ω_r being the frequency of the carrier frequency and the FCF being the frequency of the modulating signal. The demodulated signal is called the envelope. One of the most commonly used demodulation methods is the Hilbert transform. For a signal $x(t)$, the Hilbert transform is defined as [20]

$$H[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \frac{1}{t-\tau} d\tau \quad (14)$$

With the envelope, it shows peaks at the FCF and its harmonics in the envelope spectrum. However, for the time-varying speed cases, T_p is also varying. In addition, the signal is always masked by noise. Also, the signal is considered further amplitude modulated with the ISRF as the frequency of the modulating signal if the bearing has an inner race fault [2]. Moreover, the collected vibration signal usually comprises the signal with the shaft rotational frequency as the fundamental frequency caused by misalignment, eccentric mass or imbalance [6]. To account for these four factors, the model can be modified to

$$x(t) = A(t) \sum_{m=1}^M L_m e^{-\beta(t-t_m)} \sin[\omega_r(t-t_m)] u(t-t_m) + s(t) + n(t) \quad (15)$$

where $A(t) = 1 + \alpha \cos(2\pi f_r(t)t)$ is the factor used to reflect the amplitude modulation of the responses with the ISRF ($\alpha > 0$ for inner race fault and $\alpha = 0$ for outer race fault), $f_r(t)$ being the frequency of the modulating signal; α is a constant; $L_m = L_0 + \eta f_r(t_m)$ is the amplitude which is proportional to the rotational speed; L_0 and η are constants, t_m is the occurrence time of the m th impulse, $s(t)$ is the sinusoidal signal with the shaft rotational frequency as the fundamental frequency, and $n(t)$ is random noise, SNR can be used to describe the intense of the noise. The occurrence time of the impulses can be computed by

$$\begin{cases} t_1 = (1 + \mu)[1/f_c(0)] \\ t_m = (1 + \mu)[1/f_c(0) + 1/f_c(t_1) + \dots + 1/f_c(t_{m-1})] \end{cases} m = 2, 3, \dots, M \quad (16)$$

where μ is the slippage ratio varying from 0.01 to 0.02, $f_c(t)$ denotes the IFCF, the time interval between two impulses is $1/f_c(t)$ and $f_c(0)$ cannot be zero. $s(t)$ can be simulated by

$$s(t) = \sum_{i=1}^{N_s} B_i \cos(2\pi f_r(t) \cdot t) \quad (17)$$

where N_s is the number of harmonics of $s(t)$ and B_i is the amplitude of i th harmonic of $s(t)$. With the demodulated signal, the IFCF and its harmonics show curves in the TFR of the envelope signal if $\alpha \neq 0$. The rotational frequency and sidebands $f_c(t) \pm n f_r(t)$, $n = 0, 1, \dots$ also show curves in the TFR.

To test the performance of the proposed method for bearing fault diagnosis under time-varying speed conditions, different types of signals are simulated. First, both linear and non-linear speed variations are considered. For linear variation cases, speed-up is utilized for the simulation since speed-up and speed-down follow the same pattern. For the non-linear cases, exponential decay speed and quadratic speed with speed-up plus speed down are considered. Additionally, to test the fault diagnosis strategy, two conditions featuring respectively dominant ISRF and non-dominant ISRF in the TFR are simulated for the linear speed variation cases with the different levels of clearness of the ISRF in the TFR.

4.1. Simulated signal with linear speed-up ISRF

4.1.1. Case 1: ISRF is dominant in the TFR

A signal with linearly speed-up ISRF is simulated to test the performance of the proposed method, where $f_r(t) = 20 + 3t$. The parameters in Eq. (15) are set as $\alpha = 0.9$, $L_0 = 1$, $\eta = 0.05$, $\beta = 500$, $\omega_r = 2\pi \times 4000 \text{ rad/s}$, $\mu = 0.01$, $N_s = 2$, $B = [0.8, 0.2]$, and

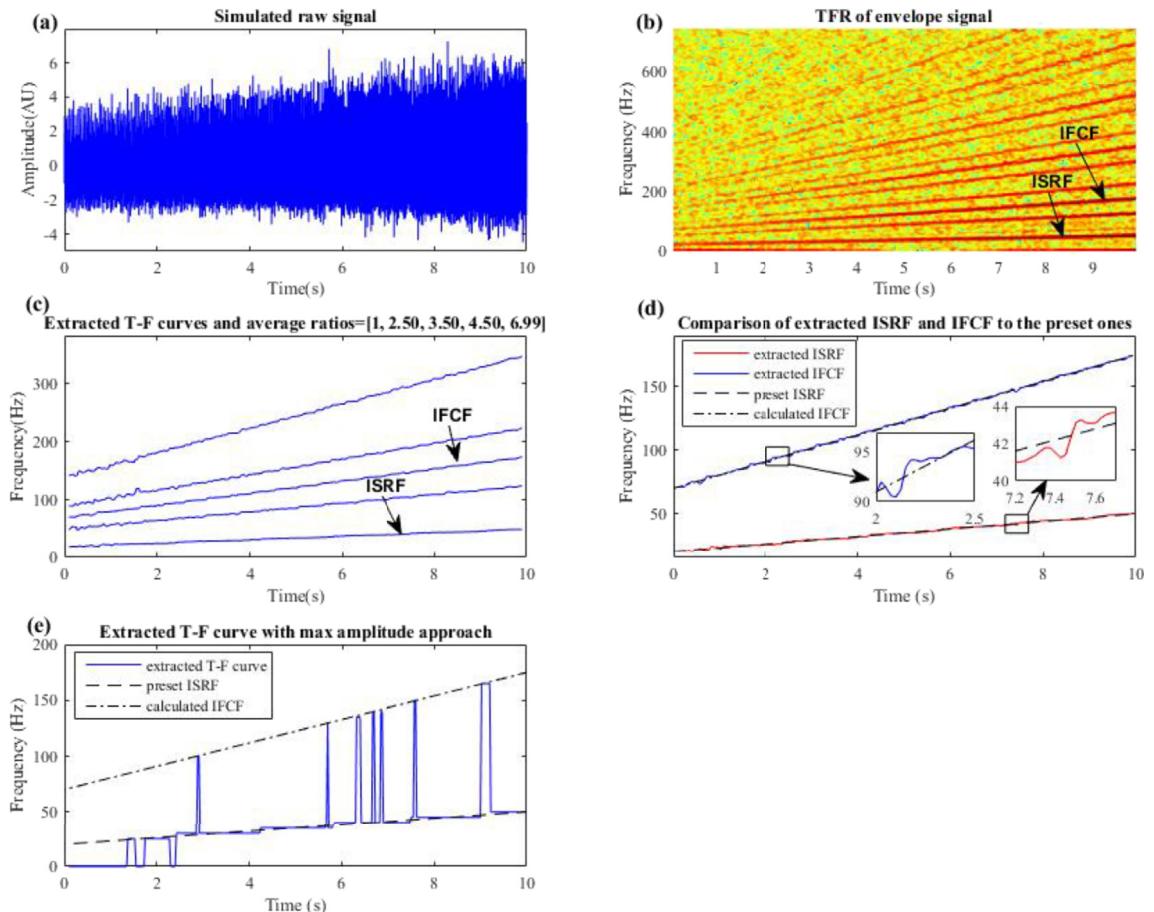


Fig. 5. Results of the proposed method applied to the simulated signal with linear speed-up ISRF and dominant ISRF in the TFR. (a) Simulated raw signal; (b) TFR of envelope signal; (c) extracted T-F curves; (d) comparison of extracted ISRF and IFCF to their preset curves; (e) extracted T-F curve by max amplitude approach.

$\text{SNR} = -5 \text{ dB}$. Additionally, the FCC is set as 3.5, i.e. $f_c(t) = 3.5f_r(t)$. The sampling frequency of the signal is 20 kHz and the signal length is 10s.

