This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A Concentrated Time–Frequency Analysis Tool for Bearing Fault Diagnosis

Gang Yu

***Abstract*— In industrial rotating machinery, the transient signal usually corresponds to the failure of a primary element, such as a bearing or gear. However, faced with the complexity and diversity of practical engineering, extracting the transient signal is a highly challenging task. In this paper, we propose a novel time–frequency analysis method termed the transient-extracting transform, which can effectively characterize and extract the transient components in the fault signals. This method is based on the short-time Fourier transform and does not require extended parameters or a priori information. Quantized indicators, such as Rényi entropy and kurtosis, are employed to compare the performance of the proposed method with other classical and advanced methods. The comparisons show that the proposed method can provide a much more energy-concentrated time–frequency representation, and the transient components can be extracted with a significantly larger kurtosis. The numerical and experimental signals are used to show the effectiveness of our method.**

***Index Terms*— Empirical mode decomposition (EMD), spectral kurtosis (SK), synchrosqueezing transform, time–frequency analysis, transient-extracting transform (TET).**

I. Introduction

IN THE field of fault diagnosis for rotating machinery, a signal processing method is widely applied to find the features that are closely related to the mechanical fault [1], [2]. In the recorded vibration and sound signals, the faults usually show the transient features that appear in a short period of time [3]. Considering that the different fault signals occupy distinct frequency bands, the joined time–frequency (TF) analysis (TFA) is an effective tool for characterizing transient faults that have nonstationary TF features [4]. Although the direct applications of TFA methods in fault diagnosis have been reported in many studies, the inherent shortcomings of the classical TFA method have never been resolved effectively. The linear TFA methods, e.g., short-time Fourier transform (STFT) and wavelet transform (WT), are used to calculate the inner product between the signal and the basis function that has the ability to locate the TF features. However, since there are no TF basis functions that can be compactly supported in the TF domain simultaneously, the linear TFA methods show a poor ability to characterize the precise TF features. Bilinear TFA methods, such as the Wigner–Ville distribution and the Cohen class distribution, are to calculate the Fourier transform of the local signal correlation. However, the unexpected cross terms restrict the application of bilinear TFA methods greatly. These shortcomings in the classical TFA methods can decrease the sensitivity of the diagnosis system to some unobvious faults, such as weak faults in their early stage and faults surrounded by strong noises. To enhance the ability of the TFA methods for detecting faults in a complex environment, some advanced approaches have been proposed and introduced in the past decade, e.g., empirical mode decomposition (EMD) [5], the spectral kurtosis (SK) method [6], [7], and the synchrosqueezing transform (SST) [8]–[10].

EMD is a data-driven method to decompose the 1-D signal into a series of intrinsic mode functions (IMFs). Since different IMFs occupy distinct frequency bands, the transient features of the IMF that contains the fault band can be highly enhanced when compared to the original signal. Due to this superiority, many EMD-based fault diagnosis methods have been developed, and a comprehensive review can be found in [11]. Although we cannot understand the mathematical foundation of this method very well, some studies show that the EMD behaves as a dyadic filter bank when dealing with Gaussian noise. It denotes the factor that while performing the time-series signal processing, the EMD decomposes a signal using fixed dyadic filter banks. Since the frequency band of the fault component in a real-world signal cannot be known in advance, the decomposition results are such that some IMFs may contain the expected fault component or a fault component may be decomposed into several IMFs, which is usually called mode mixing. Because the processing behaviors of the EMD are difficult to control, sometimes the EMD-based fault diagnosis methods are unpredictable and unstable. Recently, more advanced methods are established to improve the performance of the EMD, e.g., local mean decomposition [12], ensemble EMD [13], and extreme-point weighted mode decomposition [14].

The SK method is a technique to extract the most transient component based on the kurtosis indicator. The kurtosis is a statistical variable used to measure the temporal dispersion of a time-series signal, and it can also be used to detect the transients included in the fault signal. The SK method first needs to expand the 1D signal into the 2-D TF plane based on the STFT or the bandpass filter, and then one

Manuscript received October 16, 2018; revised January 27, 2019; accepted February 11, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61374074. The Associate Editor coordinating the review process was Mohamed Abou-Khousa.

The author is with the School of Electrical Engineering, University of Jinan, Jinan 250022, China (e-mail: yugang2010@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2019.2901514

can reconstruct or select the component that is most related to the fault with the largest kurtosis. Benefiting from the sensitivity of the kurtosis indicator to the transient fault, the SK methods show its effectiveness in diagnosing mechanical faults [15], [16].

The SST method was introduced as a postprocessing tool for the linear TFA methods and has been applied in the fault diagnosis of rotating machinery [10]. The SST is intended to obtain a sharper TF representation, which can characterize the faults in high TF resolution. Meanwhile, the transient components can be extracted from the sharper TF result. To extract a signal from the SST result, it is necessary to first estimate the IF trajectory corresponding to the transient component. However, it is challenging to estimate the IF of the transient component precisely, because the fault signals usually do not meet the weakly time-varying requirement of the SST framework. Furthermore, unexpected background noises will introduce serious interference into the SST result, which may lead to the IF being unable to be accurately characterized. To further improve the performance of the SST, some advanced methods are proposed, e.g., demodulated SST [17], matching SST [18], high-order SST [19], and synchroextracting transform (SET) [4].

From the above introduction, we can see that many advanced technologies have been introduced to extract the transient component from the original signal, which is an essential issue for improving the fault detection ability of a diagnosis system. In this paper, we propose a novel TFA method that can characterize the transient features in the TF plane precisely and extract it in the time domain. Comparisons are made between the proposed method and the advanced fault diagnosis methods, which include SK, EMD, SST, and their improved versions. The rest of this paper is organized as follows. Section II details the theory of our proposed method. In Section III, the Rényi entropy and kurtosis indicator are used to illustrate the quantified comparisons of the TF results generated by different TFA methods. Experimental validations are provided in Sections IV and V. The conclusions are drawn in Section VI.

II. Transient-Extracting Transform

A. Background of Time–Frequency Analysis

For a time-varying signal $s(t)$, the linear TFA method functions by correlating it with a dictionary of waveforms that are concentrated in time and in frequency, i.e., $\mathrm{TFA}(t,\omega) = \langle s(u), \psi_{t,\omega}(u)\rangle$, where $\langle , \rangle$ denotes the inner product operator. The basis function $\psi_{t,\omega}(u)$ is usually considered the TF-shift waveform. However, the basis functions are support-limited in the TF domain, such that the TF features of the signal must be projected on a square region called the Heisenberg box. Therefore, linear TFA methods provide a blurry TF explanation for the time-varying signal. Especially for fault signals that have a rapidly changing status, it is impossible to use the linear TFA methods to characterize the transient feature precisely. A numerical fault signal is modeled in Fig. 1(a) according to the mechanical theory [20]–[23], and the analysis results generated by the STFT and the WT are shown in Fig. 1(b) and (c). For the STFT, we select the Gaussian function as the moved window and the window length is 100 samples. For the WT, we employ the Morlet function to address this transient signal. It can be observed that in the time direction, the transient component in the TF plane has a much larger spread than that in the original time-series domain. Although the time-varying TF information can be obtained by the linear TF transform, it must sacrifice the ability of locating the transient event precisely.

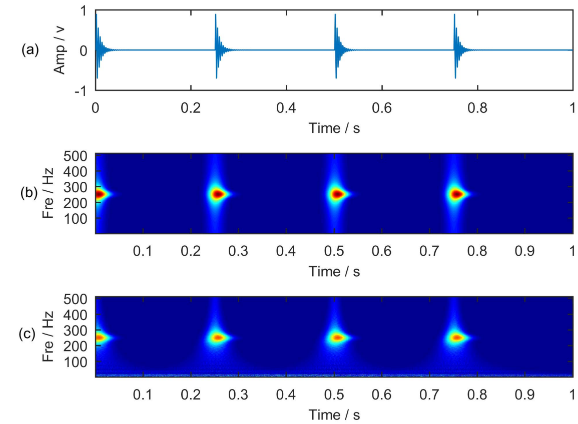

Fig. 1. (a) Signal waveform. (b) STFT result. (c) WT result.

