Morphology-based Fault Feature Extraction and Resampling-free Fault Identification Techniques for Rolling Element Bearing Condition Monitoring

by

Juanjuan Shi

A Dissertation Submitted to the Faculty of Graduate and Postdoctoral Studies in Fulfillment of the Requirement for the Degree of

Doctorate of Philosophy

in Mechanical Engineering

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering
University of Ottawa

Abstract

As the failure of a bearing could cause cascading breakdowns of the mechanical system and then lead to costly repairs and production delays, bearing condition monitoring has received much attention for decades. One of the primary methods for this purpose is based on the analysis of vibration signal measured by accelerometers because such data are information-rich.

The vibration signal collected from a defective bearing is, however, a mixture of several signal components including the fault-generated impulses, interferences from other machine components, and background noise, where fault-induced impulses are further modulated by various low frequency signal contents. The compounded effects of interferences, background noise and the combined modulation effects make it difficult to detect bearing faults. This is further complicated by the nonstationary nature of vibration signals due to speed variations in some cases, such as the bearings in a wind turbine. As such, the main challenges in the vibration-based bearing monitoring are how to address the modulation, noise, interference, and nonstationarity matters. Over the past few decades, considerable research activities have been carried out to deal with the first three issues. Recently, the nonstationarity matter has also attracted strong interests from both industry and academic community. Nevertheless, the existing techniques still have problems (deficiencies) as listed below:

(1) The existing enveloping methods for bearing fault feature extraction are often adversely affected by multiple interferences. To eliminate the effect of interferences, the prefiltering is required, which is often parameter-dependent and knowledge-demanding. The selection of proper filter parameters is challenging and even more so in a time-varying environment.

(2) Even though filters are properly designed, they are of little use in handling in-band noise and interferences which are also barriers for bearing fault detection, particularly for incipient bearing faults with weak signatures.

(3) Conventional approaches for bearing fault detection under constant speed are no longer applicable to the variable speed case because such speed fluctuations may cause "smearing" of the discrete frequencies in the frequency representation. Most current

methods for rotating machinery condition monitoring under time-varying speed require signal resampling based on the shaft rotating frequency. For the bearing case, the shaft rotating frequency is, however, often unavailable as it is coupled with the instantaneous fault characteristic frequency (IFCF) by a fault characteristic coefficient (FCC) which cannot be determined without knowing the fault type. Additionally, the effectiveness of resampling-based methods is largely dependent on the accuracy of resampling procedure which, even if reliable, can complicate the entire fault detection process substantially.

(4) Time-frequency analysis (TFA) has proved to be a powerful tool in analyzing nonstationary signal and moreover does not require resampling for bearing fault identification. However, the diffusion of time-frequency representation (TFR) along time and frequency axes caused by lack of energy concentration would handicap the application of the TFA. In fact, the reported TFA applications in bearing fault diagnosis are still very limited.

To address the first two aforementioned problems, i.e., (1) and (2), for constant speed cases, two morphology-based methods are proposed to extract bearing fault feature without prefiltering. Another two methods are developed to specifically handle the remaining problems for the bearing fault detection under time-varying speed conditions. These methods are itemized as follows:

(1) An efficient enveloping method based on signal Fractal Dimension (FD) for bearing fault feature extraction without prefiltering,
(2) A signal decomposition technique based on oscillatory behaviors for noise reduction and interferences removal (including in-band ones),
(3) A prefiltering-free and resampling-free approach for bearing fault diagnosis under variable speed condition via the joint application of FD-based envelope demodulation and generalized demodulation (GD), and
(4) A combined dual-demodulation transform (DDT) and synchrosqueezing approach for TFR energy concentration level enhancement and bearing fault identification.

With respect to constant speed cases, the FD-based enveloping method, where a short time Fractal dimension (STFD) transform is proposed, can suppress interferences and highlight the fault-induced impulsive signature by transforming the vibration signal into a STFD representation. Its effectiveness, however, deteriorates with the increased complexity of the interference frequencies, particularly for multiple interferences with high frequencies. As such, the second method, which isolates fault-induced transients from interferences and noise via oscillatory behavior analysis, is then developed to complement the FD-based enveloping approach. Both methods are independent of frequency information and free from prefiltering, hence eliminating the tedious process for filter parameter specification. The in-band vibration interferences can also be suppressed mainly by the second approach. For the nonstationary cases, a prefiltering-free and resampling-free strategy is developed via the joint application of STFD and GD, from which a resampling-free order spectrum can be derived. This order spectrum can effectively reveal not only the existence of a fault but also its location. However, the success of this method relies largely on an effective enveloping technique. To address this matter and at the same time to exploit the advantages of TFA in nonstationary signal analysis, a TFA technique, involving dual demodulations and an iterative process, is developed and innovatively applied to bearing fault identification.

The proposed methods have been validated using both simulation and experimental data collected in our lab. The test results have shown that the first two methods can effectively extract fault signatures, remove the interferences (including in-band ones) without prefiltering, and detect fault types from vibration signals for constant speed cases. The last two have shown to be effective in detecting faults and discern fault types from vibration data collected under variable speed conditions without resampling and prefiltering.

Acknowledgements

First and foremost I would like to thank my primary supervisor Dr. Ming Liang. He is an expert in my research area and has been a great role model for me. I and my work have greatly benefited from his expert guidance, professional mentorship and partial financial support. I really thank for his time, idea, help and encouragement. The joy and enthusiasm he has for the research have been motivational for me during the Ph.D pursuit, even during tough time in my PhD study. I would also like to express my great appreciation to my Co-supervisor Dr. Dan-Sorin Necsulescu. It is a real privilege for me to share his exceptional scientific knowledge as well as his humour. Being their Ph.D student has been a great honor for me. They have taught me, both consciously and unconsciously, how to become a good researcher. It would never have been possible for me to have this work completed without their incredible suggestions and support.

I would also like to thank Dr. Buddour, Dr. Robitaille and Dr. Lanteigen for being the examiners of my Ph.D comprehensive exam. Special appreciation also goes to Dr. Marc Doumit, Dr. Balbir Dhillon, Dr. Jie Liu of Carleton University, and Dr. Viliam Makis of University of Toronto for being the committee members of my Ph.D dissertation defence. I thank them for their time spent on examining my thesis. They asked questions and offered insights that helped me improve the work greatly.

The members of my lab group have contributed immensely to my personal and professional time at the University of Ottawa. The group has been a source of friendship as well as good advice and collaboration. I am also grateful for all my friends, particularly those who have been with me for so many years. Their company and encouragement helped me go through the tough times in the course of my Ph.D study.

I gratefully acknowledge the funding support for my research study during the four-year period from the University of Ottawa and Chinese Scholarship Council.

Last but not least, I can never have enough words to express how grateful I am for my parents for all their immeasurable love and support. They have always encouraged me to explore my potential and pursue my dreams. I could not reach this stage in my life without their faith in me. I would also like to thank my younger sister and brother for the love and concern for me. Finally, I would especially like to thank Mr. Fu for his support and motivation throughout the work. Thank you all.

Table of Contents

List of Figures

List of Tables

Acronyms

AM-FM	Amplitude Demodulation-Frequency Demodulation
ASTFT	Adaptive Short Time Fourier Transform
BPOF	Ball Pass Outer Raceway Frequency
BPIF	Ball Pass Inner Raceway Frequency
CWT	Continuous Wavelet Transform
DRS	Discrete Random Separation
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
EO	Energy Operator
FCC	Fault Characteristic Coefficient
FCF	Fault Characteristic Frequency
FD	Fractal Dimension
FT	Fourier Transform
GD	Generalized Demodulation
GFT	Generalized Fourier Transform
HHT	Hilbert-Huang Transform Spectrum
HRF	High Resonance Frequency
IF	Instantaneous Frequency
IFCF	Instantaneous Fault Characteristic Frequency

IOBSD	Iterative Oscillatory Behavior-based Signal Decomposition
KPSA	Kurtosis-based Peak Search Algorithm
LMD	Local Mean Decomposition
MA	Morphological Analysis
MCA	Morphological Component Analysis
MHD	Margenau-Hill distribution
MMA	Multi-scale Morphological Analysis
MSE	Mean Squared Error
OBSD	Oscillatory Behavior-based Signal Decomposition
RS-SES	Reversed Sequence Squared Envelope Spectrum
SANC	Self-adaptive Noise Cancellation
SE	Structure Element
SIR	Signal-to-Interference Ratio
SK	Spectral Kurtosis
SNR	Signal-to-Noise Ratio
ST	Synchrosqueezing Transform
STFD	Short Time Fractal Dimension
STFT	Short Time Fourier Transform
SVD	Singularity Value Decomposition
SVM	Support Vector Machine

TEO	Teager Energy Operator
TFA	Time-Frequency Analysis
TFR	Time-Frequency Representation
TQWT	Tunable Q-factor Wavelet Transform
T-SES	Traditional Squared Envelope Spectrum
VPMCD	Variable Predictive Model Based Class Discriminate
WPT	Wavelet Package Transform
WPT	Wavelet Packet Transform
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

Nomenclature

a	Scale factor
$\mathbf{a}$	Coefficient vector
$\mathrm{A}(t)$	Instantaneous amplitude
b	Translation factor
C_{max}	Maximum mutual coherence between bases
ΔC_{max}	Gain of C_{max}
d	Planar extent of a waveform
$d(t)$	Demodulator
$d(t,u)$	Composite demodulator of DDT
$d^{b}(t)$	Backward demodulator
$d^{f}(t)$	forward demodulator
dist( )	Euclidean distance
Envelope( )	Envelope of a signal
$f_0(f_c)$	Constant frequency
$f(t)$	Function of IFCF
$f_e(t)$	Function of estimated IFCF
$f_I(t)$	Function of IFCF of inner race fault
$f_o(t)$	Function of IFCF of outer race fault

$f_x(t)$	Function of IF of signal x
$H_0(\ )$	Frequency response of low-pass filter
$H_1(\ )$	Frequency response of high-pass filter
$i(t)(\mathrm{I}(t))$	Interferences
$J(\ )$	Objective function
J_{max}	Maximum number of stage
$kurt(\ )$	Kurtosis value
L	Sum of Euclidean distance
$\max(\ )$	Maximum value
$mod(\ )$	Modulus maxima
$\mathrm{n}(t)$	Noise
$O(\ )$	Big O-notation
$s(t)$	Simulated impulsive signal
soft()	Soft-thresholding rule
t	Time instant
t_m	Occurrence time of the mth impulse
$u(t)$	Step function
$\mathbf{w}$	Wavelet transform coefficient vector
$W_x(a,b;\omega)$	Wavelet transform of signal x at scale a, time b and mother wavelet ω
$x(t)$	Signal mixture

$x_d(t)$	Dual demodulated signal
$X(t)$	Analytical version of signal x
$X(f)$	Fourier transform of signal x
$X_G(f)$	Generalized Fourier Transform
β	A coefficient used to count slippage effect
γ	Regularized parameter
δ	Mean square error difference
$\delta(t)$	Impulse function
λ	Coefficient of $d(t)$
μ	Penalty parameter
$\Im()$	Augmented Lagrangian function
σ	Standard deviation
$\phi(t)$	Instantaneous phase
$\mathbf{\Phi}$	Basis
$\psi(t)$	Modulating source
$\psi_x(t)$	Angular frequency of signal x
$\omega_{a,b}(t)$	Mother wavelet transform

Chapter 1 Introduction

1.1 Overview

Rolling element bearings are among the most commonly used mechanical components. They are often used in a harsh environment carrying heavy load and are hence one of the most failure-prone mechanical elements. For example, 44% of the total amount of failures of large induction motors is related to bearing breakdown (Zhang et al., 2011). According to Mortazavizadeh and Mousavi (2014), common machine faults in electric rotor are :

1) Bearing failure;
2) Rotor broken bars;
3) Rotor body failure;
4) Bearing misalignment;
5) Rotor misalignment;
6) Bearing loss of lubrication;
7) Rotor mechanical or thermal unbalanced.

As shown on the above list, of all causes for rotor failures, three were directly related to bearing. A bearing failure could result in a stall of a mechanical system with a catastrophic consequence, leading to productivity and financial losses. Therefore, assessment of the running conditions of bearings is of great importance to prevent rotating machinery from unexpected and catastrophic failures. The issue of providing timely and effective detection of bearing faults has been attracting much interest in both industry and academia (Benbouzid and Kliman, 2003).