The simulated signal is shown in Fig. 5(a). It can be seen that the amplitude of the signal increases with the increase of the ISRF. Also, the spacing of the impulses becomes smaller due to the increasing ISRF. By applying the Hilbert transform and the STFR to the simulated signal, the TFR of the envelope signal is obtained as shown in Fig. 5(b). From the TFR, many ridges can be observed, including the ISRF, IFCF, its harmonics, as well as the sidebands. The ridge of the ISRF is almost as clear as the ridge of the IFCF.

According to the proposed procedure for fault diagnosis, the multiple T-F curve extraction algorithm is applied to the TFR first. The number of curves to be extracted, Q , is set as 5 in this simulation, thus five curves are extracted shown in Fig. 5(c). The corresponding average ratio R_a of each curve is calculated as 1, 2.50, 3.50, 4.50 and 6.99, respectively. Among the average ratios, 3.50 matches the FCC (3.5) completely and 6.99 matches 2^*FCC . Additionally, 2.50 and 4.50 match FCC-1 (sideband) and FCC+1 (sideband), respectively. Therefore, the fault diagnosis is completed and the extracted curve of the lowest frequency is the ISRF and the one of the third lowest is the IFCF.

Comparisons of the extracted ISRF and IFCF to their pre-set curves are shown in Fig. 5(d). It can be seen that both the extracted ISRF and IFCF are well matched with the pre-set ones. The average relative error of the extracted IFCF to the calculated IFCF is 0.51% and the average relative error of the extracted ISRF to the preset ISRF is 1.52%, respectively.

To compare the fast path-optimization approach with the maximum amplitude approach for T-F curve extraction, the extracted T-F curve by extracting the ridge point of the maximum amplitude at each moment is shown in Fig. 5(e). The preset ISRF and calculated IFCF are also shown in the same figure. It can be seen that the extracted T-F curve is quite messy, jumping from the bottom to the IFCF, and then back and forth between the IFCF and ISRF. Obviously, the result is not as reliable as the curves shown in Fig. 5(c) obtained via the proposed multiple T-F curve extraction algorithm.

The performance of the proposed method to the simulated signal blended with different levels of noise is also tested. Due to space limit, only the average relative errors of the IFCF and the average relative errors of the ISRF are reported in this paper as shown in Table 1, where SNR varies from $\text{SNR} = 0 \text{ dB}$ to $\text{SNR} = -10 \text{ dB}$. For each SNR, the average relative error of the IFCF is smaller than the average relative error of the ISRF. The reason is that the IFCF of the simulated signal is more significant than the ISRF in the TFR. In addition, with the increase of the noise level, the average relative errors of the IFCF and the ISRF both increase. The average relative error of the IFCF is increased by 0.41% from 0.44% to 0.85% and the average relative error of ISRF is increased by 2.5% from 1.41% to 3.91% when SNR is reduced from 0 dB to -10 dB. The results when $\text{SNR} = -10 \text{ dB}$ are still acceptable. Also, the increase rates of the average relative errors for both the IFCF and the ISRF become larger with the decrease of SNR.

4.1.2. Case 2: ISRF is weak in the TFR

To test the performance of the proposed procedure in the case that the ISRF is not extracted from the TFR by the multiple T-F curve extraction algorithm, a signal with a weak ISRF is simulated in this section. The rotational frequency is kept the same $at_r(t) = 20 + 3t$, and the parameters are set as $\alpha = 0$, $L_0 = 1$, $\eta = 0.05$, $\beta = 500$, $\omega_r = 2\pi^*4000 \text{ rad/s}$, $\mu = 0.01$, $N_s = 2$, $B = [0.7, 0.2]$, $\text{SNR} = -5 \text{ dB}$, and $\text{FCC} = 3.5$, i.e. $f_c(t) = 3.5f_r(t)$. Compared to the settings for case 1, α is smaller, which leads to less visible ridges of the ISRF in the TFR. The sampling frequency of the signal is 20 kHz and the signal length is 10s.

The simulated signal is shown in Fig. 6(a). The amplitude of the signal is increasing with the increase of the rotational frequency. By applying the STFR to the demodulated signal, the TFR is obtained as shown in Fig. 6(b). The IFCF and its harmonics can be observed from the TFR. Unlike the TFR shown in Fig. 5(b), the ISRF is not very clearly shown in the TFR, especially in the low-frequency portion.

The extracted T-F curves are shown in Fig. 6(c) by applying the proposed procedure to the simulated signal. First, the multiple T-F curve extraction algorithm is used to extract curves. The Q is again set as 5, thus five curves are extracted, shown in blue in Fig. 6(c). The average ratios of the curves to the one of lowest frequency are calculated as 1, 2, 3, 4, and 5, respectively. None of the ratios matches the FCC (3.5). Therefore, according to the fault diagnosis strategy, an additional curve is extracted from the TFR below the ridge of the lowest frequency previously extracted. Here, P_{\max} is set as 3 for the fault diagnosis strategy. The newly extracted curve is shown in red in Fig. 6(c). With the new curve, the average ratios are recalculated as 1, 3.51, 7.01, 10.52, 14.03 and 17.54. Among the recalculated average ratios, 3.51 matches the FCC (3.5) within 5% relative error. Also, average ratios 7.01, 10.52, 14.03 and 17.54 are matched with 2^*FCC , 3^*FCC , 4^*FCC and 5^*FCC , respectively, which implies the 2nd, the 3rd, the 4th and the 5th harmonics of the IFCF. Therefore, the fault diagnosis can be completed.

A comparison diagram of the extracted IFCF and the extracted ISRF to the preset curves is shown in Fig. 6(d). The average errors of the IFCF and the ISRF are calculated as 0.46% and 2.78%, respectively. It can be seen that the extracted IFCF matches the calculated IFCF. The extracted ISRF fits the pre-set ISRF overall, however, does not match well in the low-frequency portion. The reason is that the ridge of the ISRF in the relatively low-frequency portion is less clear in the TFR.

Table 1

Average relative errors of IFCF and ISRF with different level of noise for case 1.

	$\text{SNR} = 0 \text{ dB}$	$\text{SNR} = -3 \text{ dB}$	$\text{SNR} = -5 \text{ dB}$	$\text{SNR} = -8 \text{ dB}$	$\text{SNR} = -10 \text{ dB}$
Average relative error (IFCF)	0.44%	0.49%	0.51%	0.62%	0.85%
Average relative error (ISRF)	1.41%	1.48%	1.52%	2.39%	3.91%