B. Theory of Our Proposed Method

In this section, we start our study from the framework of the STFT in analyzing the Dirac delta function that has the perfect time location property. The STFT expression is given as

$$G(t,\omega) = \int_{-\infty}^{+\infty} g(u-t)\cdot s(u)\cdot e^{-i\omega u}\,du \tag{1}$$

where $g(u-t)$ is the moved window and $s(u)$ is the signal. The STFT basis function $g(u-t)e^{-i\omega u}$ is moved in the time domain and modulated in the frequency domain to detect any time-varying changes that occurred in a signal. In mathematics, the Dirac delta function $\delta(t)$ is a generalized function on the real number line that is zero everywhere except at zero, with an integral of one over the entire real line. In the time domain, the Dirac function $\delta(t)$ should have the best time location, since it only appears at one time point. Such that the Dirac function can be regarded as an ideal model of a signal that has transient features. Usually, a Dirac signal can be expressed as

$$s_\delta(t) = A\cdot\delta(t-t_0). \tag{2}$$

Herein, a discrete Dirac signal being sampled at 200 Hz and $A = 1, t_0 = 0.5$ s is employed to illustrate the deduced procedure more intuitively. The time waveform and frequency spectrum of the Dirac signal are shown in Fig. 2(a) and (b). It can be seen that the ideal TF feature of the Dirac signal should have the best time location and the worst frequency location. The processed result of the Dirac signal by the STFT is shown in Fig. 2(c). It can be observed that the energy corresponding to the Dirac signal, which should have an ideal time location, is expanded heavily in the TF plane. Restricted

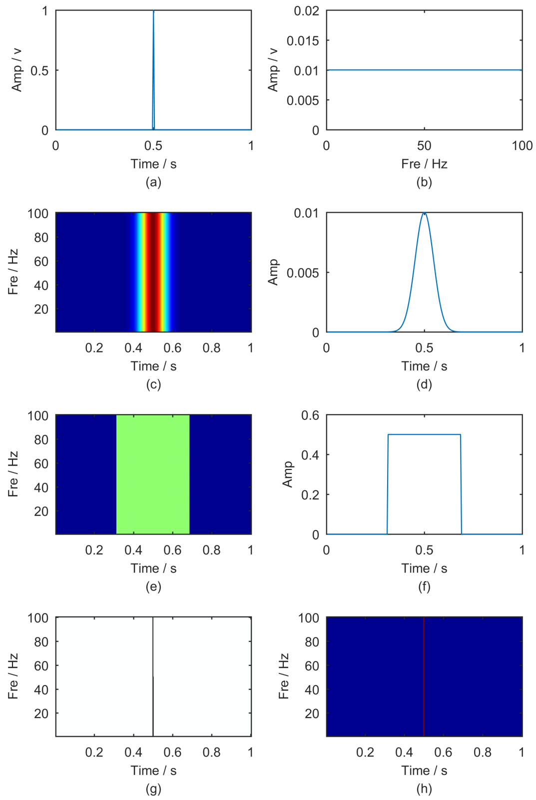

Fig. 2. (a) Dirac function signal with $t_0 = 0.5$ s. (b) Frequency spectrum. (c) STFT spectrogram. (d) Slice of the spectrogram at $\omega_0 = 50$ Hz, i.e., $|G(t,\omega_0)|$. (e) GD $t_0(t,\omega)$. (f) Slice of the GD at $\omega_0 = 50$ Hz, i.e., $t_0(t,\omega_0)$. (g) TEO. (h) TET result.

to the Heisenberg uncertainty principle, for the STFT method, it is impossible to achieve the ideal description even for a simple Dirac signal. To explore the TF energy distribution of the STFT result more specifically, we substitute (2) into (1), then we can have

$$\begin{aligned} G(t,\omega) &= \int_{-\infty}^{+\infty} g(u-t)\cdot A\cdot\delta(u-t_0)\cdot e^{-i\omega u}du \\ &= A\cdot g(t_0-t)\cdot e^{-i\omega t_0}. \end{aligned} \tag{3}$$

Because $|e^{-i\omega t_0}| = 1$, the energy distribution of the STFT result of Dirac delta function can be formulated as

$$|G(t,\omega)| = A\cdot g(t_0-t). \tag{4}$$

By (4), we can see that, because the window function $g(\cdot)$ is compact in the time domain, the energy distribution of $|G(t,\omega)|$ concentrates on the time $t = t_0$ and reaches the maximum value $A\cdot g(0)$ at this time point. To illustrate this analysis clearly, a slice of the STFT spectrogram of this signal at the frequency bin $\omega_0 = 50$ Hz is shown in Fig. 2(d). It can be seen that, in the time direction, the TF energy spreads over the region of window support $[t_0-\Delta, t_0+\Delta]$, where Δ denotes the window function time support.

Another point that should be demonstrated is that the STFT of a Dirac function is constituted by a series of Dirac functions with the same group delay (GD), in which the GDs are all equal to t_0. To estimate the GDs of each Dirac function precisely, it is suggested to first calculate the derivative of $G(t,\omega)$ with respect to the frequency variable. This results in the following equation:

$$\begin{aligned} \partial_\omega G(t,\omega) &= \partial_\omega(A\cdot g(t_0-t)\cdot e^{-i\omega t_0}) \\ &= -it_0\cdot A\cdot g(t_0-t)\cdot e^{-i\omega t_0} \\ &= -it_0\cdot G(t,\omega). \end{aligned} \tag{5}$$

The expression (5) leads to (6). For any (t,ω) and for which $G(t,\omega)\neq 0$, a 2D GD $t_0(t,\omega)$ for the STFT result (3) can be formulated as

$$t_0(t,\omega) = i\cdot\frac{\partial_\omega G(t,\omega)}{G(t,\omega)}. \tag{6}$$

To explain the GD more clearly, it is shown in Fig. 2(e), and the slice of the GD at the frequency bin ω_0 is shown in Fig. 2(f). It can be observed that in the TF region $t\in[t_0-\Delta, t_0+\Delta]$, all values of the 2D GD are equal to $t_0 = 0.5$ s. For the ideal TF representation of the signal (2), the energy should only appear at the time t_0 instead of being spread over a large region. This motivates us to remove the smeared TF coefficients and only to retain the TF coefficient at the time t_0. To achieve this goal, a postprocessing procedure called the transient-extracting operator (TEO) is proposed as

$$\mathrm{TEO}(t,\omega) = \delta(t-t_0(t,\omega)). \tag{7}$$

This considers that

$$t_0(t,\omega) = \begin{cases} t_0, & t\in[t_0-\Delta, t_0+\Delta], \omega\in\mathrm{R}^+ \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Then, we can obtain

$$\delta(t-t_0(t,\omega)) = \delta(t-t_0). \tag{9}$$

Equation (9) means that the TEO is a 2D binarization representation that has a value of one only at time t_0 [see Fig. 2(g)], such that it can be used to extract the TF coefficient of $G(t,\omega)$ at this time point. Considering that (7) has a transient-extracting behavior that can make the STTF result much sparser, the new TFA method employing the TEO is termed the transient-extracting transform (TET) and formulated as

$$Te(t,\omega) = G(t,\omega)\cdot TEO(t,\omega). \tag{10}$$

By (4), it is known that the STFT energy spreads in a large region. Then, it is necessary to explore the energy distribution of the proposed TET method in dealing with the Dirac function signal. It is known that the Dirac delta function $\delta(x)$ has the following property:

$$f(x)\cdot\delta(x-x_0) = f(x_0)\cdot\delta(x-x_0). \tag{11}$$

Therefore, the energy distribution of the novel TF representation can be calculated as

$$\begin{aligned} |Te(t,\omega)| &= |G(t,\omega)\cdot TEO(t,\omega)| \\ &= |A\cdot g(t_0-t)\cdot e^{-i\omega t_0}\cdot\delta(t-t_0(t,\omega))| \\ &= |A\cdot g(t_0-t)\cdot e^{-i\omega t_0}\cdot\delta(t-t_0)| \\ &= |A\cdot g(0)\cdot e^{-i\omega t_0}\cdot\delta(t-t_0)| \\ &= A\cdot g(0)\cdot\delta(t-t_0). \end{aligned} \tag{12}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