Over the past few decades, much work has been carried out and various methods have been explored for bearing fault detection. These methods mainly rely on four types of information sources, i.e., acoustics, temperature, wears debris and vibration signals (Garner and Leopard, 1985; Mba, 2003; Dempsey et al., 2006; Patil et al., 2008; He et al., 2009). Among these, it has been found out that vibration signals resulting from rolling element bearings carry rich information related to bearing conditions (Howard, 1994; Wang and Liang, 2011). The availability of a variety of affordable vibration sensors and signal conversion devices, together with the advances in high speed data acquisition provide a unique opportunity for vibration signal collection. Therefore, analysis based on vibration signals has been established as one of the most common and reliable ways for bearing fault detection (Nikolaou and Antoniadis, 2002a). However, vibration signals are indirect sources of bearing health condition and the bearing condition cannot be directly assessed from the acquired raw signals. To reliably detect bearing faults, advanced vibration signal analysis methods have to be developed. The main factors (matters) affecting the *quality* of fault detection results include the modulation effect, background noise and various interferences contained in the vibration signal and the non-stationary nature of certain bearing signals. The causes of these factors are itemized as follows (McFadden and Smith, 1984; Howard, 1994; Antoni and Randall, 2002; Randall, 2004; Antoni, 2006; Zhang and Randall, 2009; Zhang et al., 2013):

1) Modulation effect: A rolling element bearing with localized defects generates a series of impulses through making contact with another surface in the bearing. The fault impulse is short and transient in nature and its frequency content, hence, spreads over a wide frequency range, which usually excites a resonance in the system at a high frequency. The resonant frequency is modulated by the impulse frequency of the bearing fault. Moreover, due to the usual slip of the rolling element, the rate of repetition of the impulses experience a certain degree of randomness. Such variations between the consecutive impulses cause smearing of the spectral energy. Thus, to effectively reveal the frequency of impulse repetition, i.e., the fault characteristic frequency (FCF), demodulation is often required (see Fig. 1.1 (a)). However, most the available demodulation methods are adversely affected by noise and interferences vibration. Therefore, the enveloping spectrum, i.e., the spectrum of the demodulated signal, may be still clouded with artifacts of shaft, gear meshing and other unwanted signal components, leading that the FCF and its harmonics may not be readily revealed (Afshari and Loparo, 1998) (see Fig. 1.1 (b) and (c)).
2) Strong noise: In an industrial setting, the collected vibration signal contains strong background noise from various sources such as data acquisition devices, electric cables, wiring, and power line. Moreover, many rotating machinery contain various mechanical components such as gears, turbines, injectors, and pumps, etc., which would produce additional random noise. The signal of a bearing fault, especially at its early stage, can be masked by such noise, making it difficult to detect a bearing fault (see Fig. 1.1 (b)).

Filtering the signal via a bandpass filter with center frequency around the resonant frequency would improve the signal-and-noise ratio (SNR); nevertheless, the filter design is often knowledge-demanding and time-consuming. Moreover, even if the appropriate filter is designed initially, it may become ineffective later in a changing environment.

3) Various interferences: Except the strong background noise, there are frequency disturbances together with their harmonics caused by mass unbalance, power line system, shaft misalignment and other mechanical/electrical components. These interferences can easily mislead the detection results (see Fig. 1.1 (c)). Though some interferences can be removed by bandpass filtering, if designed properly, the in-band interferences (i.e., the interferences residing in the resonant frequency band) cannot be easily and quickly removed by most of the existing methods.
4) Non-stationarity: Apart from the constant speed condition, bearings are always subject to speed fluctuations, such as bearings in wind turbine, mining equipment etc. Such speed variations would give rise to severe ‘smearing’ on the spectrum and lead the discrete frequency lines to being unobservable even on the envelope spectrum (see Fig. 1.1 (d)). Therefore, the methods developed for bearing fault detection under constant running conditions are no longer applicable to time-varying running conditions.

The graphical illustration of the effects of modulation, noise, interferences and nonstationarity is presented in Fig. 1.1 and explained as follows. The first column sub-figures show the raw signals which are the clean impulsive signal incorporating roller slippage, impulses mixed with noise, impulses polluted with noise and interferences, and nonstationary signal (vibration under variable speed condition), respectively. The corresponding frequency spectra and envelope spectra of these signals are respectively presented in second and third columns. According to the spectrum of the clean impulsive signal, the frequency content spreads over a wide frequency band as shown in the graph in middle of the first row in Fig. 1.1. Furthermore, the harmonics of FCF in the high frequency region cannot be observed due to the smearing effect caused by roller slippage. The impulsive signal is then demodulated. Frequency analyzing the demodulated signal yields the envelope spectrum in the sub-figure of the first row on the right hand side. As can be seen, the frequency of the impulse repetition can be more easily identified.

However, bearing vibrations are often tainted by strong noise and interferences which mask the FCF and its harmonics. As we can see in Fig. 1.1 (b) and (c), neither spectrum nor envelope spectrum can reveal the FCF and its harmonics, which illustrates that the effectiveness of classical enveloping method is undermined by noise and interferences. Fig. 1.1 (d) presents the simulated vibration signal from a time-varying speed case. The ‘smearing’ phenomenon on the spectrum is much more severe because of the speed fluctuations. As a result, the discrete frequency lines cannot be recognized on the envelope spectrum, as seen in Fig. 1.1 (d) on the right, even if the signal is clean. The spectral-analysis-based fault detection hence becomes more complicated.

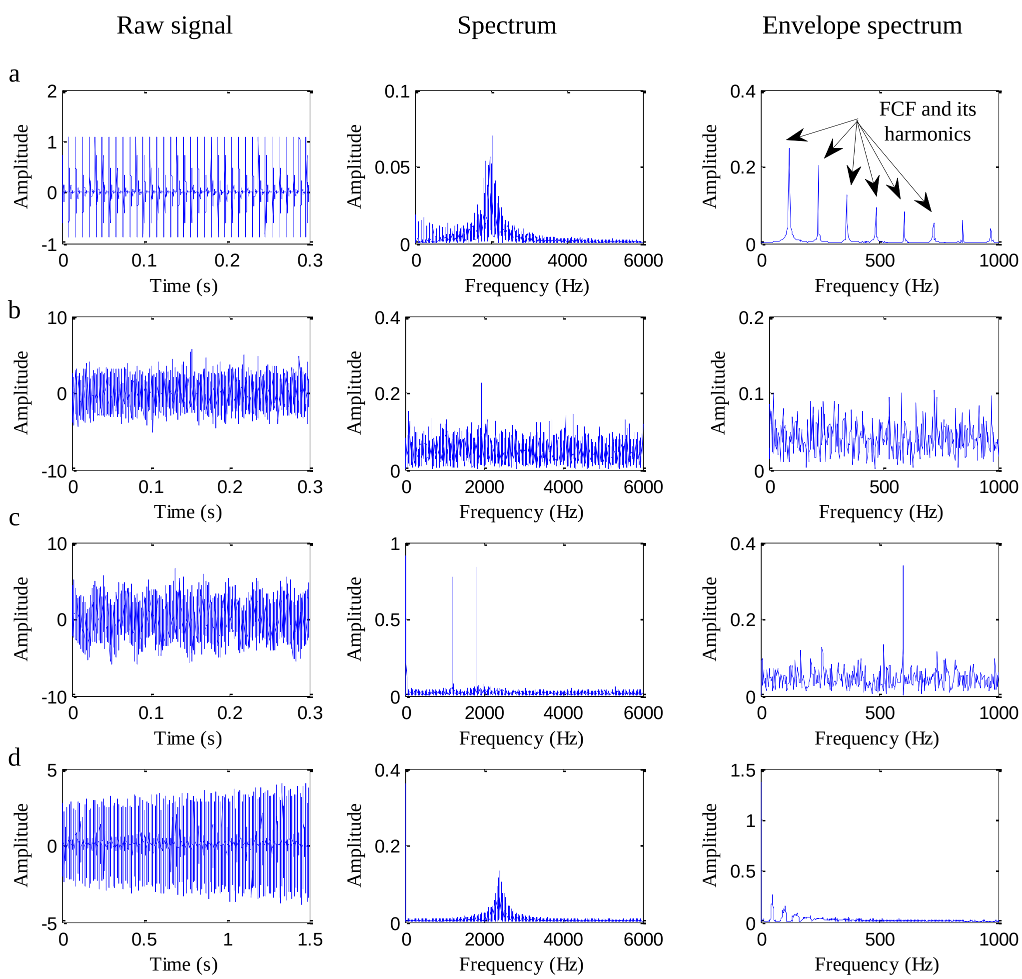


Fig. 1.1 The graphic representations of the four factors listed above: (a) the signal modulation effect, (b) strong background noise, (c) interference vibration, and (d) the nonstationary signal (from left to right for each row: raw signal, frequency spectrum and enveloping spectrum; the resonance frequency : 2 KHz; and the rate of impulse repetition: 125 Hz).

In view of the above, research has been carried out to address the issues caused by these factors. The modulation effect of the fault impulses has inspired the development of envelope extraction methods. Hilbert Transform is most commonly used for this purpose in bearing condition monitoring. However, the Hilbert transform is vulnerable to noise and interferences. To obtain a reliable detecting result, pre-processing like filtering is often required. A typical example of this is the high resonance frequency (HRF) technique (McFadden and Smith, 1984; Tandon and Choudhury, 1999). The main steps of this method include bandpass filtering the vibration signal around the resonant frequency, enveloping the filtered signal and then Fourier transforming the demodulated signal. However, the bandpass filter design is based on the resonant frequency which is often unknown, thus it is a challenging task to select a proper center frequency and bandwidth. Another two widely used enveloping techniques in the literature are multi-scale morphological analysis (MMA) and energy operator (EO). The former extracts impulses by using structure elements at different scales to dilate or erode vibration signals (Zhang et al., 2008). One of the deficiencies of MMA is that it has to empirically select structure elements. Moreover, the MMA and HFR technique may become ineffective in analyzing signals containing strong noise or/and multiple interferences. EO has shown to be capable of extracting signal envelope. Various EO variants have been proposed over these years. However, similar to the MMA and HFR techniques, the EO is vulnerable to multiple interferences and noise. Thus, it is often required to be combined with other methods, such as EMD and filter-based processes, for many applications (Cheng et al., 2007; Rai and Mohanty, 2007; Li et al., 2010; Liang and Bozchalooi, 2010).

To successfully extract envelope in a noisy environment, many existing methods require a denoising step. For instance, as mentioned above, the commonly used HRF technique involves a denoising procedure to bandpass filter the signal. Two major denoising techniques, namely filter-based method and decomposition-based method, have been extensively used to clear vibration signals from defective bearings (Altmann and Mathew, 2001; Lin et al., 2004; Qiu et al., 2006). The success of the former approach largely relied on the proper selection of filter parameters like center frequency and bandwidth, which is often problem-dependent and running condition-sensitive. Likewise, the performance of the latter that uses the characteristic of the wavelet transform to clean signals is influenced by the wavelet parameter determination strategy and mother wavelet selection. Furthermore, such denoising methods are developed for Gaussian white noise reduction and thus incapable of removing interferences. In addition to the above, the self-adaptive noise cancelling (SANC) method and its variant, discrete random separation (DRS) are successfully used to remove noise in bearing condition monitoring (Antoni and Randall, 2004). However, for the success of such methods, parameters such as filter length and delay duration should be selected properly.

Apart from the noise, overcoming the adverse effects of interferences in fault detection is also important as they can result in confusing spectrum and hence misleading detecting results. However, reported research on this topic has been very limited. In a recent study by Zhang and Liang (2013), bandpass and bandstop filtering approaches are jointly used to successfully remove interferences from vibration signals. However, this method requires an iterative auto-correlation module to suppress noise. The other filtering methods may also remove some interferences and its harmonics to some extent but they are of little use when frequencies of interferences reside in the band of resonant frequency. Moreover, linear frequency filtering can no longer be used in the cases where the interaction between interferences and bearing signals has been found to be multiplicative (Randall, 2004).