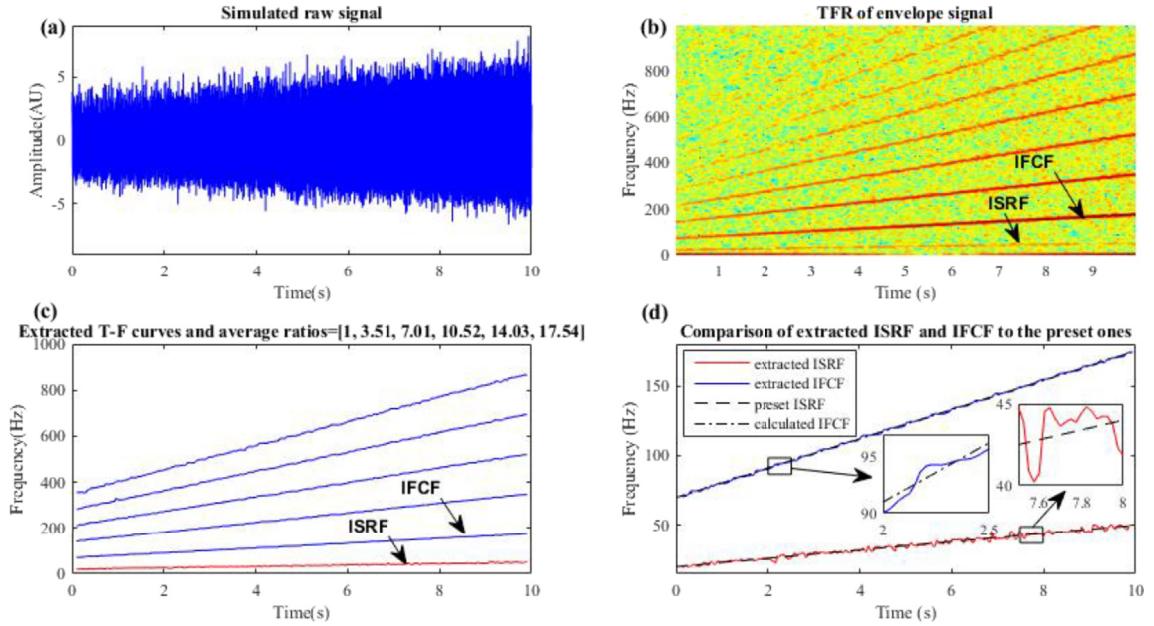


Fig. 6. Results of the proposed method applied to the simulated signal with linear speed-up ISRF and weak ISRF in the TFR. (a) Simulated raw signal; (b) TFR of envelope signal; (c) extracted T-F curves; (d) comparison of extracted ISRF and IFCF to their preset curves.

Table 2

Average relative errors of ISRF and IFCF with different level of noise for case 2.

	SNR = 0 dB	SNR = -3 dB	SNR = -5 dB	SNR = -8 dB	SNR = -10 dB
Average relative error of IFCF	0.39%	0.44%	0.46%	0.54%	0.65%
Average relative error of ISRF	1.71%	1.99%	2.78%	6.98%	9.85%

The proposed method is also applied to the signal with different levels of the noise. The average relative errors of the IFCF and the average relative errors of the ISRF are given in Table 2. Similar to case 1, for each SNR, the average relative errors of the ISRF are higher than those of the IFCF. The reason is that the ridge of the IFCF is significantly clearer than that of the ISRF in the TFR. The average relative errors of both the IFCF and the ISRF increase with the increase of the noise level. The average relative error of the IFCF increases by 0.26% from 0.39% to 0.65% and the average relative error of ISRF increases by 8.41% from 1.71% to 9.85% when the SNR is reduced from 0 dB to -10 dB. The results when SNR = -10 dB are still acceptable. Additionally, compared to the results given in Table 1, the average relative errors of the IFCF for case 2 is smaller than those for case 1. However, the average relative errors of the ISRF for case 2 are higher than those for case 1. The reason can be found from the comparison of the clearness of the ISRF and the IFCF in Fig. 5(b) to the clearness of the ISRF and the IFCF in Fig. 6(b), respectively. It can be seen that the ridge of the IFCF in Fig. 6(b) is clearer than the one in Fig. 5(b), however, the ridge of the ISRF in Fig. 6(b) is less clear than the one in Fig. 5(b).

4.2. Simulated signal with non-linear ISRF

4.2.1. Case 3: Signal with exponentially decaying ISRF

To test the performance of the proposed method using a bearing signal with non-linearly varying rotational speed, the signal with an exponentially decaying ISRF is simulated in this section. The rotational frequency is set as $f_r(t) = 80e^{-0.35t}$ and the other parameters are $\alpha = 0.9$, $L_0 = 1$, $\eta = 0.05$, $\beta = 500$, $\omega_r = 2\pi \cdot 4000 \text{ rad/s}$, $\mu = 0.01$, $N_s = 2$, $B = [0.8, 0.2]$, SNR = -5 dB, and FCC = 3.5, i.e. $f_c(t) = 3.5f_r(t)$. The sampling frequency of the signal is 20 kHz and the signal length is 5s.

The simulated signal with the given settings is shown in Fig. 7(a). It can be seen that with the decreasing speed, the density of the signal reduces and the amplitude of the signal also reduces. The TFR of the envelope of the simulated signal is shown in Fig. 7(b), obtained via the STFT. Similar to case 1, the IFCF and its harmonics, the ISRF and the sidebands show ridges in the TFR. Additionally, the ISRF is almost as clear as the IFCF.

Applying the proposed multiple T-F curve extraction algorithm to the TFR with $Q = 5$, five T-F curves are extracted, as shown in Fig. 7(c). The average ratios are calculated as 1, 2.51, 3.51, 4.51 and 7.01, respectively. Obviously, among those average

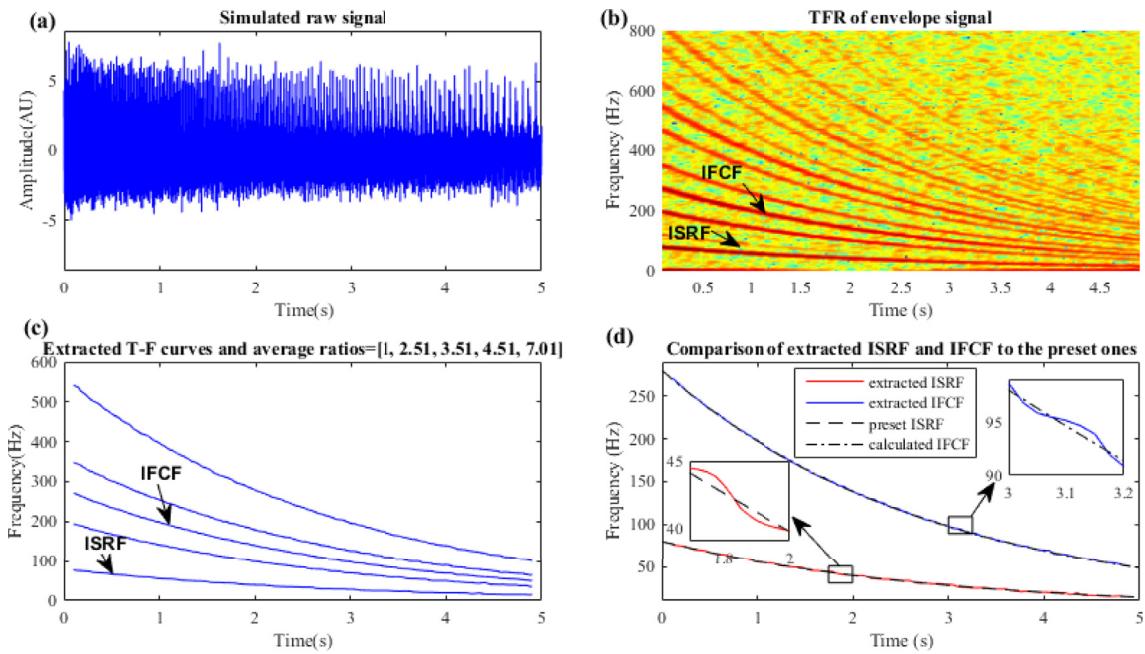


Fig. 7. Results of the proposed method applied to the simulated signal with exponentially decaying ISRF. (a) Simulated raw signal; (b) TFR of envelope signal; (c) extracted T-F curves; (d) comparison of extracted ISRF and IFCF to their preset curves.

ratios, 3.51 matches the FCC of (3.5) within a 5% relative error and 7.01 matches 2*FCC. In addition, 2.51 and 4.51 match the coefficient for the sidebands (FCC-1 and FCC+1). Therefore, according to the proposed procedure, the fault diagnosis is completed.

Compared to the curve at which the average ratio is 1 to the pre-set ISRF and the curve at which the average ratio is 3.51 to the calculated IFCF, they are closely overlapped, seen in Fig. 7(d). The average relative errors of the IFCF and the ISRF are 0.46% and 1.69%, respectively.