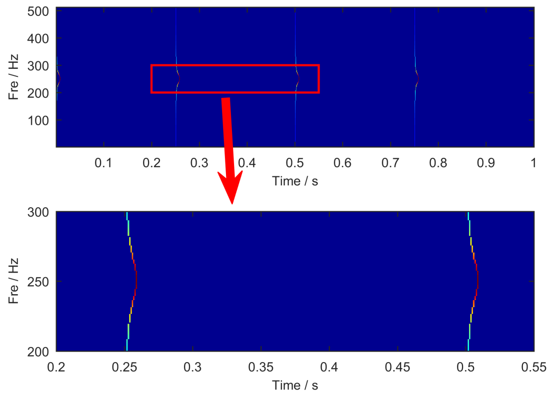

Fig. 3. TET representation of the numerical signal and the zoom.

Compared with the blurry energy distribution of the STFT result (4), the energy of the TET result (12) is highly concentrated and only appears at time t_0, which is equivalent to the definition of the ideal TF representation of the Dirac delta function. The TET result of the Dirac signal is shown in Fig. 2(h). It can be observed that the smeared TF energy disappears, and the retained TF coefficient has the same ideal time location with the Dirac function signal. The numerical signal of Fig. 1(a) is processed by the TET method. The TF result and a zoomed-in view of it are shown in Fig. 3. It can be seen that the energy concentration of the newly generated TF representation is highly improved, and the transient features are characterized much more precisely. From the theory analysis and numerical validation, it can be concluded that the proposed TET is more suitable for processing the transient signals than the STFT.

Although the time-varying signal is represented in the TF plane precisely, sometimes it is necessary to recover the time waveform of the signal to obtain more information that is essential for fault diagnosis. It is known that the STFT has a signal reconstruction expression

$$s(t) = (2\pi g(0))^{-1} \cdot \int_{-\infty}^{+\infty} G(t,\omega) \cdot e^{i\omega t} d\omega \tag{13}$$

which can be interpreted to summarize all TF coefficients along the frequency direction. It motivates us to extract the transient component from the TET result with a similar reconstruction expression, given as

$$s(t) = (2\pi g(0))^{-1} \cdot \int_{-\infty}^{+\infty} Te(t,\omega) \cdot e^{i\omega t} d\omega. \tag{14}$$

One can imagine that by means of integrating the more concentrated TF result, the TET reconstruction is more suitable to extract the transient component than the STFT reconstruction.

C. Algorithm Implementation

In this section, we focus on the practical implementation of the proposed algorithm. In discrete data processing, the partial derivative is commonly implemented approximately by a finite-difference operator, such that $\partial_\omega G(t,\omega)$ should be

$$\partial_\omega G(t,\omega) \approx (G(t,\omega+\omega) - G(t,\omega))/\omega. \tag{15}$$

However, (15) is a bias estimate for the parameter $\partial_\omega G(t,\omega)$. A more precise parameter estimation of $\partial_\omega G(t,\omega)$ can be calculated by

$$\begin{aligned}\partial_\omega G(t,\omega) &= \partial_\omega \left(\int_{-\infty}^{+\infty} g(u-t) \cdot s(u) \cdot e^{-i\omega u} du \right) \\ &= -i \cdot \int_{-\infty}^{+\infty} g(u-t) \cdot u \cdot s(u) \cdot e^{-i\omega u} du \\ &= -i \cdot \int_{-\infty}^{+\infty} g(u-t) \cdot (u-t) \cdot s(u) \cdot e^{-i\omega u} du \\ &\quad - it \cdot \int_{-\infty}^{+\infty} g(u-t) \cdot s(u) \cdot e^{-i\omega u} du \\ &= -i(G^{tg}(t,\omega) + t \cdot G(t,\omega)).\end{aligned} \tag{16}$$

where $G^{tg}(t,\omega)$ can be calculated in the STFT framework using alternative windows $t \cdot g(t)$. Substituting (16) into (7), we have

$$\delta(t - t_0(t,\omega)) = \delta\left(\frac{G^{tg}(t,\omega)}{G(t,\omega)}\right). \tag{17}$$

Additionally, (10) can be rewritten as

$$Te(t,\omega) = G(t,\omega) \cdot \delta\left(\frac{G^{tg}(t,\omega)}{G(t,\omega)}\right). \tag{18}$$

Herein, we consider the practical implementation of the discrete data $s[n]$, $n = 0, 1, \ldots N-1$, where N is the number of samples and the data $s[n]$ correspond to a uniform discretization of $s(t)$ taken at the time $t_n = t_0 + nT$, where T is the sampling interval. The Fourier transform of data $s[n]$ is calculated by$S[k] = \sum_{n=0}^{N-1} s[n] \cdot e^{-i(2\pi/N)nk}$, where $k = 0, 1, \ldots N-1$. First, the TET algorithm needs to calculate two STFTs ($G[n,k]$ and $G^{tg}[n,k]$) with respect to the windows $g[n]$ and $n \cdot g[n]$. The corresponding expressions are, respectively, written as

$$G[n,k] = \sum_{m=0}^{N-1} s[m] \cdot g[m-n] \cdot e^{-i\frac{2\pi}{N}mk}. \tag{19}$$

$$G^{tg}[n,k] = \sum_{m=0}^{N-1} s[m] \cdot [m-n] \cdot g[m-n] \cdot e^{-i\frac{2\pi}{N}mk}. \tag{20}$$

Then, it is necessary to estimate the TEO. In a discrete data process, the Dirac delta function can be obtained by

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0. \end{cases} \tag{21}$$

Considering that the practical data are real-valued, the operator should be taken with the real part of the equation. At the same time, the practical calculation errors cannot be neglected. Therefore, the TEO can be obtained by

$$\delta\left[\frac{G^{tg}[n,k]}{G[n,k]}\right] = \begin{cases} 1, & \left|\mathrm{Re}\left[\frac{G^{tg}[n,k]}{G[n,k]}\right]\right| < \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

where $\varepsilon = (t_l - t_{l-1})/2 = T/2$ and $Re[]$ denotes taking real part. Then, the TET representation can be obtained by

$$Te[n,k] = \begin{cases} G[n,k], & \left|\mathrm{Re}\left[\frac{G^{tg}[n,k]}{G[n,k]}\right]\right| < \frac{T}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

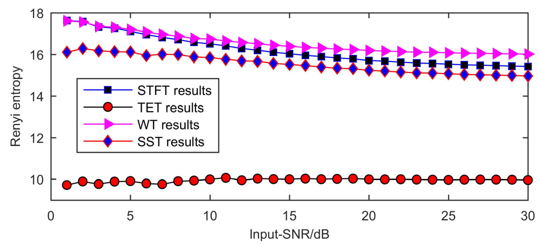

Fig. 4. Rényi entropies of the TF results by different TFA methods under different noise levels (SNR of 1–30 dB).

Eventually, according to the signal reconstruction expression (14), the discrete transient component can be extracted by

$$s[n] = (2\pi g(0))^{-1} \cdot \mathrm{Re}\left[\sum_{k=0}^{N-1} \mathrm{Te}[n,k] \cdot e^{i\frac{2\pi}{N}nk}\right]. \quad (24)$$

From the discrete formulation (23), the computational complexity can be known that the TET has the same level of burden with the STFT. Meanwhile, the transient component can be recovered conveniently. Considering that the STFT has been widely applied in real-time engineering, the TET method has the potential for use in real-time applications as well.