In addition to the three problems mentioned above, nonstationary vibration signals collected from bearings working under variable speed condition even make bearing diagnosis more challenging. One of the common yet powerful methods for nonstationary signal study is TFA as it exploits both time and frequency information simultaneously. The wavelet transform based TFA has emerged as a powerful tool and led to significant developments in nonstationary signal processing for fault detection (Satish, 1998; Peng and Chu, 2004; Peng et al., 2005). Unlike the traditional Fourier transform (Wald et al.) (Wald et al.) which is blind to time information, wavelet transform (WT) can be used for multi-resolution analysis, providing high time resolution for high-frequency events, and high frequency resolution for low-frequency events. Therefore, the WT is better suited to bearing condition monitoring under variable speed. However, the mother wavelet and wavelet parameters should be selected appropriately, as revealed earlier. Another popular method for TFA is windowed FT, such as short-time Fourier transform (STFT) and Gabor transform (Russell et al., 1998; Cocconcelli et al., 2012). However, the Heisenberg-Gabor inequality makes it impossible to obtain desirable time and frequency resolutions simultaneously in the time-frequency (TF) plane. The bilinear TFR, like the Wigner-Ville distribution (WVD) and Margenau-Hill distribution (MHD), is also commonly used (Li et al., 2006) as they have good concentration in the TF plane. Unfortunately, the problem with the WVD and MHD is the cross-term interference because the support areas of the signal overlap each other. Most of these TFA methods have been used for rotational machine health monitoring. However, a majority of them are customized for gearbox application, where the instantaneous frequency (IF) of the shaft can be easily obtained because the relationship between the IF of shaft and the fault-related TF component on the TF plane is easily observable. This is nevertheless not the case in the context of bearing fault detection because the IF of shaft for bearing cannot be directly attained from the extracted fault-related TF component as the two are coupled by the fault type. Another class of widely used techniques for bearing condition monitoring under variable speed condition involves order tracking analysis. This class of methods is largely dependent on the shaft IF or IFCF extraction. Even though either shaft TF or IFCF can be obtained successfully, it does not provide direct detection result of bearings under unstable speed. This is usually solved by resampling the collected vibration signals to eliminate influence of speed variations (Fyfe and Munck, 1997; Wang et al., 2014), followed by order spectrum analysis. Only then can the fault information of bearings be revealed. However, the overall accuracy of order tracking is affected by many factors, such as filtering, rotational speed, interpolation method and noise (Fyfe and Munck, 1997). For example, the order tracking inevitably incorporates error to the result as the resampling is achieved via polynomial interpolations, while vibration signals are generated by cyclic phenomena and thus sinusoidal rather than polynomial. Besides, the resampling algorithm itself could considerably complicate proposed methods.

Except the above issues related to the *quality* of the detection result, another issue crucial to on-line monitoring is computational *efficiency*. Good quality of detection result alone does not guarantee industrial acceptance if the method is too computationally demanding.

1.2 Motivation and Proposed Study

1.2.1 Motivation

As introduced above, a variety of approaches have been developed to address the four mentioned matters (factors), i.e., modulation effect, background noise, interferences and nonstationarity. The diagnosis quality related *deficiencies* (*problems*) of these existing approaches listed above can be summarized in the following:

1) Modulation effect

The existing enveloping techniques are susceptible to noise and interferences, therefore parameter-dependent pre-treatments are often required.

2) Denoising

Filter-based denoising methods require the determination of center frequency and bandwidth, which is often knowledge-demanding and case environment-sensitive. Even though the filter may work initially it may become ineffective later as the running condition changes. In terms of wavelet-based de-noising, a suitable mother wavelet and shape factor also has to be selected in advance. More importantly, even if filters are well-designed, they may not be effective in handling in-band noise.

3) Interferences removal

Interferences can be removed via prefiltering; however, filter parameter determination is a challenging issue. Though filters with properly selected parameters can eliminate some out-band interferences, the in-band interferences cannot be removed.

4) Nonstationary signals

Since the shaft IF undergoes changes with the time, the classical frequency analysis methods cannot reveal the fault information in this case. Most of the existing TFA methods are effective for *gear* fault detection, but are not suited for bearings. This is because that the fault-related TF components on the TF plane only show the existence of a fault, if any, however the fault type cannot be revealed. Therefore, it is more challenging when using TFA methods for bearing fault diagnosis and the literature in this direction is very limited. Since the fault type cannot be determined, the shaft IF cannot be calculated, making it impossible to do order-tracking analysis as its resampling procedure is based on the shaft IF. Even if the shaft IF can be accurately obtained, the order tracking inevitably propagates errors as the resampling is achieved via polynomial interpolations, while vibration signals are generated by cyclic phenomena which is not polynomial in nature. In addition, the resampling procedure may substantially complicate proposed methods.

According to the above description, these problems can be classified into two types:

a) Bearing fault detection under constant speed condition: the main challenges in this case are to remove noise and interferences, and then demodulate the processed signal to extract fault feature, and

b) Bearing fault detection under variable speed condition: the main tasks in this case are to identify faults as well as discern fault types without a resampling procedure and explore the potential of the TFA methods in beating fault identification.

The Type a) issues can be addressed by developing an enveloping method which can suppress noise and interferences without pre-filtering, or a noise and interference removal approach independent of frequency information. In this way, the parameter-dependent filtering process is not involved and in-band noise and interferences can be removed. The solutions to Type b) issues require a resampling-free technique for bearing fault identification without prefiltering, and proper tailored TFA methods for bearing condition monitoring.

As such, the objectives of this thesis work are to:

1) Develop a parameter-independent enveloping method for fast and direct extraction of bearing fault features from signals without prefiltering,
2) Design a frequency-independent technique for interferences and noise removal regardless of their frequency contents,
3) Devise a resampling-free diagnostic tool for bearing fault detection and fault type identification from nonstationary vibration signals acquired under variable speed, and
4) Propose a novel TFA method and explore its application to bearing fault identification under time-varying speed condition.

The issues, challenges, motivations and proposed approaches for bearing fault detection as discussed above are summarized in Fig. 1.2.

1.2.2 Proposed Study

The research plans and their rationales in accordance with the objectives of this thesis are outlined as follows.

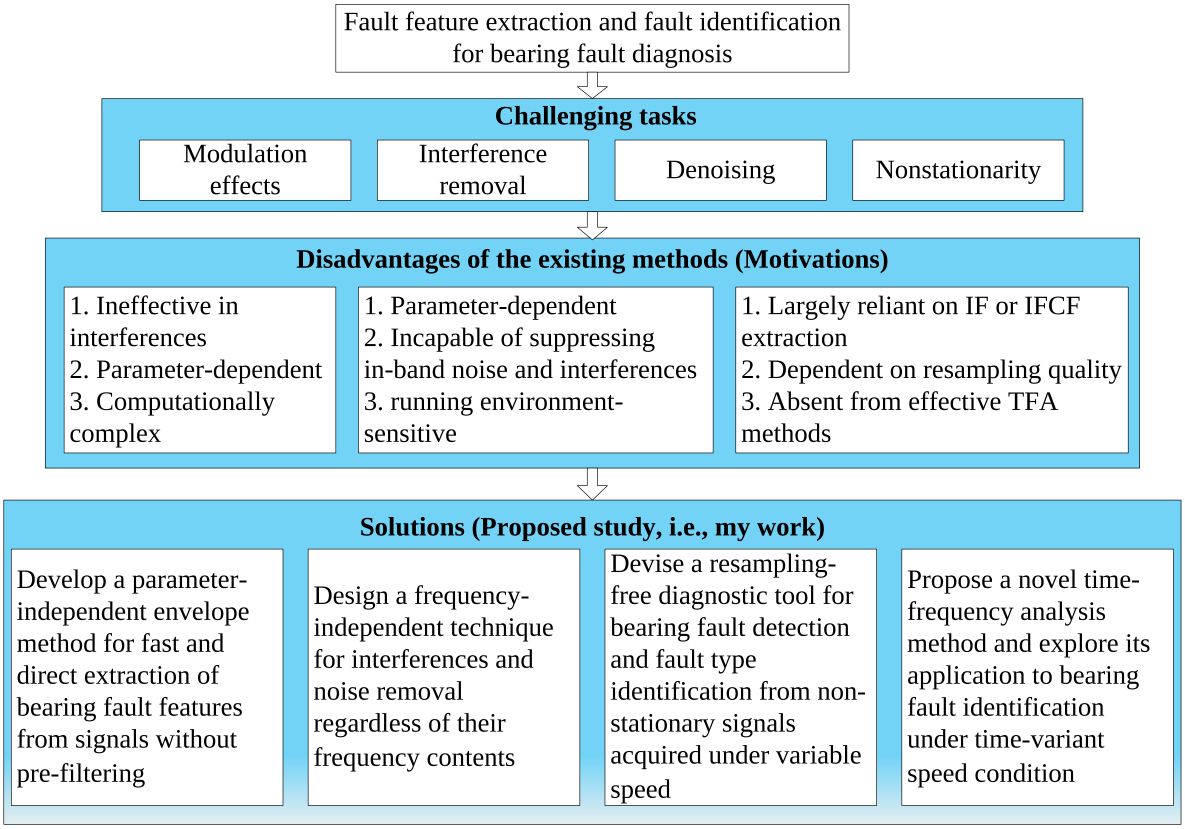


Fig. 1.2 Summary of the issues, challenges, motivations and scope of the thesis.

1) Development of an enveloping method for fast and direct extraction of bearing fault signatures without prefiltering

Aiming at the problem in previous enveloping methods such as their vulnerability to interferences and the difficulty in filter parameter selection, a novel envelope extraction approach is proposed. This demodulation method is based on the signal complexity. Unlike the existing morphological approaches that use structure elements to dilate and erode raw signals (Nikolaou and Antoniadis, 2003; Chen et al., 2013), the proposed method is based on the fractal dimension (FD) concept. The FD is a measure of how ‘complicated’ a self-similar figure is (Hadjileontiadis and Rekanos, 2003). In the context of signal analysis, the FD can assess the signal complexity in the time domain. The complexity of a signal is linked to changes over its time span. Therefore, by means of FD, the sudden changes of signals, such as impacts generated by bearing faults, can be tracked. To detect each impact, the short time fractal dimension (STFD) transform is proposed in this method, and then a STFD sequence can be obtained.

As the cyclic interferences from shaft misalignment and rotor unbalance etc. change more smoothly in comparison with fault-generated transient signals, the FDs of interference components are far weaker than that of fault induced impacts. As a result, the fault-induced impacts are highlighted in a STFD representation and on the contrary interferences are weakened. The envelope of the fault impulses can, therefore, be more easily identified without using any additional filtering process. Hence the difficulty involved in filter parameter selection can be avoided.

Another observation of the proposed method is that a STFD sequence can be easily obtained in time domain via simply mathematical operations. This leads to a very efficient algorithm suited for on-line bearing condition monitoring. To quantify the computational efficiency of the FD-based algorithm, its computational complexity is analyzed in this study.

2) Development of a frequency-independent approach for the reduction of both in-band and out-band interferences and noise based on signal oscillatory behaviors

The above method addresses the issues from the enveloping point of view. The performance of the enveloping method, however, deteriorates with the increase of interference frequencies because the FD values increases with the complexity of interference frequencies. To solve the problem, the second method is proposed from the perspective of interference and noise removal.

The main purpose of this method is to remove interfering vibrations and noise regardless where they are located, in or out of frequency band covered by the resonance frequency. To achieve this, a signal oscillatory-behavior-based method is proposed. Signal oscillating degree is used to describe oscillatory behaviours of signals (Selesnick, 2011a). Signal components that present different degrees of oscillation can be decomposed using morphological component analysis (MCA) as long as proper bases with different oscillatory behaviors can be found for the representation of each component. As such, this approach is no longer based on the frequency differences among signal components and can separate the fault-induced transients from resonant parts and other signal components. Hence, the proposed method is free from filters and can remove interferences and noise of any frequencies in the frequency band of interest.

Each time a rolling element moves over a damaged area, an impulse is generated, in the form of a very short burst (presenting no oscillations) and several oscillations whose amplitudes decay exponentially. The harmonic interferences (manifested as continuous oscillations) often feature sustained oscillations. In this way, the bursts generated by bearing faults and the interferences are classified into two distinct signal component groups, i.e., low-oscillation component (with no oscillations) and high-oscillation component (with sustained oscillations), based on the differences of oscillatory behaviors. The objective function can be constructed based on the MCA and these signal components presenting different levels of oscillation can then be obtained by solving the objective function.

3) Development of a resampling-free technique for bearing fault identification under time-varying speed condition

The previous methods are devised for constant speed. However, bearings often work under time-varying rotational speed condition in practice. Vibration signals collected in these cases are nonstationary, where the traditional frequency analysis methods are no longer effective. New methods for bearing condition monitoring under variable speed are therefore required. Considering the drawbacks of the existing methods for detecting bearing faults under variable speed conditions, this thesis proposes a new method free from prefiltering and resampling for bearing fault identification under time-varying speed conditions.

This method is developed via the joint application of the STFD and GD. The STFD is proposed to extract the envelope from the raw signal, paving the way to obtain a clear TFR. The IFCF of the defective bearing is then estimated via a peak search algorithm based on the TFR. With the IFCF (demodulator), the GD can map the IF trajectory of a TF component into a linear path parallel to the time axis and at the same time the energy concentration level of the TF component is improved. From the frequency spectrum of the generalized demodulation signal, it can be seen that the frequency line where the linear path parallel to time axis is located dominates the frequency spectrum. Different demodulators proportional to the extracted IFCF are adopted to iteratively demodulate the signal, followed by analyzing frequency spectrum and searching the amplitude of the constant frequency where the linear path is situated. Assembling these pairs of frequency and amplitude, a new spectrum named resampling-free order spectrum is obtained. The detection decision can be made based on the resulting spectrum.