The proposed procedure is also applied to the signal with different levels of the noise. The average relative errors of the IFCF and the average relative errors of the ISRF are given in Table 3, where the SNR varies from 0 dB to –10 dB. The average relative errors of the ISRF are higher than those of the IFCF for each SNR except SNR = –10 dB. Additionally, the average relative errors of both the ISRF and the IFCF increase with the decrease of the SNR. When SNR decreases from 0 dB to –8 dB, the average relative error of the IFCF increases by 0.12% from 0.41% to 0.53% and the average relative error of the ISRF increases slightly (0.26%) from 1.49% to 1.75%. However, when the noise further increases to SNR = –10 dB, average relative errors of the IFCF and the ISRF become 15.38% and 7.82%, which is beyond the acceptable limits. This implies that the proposed method for an exponentially decaying ISRF case has less tolerance for noise compared to the cases with linearly varying ISRF.

4.2.2. Case 4: Signal with quadratic ISRF

To further test the performance of the proposed method in processing the bearing signal with non-linearly varying rotational speed, a signal with quadratic ISRF is simulated in this section. The rotational frequency is set as $f_r(t) = 9.6t^2 + 48t + 20$ and the other parameters are $\alpha = 0.9$, $L_0 = 1$, $\eta = 0.1$, $\beta = 500$, $\omega_r = 2\pi \times 4000$ rad/s, $\mu = 0.01$, $N_s = 2$, $B = [0.8, 0.2]$, SNR = –5 dB, and FCC = 3.5, i.e. $f_c(t) = 3.5f_r(t)$. The sampling frequency of the signal is 20 kHz and the signal length is 5 s.

The simulated signal with the given settings is shown in Fig. 8(a). With quadratically varying rotational speed, the amplitude of the signal increases in the left half time span and then reduces in the right half time span. The TFR of the envelope of the simulated signal is shown in Fig. 8(b), obtained via the STFT. It can be seen that the IFCF and its harmonics, ISRF and sidebands show ridges in the TFR. Similarly, the ISRF is almost as clear as the IFCF. However, for high order harmonics, the ridges at the low-frequency portion are unclear.

According to the proposed procedure, multiple T-F curve extraction is applied to the TFR with results shown in Fig. 8(c). It can be observed that extracted curves of the highest frequency and the second highest frequency exhibit some meandering pattern at the low-frequency portion, i.e. at the beginning and at the end. This is due to the unclear ridges at the corresponding portion in the TFR. The average ratios of the five extracted T-F curves are calculated as 1, 2.50, 3.49, 4.49, and 6.99, respectively. Among these average ratios, 3.49 matches the FCC (3.5) within 5% relative error and 6.99 matches 2*FCC. In addition, 2.50 and 4.49 match FCC-1 and FCC+1 (i.e. sidebands), respectively. Then, the extracted curve of the lowest frequency is considered as the ISRF and the curve with average ratio 3.49 is identified as the IFCF. With these results, the fault can be diagnosed even though two of the extracted curves are less ideal.

Table 3

Average relative errors of IFCF and ISRF with different level of noise for case 3.

	SNR = 0 dB	SNR = -3 dB	SNR = -5 dB	SNR = -8 dB	SNR = -10 dB
Average relative error of IFCF	0.41%	0.44%	0.46%	0.53%	15.38%
Average relative error of ISRF	1.49%	1.53%	1.69%	1.75%	7.82%

Comparisons of the extracted IFCF and the extracted ISRF to the preset curves are shown in Fig. 8(d). Also, two local comparisons (in the time duration of 0–1 s) between the extracted IFCF and the calculated IFCF, and between the extracted ISRF and the preset ISRF are provided in Fig. 8(e) and (f), respectively, to reveal the close-up views. It can be seen that the extracted ISRF and the extracted IFCF are almost identical to the pre-set curves. The average relative error of the IFCF is 0.42% and the average relative error of the ISRF is 1.34%, respectively.

The performance of the proposed method is also tested on the signals with different SNR for the quadratic ISRF. The results of the average relative errors of the IFCF and the ISRF are given in Table 4 when SNR varies from 0 dB to -10 dB. Similar to case 3, the average relative errors of the ISRF are higher than those of the IFCF for each SNR except SNR = -10 dB. Additionally, the average relative errors of the IFCF and the ISRF both increase with a decrease of the SNR. When the SNR reduces from 0 dB to -8 dB, the average relative error of the IFCF increases 1.71% from 0.35% to 1.06% and the average relative error of the ISRF increases 1.01% from 1.06% to 2.07%. The increase rate is higher than for case 3 but the results are still acceptable. However, when the noise further increases to SNR = -10 dB, the results become unacceptable with average relative errors of the IFCF 12.35% and the ISRF 8.24%, which implies a narrower tolerance range of the proposed method when applied to the signal with quadratic ISRF.

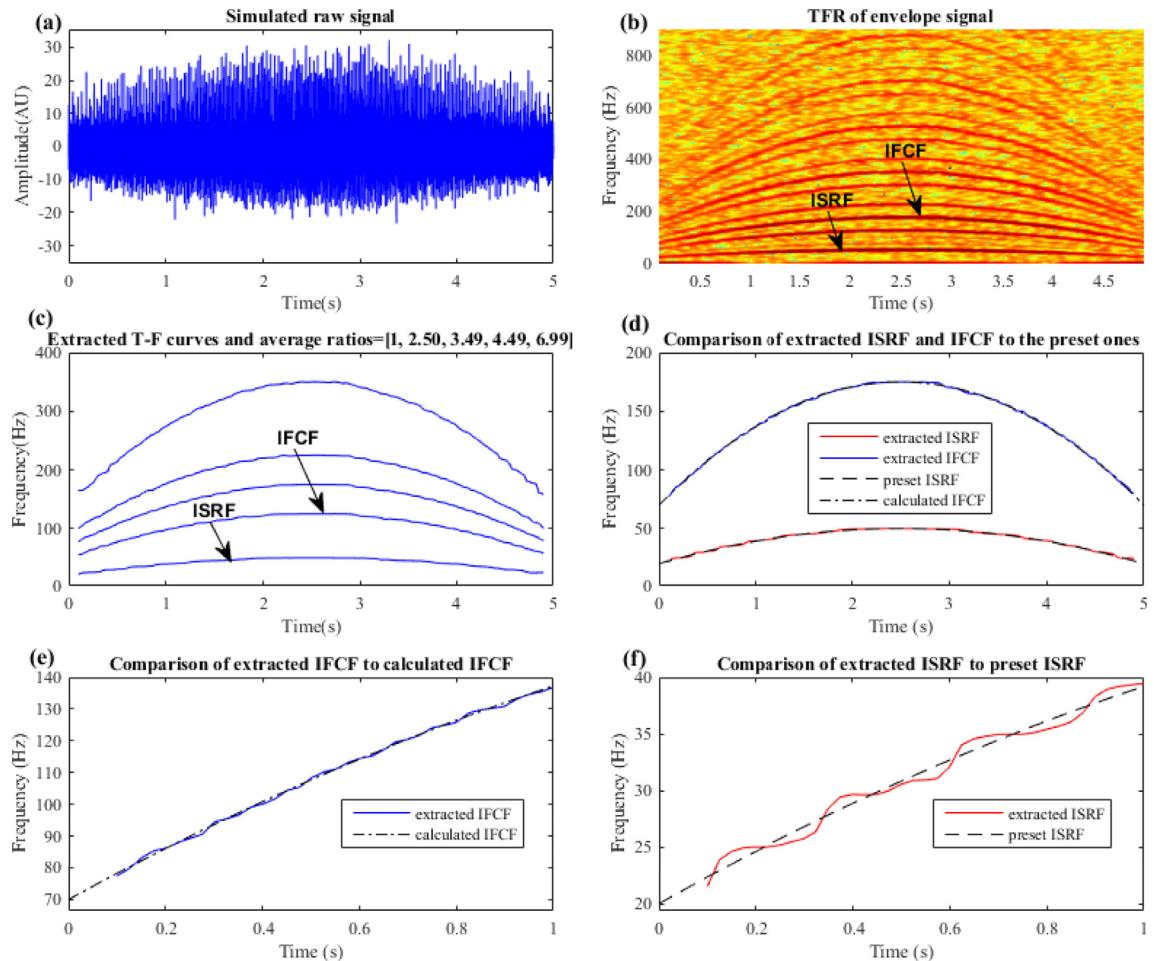


Fig. 8. Results of the proposed method applied to the simulated signal with quadratic ISRF. (a) Simulated raw signal; (b) TFR of envelope signal; (c) extracted T-F curves; (d) comparison of extracted ISRF and IFCF to their preset curves; (e) comparison of extracted IFCF to calculated IFCF (local view); (f) comparison of extracted ISRF to preset ISRF (local view).