III. Numerical Validation

To evaluate the performance of the generated TF representation and to extract the transient component by different methods objectively, we employ two indicators to quantify the analysis results, which are the Rényi entropy and kurtosis. Information entropy is a common measure used to estimate the dispersion of information content, such that we adopt the Rényi entropy to evaluate the energy distribution of the TF representation. The Rényi entropy of order α for the TF representation is defined as

$$R^{\alpha} = \frac{1}{1-\alpha}\log_2 \frac{\int\int \mathrm{TFR}(t,\omega)^{\alpha}dtd\omega}{\int\int \mathrm{TFR}(t,\omega)dtd\omega} \quad (25)$$

where the order is usually set as $\alpha = 3$. A lower Rényi entropy of the TF result denotes that the TFA method in question can generate a more concentrated TF representation and provide a more accurate characterization of the time-varying TF features. The numerical signal in Fig. 1 is added with Gaussian white noises ranging from 1 to 30 dB of SNR. The TFA methods, STFT, WT, SST, and our proposed TET, are employed to address these noisy signals. The Rényi entropies of these TF representations generated by different TFA methods are calculated and listed in Fig. 4, which shows that the TET results have the smallest Rényi entropies at each noise level.

The TF representations of the noisy signal with an SNR equal to 3 dB are shown in Fig. 5. It can be observed that restricted to the Heisenberg uncertainty principle, the STFT and WT cannot provide an energy-concentrated TF representation. The SST technique is designed to enhance the TF energy concentration of the signal with a weakly modulated frequency and asymptotic modulated amplitude [8]. However, when addressing a signal with a strongly modulated amplitude, e.g., this numerical signal, the SST fails to give an energy-concentrated TF result [see Fig. 5(d)]. In Fig. 5(b), the TET result looks much sparser than in the other TFA methods.

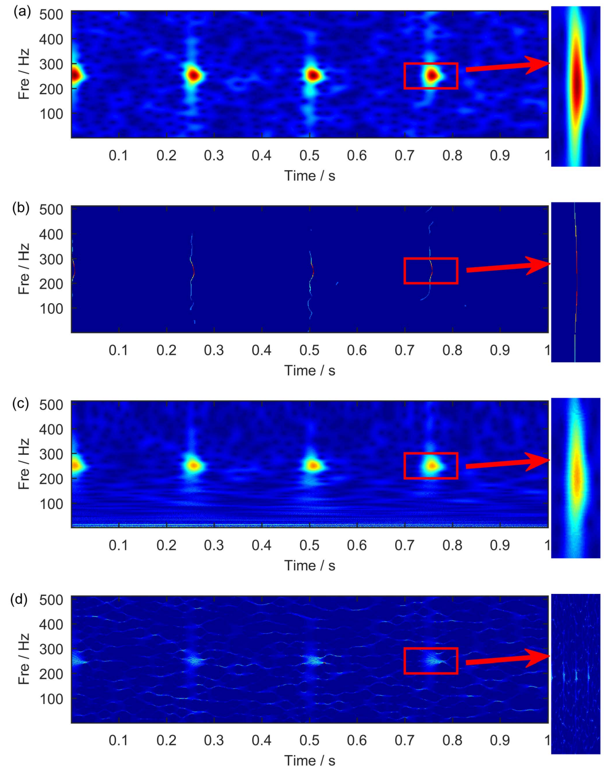

Fig. 5. (a) STFT result. (b) TET result. (c) WT result. (d) SST result.

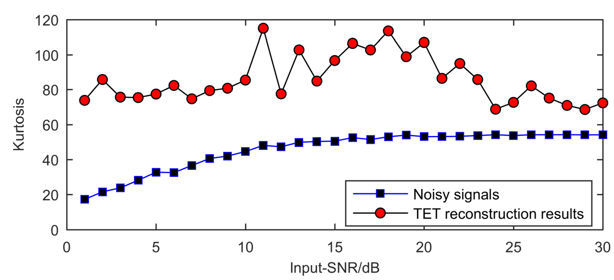

Fig. 6. Under different noise levels (SNR of 1–30 dB), the kurtosis of the reconstructed results from the TET is compared to the original noisy signals.

Kurtosis is a measure of peakedness and, hence, has been accepted as a sensitive indicator to reflect the transient features of a fault signal. Herein, the kurtosis is employed to evaluate the analysis results. A larger kurtosis of the results reconstructed by the TFA method means that a given method has a better ability to extract the transient component.

Due to the fact that the TET allows for transient component reconstruction, the kurtosis of the original signal and the signals recovered by the TET are calculated and shown in Fig. 6. It can be seen that with an increase of the added noises, the kurtosis of the original signals decreases, which means the noises can destroy the transient features of the signal. At each noise level, the TET results have the larger

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

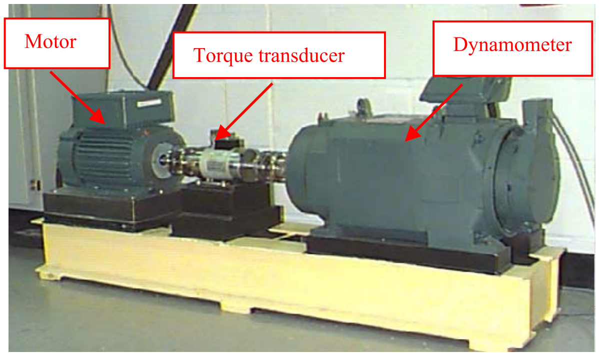

Fig. 7. Structural sketch of the machine set.

kurtosis, which denotes that the TET is effective to extract the transient component when compared to the original signal

IV. Bearing Fault Analysis of Rotating Machinery Under Constant Speed

Rolling element bearings are one of the most prevalent components in rotating machines, and their failure is one of the most frequent reasons for machine breakdown. The vibration signal recorded from the rotating machines is an effective way to diagnose the current operation of the bearing. In this section, the data set with two-type faults provided by the Case Western Reserve University Bearing Data Center is used as the experimental data to validate the proposed method. The structural sketch is shown in Fig. 7, which mainly contains a motor, a torque transducer, and a dynamometer [14]. The test bearing (6105-2RS JEM SKF) is used to support the motor shaft. Two heavy faults are introduced to the bearing outer race and inner race via electrodischarge machining. Vibration signals are recorded by an accelerometer, which is placed at the drive end of the motor housing. The current rotating speed is 1796 r/min. According to the bearing parameters and the rotating speed, we can calculate that the fault characteristic frequencies of the outer race fault and the inner race fault are 107.3 and 162.1 Hz, which also means that the interval between two adjacent transients should be 9.3197 and 6.169 ms, respectively.