4) Proposal of a novel TFA method and its application to bearing fault identification under *unknown* rotational speed condition

The TFA provides time and frequency information simultaneously; it is, therefore, a powerful tool to characterize the vibration signals of bearings working in variable speed conditions. Reported studies in this direction are, however, still inadequate.

As such, this dissertation proposes a new TFA method named dual demodulation transform (DDT) and applies it to bearing fault diagnosis. It is termed DDT as it contains two demodulations, i.e., one forward and the other backward. The forward demodulation can transform the curved IF track into a line parallel to time axis, and hence the spectrum of the demodulated signal is more concentrated than the original signal. However, the real IF is not reflected by the spectrum of forward demodulated signal. The backward demodulation is therefore implemented to recover the real frequency of each time instant. Via such demodulations, the actual frequency at each time instant can be restored and the energy concentration level of the TFR around the time instant is improved. Thus, the TFR of the entire signal can be correspondingly enhanced. Based on the DDT, the iterative DDT (IDDT) is developed to handle nonstationary multicomponent signals where each signal component is subject to different modulating sources. An iterative procedure is involved in the IDDT to decompose the signal into monocomponents of constant frequency, thereby satisfying the monocomponent requirement of the DDT. With such an improvement, the proposed method can be extended to analyze nonstationary multicomponent signals, and the derived TFR has satisfactory energy concentration level and is free from cross-term. Besides, as the IF estimation has direct effects on demodulation results, a step for estimated IF refinement is employed to improve the IF accuracy. The proposed method is then applied to bearing condition monitoring, which at the same time validates the effectiveness of the proposed TFA method.

1.3 Organization of the thesis

The thesis is organized as follows. A literature review on bearing fault diagnosis methods is given in Chapter 2. This chapter is organized based on the matters revealed previously, thus presenting a review covering previous work of demodulation, denoising, interferences removal and nonstationary signal process for bearing condition monitoring. Chapters 3 and 4 present the proposed methods for bearing fault feature extraction under constant speed conditions. More specifically, Chapter 3 details the FD-based enveloping method and its applications to fault feature extraction from both simulated and experimental data. In Chapter 4, the proposed oscillatory-behavior-based method is first presented and then validated by simulation and experiments. Chapter 5 proposes a resampling-free diagnosis strategy for bearing fault identification under time-varying speed conditions. A novel TFA technique and its application to bearing condition monitoring are reported in detail in Chapter 6. Conclusions and recommended future work are presented in Chapter 7.

Chapter 2 Literature Review

This chapter presents a review of previous work that has been done on the four factors (matters) mentioned earlier, namely the analyses of the modulation effect, noise deduction, interference removal and time-varying signals in the context of bearing condition monitoring.

2.1 Envelope demodulation methods

As mentioned earlier, a faulty bearing, when in operation, modulates the carrying frequency (i.e., the fault-excited resonant frequency). Therefore, demodulation analysis has been established for extracting fault feature, in which envelope demodulation, also named amplitude demodulation, has been used widely (Yang et al., 2007).

Ho and Randall (2000) utilized the Hilbert enveloping method with SANC to extract amplitude of the modulation. A recent study on bearing failure detection in wind turbines also used the Hilbert transform (Amirat et al., 2011). The Hilbert transform in combination with other approaches such as empirical mode decomposition (EMD) and WT have been extensively applied (Peter et al., 2001; Yu et al., 2005; Zhang and Ai, 2008; Georgoulas et al., 2013). For example, the EMD with the Hilbert transform can give a full energy-frequency-time distribution of a signal. Such a method is designated as the Hilbert-Huang transform (HHT) which is effective for nonstationary time series analysis (Huang et al., 1998). These methods involve signal pre-processing followed by Hilbert transform. This means that the Hilbert transform should be complemented by other methods, which indicates that using Hilbert enveloping alone is inadequate because it can be easily affected by noise and interfering signals.

Recently, the energy operator (EO) has shown to have a great potential for envelope extraction. Liang and Bozchalooi (2010) exploit a parameter-free approach based on the EO. A study proposed by Tang et al. (2011) employs energy operator demodulation together with wavelet packet transform (WPT) for bearing condition monitoring. Their study also compares the Hilbert demodulation and energy operator demodulation, showing that the envelope obtained by EO demodulation has less end effect compared with the Hilbert transform. However, a high-pass filter is required by this method before energy operator demodulation. The Teager energy operator (TEO) is also favourably used for bearing fault detection as it has a good adaptability to the instantaneous changes in signals and an excellent ability to resolve transient events. Liu et al. (2013) develop an approach to bearing fault diagnosis based on the combination of TEO and Elman neural network. In their work, the TEO is utilized to extract the cyclical impact caused by bearing failure but the WPT is required to reduce the noise of the Teager energy signal. The joint EO and EMD, namely Teager-Huang Transform, are presented in several studies (Cheng et al., 2007; Li and Zheng, 2008; Li et al., 2009). Similar to the Hilbert envelope demodulation, in most of related research, the EO technique pins its success to pre-processing.

Another alternative enveloping method is morphological analysis (MA), a signal processing method that extracts the local morphological features of a signal by intersecting it with a structure element. With the MA, a complex signal can be separated from the background noise, and decomposed into various components while reserving morphological features of the signal by basic morphological operators, including dilation, erosion, opening and closing. Nikolaou and Antoniadis (2003) examined the effect of the noise on the length of the structuring element and on the shape of the produced envelope, concluding that a structure element length of 0.6-0.7 times the pulse repetition period can minimize the noise effects. However, this method requires the initial impulse repetition period of the original signal and it fixes the structure element (SE) in the single-scale morphology analysis. To overcome the disadvantages, multi-scale morphology analysis (MMA) has been proposed for bearing failure detection (Zhang et al., 2008). The definition of the length and height scales of structure element is introduced and an adaptive algorithm determining SE at different scales is also presented. Although this method can enhance the SNR, a triangular SE does not well match the impulsive feature in many applications. In order to construct the SE that is similar to impulses in shape, an impulsive attenuation function is applied to create the SE (Wang et al., 2009). A new criterion is put forward to optimize the structure element. Nevertheless, they did not clarify how to determine the search range for the maximum amplitude, decay coefficient and natural frequency which are important to the success of this method. An integrated MMA and support vector machine (SVM) approach is suggested by Hao et al. (2011). Though this method is effective in distinguishing different fault features, the SVM part in the method requires a set of training data for fault pattern recognition. Such data are not always available particularly in an environment where the operating condition varies frequently. Moreover, the SVM is trained by a specific bearing under a particular environment, which severely limits its application scope. More recently, a morphogram-based method is proposed by (Wang et al., 2012). An index called construct index is developed for this method, which is used to automatically determine the optimal length range of the SE based on the most likely fault interval observed by the morphogram. However, this paper did not examine noise and interference handling capability of the method. In addition, the signals analyzed in this study are quite clean.

In summary, there are two issues associated with the previous enveloping methods. First, the effectiveness of most of such methods is dependent on the quality of the pre-treated data, like Hilbert demodulation which requires band-pass filtering. Next, some methods, like MA, are dependent on parameter selection.

2.2 Denoising techniques for bearing fault detection

As reviewed above, noise and interferences have profound effects on various envelope demodulation techniques. Vibration signals are often heavily clouded by strong noise from a complex mechanical system where bearings are installed. Consequently, a variety of noise reduction methods have been developed to purify vibration signals. These denoising methods can in general be grouped into two classes, i.e., the filter-based and decomposition-based. A review of the two types of denoising methods is given in the following.

2.2.1 Filter-based denoising methods

In this subsection, two commonly used filters, namely classical filters including low-pass, high-pass, band-pass, band-stop filters and wavelet filters will be reviewed.

1) Classical filter-based noise reduction

The key point of classical filter-based denoising methods is to find the proper center frequency and bandwidth so as to keep the impulse-like signal component as much as possible but discard other noise- and interference-related signal components. However, this often requires prior knowledge of the system and signals to be studied, which in reality is unknown beforehand. Even if the desirable parameters can be selected initially, they may work ineffectively afterwards. This is because the working condition may change, e.g., in a wind turbine. To address this issue of parameter selection, many methods have been put forward. Recently, a method called spectral kurtosis (SK) proposed by Antoni (2006) has found to be very useful for characterizing nonstationary signals. Right after this, a concept of kurtogram which displays the SK as a function of the frequency and spectral resolution is proposed (Antoni and Randall, 2006). The maximum of the kurtogram provides the optimal parameters with which a band-pass filter can be designed. The SK and kurtogram ideas have spawned a large number of studies. For example, Combet and Gelman (2009) proposed a methodology for gearbox fault diagnosis by applying the optimal denoising filter based on the SK. A fast kurtogram has been used to extend the application of SK method by a study of Antoni (2007). However, as pointed out by (Wang and Liang, 2011), this method may not be practical because it is unrealistic to examine all window lengths used in the SK technique to search the optimal center frequency and bandwidth of filters. Moreover, the kurtogram tiling pattern is not adaptive to the signal under analysis. In addition, as stated by Sawalhi et al. (2007), a high value of SK requires no overlap between individual transients. It in turn means that their damping must be sufficiently high such that each impulse dies out before the coming of the next particularly when the impulse repetition frequency is high. Aiming at this issue, an algorithm is developed using the minimum entropy de-convolution approach (Sawalhi et al., 2007). The minimum entropy de-convolution efficiently eases the effect of the transmission path and clarifies the impulse-like signals. A study of Wang et al. (2013) replaced kurtosis of the temporal signals with that of power spectrum of the envelope of the signals extracted from wavelet packet nodes at various scales.

2) Wavelet filter-based denoising

The above is all about the classical filter-based denoising. The other filter-based denoising is based on WT, which is detailed in the following. Over the past two decades, the wavelet transform has emerged as an effective signal processing tool. The concept of wavelet was first put forward by Grossmann and Morlet (1984). The WT is defined as the inner product of a signal and a family of wavelets, which are created by scaling and translating the mother wavelet function in the following way

$$\omega_{a,b}(t) = \frac{1}{\sqrt{a}}\,\omega\left(\frac{t-b}{a}\right) \tag{2.1}$$

where $\omega(t)$ is the mother wavelet function with a zero average $\int_{-\infty}^{\infty} \omega(t)dt = 0$, the parameters a and b are scale factor and translation factor respectively. Morlet (1984) expressed the continuous wavelet transform (CWT) of a signal $x(t)$ at time b and scale a as

$$W_x(a,b;\omega) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{a}}\omega^*(\frac{t-b}{a})dt \tag{2.2}$$

where $\omega^*(t)$ is the complex conjugate of $\omega(t)$.

However, the CWT is computation-demanding which limits its applications in reality. To improve computational efficiency and facilitate inversion, the discrete wavelet transform (DWT) has been developed. The idea is to replace the CWT by a transform with discrete scale factor and translation factor. The discretized two factors are as follows

$$a = a_0^m,\ \ b = na_0^m b_0 \tag{2.3}$$

where m and n are integers. Then the wavelet function accordingly becomes discrete, i.e.,

$$\omega_{m,n}(t) = a_0^{-m/2}\omega(a_0^{-m}t - nb_0) \tag{2.4}$$

Based on the discrete wavelet function, the DWT was defined as

$$W_x(m,n;\omega) = a_0^{-m/2}\int x(t)\omega^*(a_0^{-m}t - nb_0)dt. \tag{2.5}$$

The WT can be used for multi-resolution analysis of a signal by means of dilation and translation, and thus it can extract time-frequency features effectively.

The WT can be used as a filter due to an important property of the FT, i.e., the convolution in one domain corresponding to multiplication in the other domain. Therefore, Eq.(2.2) can take the following alternative form:

$$W(a,b) = \sqrt{a}\mathrm{F}^{-1}\{X(f)\omega^*(af)\} \tag{2.6}$$

where $X(f)$ and $\omega(f)$ are the FTs of $x(t)$ and $\omega(t)$, respectively, and F^{-1} stands for the inverse FT. Eq.(2.6) shows that the convolution process in the WT is a filtering operation if the daughter wavelet is treated as a filter kernel. Hence the WT can also be considered as a special filtering operation. The frequency response of the wavelet filter varies as the scale factor and shape factor change. In this way, low-pass, high-pass, band-pass or multi-band pass filters can be obtained.

Signals filtered by wavelet filters can be reconstructed by wavelet coefficients at selected scales. The process is equivalent to that of classical filters.