Table 4

Average relative errors of IFCF and ISRF with different level of noise for case 4.

	SNR = 0 dB	SNR = -3 dB	SNR = -5 dB	SNR = -8 dB	SNR = -10 dB
Average relative error of IFCF	0.35%	0.40%	0.42%	1.06%	12.35%
Average relative error of ISRF	1.06%	1.16%	1.34%	2.07%	8.24%

Table 5

Processing time required for simulation data analysis.

	Sampling rate	Signal length	Number of extracted curves	Processing time	Average processing time for 1 s of data
Case 1	20 kHz	10s	5	21.08s	2.11s
Case 2	20 kHz	10s	6	22.60s	2.26s
Case 3	20 kHz	5s	5	11.90s	2.38s
Case 4	20 kHz	5s	5	10.00s	2.00s

The results of simulation illustrate that the proposed procedure for bearing fault diagnosis under unknown time-varying rotational speed conditions is effective regardless whether the varying rotational speed is either dominant or non-dominant in the TFR, and it is also applicable to both linear and non-linear speed variation patterns. However, its performance is less robust to noise in the non-linearly varying rotational speed cases in comparison to the linearly varying rotational speed cases. Furthermore, the results of the T-F curve extraction are also affected by the quality, i.e., clearness of the ridges in the TFR.

4.3. Data processing time

To provide more information about the computational cost of this method, the processing times required in the simulations are given in **Table 5**. All the simulations were performed using MATLAB R2015b, on a computer running Windows 7 with an Intel Core i5-4690 3.5 GHz processor and 8.0 GB of RAM. The results in **Table 5** show that the computational cost in processing simulation data is affected by the signal sampling rate, signal length, number of extracted curves, and the time-varying pattern of the ISRF.

5. Experimental validation

To further validate the effectiveness of the proposed method for bearing fault diagnosis under unknown time-varying speed conditions, experiments are conducted to collect vibration signals from bearings operating under varying speed. The experimental set-up is shown in **Fig. 9**. The shaft is driven by a motor which is running at variable speed, controlled by the AC driver. Two bearings are mounted to support the shaft, the one on the right side is healthy and the one on the left side is changed with three different bearings for separate experiments, i.e. a bearing with an outer race defect, a bearing with an inner race defect and a healthy bearing. An accelerometer is installed on the housing of the bearing on the left side to collect

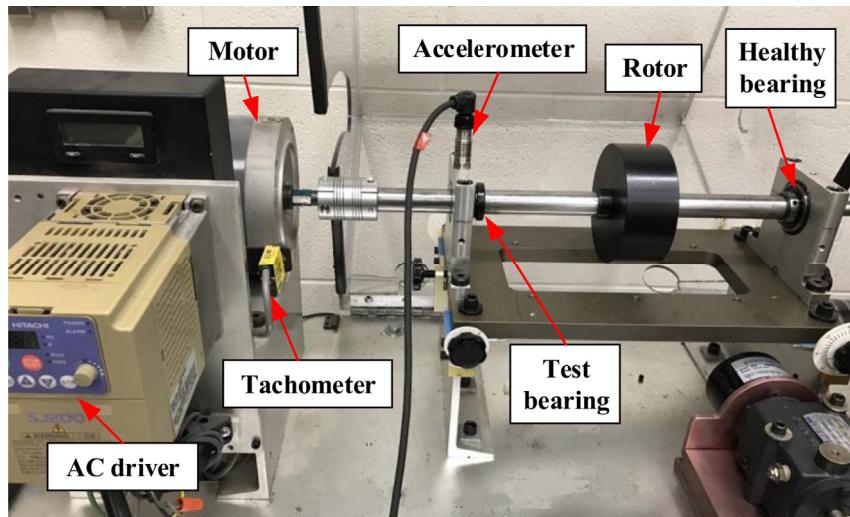


Fig. 9. Experimental set-up.

Table 6

Parameters of bearings used for outer race fault test.

Bearing type	Pitch diameter	Ball diameter	Number of balls	FCC
ER10K	33.50 mm	7.94 mm	8	$3.05f_r$

vibration signals. To validate the result of the proposed method, a tachometer is used to measure the shaft rotating speed. The measured signals are collected by a DAQ card and sampled using LABVIEW.

5.1. Bearing with outer race fault

The experiment is first conducted on a bearing with an outer race fault, i.e. the left bearing shown in Fig. 9 is a bearing with an outer race defect. The parameters of the bearing used for the experiment are given in Table 6. According to Eq. (9), the FCC of the bearing for outer race fault is calculated as 3.05. The sampling frequency of the signal is set as 24 kHz, and the length of the signal is 5.3 s.

The collected vibration signal and shaft rotational speed (i.e. ISRF) are shown in Fig. 10(a) and (b), respectively. The rotational speed is approximately linearly increasing from 25 Hz to 40 Hz. The IFCF can be calculated according to $IFCF = FCC * ISRF$. It can be seen from Fig. 10(a) that the spacing of the impulses becomes smaller and the amplitude of the signal becomes larger with the increase of the shaft rotational speed. The TFR of the amplitude demodulated signal is shown in Fig. 10(c), obtained via the STFT. Many ridges are shown in the TFR, including the ISRF. However, the IFCF cannot be clearly observed.