A. Outer Race Fault Data Analysis

The waveform of the vibration signal with the outer race fault is shown in Fig. 8(a), which shows somewhat repetitive transients. The TF representations generated by the STFT, TET, WT, and SST are shown in Fig. 8(b)–(e), and the zoom-in view of the TF results is shown in the right-hand side. With the aid of TFA technology, the TF features of the repetitive transients can be expanded into the TF plane to provide more essential information than the individual time-domain analysis. It can be observed that the STFT provides a coarse description of the transients that makes it difficult to locate the TF information precisely. In Fig. 8(c), the proposed TET method generates a much sparser TF result, which can provide better TF location ability. The WT provides a similarly coarse TF result as that of the STFT [see Fig. 8(d)], which

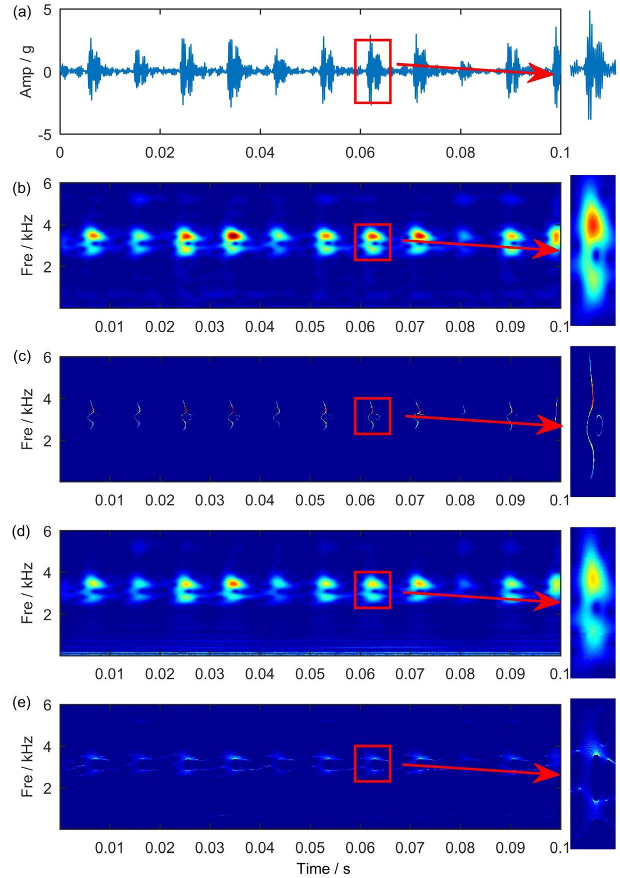

Fig. 8. (a) Fault signal waveform. (b) STFT result. (c) TET result. (d) WT result. (e) SST result.

TABLE I
Rényi Entropies of the TF Results in Figs. 8 and 9

TFA	STFT	TET	WT	SST
Rényi Entropy	16.844	10.9069	17.1492	15.1934
TFA	2nd-SST	3rd-SST	4th-SST	SET
Rényi Entropy	13.2646	14.1363	13.8667	12.2314

TABLE II
Required Computational Time by Several Methods

TFA	STFT	TET	WT	SST
Time (s)	0.08	0.33	0.35	0.67
TFA	2nd-SST	3th-SST	4th-SST	SET
Time (s)	0.79	0.81	0.95	0.46

is because the inner product-based TFA methods must be restricted by the Heisenberg uncertainty principle. Although the SST is designed to obtain a sharper TF result, the SST result in Fig. 8(e) is still blurry and divergent.

For more comparisons, we utilize the high-order SST (e.g., second-SST, third-SST, and fourth-SST) and SET methods to address this signal, and the TF results are shown in Fig. 9. It can be observed that the SST and SET methods seem to

TABLE III
KURTOSIS OF THE DECOMPOSED RESULTS BY DIFFERENT METHODS

TFA	TET	SK	IMF1	IMF2	IMF3	IMF4
Kurtosis	38.854	4.289	4.469	4.398	2.431	0.745

TABLE IV
RÉNYI ENTROPIES OF THE TF RESULTS IN FIG. 14

TFA	STFT	TET	WT	SST
Rényi Entropy	17.8893	11.4630	17.9287	15.8151

TABLE V
KURTOSIS OF THE DECOMPOSED RESULTS BY DIFFERENT METHODS

TFA	TET	SK	IMF1	IMF2	IMF3	IMF4
Kurtosis	111.615	2.143	2.394	1.607	0.205	1.181

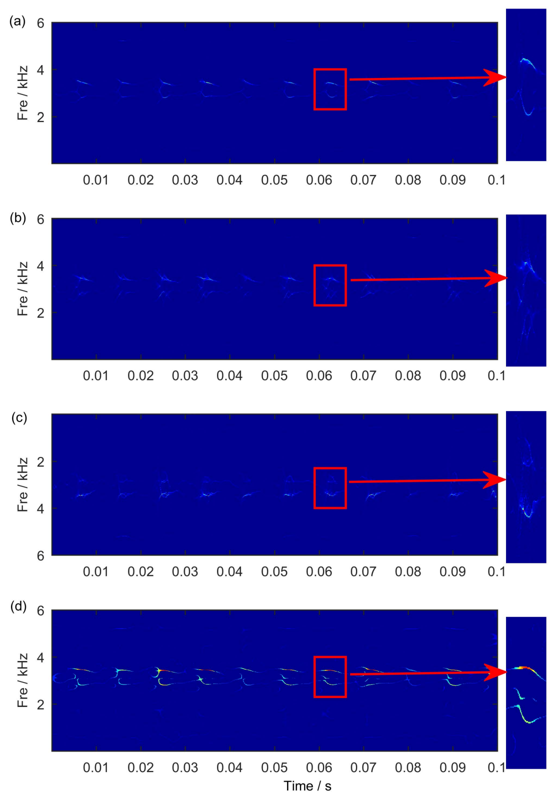

Fig. 9. (a) Second-SST result. (b) Third-SST result. (c) Fourth-SST result. (d) SET result.

be not suitable for dealing with such a signal with strong amplitude-modulated law. Moreover, the corresponding Rényi entropies of all TF results are listed in Table I, which quantitatively illustrate that the TF location ability of our proposed method is the best option. The computation efficiency of a TFA method is also essential in real-time applications. Herein, we test the computational time required for the above-mentioned TFA methods in addressing this vibration signal. The computation times are listed in Table II. It can be seen that the TET method and other methods can finish the processing

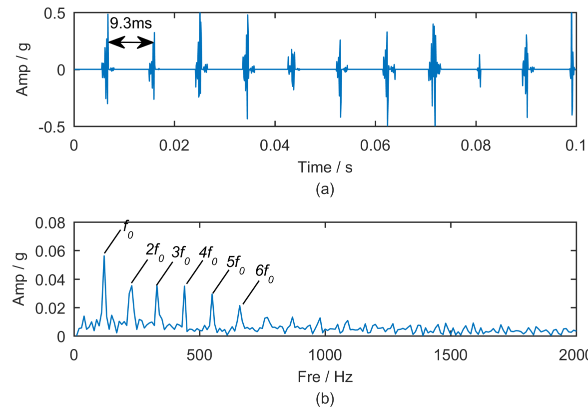

Fig. 10. Extracted transient components by (a) TET method and (b) envelop spectrum.

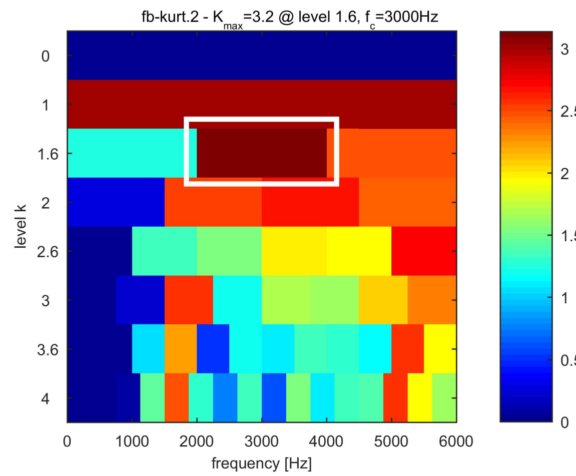

Fig. 11. Kurtogram generated by the SK.

within 1 s, which denotes that the proposed method has a high efficiency level.

Meanwhile, the transient component in the TET result is extracted and shown in Fig. 10(a). Compared with the original signal, the transient features of the extracted signal are obviously improved, making this reconstruction more suitable for diagnosing the fault type. In Fig. 10(b), the outer race fault frequency (denoted as $f_0 = 107$ Hz) derived from the envelop spectrum is clearly characterized.

We also consider using the SK and EEMD techniques to reconstruct the transient components from the original vibration signal. According to the framework of the SK method, we first display the kurtogram in Fig. 11, which can indicate the best bandwidth and the central frequency. The reconstructed component is shown in Fig. 12, which shows some transients.