Based on the property explained above, wavelet filter based methods have been carried out for weak signature enhancement. For example, this method has been tailored to detect weak signature from defective bearings (Qiu et al., 2003). A study of Lin and Zuo (2003) utilizes Morlet wavelet function to construct wavelet filter, where the parameters in the Morlet wavelet are optimised based on the kurtosis maximisation principle. This method searches the optimal wavelet filter that yields largest kurtosis but it does not address the periodicity of the signal which is important in bearing fault detection. A technique that can detect signal periodicity is then proposed based on the singularity value decomposition (SVD) (Qiu et al., 2006). However, the performance of this de-noising technique is greatly affected by the wavelet parameter selection strategy. Though a recursive scheme could be used to select wavelet parameters, it leads to taxing computation. Another example of wavelet filter application is the work of Bozchalooi and Liang (2008) in which they propose a more sophisticated algorithm for bearing fault detection based on the spectrum subtraction and wavelet filtering. Nevertheless, this method requires properly selected parameters. Luo et al. (2013) proposed a kurtosis method to guide the selection of band-pass filters from those obtained by the tunable-Q wavelet transform (TQWT). This method has its own drawback, i.e., the possible loss of a large portion of useful signal contents.

2.2.2 Decomposition-based denoising methods

Apart from denoising by filters, decomposition-based denoising is also widely used for this purpose. This subsection reviews wavelet decomposition-based denoising, one of the most widely used denoising methods.

Let the signal be $s(n)$, $n \in 1, 2 \ldots$, N, where N is the signal length. Suppose the signal has been corrupted by addictive noise and the signal-noise mixture is

$$x(t) = s(t) + s_n,\ t = 1, 2..., N \tag{2.7}$$

where s_n is the standard Gaussian white noise. The goal is to remove the noise and to obtain an estimate $\hat{s}(n)$ of $s(n)$ which minimize the mean squared errors

$$\mathrm{MSE}\left(\hat{s}\right)=\frac{1}{N^2}\sum_{i=0}^{N}\left(\hat{s}_i-s_i\right)^2 . \tag{2.8}$$

The theoretical foundation of wavelet decomposition de-noising is the principle of multi-resolution analysis (Mallat, 1989; Donoho and Johnstone, 1995). As stated above, the WT is the inner product of a signal and a family of wavelets. In other words, wavelet coefficients represent the degree of correlation between the signal and the wavelet family. Based on this explanation, it is expected to see larger wavelet coefficients in a wavelet window where the signal is strongly correlated with the daughter wavelet. By multi-level wavelet decomposition, the detailed coefficients and approximation coefficients can be obtained. If larger wavelet coefficients are selectively reserved and small ones are discarded, the noise can be reduced.

In general, wavelet decomposition-based de-noising is implemented in three steps: signal decomposition, coefficient thresholding and signal reconstruction. Among them, coefficient thresholding is the most challenging task.

There are two thresholding strategies frequently used in the literature. One is soft-thresholding which either removes the input wavelet coefficient if it is below the threshold or shrinks it by T if the input is greater than T. The other is hard-thresholding, which removes the input coefficient if it is below the threshold or simply keep it if it is greater than T. (Donoho, 1995; Chang et al., 2000). Each of the two strategies has its own drawback. As the soft-thresholding strategy preserves only the portion that is above the threshold. This will lead to serious signal distortion for a very noisy signal and part of the fault related signal content can be removed. On the other hand, hard-thresholding may treat noise as fault related impulse for wavelets above the threshold. In addition, both strategies tend to over-truncate the wavelet coefficients that are below the threshold and hence signal components with weak impulsiveness can be easily excluded.

The hard-thresholding method have been used to de-noise the dynamic transient signals (Chen et al., 1999; Wang et al., 1999). Lin and Qu (2000) exploit an advanced version of the soft-thresholding de-noising method to extract feature for mechanical vibration signals. Selecting the wavelet coefficients with the aid of a statistics-based criterion is proposed for vibration signal analysis (Yen and Lin, 1999). Yang and Liao (2001) apply the wavelet de-noising method to a power quality monitoring system, in which the threshold is adjusted adaptively in accordance with background noise. The flexibility of the WPT and the systematic parameter selection criterion is investigated by Nikolaou and Antoniadis (2002b) with the assistance of hard-thresholding function. Lin et al. (2004) develop a new method for wavelet threshold de-noising. They employ the Morlet wavelet as the base wavelet to match the impulse and use the maximum likelihood estimation for threshold. It is reported that this method performed very well in noise reduction. The shortcoming is that it requires prior information. Hong and Liang (2007) proposed a kurtosis-based hybrid thresholding method to de-noise mechanical signals. In their method, the threshold is on-line dynamically optimized based on a "cost" function incorporating both the mean squared error and false identification criteria. Along this line, Li et al. (2008) suggests customizing wavelet de-noising using intra- and inter-scale dependency of wavelet coefficients for bearing fault detection. The results of the application to bearing condition monitoring seem to be effective. However, it is questionable whether such threshold optimization or customization approach is suited to on-line applications due to amount of computations required for the optimization or customization process. In addition, such methods are developed for Gaussian white noise reduction, hence of little use for removing discrete interferences.

To conclude, the effectiveness of the traditional filter-based denoising methods is subject to the proper parameter selection of filters. Even if the "optimum" parameters can be obtained using methods such as the SK-based approaches, the traditional filter-based denoising methods are ineffective in removing in-band interferences and noise. For wavelet filter-based approaches, most of them are also parameter-dependent and parameter selection is usually more complex. In relation to wavelet decomposition-based de-noising, it requires sophisticated thresholding strategies to maintain the quality of the thresholding results. However, such "sophisticated" strategies are often computation-demanding and hence may not be practical for on-line applications. Another disadvantage of the wavelet decomposition-based denoising methods is that their performance deteriorates when applied for impulsive signal processing as the wavelet coefficients are not sparse (Qiu et al., 2006), though they are effective for smooth signals.

2.3 Interference removal for bearing fault detection

Interference removal can be achieved to a certain extent by some of the aforementioned filter-based de-noising methodologies. For example, the SK technique is used to choose the frequency band with a higher kurtosis value, which can eliminate out-band cyclic interferences as well. However, when the interfering signal becomes stronger, the relative contribution of the impulses to the kurtosis value will decrease. Therefore, it is difficult to distinguish interferences and fault impulses by kurtosis value. As a result, it becomes more challenging to determine the optimal frequency band in the presence of cyclic interferences. The wavelet decomposition-based methods are effective in removing noise because the energy of noise is scattered throughout the all wavelet scales with small coefficients. However, they are ineffective in eliminating cyclic interferences because the energy of each interference is concentrated within a single or a very few wavelet scales, leading to large wavelet coefficients (Qiu et al., 2006; Zhang et al., 2013).

Though the interference issue has been one of the main obstacles for reliable bearing fault detection, research dedicated to this issue has been limited. The study that explicitly addresses this issue is reported by Zhang et al. (2013). In their work, a kurtosis-based adaptive band-stop filtering approach is developed for interference removal. Their simulation and experimental results demonstrate that this method can effectively remove discrete interferences and can handle signals with low SNR and SIR. However, this method requires an additional de-noising method. Furthermore, the bandwidth of band-stop filters in this method is dependent on the threshold. The selection of the threshold value is based on the noise level of the raw signal, which in turn means that the prior knowledge of the raw signal is required to determine an effective threshold.

The review of bearing fault detection under constant speed is presented above. Much work has been done to enrich the literature in this area. To conclude, the objective is to extract fault feature by demodulation and the main problems (deficiencies) of the existing methods for this purpose are that 1) most existing demodulation methods requires prefiltering which is parameter-dependent, and 2) they are of little use when removing in-band interferences and noise. To address these issues, we propose to demodulate signals without the involvement of prefiltering or remove interferences and noise by a method independent of frequency information, to relieve the difficulty of filter parameter specification and enable in-band interferences and noise to be removed. Based on the two different viewpoints, methods are developed in the thesis. They solve the main problems from two different ways, and at the same time each one has its own advantages and complements the other.

2.4 Bearing fault detection under time-varying speed condition

Most of the above reviewed methods can detect faults for bearings working under constant rotational speed. In reality, bearings, however, often serve in variable speed applications, such as wind turbine, mining equipment, fans, pumps and ordinary machines during speed-up and ramp-down processes. When the rotating speed is variable, such speed fluctuation may cause "smearing" of the discrete frequencies in the frequency representation, resulting in these frequencies no longer observable as discrete frequency lines (Urbanek et al., 2011). Therefore, the signal processing approaches that are developed for signal analysis of bearings at a constant speed become ineffective in variable speed cases.

Recently, fault diagnosis of bearings under variable speed has become a new challenge and attracted researchers' attention. In a study by Teotrakool et al. (2009), bearing fault detection in adjustable-speed drives is addressed with the aid of motor current signature analysis and wavelet packet decomposition. A novel method based on angular measurement of true instantaneous angular speed is reported in (Renaudin et al., 2010). The main contributions of this paper lie in the demonstration that localized bearing faults create small angular speed fluctuations which can be measured by optical or magnetic encoders. It is found that the insensitivity to speed fluctuation can be ensured by measuring instantaneous angular speed with the pulse timing method. Gong and Qiao (2013) studied the bearing fault diagnosis for direct-drive wind turbine operating at unstable rotating speed conditions by exploring the joint application of appropriate current frequency, amplitude demodulation algorithm and a 1P-invariant power spectrum density algorithm. Yang et al. (2013) developed a method based on the variable predictive model based class discriminate (VPMCD), order tracking technique and local mean decomposition (LMD) for analyzing bearing vibration signal measured under variable speed condition. It has been reported that the proposed method can monitor roller bearing conditions and distinguish fault types accurately. An approach named reversed sequence squared envelope spectrum (RS-SES) is described in (Borghesani et al., 2013), which is exploited to address the inherent problem of the prevailing traditional squared envelope spectrum (T-SES) for real applications. This approach consists of band-pass filtering, squaring enveloping, resampling and discrete FT. The rotating speed in this paper is picked up by a tachometer. Zhao et al. (2013) suggests a tacholess envelope order method for bearing fault detection under unstable speed, where shaft IF is picked up from the spectrum obtained by adaptive short time Fourier transform (ASTFT) and then the raw signal is resampled. More recently, a novel technique for bearing fault detection under time-varying condition is proposed by Wang et al (Wang et al., 2014). They proposed an amplitude-based peak search algorithm for STFT to extract the IFCF and then resampled the nonstationary signal in accordance with the extracted IFCF rather than shaft IF. In this way, the non-stationary signal in time domain is transferred into stationary signal in the angle domain, thereby the classical technique like Fourier transform (Wald et al., 2010) becoming applicable. The final diagnosis decision can be made based on the fault characteristic order spectrum. The simulating and experimental validations in this paper show that the proposed method can reveal the existence and pattern of faults.

The aforementioned work has fertilized the literature on bearing fault diagnosis under non-stationary working conditions. Meanwhile, it can be seen that order tracking, which can remove the effect of speed variations and convert the vibration signal into a stationary one in angular domain (Fyfe and Munck, 1997; Li et al., 2009; Luo et al., 2012; Borghesani et al., 2014), is the key step for these methodologies. To use order tracking analysis, the vibration signal must be resampled at a rate proportional to the shaft rotational speed, which in turn means that the shaft rotational speed have to be known beforehand. The straightforward method of obtaining the shaft speed is to use tachometers, encoders etc. As reported in (Borghesani et al., 2013), with the aid of tachometers, the vibration signal is resampled and the effect of small speed variation is hence eliminated, making the squared envelope spectrum applicable. However, speed measuring devices are not always allowed to be installed in every case due to the design reason and cost concerns. Consequently, it is highly desirable to extract instantaneous rotational speed information from vibration signals. Combet and Zimroz are among the first who originally develop the method for this purpose (Combet and Zimroz, 2009). Recently Urbanek et al. (Urbanek et al., 2013) advanced this idea by proposing a two-step procedure to predict instantaneous speed with large fluctuations based on phase demodulation and TFR. However, the estimated IF information may not always be accurately extracted from the original signal. Even if the IF information can be properly extracted directly from the vibration signal, the resampling procedure itself complicates the detection technique. Moreover, order tracking inevitably introduces error to the result as the resampling is achieved via polynomial interpolations, while vibration signals are generated by cyclic phenomena but sinusoid is not polynomial in nature.

Therefore, an approach that directly extracts fault features from the non-stationary signal without resampling will be better suited to real applications due to the simplified process and improved computational efficiency, i.e., a technique without resampling process is preferred. However, such an approach has not yet been reported in the accessible literature.

It is well-known that time-frequency analysis has been widely used in analyzing nonstationary signal as it helps to gain insights into the complex structure of a signal containing multiple components (Qian and Chen, 1999; Sejdić et al., 2009). Various TFA methods have been investigated in the literature. The most common one is the linear TFR (Hlawatsch and Boudreaux-Bartels, 1992) that is characterized by the inner product with a dictionary of TF atoms, such as the STFT and WT. However, the effectiveness of this class of TFA methods is often undermined by the ill-selected dictionary and resolution paradox reflected by the Heisenberg uncertainty principle. More specifically, the performance of such methods is highly dependent on the selected basis, and the accurate time localization is compromised by low frequency resolution and vice versa. Another category of TFA methods is the quadratic (bilinear) transform, which includes the well-known Wigner-Ville distribution (WVD) and its variants (Bertrand and Bertrand, 1992). Though such methods have desirable theoretical characteristics, the interference terms in the TF plane have made it very challenging to apply in real applications (Gonçalves and Baraniuk, 1998).