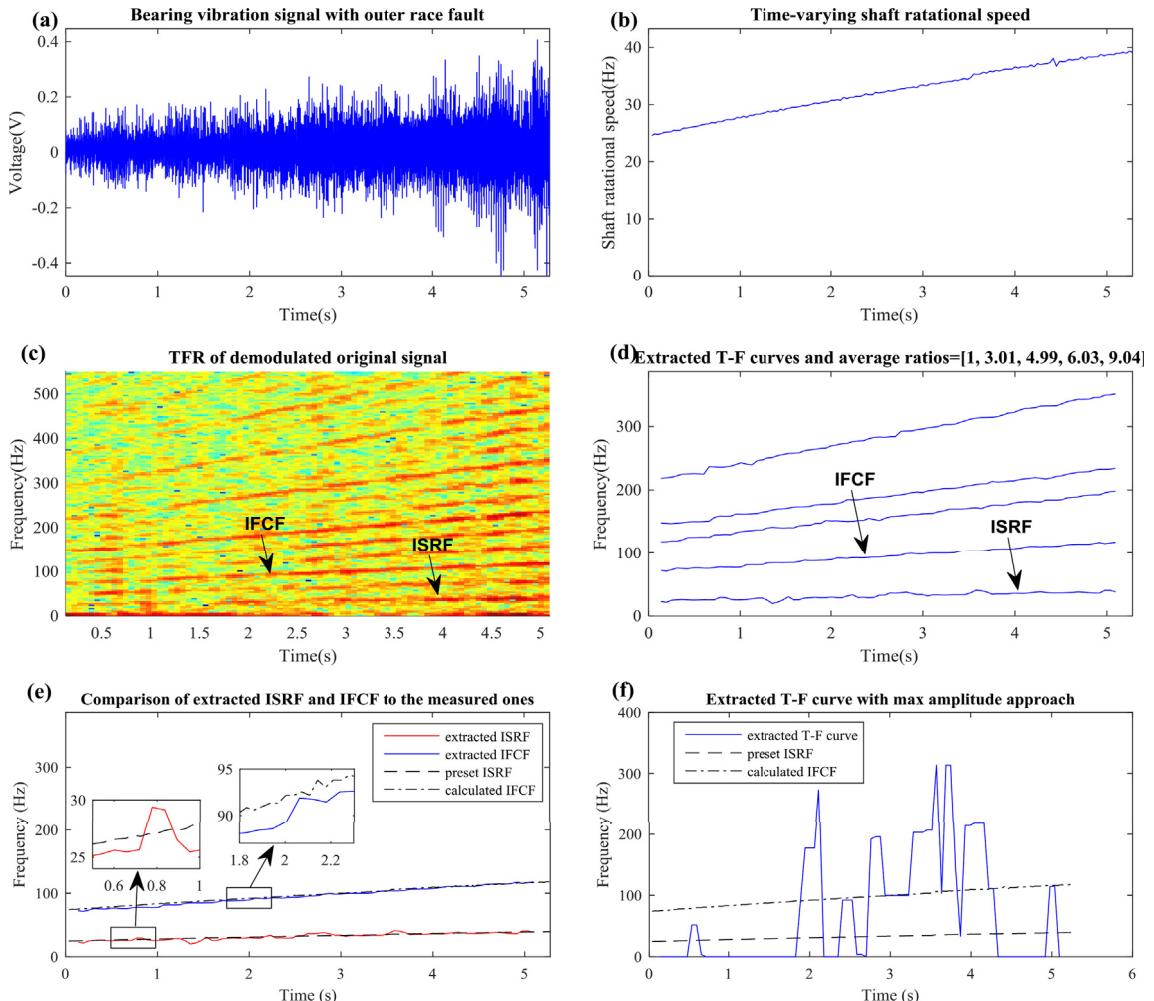


Fig. 10. Results of bearing with outer race fault. (a) Bearing vibration signal with outer race fault; (b) time-varying shaft rotational speed; (c) TFR of demodulated original signal; (d) extracted T-F curves and average ratios; (e) comparison of extracted ISRF and IFCF to the measures curves; (f) extracted T-F curve with max amplitude approach.

Applying the multiple T-F curve extraction algorithm to the TFR, the five curves are obtained, shown in Fig. 10(d). According to the proposed fault diagnosis procedure, the average ratios of the curves to the curve of the lowest frequency are calculated as 1, 3.01, 4.99, 6.03, and 9.04, respectively. Among these average ratios, 3.01 matches the FCC (3.05) with 1.3% relative error (within 5%). Additionally, 6.03 and 9.04 match 2*FCC and 3*FCC with relative error 1.1% and 1.2%, respectively. Also, 4.99 matches 2*FCC-1 (sideband) with relative error 2.2%. Therefore, an outer race fault can be diagnosed according to the proposed bearing fault diagnosis procedure. In addition, the extracted curve of the lowest frequency is revealed as the ISRF and the curve of the average ratio 3.01 is considered as the IFCF.

Comparisons of the extracted IFCF to the calculated IFCF and the extracted ISRF to the measured ISRF are shown in Fig. 10(e). It can be seen that the extracted ISRF and the extracted IFCF are in agreement with the measured ISRF and the calculated IFCF, respectively. The average relative errors of the IFCF and the ISRF are 2.11% and 5.45%, respectively.

To compare the fast path-optimization approach with the maximum amplitude approach for T-F curve extraction, the extracted T-F curve by extracting the point of the maximum amplitude at each moment is shown in Fig. 10(f). It can be seen that the curve is sharply discontinuous, which matches neither the IFCF, nor the ISRF. Obviously, the quality of extracted T-F curves shown in Fig. 10(d) is much better than that in Fig. 10(f).

Table 7

Parameters of bearings used for inner race fault test.

Bearing type	Pitch diameter	Ball diameter	Number of balls	FCC
ER16K	38.52 mm	7.94 mm	9	$5.43f_r$

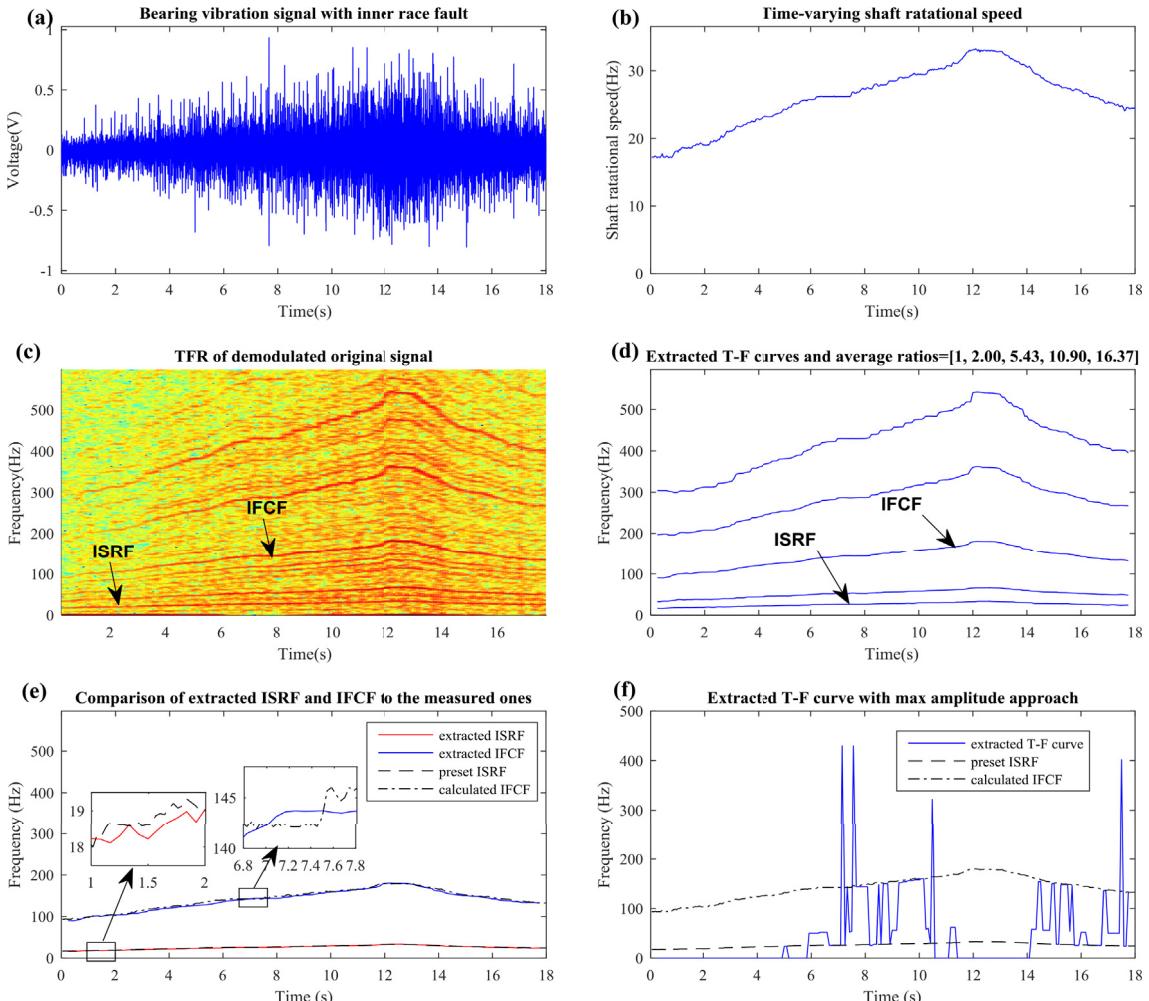


Fig. 11. Results of bearing with outer race fault. (a) Bearing vibration signal with inner race fault; (b) time-varying shaft rotational speed; (c) TFR of demodulated original signal; (d) extracted T-F curves; (e) comparison of extracted ISRF and IFCF to the measures curves; (f) extracted T-F curve with max amplitude approach.