Considering that the EMD may suffer from drawbacks such as mode mixing and the end effect, an EEMD method that can improve the performance of the original EMD is employed to decompose the IMFs. In the operation of EEMD, the parameters are set to an ensemble member of 100 and added noise with a standard deviation of 0.2. The first five

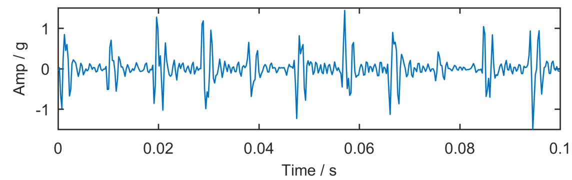

Fig. 12. Signal filtered by the SK method.

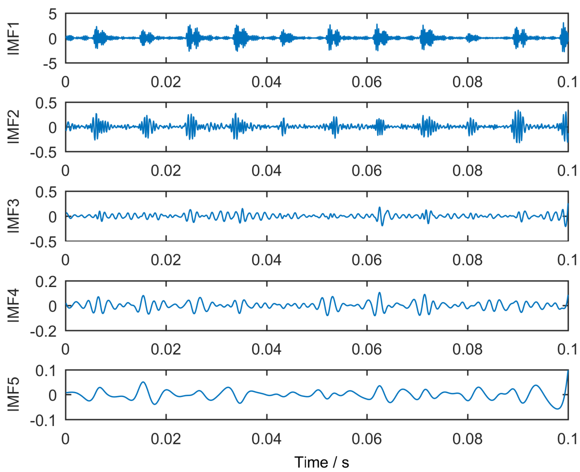

Fig. 13. IMF results decomposed by the EEMD.

IMFs are shown in Fig. 13. To compare the decomposed results quantitatively, the kurtosis of the results processed by the TET, SK, and EEMD is calculated and listed in Table III. It can be seen that the proposed TET method provides a decomposed result with a significantly larger kurtosis than the other methods. It can be concluded that our proposed method is more suitable for extracting the transient components.

B. Inner Race Fault Data Analysis

In this section, we focus on analyzing the vibration signal with an inner race fault, whose waveform is shown in Fig. 14(a). It can be seen that this signal is surrounded by a heavy amount of noise and does not show obvious fault transients. The TF representations created by different methods are shown in Fig. 14(b)–(e). The Rényi entropies are listed in Table IV. It can be seen that the proposed TET method provides the sparsest TF representation and characterizes the fault transients with the best TF resolution when compared to other TFA methods. The extracted result and the envelop spectrum are shown in Fig. 15 and show the obvious inner race fault frequency (denoted as $f_{\mathrm{i}} = 162$ Hz).

The SK and EEMD methods are used to analyze the vibration signal. According to the kurtogram (see Fig. 16), the component with the largest kurtosis is decomposed and displayed in Fig. 17. It can be seen that the decomposed results of SK do not show obvious fault transients. This is because the selected bandwidth may not cover the significant portion of the fault-excited frequency region, such that the kurtogram fails to indicate the best bandwidth and central frequency.

The first five IMFs obtained by EEMD are shown in Fig. 18. It can be observed that the fault transients are still surrounded

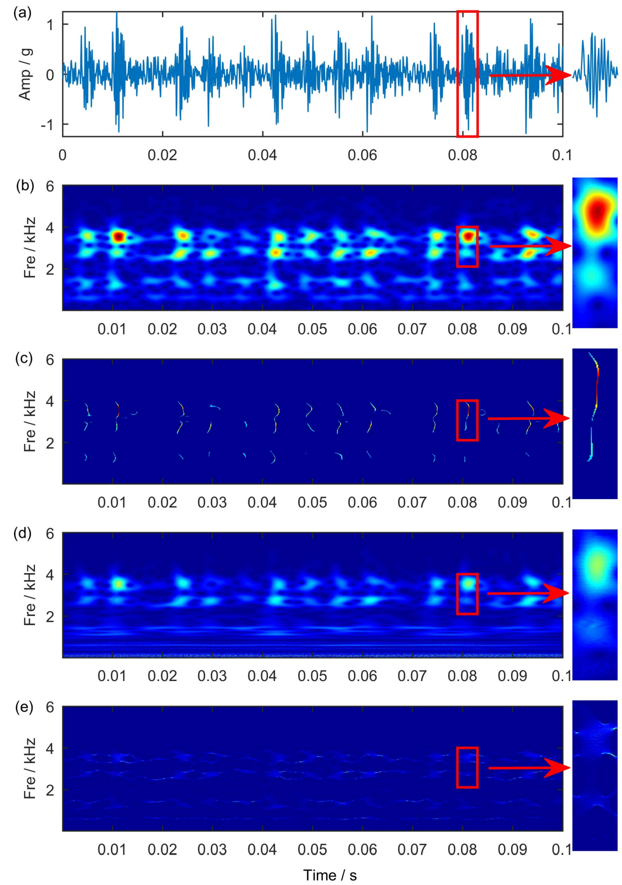

Fig. 14. (a) Fault signal waveform. (b) STFT result. (c) TET result. (d) WT result. (e) SST result.

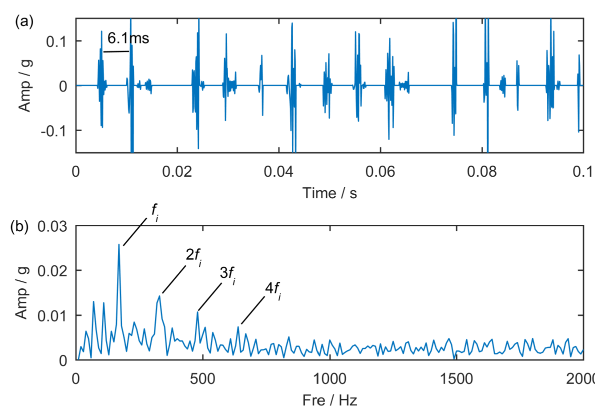

Fig. 15. Extracted transient components by (a) TET method and (b) envelop spectrum.

by heavy amounts of noise. The kurtosis of all decomposed results is listed in Table V, which demonstrates that the proposed TET method provides a result with a significantly larger kurtosis than other methods.

V. Bearing Fault Analysis of Rotating Machinery Under Time-Varying Speed

In this section, we focus on analyzing the bearing faults of a rotating machinery under time-varying speed. The structural

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

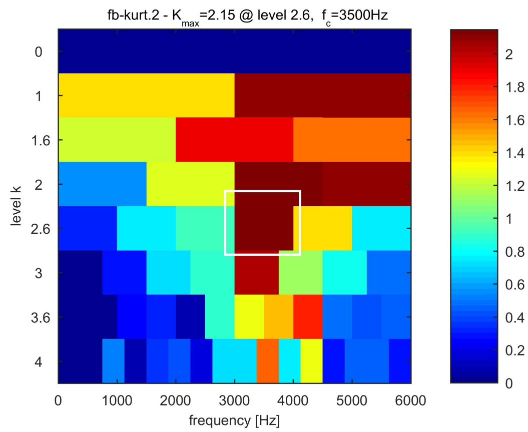

Fig. 16. Kurtogram generated by the SK.

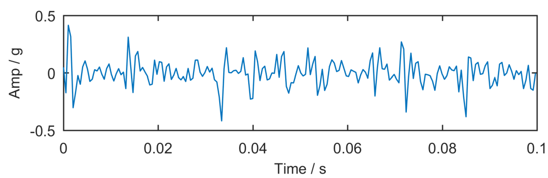

Fig. 17. Filtered signal by the SK method.