The effectiveness of the TFA methodologies lies in their capability of concentrating the energy of TFR at or around the true IF in the TF plane. Targeting at this, the synchrosqueezing wavelet transform is proposed in (Daubechies et al., 2011). This technique eliminates the ambiguity in frequency (scale) dimension via squeezing the TFR along the frequency (scale) axis, thereby acquiring a better frequency resolution. However, the synchrosqueezing transform (ST) is restricted to frequency (scale) variables of the continuous wavelet transform (CWT) (Daubechies and Maes, 1996; Li and Liang, 2012a). There exists a time offset corresponding to time variable in addition to frequency (scale), which could also cause diffusion in time dimension and cannot be ignored when dealing with noisy and amplitude-frequency modulated (AM-FM) signals. To resolve the bi-dimensional smear problem, Li and Liang (Li and Liang, 2012a; Li and Liang, 2012b) proposed the generalized ST based on the GD which paves the way towards achieving a much more concentrated TFR by a combination of multiple TF plane manipulations in one operation. This technique significantly improves the energy concentration in TF plane. However, many signals often consist of nonstationary multiple components and each of the components can be subjected to different modulating sources. Thus, one time application of the GD cannot demodulate all components and map their IFs into constant frequencies. As a result, the TF blur still occurs for such multicomponent signals. To overcome the problem of the GD in handling multicomponent signals, an iterative generalized synchrosqueezing transform is developed by Feng et al (Feng et al., 2015) to decompose a complex multi-component signal into monocomponent ones and then transform all constituent components with time-variant frequencies into constant ones. In doing so, the requirement of monocomponent as well as the constant frequency required by ST can be satisfied. However, this method has a drawback of strong dependence of the initial IF estimation. The inaccurate IF estimation would lead to confusing or even wrong results. Moreover, this method is tailored for planetary gearbox fault diagnosis under time-varying speed. The capability of such TFA methods for bearing fault diagnosis under variable speed has not been adequately explored.

In summary, the problems of available methods in literature for bearing fault identification under time-varying speed condition are listed as follows: 1) the shaft rotating IF is unavailable as the fault type is unknown, 2) resampling and/or demodulation procedure is significantly affected by the accuracy of estimated IFCF, and 3) even if the IFCF is accurately predicted, the resampling procedure could be error-propagating and complicate the process substantially. In addition, the TFA that is supposed to be a powerful tool in analyzing nonstationary signal has quite limited applications to bearing condition monitoring under unstable condition. Therefore, this thesis proposes resampling-free techniques to address these issues and also explores the TFA application to bearing fault identification. The satisfactory IFCF accuracy of the first proposed technique is realized by a clear TFR which is obtained based on an effective enveloping method and the resampling-free strategy is implemented via iteratively generalized demodulating the envelope signal. The second proposed method aims to develop a novel TFA method to improve the readability of TFR and explore the TFA application to bearing fault detection.

Part I Bearing Fault Feature Extraction under Constant Speed Condition

Chapter 3 A Fractal-Dimension-based Envelope Demodulation for Bearing Fault Feature Extraction without Prefiltering[1]

In this chapter, a fast envelope detection technique based on FD is proposed for bearing fault feature extraction without pre-treatment. To allow for the easier detection of bearing fault feature, the envelope demodulation has been used together with spectrum analysis. Although various methods have been developed for envelope extraction in bearing fault detection, the most common practice is to filter the raw signal at the bearing resonant frequency located in the high frequency range and then envelope demodulate the filtered signal. However, the main problem is that the filtering procedure in these envelope methods requires pre-specification of filter parameters which is often time-consuming and knowledge-demanding. Even if the filter initially performs properly, it may become ineffective with the changes of running condition. Thus, an enveloping method which can suppress interferences and noise without prefiltering is preferred. However, this type of methods has not been adequately studied and not been reported in the literature pertinent to bearing fault detection. Motivated by the above, an envelope demodulation method which can detect the envelope of impulsive signal and suppress interferences and noise is proposed in this chapter.

The proposed envelope demodulation is termed short time fractal dimension (STFD) transform for fault feature extraction from vibration signal mixture including the fault feature (i.e., fault-induced impulses), periodic interferences from other mechanical/electrical components, and background noise. FD can be used to measure a signal complexity. In order to extract the envelope as well as track the position of each impulse, the STFD transform is proposed. The issues related to STFD transform calculation are first addressed in this chapter. Then, by STFD, the original signal can be quickly transformed into a STFD representation, where the envelope of fault-induced impulses becomes more pronounced whereas interferences are partly weakened due to their morphological appearance differences. It has been found that the lower the interference frequency, the less effect the interference has on STFD representation. When interference frequency complexity keeps increasing, more effect on STFD representation

---

[1] Published in Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, doi:10.1177/0954406215608894.

will be resulted. Such effect can be reduced by the proposed kurtosis-based peak search algorithm (KPSA). Therefore, bearing fault signature is kept and interferences are further weakened in the STFD-KPSA representation. The proposed method has been favourably compared with two widely used enveloping methods, i.e. multi-morphological analysis (MMA) and energy operator (EO) in extracting impulse envelopes from vibration signals obscured by multiple interferences. Its performance has also been examined using both simulated and experimental data. In addition, the proposed approach is highly efficient, making it a good candidate for on-line applications.

3.1 A Fractal-Dimension-based enveloping method for bearing fault diagnosis

FD is a useful concept in describing natural objects. There are a variety of closely related notions of FD, such as the Ausdorff dimension, the packing dimension, the information dimension, the box counting dimension and the correlation dimension (Moon, 1992). These notions are based on the theoretical perspective. In the fractal geometry, the FD is a statistical quantity that gives an indication of how complicated a self-similar figure is. From this point of view, the FD can be considered as a measure of the number of basic building blocks that form a pattern (Falconer, 2013). As a result, the FD can reflect the signal complexity in the time domain. This structural complexity could vary with sudden occurrence of transient events, such as bursts in a signal. Therefore, a signal with or without bursts can be reflected by FD.

3.1.1 FD estimation

Several techniques have been reported to approximate the FD of a waveform in the time domain. These include the Katz, Sevik, and Higuchi methods, as well as the box-counting method and its variants (Higuchi, 1988; Katz, 1988; Tricot, 1995; Sevcik, 2010). In this study, the Katz method is chosen for computing the FD for two reasons. First of all, it is sensitive to the morphological changes of a waveform. As morphological appearance changes, the FD changes accordingly (Raghavendra and Dutt, 2010). Therefore, the FD is effective in detecting sudden changes in signal strength. Next, for on-line bearing condition monitoring, a fast method is required for the FD computation. The Katz method, which is based on an operation directly applied to the signal and not to any state space (Hadjileontiadis, 2005), provides a fast computational implementation of FD. This means that the FD estimation no longer needs the signal to be embedded into higher dimensional space, which makes it computationally efficient.

The Katz FD is described in the following. Consider a signal $s = \{s_1, s_2, \ldots, s_N\}$, where N is the total number of sampling points in the signal. One arbitrary point in the signal is represented by s_i and its coordinate in the graph is (x_i, y_i), $i = 1, 2, \ldots, N$. Then the Euclidean distance between two consecutive points is

$$\mathrm{dist}(s_i, s_{i+1}) = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}. \tag{3.1}$$

The Katz FD for a signal s with N points is mathematically defined as (Katz, 1988)

$$\mathrm{FD}_s = \frac{\log(L)}{\log(d)} \tag{3.2}$$

where FD_s is the fractal dimension of the signal s, L is the total length of the signal, derived from the sum of Euclidean distances of all adjacent points, i.e.,

$$L = \sum_{i=1}^{N-1} (\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}), \tag{3.3}$$

and d, the planar extent of the waveform, can be usually considered as the farthest of all distances between the first point (point 1) and all subsequent points of the signal. Mathematically, d can be written as

$$d = \max(\mathrm{dist}(s_1, s_i)) = \max(\sqrt{(x_1 - x_i)^2 + (y_1 - y_i)^2}),\ i = 2, 3, \ldots, N. \tag{3.4}$$

The Katz's FD compares the actual number of units that compose a curve with the minimum number of units required to build the simplest pattern of the same spatial extent. However, the FD value obtained from Eq. (3.2) is affected by measurement unit. To overcome this difficulty, Katz proposed to normalize L and d by the average distance (denoted by a) between successive points. The average distance a is defined as $a = L/k$, where k is the number of steps in the curve, i.e., $k = N$-1. Dividing L and d by a in Eq. (3.2) leads to

$$\mathrm{FD}_s = \frac{\log(k)}{\log(d / L) + \log(k)}\ . \tag{3.5}$$

It should be noted that, in the STFD to be discussed later, N is the window length and hence k is the number of data intervals in a window.

3.1.2 STFD transform and its application to bearing fault feature extraction

In this subsection, the FD will be exploited to track the locations of impacts generated by bearing faults and also extract the envelope of the impacts. A typical vibration signal produced by a bearing fault is periodic, comprising a sharp rise that corresponds to the impact between the surfaces at the location of a defect (Nikolaou and Antoniadis, 2003). To extract the feature of periodic impacts using FD, the STFD transform is therefore developed in this study.

A moving window is used to truncate the original signal in order to obtain STFD sequence, among which each STFD value corresponds to one windowed signal portion. The length of the window is selected as N_{win} = int(αf_s), where int($\cdot$) stands for the integer part of the value, α is a constant pre-specified by the user and f_s is the sampling frequency of the waveform (Hadjileontiadis and Rekanos, 2003). The constant α is empirically set. Some researchers have used a window length of $0.006 f_s$ for lung and bowel sound signal processing as this kind of signals are much clearer than mechanical vibration signals (Hadjileontiadis and Rekanos, 2003; Hadjileontiadis, 2005). In our case, as the fault-induced impulses are often of low energy and masked by background noise and vibration interferences, the window has to be shorter to obtain a satisfactory result. This is because too long a window may make the STFD sequence over smoothened. Therefore, in the application to bearing fault detection, the constant α should be smaller than the one used for lung and bowel sound signal processing to ensure that the incipient impulses are not missed. However, it should be noted that N_{win} cannot be too short. Too short a window may lead to many fictitious FD peaks, where obviously impacts cannot be identified correctly. In the following subsections, the constant α is empirically set to 0.002 – 0.003 and is justified by the performance of the algorithm. The window is shifted by one sample step each time along the input vector in order to acquire the point-to-point time series of STFD values.

Without affecting the dynamic range of the STFD sequence, the following transform is applied in the subsequent subsections

$$\mathbf{STFD} = \mathbf{STFD}^{\mathbf{original}} - \mathbf{MIN} + \mathbf{1} \tag{3.6}$$

where the vector $\mathbf{STFD^{original}}$ is the STFD sequence calculated using Eq. (3.5) for all signal segments, **MIN** is a vector of the same length (dimensions) as $\mathbf{STFD^{original}}$ and all its components equal to the minimum component of $\mathbf{STFD^{original}}$, **1** represents a unit vector also with an identical length to that of $\mathbf{STFD^{original}}$, and **STFD** denotes the obtained STFD sequence. According to Eq. (3.6), the minimum of STFD sequence is always limited to 1.

To transform the raw vibration signal into a STFD sequence, the STFD value of each signal segment is computed by Eq. (3.5) from each window of the raw signal and assigned to the midpoint of the signal segment in the STFD representation of the signal. However, this would make the length, N-N_{win}+1, of the obtained STFD sequence shorter than the length, N, of the raw signal. To preserve the signal length, the raw signal is extended by duplicating: a) the first half number of points, i.e., (N_{win}-1)/2 points for odd N_{win}, (N_{win} /2, for even N_{win}) in the beginning, and b) the last (N_{win}-1)/2, for odd N_{win}, ((N_{win}-2) /2, for even N_{win}) points at the end of the raw signal, respectively. The length of the signal is then extended to $N + N_{win} - 1$. In this way, the length of obtained STFD sequence is identical to the raw signal. This step is necessary because of the iterative KPSA procedure to be described later.

To illustrate, consider an impulsive signal $s(t)$. It is the simulated vibration of faulty bearing under constant speed and generated by following model (Liang and Bozchalooi, 2010):

$$s(t) = \sum_{m=-M}^{M} L_m e^{-\beta(t-mT_p-\sum_{i=-M}^{m}\tau_i)} \sin(\omega_r(t-mT_p-\sum_{i=-M}^{m}\tau_i))u(t-mT_p-\sum_{i=-M}^{m}\tau_i) \tag{3.7}$$

where M is half of the number of the fault generated impulses, L_m is the amplitude of mth fault impulse, β the decay parameter, T_p is the time period corresponding to the fault characteristic frequency, τ_i presents the effect of random slippage of the rollers and is modeled as a uniformly distributed random variable with a zero mean and a standard deviation of 0.01T_p, ω_r is the excited resonance frequency, and $u(t)$ is a unit step function. The values of the parameters used in the simulation are listed in Table 3.1.