5.2. Bearing with inner race fault

The experiment is also conducted by replacing the left bearing shown in Fig. 9 with one having an inner race defect. The parameters of the bearings used in the test are given in Table 7. According to Eq. (10), the FCC of the inner fault is calculated as 5.43 with the given parameters. The sampling frequency of the signal is set as 12 kHz and the length of the signal is 17.9s.

The collected vibration signal is shown in Fig. 11(a) and the measured shaft rotational speed (ISRF) is shown in Fig. 11(b). The shaft rotational frequency speeds up from 17.3 Hz to 33 Hz at 0–12s and then speeds down from 33 Hz to 24.4 Hz for the rest of the time span. Correspondingly, the amplitude of the vibration signal increases at the beginning and then reduces in Fig. 11(a). The TFR of the demodulated signal is obtained via the STFT shown in Fig. 11(c). Many ridges can be observed in the TFR, including the ISRF and the IFCF and its harmonics. The ridges appear to be cleaner than those in Fig. 10(c) for the outer race fault test.

According to the proposed procedure, the multiple T-F curve extraction algorithm is applied to the TFR. The extracted five T-F curves are shown in Fig. 11(d). The average ratios of the curves to the curve of the lowest curve are calculated as 1, 2.00, 5.43, 10.90, and 16.37, respectively. Among the calculated average ratios, 5.43 matches the FCC (5.43) perfectly. Additionally, 10.90 and 16.37 match 2*FCC and 3*FCC with relative error 0.4% and 0.5%, respectively. Also, 2.00 indicates a double of the curve of the lowest frequency. Therefore, according to the proposed fault diagnosis strategy, an inner race fault is detected. In addition, the extracted curve of the lowest frequency is regarded as the ISRF and the extracted curve of the average ratio 5.43 is regarded as the IFCF.

Comparisons of the extracted IFCF and the extracted ISRF to the measured curves are shown in Fig. 11(e). It can be seen that the extracted ISRF and the extracted IFCF well correspond to the measured ISRF and the calculated IFCF, respectively. Moreover, the relative error of the IFCF and the ISRF are calculated as 2.40% and 2.59%, respectively.

To compare the fast path-optimization approach with the maximum amplitude approach for T-F curve extraction, the extracted T-F curve by extracting the point of the maximum amplitude at each moment is shown in Fig. 11(f). Similar to the result shown in Fig. 10(f), the extracted curve is discontinuous with random jumps. Obviously, the proposed multiple T-F curve extraction algorithm based on the fast path-optimization approach is superior.

5.3. Healthy bearing

To illustrate the applicability of the proposed bearing fault diagnosis method for healthy bearings, the left bearing shown in Fig. 9 is replaced by a healthy bearing. The parameters of the bearing used for the test are given in Table 7. For this type of bearing, the FCCs of outer race fault and inner race fault are calculated as 3.57 and 5.43, respectively. The sampling frequency of the signal is set as 12 kHz and the signal length is 8.65s.

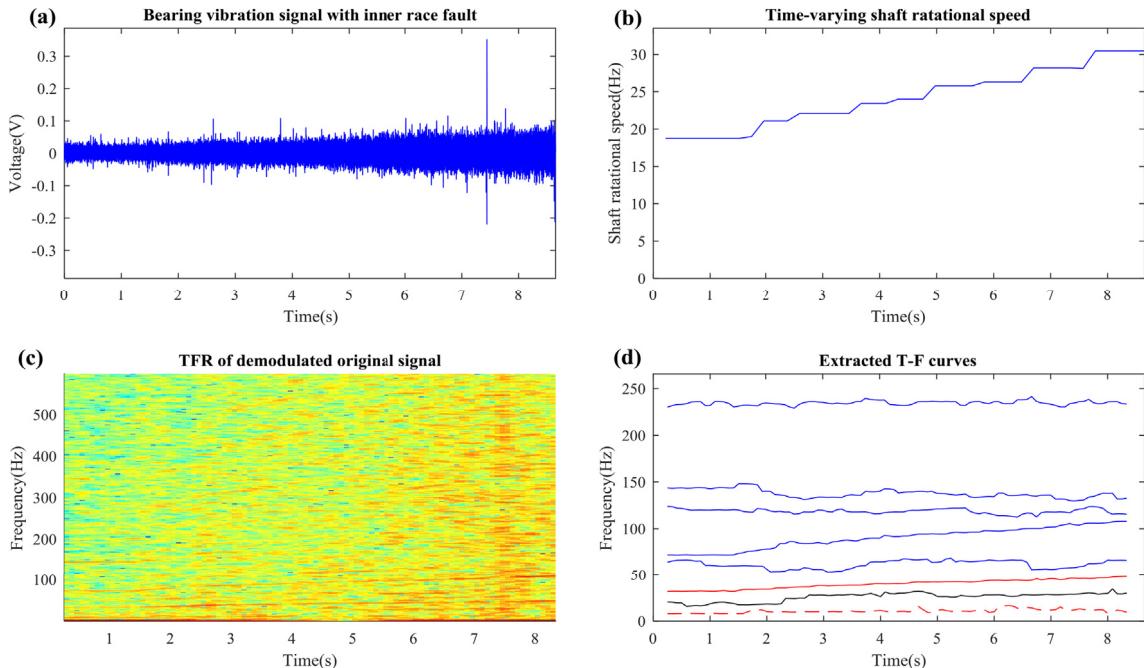


Fig. 12. Results of healthy bearing. (a) Bearing vibration signal; (b) time-varying shaft rotational speed; (c) TFR of demodulated original signal; (d) extracted T-F curves.

Table 8

Processing time required for experimental data analysis.

	Sampling rate	Signal length	Number of extracted curves	Processing time	Average processing time for 1sec of data
Experiment 1	24 kHz	5.3s	5	18.37s	3.47s
Experiment 2	12 kHz	17.9s	5	31.81s	1.78s
Experiment 3	12 kHz	8.65s	8	22.62s	2.62s

The collected vibration signal is shown in Fig. 12(a) and the shaft rotational speed (i.e. ISRF) varies from 18.75 to 30.46 Hz with a pattern shown in Fig. 12(b). The amplitude of the vibration signal correspondingly increases slightly. The TFR of the demodulated signal is obtained via the STFT as shown in Fig. 12(c). In the TFR, no ridges can be clearly observed.