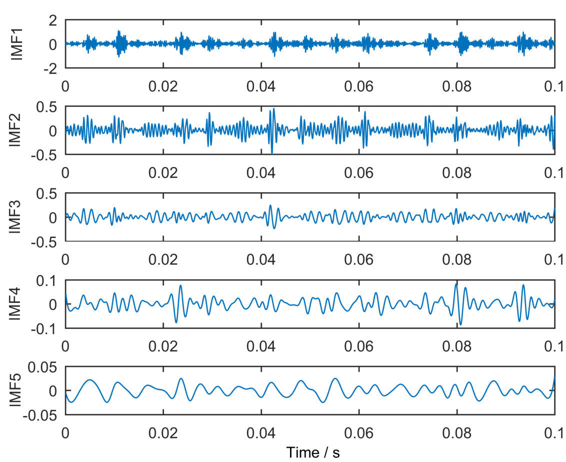

Fig. 18. IMF results decomposed by the EEMD.

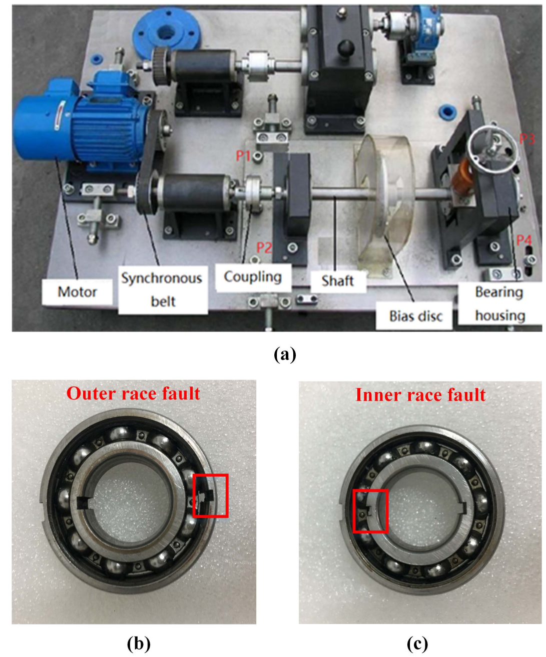

Fig. 19. (a) Structural sketch of the machine set. (b) Outer race fault. (c) Inner race fault.

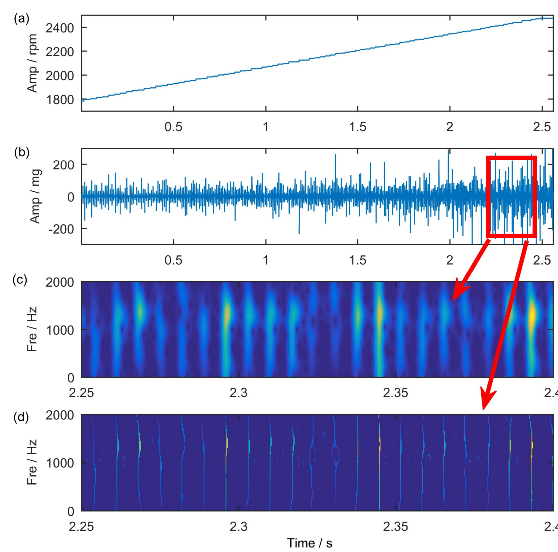

Fig. 20. (a) Rotating speed. (b) Waveform of vibration signal. (c) STFT result. (d) TET result.

sketch is shown in Fig. 19(a). The type of the test bearing is SKF 6205. Two faults are introduced to the bearing outer race and inner race via wire cutting, which are shown in Fig. 19(b) and (c). Vibration signals are recorded by the accelerometer that is placed at the bearing housing. Meanwhile, the rotating speed is also recorded by an inductive sensor (tachometer). According to the bearing parameters, it can be calculated that the ratio of the fault characteristic frequencies of the outer race fault and the inner race fault with respect to the rotating frequency should be 3.584 and 5.416, respectively.

A. Outer Race Fault Data Analysis

In this section, we first deal with the vibration signal of the bearing with outer race fault. The vibration signal is collected during a speed-up procedure of about 1800–2450 r/min (30–40.8 Hz). The rotating speed and the vibration signal are shown in Fig. 20(a) and (b). It can be seen that, with the increasing of rotating speed, the amplitude of the vibration signal becomes much larger. Then, this signal is first addressed by the STFT and TET methods. In Fig. 20(c) and (d), to show much clearer TF features, we only display the TFA results at a small TF region (2.25–2.4 s and 0–2000 Hz). It is obvious that the energy of TET result is more concentrated than that of the STFT result, which is more helpful for precisely locating the occurrence of each transient.

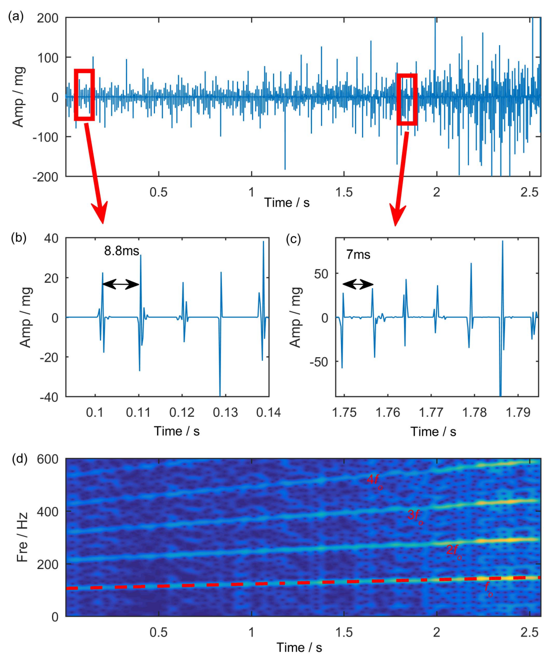

Fig. 21. (a) Reconstructed signal by TET result. (b) Zoomed-in view waveform. (c) Zoomed-in view waveform. (d) TF spectrogram.

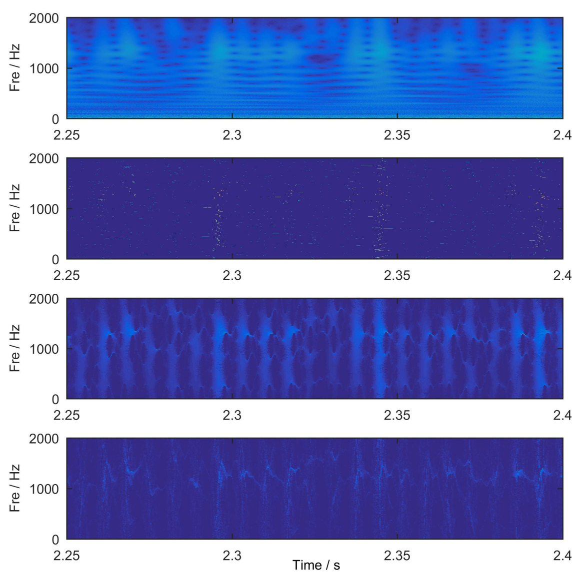

Fig. 22. (a) WT result. (b) SET result. (c) SST result. (d) Fourth-SST result.

Then, we reconstruct the signal from the TET result, which is plotted in Fig. 21(a). Meanwhile, two fragments of the reconstructed signal at a short time are shown in Fig. 21(b) and (c). It is obviously shown that each transient has a very short duration and the interval between two adjacent transients in the low-speed stage is larger than that in the high-speed stage. Moreover, in Fig. 21(d), we display the time-varying trajectory of outer race fault characteristic frequency (red dotted line) together with the TF spectrogram of the envelope of this reconstructed signal. It can be seen that the time-varying fault characteristic frequency and its high-order components are well characterized in the TF spectrogram. For more comparisons, the TF results of the vibration signal generated by other TFA methods are shown

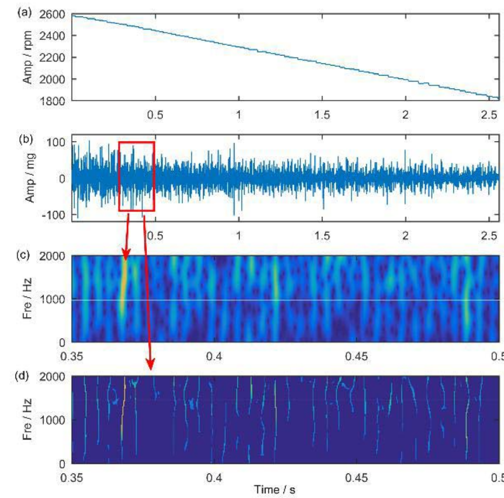

Fig. 23. (a) Rotating speed. (b) Waveform of vibration signal. (c) STFT result. (d) TET result.