Table 3.1 Sample parameters used for the simulated signal.

Parameter	L_m	β	T_p	ω_r
Value	1	800Hz	8ms	$2\pi\times$2K rad/s

To simulate the real case, white Gaussian noise is added into the signal $s(t)$ to make the SNR equal to -1 dB. The noisy signal sampled at 12 KHz is shown in Fig. 3.1 (a). Its STFD transforming result is displayed in Fig. 3.1 (b), demonstrating that the STFD can track the impulse locations more clearly in comparison with the original signal because the noise and high frequency ringing have been removed. Comparing Fig. 3.1 (a) and (b) also indicates that the STFD sequence represents a kind of envelope of the raw signal with features that can be clearly related to the initial impulse sequence.

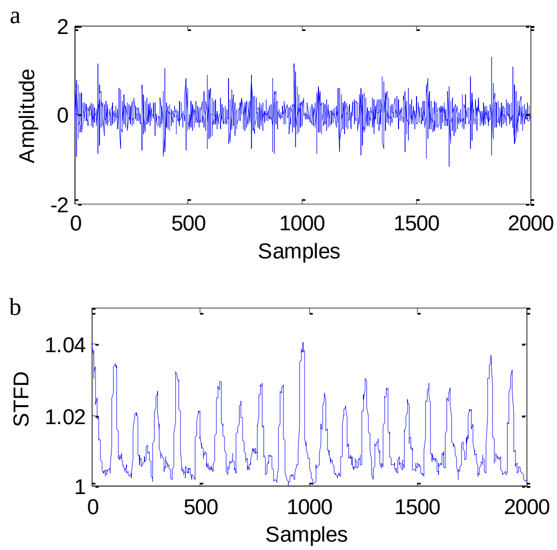


Fig. 3.1 Bearing fault feature extraction using STFD transform: (a) the noisy signal, and (b) STFD representation.

3.1.3 STFD transform calculation

It has been demonstrated that the proposed STFD transform can extract the envelope of impulses. However, as pointed out by ref. (Castiglioni, 2010), the variables describing a trajectory in the *x*-*y* plane, x_i and y_i, should be homogeneous quantities when calculating FD. In the Katz FD, x_i and y_i are different, being time and measured amplitude, which makes Katz FD dependent on the measurement unit. Even though the average distance a is used in the Katz method to normalize L and d, this is of little use, as revealed in (Castiglioni, 2010). This subsection will discuss the issues associated with STFD calculation using Katz FD and propose a modified STFD calculation method for bearing fault signature extraction.

3.1.3.1 Discussion of STFD calculation

To address the measurement-unit-dependent problem of Katz FD estimation, there is an alternative proposed in (Castiglioni, 2010) to calculate *d* and *L*. The variable *x* is neglected in FD calculation; hence, *d* and *L* are calculated in one-dimensional space using $\max(y_i)-\min(y_i)$ and $\sum_{i=1}^{N}|y_{i+1}-y_i|$, respectively. However, calculating STFD in this way is not applicable to the envelope extraction. This is because the STFD values are completely determined by the fractal component and thus a 'too-fractal' FD sequence is resulted. Consider an impulsive signal *s*(*t*) as shown in Fig. 3.2 (a). The signal *s*(*t*) is sampled at 12 KHz as well and the its STFD representation calculated in the one-dimensional space is displayed in Fig. 3.2 (b). As seen, the STFD sequence presents strong oscillations around 2 due to the reason that the STFD values are completely determined by fractal component and the non-fractal component is ignored. The envelope of the impulsive signal fails to be extracted by this means.

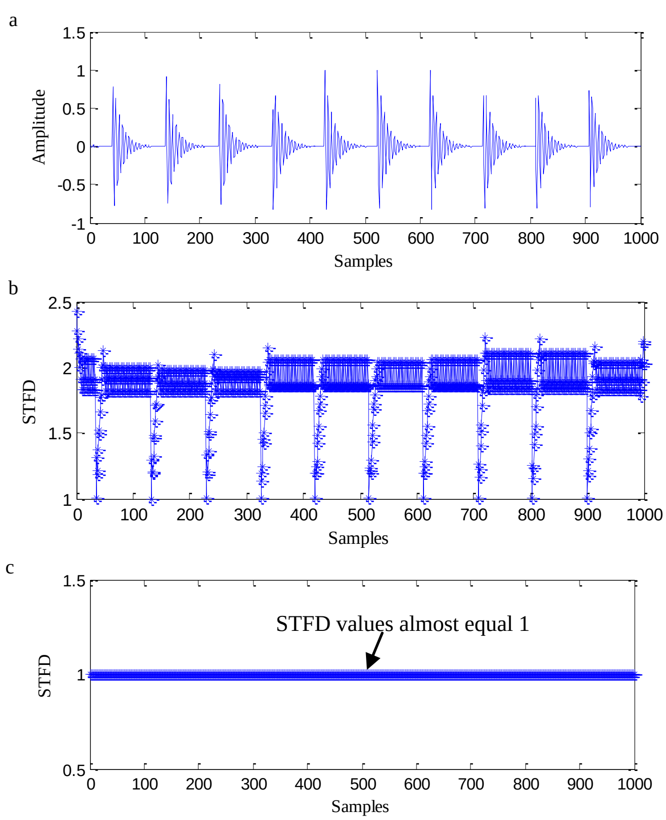


Fig. 3.2 The illustration of two STFD calculation methods in mono-dimensional space: (a) the raw signal, (b) the result of STFD transform only using vertical (amplitude) component, and (c) the result of STFD transform only using horizontal (time) component.

The other calculation method in one-dimensional space is to use time component and ignore amplitude component. The resulting STFD sequence is shown in Fig. 3.2 (c), where all STFD values are close to 1 because the time component is non-fractal. Hence, the signal envelope cannot be accurately extracted either. According to the above analyses, one can conclude that using the fractal component (i.e., amplitude component) for STFD calculation would generate a strongly oscillatory STFD representation whereas using the non-fractal component (i.e., time component) would lead to a STFD sequence with all values equal to 1. Obviously, neither can be used for reliable envelope extraction. The above analyses also explain why the FD of the signal being expressed in μV is larger than that of the signal being recorded in mV. The reason is that changing the measurement unit from μV to mV decreases the relative weight of fractal y_i amplitude component on the STFD values and thereby making STFD values decreasing to 1.

The above two calculation methods in the one-dimensional space are incapable of extracting the signal envelope. In the application to bearing envelope extraction, it is preferred that the horizontal and vertical components both contribute to STFD values. Incorporating the two components will lead to a smoother STFD sequence that is neither too-fractal (presenting strong oscillations) nor non-fractal (approaching to 1). As can be seen in Fig. 3.3 (a), the envelope of the impulses is successfully expressed by STFD representation calculated in the two-dimensional space. However, it has to be noted that the measurement unit used to express the signal influences the FD values, as mentioned previously. Fig. 3.3 (b) presents the results of STFD transform for signals which are generated via multiplying the raw signal $s(t)$ by different constants (0.5, 1, 2, 4, 6, and 8). It can be seen that the STFD values change with the amplitude components, following the pattern that STFD values increase with the growth of numerical values of amplitudes. Nevertheless, even if the STFD values change, the STFD sequence patterns of all the signals that are different only in terms of amplitude still keep unchanged. In other words, changing the measurement unit would change the numerical values of amplitude component, thereby causing the changes in the magnitude of the STFD sequence correspondingly; but the STFD sequence pattern remains unchanged. Thus, the envelope of the signal can still be extracted. It can be concluded that the variation of the STFD sequence magnitude has practically no effects on the frequency contents of the STFD sequence's spectrum.

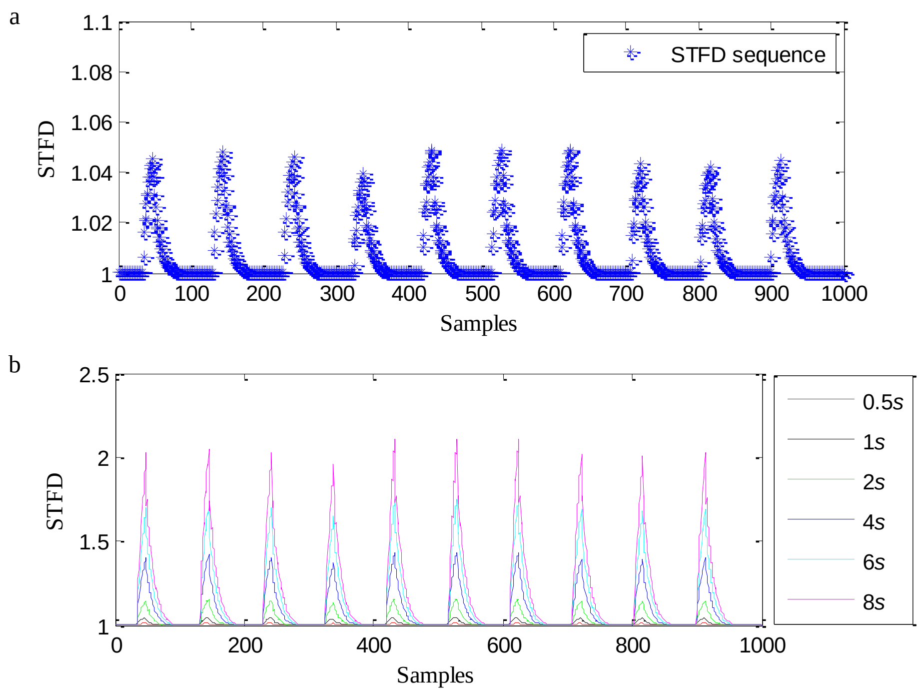


Fig. 3.3 Envelope extraction via STFD transform calculated in two-dimensional space: (a) Extracted envelope of the signal *s*, and (b) extracted envelopes of the signals 0.5*s*, *s*, 2*s*, 4*s*, 6*s* and 8*s*, to illustrate that the only difference among these FD sequences is the amplitude whereas the changing pattern is not affected.

3.1.3.2 Modified STFD calculation method for envelope extraction

However, if numerical values of the signal amplitude are increased by 1000 time, for instance, the signal is expressed in μV instead of mV; the relative weight of the amplitude component will be also increased proportionally. Further increasing values of the amplitude would make the contribution to the STFD value by the time component negligible. The STFD sequence calculated under this situation is equivalent to that calculated in the aforementioned one-component space (only using y_i items), which would produce strong oscillations. Likewise, substantial decrease of the numerical values of amplitude, e.g., expressing the signal in V instead of mV, will make the amplitude contribution insignificant and the STFD will be mainly determined by the time component.

The above two cases are undesired in the FD-based envelope extraction. To insure that the STFD values are determined by both horizontal and vertical components and then a smooth envelope can be extracted, they should have almost the same order of magnitude in calculating planar extent d and total length of the signal segment L. Therefore, we propose to use a double linear transformation to map the original signal into another embedded in an equivalent metric space. The first transformation is to normalize each point in the ordinate as:

$$y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}}, i = 1, 2, ... \tag{3.8}$$

where y_i is the original value of the ordinate, and y_{min} and y_{max} are the minimum and maximum y_i of the entire signal, respectively. The second transform normalizes the abscissa for signal segments as follows:

$$x_i^* = \frac{x_i}{x_{max}}, \ i = 1, 2, ..., N_{win} \tag{3.9}$$

where x_i is the original values of the abscissa for a signal segment, and x_{max} is the maximum x_i. The two linear transformations map the values of both abscissa and ordinate into the ones that belong to a unit square. By doing so, both time and amplitude components have almost the same order of magnitude when calculating the STFD sequence. A STFD representation that is determined by fractal and non-fractal components is then attained.

The last issue of the Katz method is that STFD values tend to approach 1 as n increases, according to Eq. (3.5). This difficulty can be mitigated in this work by using the STFD transform.

3.1.4 Periodic interference suppression via STFD transform and KPSA

As revealed earlier, the vibration signal measured from a defective bearing is often a mixture containing various interferences from the mass unbalance, shaft misalignment, gear meshing and electrical components, etc. The combined effect of interferences would cloud the signal spectrum or even lead to wrong detection results. Therefore, this subsection will detail how to use the proposed STFD transform to suppress interferences without filtering.