According to the proposed bearing fault diagnosis procedure, the multiple T-F curve extraction algorithm is applied to the TFR. The extracted five curves are plotted in the blue color shown in Fig. 12(d). The average ratios of the curves to the curve of the lowest frequency are calculated as 1, 1.47, 1.94, 2.25, and 3.83, respectively. Among these average ratios, none of them matches the FCC of the outer race fault (3.57) or the FCC of the outer race fault (5.43) within a 5% relative error. Therefore, according to the proposed fault diagnosis strategy, one more curve has to be extracted from the TFR beneath the extracted curve of the lowest frequency. The extracted curve is plotted in the red color shown in Fig. 12(d). Then, the average ratios of the six curves to the new curve are recalculated as 1, 1.58, 2.27, 3.04, 3.35, and 6.0, respectively. Similarly, none of the average ratios matches either the FCC of the outer race fault (3.57) or the FCC of the outer race fault (5.43) within a 5% relative error. Thus, according to the proposed procedure, the second extra curve has to be extracted from the TFR beneath the curve plotted in red color. The second extra T-F curve is plotted in the black color shown in Fig. 12(d). The average ratios of the seven extracted curves to the new curve are recalculated as 1, 1.57, 2.50, 3.69, 4.84, 5.63, and 9.54, respectively. Again, these average ratios do not match either the FCC of the outer race fault (3.57) or the FCC of the outer race fault (5.43) within a 5% relative error. Therefore, the third extra curve has to be extracted from the TFR beneath the curve plotted in black color. The third extra curve is extracted and plotted with red dash line shown in Fig. 12(d). The corresponding average ratios of the eight curves to the new curve are calculated as 1, 2.42, 3.77, 5.94, 8.54, 11.45, 13.31, and 22.59, respectively. Once again, none of the calculated average ratios matches the FCC of the outer race fault (3.57) or the FCC of the outer race fault (5.43) within a 5% relative error. Since the maximum number of the extra curves was set as 3, the T-F curve extraction is terminated. Therefore, according to the proposed fault diagnosis strategy, the bearing is determined to be healthy.

5.4. Data processing time

To examine the computational cost of this method in analyzing experimental data, the related processing times are listed in Table 8. Again, the method was performed using MATLAB R2015b, on the same computer used for simulations with a Windows 7 operating system, an Intel Core i5-4690 3.5 GHz processor and 8.0 GB of RAM. As shown in Table 8, the processing time required for the experimental data shows that the computational cost of the method is related to the sampling rate, signal length, and bearing health conditions.

6. Conclusions

In this paper, a multiple T-F curve extraction algorithm is proposed to extract T-F curves from the TFR for bearing fault diagnosis under time-varying speed conditions. Based on the multiple T-F curve extraction algorithm, a procedure for bearing fault diagnosis under time-varying speed conditions without tachometers is proposed.

The performance of the proposed procedure for bearing fault diagnosis under unknown time-varying speed conditions is tested with four types of simulated signals, including a signal with linearly increasing rotational speed which is dominant in the TFR, a signal with linearly increasing rotational speed which is not dominant in the TFR, a signal with an exponentially decaying rotational speed which is dominant in the TFR, and a signal with quadratically varying rotational speed which is dominant in the TFR. The results of simulations illustrate that the proposed procedure is effective regardless if the varying rotational speed is dominant or non-dominant in the TFR. At the same time, it also performed well for both linear and non-linear speed variation patterns. However, the tolerance for noise in the non-linearly varying rotational speed cases is lower than for the linearly varying rotational speed cases. Additionally, the results of the T-F curve extraction are influenced by the clearness of the ridges in the TFR.

The effectiveness of the proposed procedure is also validated with experimental data. The experimental data are collected from a bearing with an outer race fault, a bearing with an inner race fault and a healthy bearing, respectively. The results obtained by applying the proposed procedure to the experimental signals reveal that the proposed signal can be effectively used for bearing fault diagnosis under time-varying speed conditions without tachometers. For comparison, the result of the T-F curve extraction is also obtained by an existing method of extracting the maximum peak at each moment for each test. This clearly shows that the performance of the proposed T-F curve algorithm is superior.

It is worth mentioning that the results of the T-F curve extraction also depend on the resolution of the TFR. Advanced time-frequency analysis techniques developed based on continuous wavelet transform could potentially improve the accuracy of the T-F curve extraction. Additionally, the T-F curve extraction is not available for the extraction of curves that intersect each

other. Therefore, for a bearing signal contaminated by interferences which may have frequency curves that intersect, the interferences must be removed from the signal before the T-F curve extraction can take place.

Acknowledgements

This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04190, RGPIN/121433), the Ontario Centres of Excellence (OT-SE-E50622) and the China Scholarship Council. The authors would also like to thank Dr. T. Wang for the data collected in our lab at the University of Ottawa.

References

- [1] N. Tandon, A. Choudhury, Review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, *Tribol. Int.* 32 (1999) 469–480.
- [2] R.B. Randall, J. Antoni, Rolling element bearing diagnostics-A tutorial, *Mech. Syst. Signal Process.* 25 (2011) 485–520.
- [3] K.R. Fyfe, E.D.S. Munck, Analysis of computed order tracking, *Mech. Syst. Signal Process.* 11 (1997) 187–205.
- [4] P. Borgesani, P. Pennacchi, R.B. Randall, R. Ricci, Order tracking for discrete-random separation in variable speed conditions, *Mech. Syst. Signal Process.* 30 (2012) 1–22.
- [5] T. Wang, M. Liang, J. Li, W. Cheng, Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis, *Mech. Syst. Signal Process.* 45 (2014) 139–153.
- [6] Y. Wang, G. Xu, A. Luo, L. Liang, K. Jiang, An online tacholess order tracking technique based on generalized demodulation for rolling bearing fault detection, *J. Sound. Vib.* 367 (2016) 233–249.
- [7] C.J. Stander, P.S. Heyns, Transmission path phase compensation for gear monitoring under fluctuating load conditions, *Mech. Syst. Signal Process.* 20 (2006) 1511–1522.
- [8] Z. Feng, M. Liang, F. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples, *Mech. Syst. Signal Process.* 38 (2013) 165–205.
- [9] I. Daubechies, J. Lu, H.T. Wu, Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool, *Appl. Comput. Harmon. Anal.* 30 (2011) 243–261.
- [10] C. Li, M. Liang, A generalized synchrosqueezing transform for enhancing signal time-frequency representation, *Signal Process.* 92 (2012) 2264–2274.
- [11] Z. Feng, X. Chen, M. Liang, F. Ma, Time-frequency demodulation analysis based on iterative generalized demodulation for fault diagnosis of planetary gearbox under nonstationary conditions, *Mech. Syst. Signal Process.* 62 (2015) 54–74.
- [12] Z. Feng, X. Chen, M. Liang, Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions, *Mech. Syst. Signal Process.* 52–53 (2015) 360–375.
- [13] Z. Feng, X. Chen, M. Liang, Joint envelope and frequency order spectrum analysis based on iterative generalized demodulation for planetary gearbox fault diagnosis under nonstationary conditions, *Mech. Syst. Signal Process.* 76–77 (2015) 242–264.
- [14] J. Shi, M. Liang, Y. Guan, Bearing fault diagnosis under variable rotational speed via the joint application of windowed fractal dimension transform and generalized demodulation: a method free from prefiltering and resampling, *Mech. Syst. Signal Process.* 68–69 (2016) 15–33.
- [15] M. Borda, I. Naftornita, D. Isar, A. Isar, New instantaneous frequency estimation method based on image processing techniques, in: Proc. of SPIE'05, vol. 14, 2005, pp. 1–11.
- [16] D. Iatsenko, P.V.E. Mcclintock, A. Stefanovska, Extraction of instantaneous frequencies from ridges in time – frequency representations of signals, *Signal Process.* 125 (2016) 290–303.
- [17] J. Terrien, C. Marque, G. Germain, Ridge extraction from the time-frequency representation (TFR) of signals based on an image processing approach: application to the analysis of uterine electromyogram AR TFR, *IEEE Trans. Biomed. Eng.* 55 (2008) 1496–1503.
- [18] M. Liebling, T.-F. Bernhard, A.H. Bachmann, L. Froehly, T. Lasser, M. Unser, Continuous wavelet transform ridge extraction for spectral interferometry imaging, *Prog. Biomed. Opt. Imaging – Proc. SPIE* 5690 (2005) 397–402.
- [19] M. Liang, I. Soltani Bozchalooi, An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection, *Mech. Syst. Signal Process.* 24 (2010) 1473–1494.
- [20] R.N. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, New York, 1978.