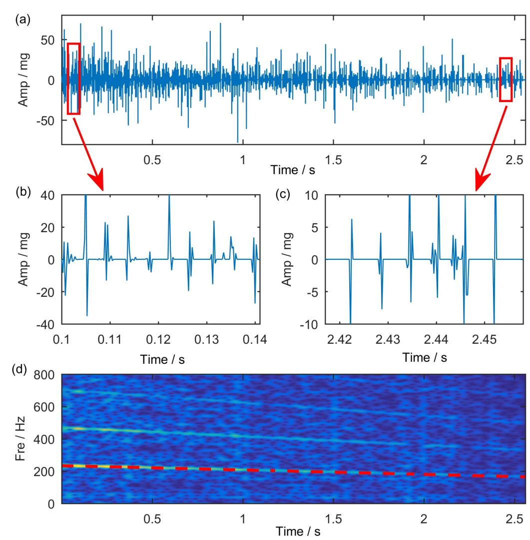

Fig. 24. (a) Reconstructed signal by TET result. (b) Zoomed-in view waveform. (c) Zoomed-in view waveform. (d) TF spectrogram.

in Fig. 22. However, all these TF results are too blurry to provide a precise TF description.

B. Inner Race Fault Data Analysis

In this section, we address the vibration signal recorded from the rotating machinery where the bearing is being with inner race fault. The vibration signal is collected during a speed-down procedure of about 2600–1800 r/min (43.3–30 Hz). The rotating speed and the vibration signal are shown in Fig. 23(a) and (b). In Fig. 23(c) and (d), the TF representations show that the TET provides a significantly concentrated result than the STFT. The reconstructed signal from the TET result and its zoomed version are shown in Fig. 24(a)–(c). It can be seen that, in the high-speed stage, the vibration signal has a relatively large amplitude and small interval between each transient. Moreover, in Fig. 24(d), we plot the time-varying trajectory of inner race fault characteristic frequency (red dotted line) together with the TF spectrogram of the envelope of this reconstructed signal. It is obvious that the time-varying fault characteristic frequency and its high-order components are clearly characterized in the TF spectrogram. It can be concluded that the proposed TET method is an effective tool for fault diagnosis of rotating machinery under the time-varying speed.

VI. Conclusion

In this paper, we propose a novel TFA method. Via a postprocessing operation for the STFT representation, most of the blurry TF energy is removed, and only the TF information closely related to signal transient features is retained, such that the TF location ability can be improved greatly compared to the STFT method. Meanwhile, the novel TF result allows for transient component reconstruction. To gauge their abilities to deal with both numerical signals and real-world signals, we focus on the quantitative comparisons of the proposed method with other classical and advanced methods. Through the comparisons, it can be concluded that our proposed method has the best performance in improving the energy concentration and can decompose the transient components with the largest kurtosis compared to the other methods. The proposed method has the same computational complexity level as the STFT, which implies that it also has potential for the real-time application. The MATLAB code of the TET can be found on: https://ww2.mathworks.cn/matlabcentral/fileexchange/70319.

References

[1] S. Wang, X. Chen, G. Li, X. Li, and Z. He, "Matching demodulation transform with application to feature extraction of rotor rub-impact fault," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1372–1383, May 2014.

[2] Z. K. Peng, G. Meng, F. L. Chu, Z. Q. Lang, W. M. Zhang, and Y. Yang, "Polynomial chirplet transform with application to instantaneous frequency estimation," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 9, pp. 3222–3229, Sep. 2011.

[3] W. Huang, G. Gao, N. Li, X. Jiang, and Z. Zhu, "Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, to be published.

[4] G. Yu, M. Yu, and C. Xu, "Synchroextracting transform," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8042–8054, Oct. 2017.

[5] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[6] J. Antoni, "The spectral kurtosis: A useful tool for characterising non-stationary signals," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, Feb. 2006.

[7] S. Wang, G. Cai, Z. Zhu, W. Huang, and X. Zhang, "Transient signal analysis based on Levenberg–Marquardt method for fault feature extraction of rotating machines," *Mech. Syst. Signal Process.*, vols. 54–55, pp. 16–40, Mar. 2015.

[8] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, Mar. 2011.

[9] S. Wang, X. Chen, G. Cai, B. Chen, X. Li, and Z. He, "Matching demodulation transform and synchrosqueezing in time-frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 69–84, Jan. 2014.

[10] J. Shi, M. L. Dan, D.-S. Necsulescu and Y. Guan, "Generalized stepwise demodulation transform and synchrosqueezing for time–frequency analysis and bearing fault diagnosis," *J. Sound Vib.*, vol. 368, pp. 202–222, Apr. 2016.

[11] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mech. Syst. Signal Process.*, vol. 19, no. 5, pp. 974–988, 2005.

[12] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *J. Roy. Soc. Interface*, vol. 2, no. 5, pp. 443–454, 2005.

[13] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 01, no. 01, pp. 1–41, Jan. 2009.

[14] J. Zheng, H. Pan, T. Liu, and Q. Liu, "Extreme-point weighted mode decomposition," *Signal Process.*, vol. 142, pp. 366–374, Jan. 2018.

[15] S. Wang, W. Huang, and Z. K. Zhu, "Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 4, pp. 1299–1320, May 2011.

[16] H. Liu, W. Huang, S. Wang, and Z. Zhu, "Adaptive spectral kurtosis filtering based on Morlet wavelet and its application for signal transients detection," *Signal Process.*, vol. 96, pp. 118–124, Mar. 2014.

[17] S. Wang, X. Chen, C. Tong, and Z. Zhao, "Matching synchrosqueezing wavelet transform and application to aeroengine vibration monitoring," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 2, pp. 360–372, Feb. 2017.

[18] S. Wang, X. Chen, I. W. Selesnick, Y. Guo, C. Tong, and X. Zhang, "Matching synchrosqueezing transform: A useful tool for characterizing signals with fast varying instantaneous frequency and application to machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 100, pp. 242–288, Feb. 2018.

[19] D.-H. Pham and S. Meignen, "High-order synchrosqueezing transform for multicomponent signals analysis—With an application to gravitational-wave signal," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3168–3178, Jun. 2017.

[20] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen, "Review on dynamics of cracked gear systems. Engineering failure analysis," *Eng. Failure Anal.*, vol. 55, pp. 224–245, Sep. 2015.

[21] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[22] D. He, H. Cao, S. Wang, and X. Chen, "Time-reassigned synchrosqueezing transform: The algorithm and its applications in mechanical signal processing," *Mech. Syst. Signal Process.*, vol. 117, pp. 255–279, Feb. 2019.

[23] G. Yu, Z. Wang, and P. Zhao, "Multi-synchrosqueezing transform," *IEEE Trans. Ind. Electron.*, to be published.

Gang Yu received the B.Eng. degree in mechanical engineering from Qingdao University, Qingdao, China, in 2010, and the Ph.D. degree in mechanical engineering from Shandong University, Jinan, China, in 2016.

He is currently a Lecturer in mechanical engineering with the University of Jinan, Jinan. His current research interests include time–frequency analysis, blind source separation, modal identification and machinery condition monitoring, and fault diagnosis.