The periodic interferences can be expressed as $i(t) = \sum_{m=1}^{M} i_m(t) = \sum_{m=1}^{M} A_{im} \cos(\omega_{im} t)$, A_{im} and ω_{im} representing amplitude and frequency for the nth interference, respectively. To assess the interference influence on the conversional envelope demodulation, a single fault-induced simulated impulse contaminated by an interference with amplitude larger than that of impulse is considered:

$$x(t) = s(t) + i(t) = Le^{-\beta t} \sin \omega_r t + A_i \cos \omega_i t \tag{3.10}$$

where L is the amplitude of the impulse, $A_i > L$, indicating that the fault-related impulse is severely masked by the interference. The analytical version of the signal, denoted by $X(t)$, is

$$X(t) = Le^{-\beta t + j\omega_r t} + A_i e^{j\omega_i t} \quad . \tag{3.11}$$

The envelope (or the instantaneous magnitude) of the analytical signal can be calculated by

$$\begin{aligned} Envelope(X(t)) &= \sqrt{(Le^{-\beta t}\cos\omega_r t + A_i\cos\omega_i t)^2 + (Le^{-\beta t}\sin\omega_r t + A_i\sin\omega_i t)^2} \\ &= \sqrt{(Le^{-\beta t})^2 + A_i^{\,2} + 2Le^{-\beta t}A_i\cos((\omega_r - \omega_i)t)} \\ &\approx Le^{-\beta t}\cos((\omega_r - \omega_i)t) + A_i \end{aligned} \tag{3.12}$$

Eq. (3.12) suggests that the dominant component, i.e., the interference, is envelope demodulated but the demodulation of impulses is unsuccessful. It can then be concluded that the traditional demodulation is adversely affected by the presence of strong interferences. Therefore, an enveloping method which has the ability of suppressing interferences is highly desirable.

3.1.4.1 Interference suppression via STFD transform

Consider a random interference ($i(t)$ = 2cos ($2\pi ft$)) as shown in Fig. 3.4 (a). Fig. 3.4 (b) and (c) are respectively the impulsive signal and the signal mixture that is composed of the impulsive signal and the interference. An important feature for Eq. (3.10) is that the resonance frequency is much higher than the frequencies of periodic interferences for most real applications, i.e., $\omega_r \gg \omega_i$ (McFadden and Smith, 1984). It, therefore, can be observed that the interference frequency is much lower than that of impulses as shown Fig. 3.4 (a) and (b). Let us look at a signal segment within a short window, i.e., nth window in Fig. 3.4 (a). The Euclidian distance of the curve within the nth window, n = 1, 2 …, is represented by L_n (defined by Eq. (3.3)) and the planar extent is denoted by d_n (defined by Eq.(3.4)). If the vibration interference frequency is low enough, L_n is almost equal to d_n in Fig. 3.4 (a) as long as the window length is properly selected, which leads to $\mathrm{STFD}(n) \approx 1,\ s.t.\ d_n \approx L_n$. It then shows that the interference has little effect on the STFD representation. For the impulsive signal, the difference between the planar extent (d_n) and Euclidian distance (L_n) of the signal segment is more substantial because of the remarkable morphological appearance change, as displayed in Fig. 3.4 (c), resulting $\mathrm{STFD}(n) > 1$ according to Eq. (3.5). Due to the feature that resonance frequency is usually much higher than that of vibration interference, the STFD values resulted by impulses are larger than these related to interferences, suggesting that the impulse-related STFD values are more pronounced on the STFD representation, compared with interference-related STFD values. Consequently, the envelope of the impulses can be highlighted and the interference is suppressed as displayed in Fig. 3.4 (d).

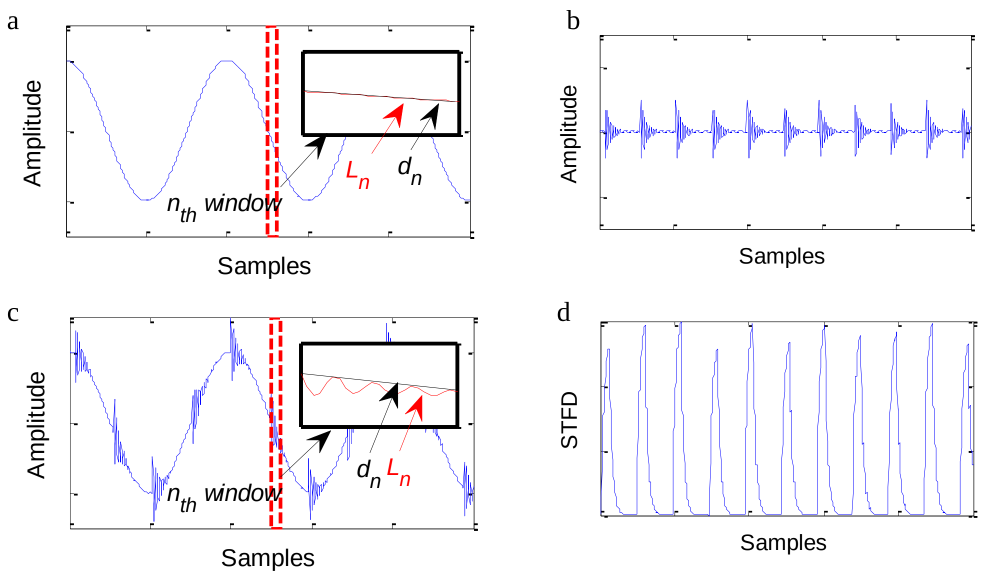


Fig. 3.4 The illustration of STFD transforming the signal mixture (impulses and a random interference): (a) the periodic interference, (b) the simulated impulses, (c) the mixture of (a) and (b), and (d) the STFD representation of the signal mixture.

3.1.4.2 Interference suppression via the joint STFD and KPSA

However, even though the effect of interference on STFD representation is less significant than that of impulses, it cannot be ignored for some cases, especially the effect of the interferences with high frequency (but still much lower than resonance frequency). For example, a synthetic signal composing of multiple interferences ($\cos(2\pi 1000t)$, $\cos(2\pi 620t)$ and $\cos(2\pi 300t)$) and the simulated impulses is shown in Fig. 3.5 (a). The STFD transform result and its spectrum are displayed in Fig. 3.5 (b) and (c), respectively. It can be seen that the compounded effect of interferences is not successfully removed as interference-related STFD values have more effects on the STFD representation due to high frequencies. The spectrum in Fig. 3.5 (c) is hence dominated by the interference frequency difference which would confuse or even mislead the detection result.

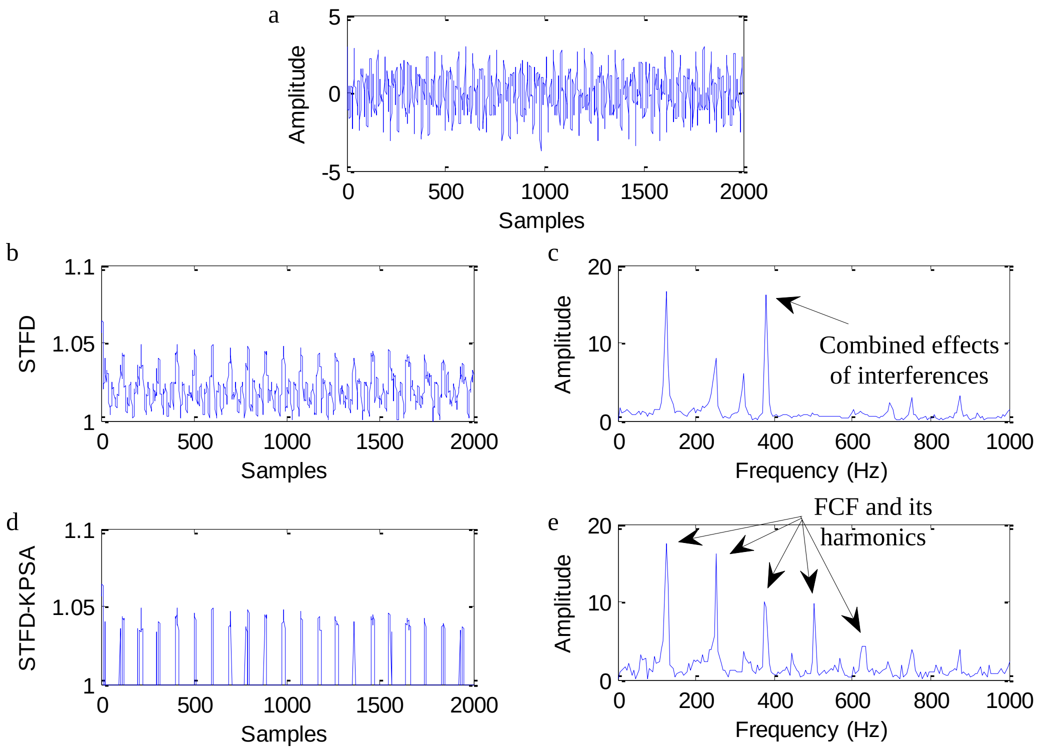


Fig. 3.5 The differences before and after the KPSA: (a) the synthetic signal consisting of the simulated faulty bearing signal and multiple cyclic interferences (cos(2π1000t), cos(2π620t) and cos(2π300t)), (b) STFD sequence without KPSA, (c) spectrum of (b), (d) STFD-KPSA sequence, and (e) spectrum of (d).

To address the problem caused by multiple interferences with relative high frequency, it is necessary to improve the proposed method so that it can further suppress the effect of interferences. Therefore, the KPSA algorithm is proposed. The idea of the KPSA is to extract the useful information and remove unwanted signal components as much as possible iteratively until a stopping criterion is reached. The criterion is kurtosis because it is a proven effective indicator of impulsiveness. It is a statistical parameter defined as the ratio between the fourth moment about the mean and the fourth power of the standard deviation of the signal as defined in the following

$$K = kurt(x(t)) = \frac{\mathrm{E}\left\{(x(t) - \mathrm{E}[x(t)]^4\right\}}{\sigma^4} \tag{3.13}$$

where σ is the standard deviation of signal $x(t)$, and $\mathrm{E}\{\cdot\}$ is the mathematical expectation operator. Considering the uncertainty in the signal, a tolerance range of a standard deviation σ^i is introduced. Hence, in each iteration of the KPSA, e.g., iteration i, the standard deviation σ^i and the kurtosis K^i of the STFD transform result (i.e., $\mathbf{STFD}^i$, a vector representing a STFD sequence) are calculated. The amplitudes of the points that are higher than 1 + σ^i are kept as is. The amplitudes of the points that are lower than 1 + σ^i are "removed" i.e., set to 1.0 (because the minimum FD value is 1 according to Eq. (6)). In this way, a new vector, named $\mathbf{STFD*}^i$ representing useful information extracted in iteration i is constructed. However, it is possible that some points of low energy, i.e., with amplitudes lower than 1 + σ^i, may also be caused by the fault. Hence a new iteration i+1 is needed to further extract the information of such low-energy impulses as long as the difference between K^i and K^{i-1} is still significant, i.e., the impulsiveness difference between the two consecutive iterations is worth pursuing. In other words, in each iteration, e.g., i+1, the above process is applied to the signal residue of iteration i, i.e., $\mathbf{STFD}^{i+1} = \mathbf{STFD}^i - \mathbf{STFD*}^i + \mathbf{1}$ and the iteration procedure continues until the impulsiveness of the signal residue becomes insignificant, i.e., $|K^i - K^{i-1}| \leq \varepsilon$ (ε is a pre-specified small positive number). Then the final signal residue is considered to be composed of noise and/or interference-induced STFD component and is hence removed. The final STFD sequence after applying the KPSA, called **STFD-KPSA** hereafter is accordingly given by

$$\mathbf{STFD\text{-}KPSA} = \sum_{i=1}^{I} \mathbf{STFD*}^i - \mathbf{I} + \mathbf{1} \tag{3.14}$$

where I is the number of iteration. The details of the KPSA algorithm are presented in Fig. 3.6.

Applying the above KPSA algorithm to the STFD sequence in Fig. 3.5 (b) leads to the result in Fig. 3.5 (d). As shown, the unwanted signal components have been removed and only the peaks that are relevant to the simulated fault are kept. The more obvious enhancement of using KPSA can be observed by looking at the spectrum which is shown in Fig. 3.5 (e), from which the interference-related frequency components are completely eliminated and only the FCF and its harmonics are kept.

3.1.4.3 Interference suppression ability of the proposed STFD-KPSA method

In this subsection, the proposed method is evaluated in terms of its capability of interference suppression. To quantify the level of interferences for a signal mixture $x(t)$ including fault-induced impulses $s(t)$ and interferences $i(t)$, the signal-to-interference ratio (Bobin et al.) can be defined as follows (Bozchalooi and Liang, 2009):

$$\mathrm{SIR}(x)=\frac{\frac{1}{T}\int_0^T s^2(t)dt}{\frac{1}{T}\int_0^T i^2(t)dt}=\frac{P_s}{P_i} \tag{3.15}$$

where T is the signal length, P_s is the power of fault impulses and P_i is the power of interferences.

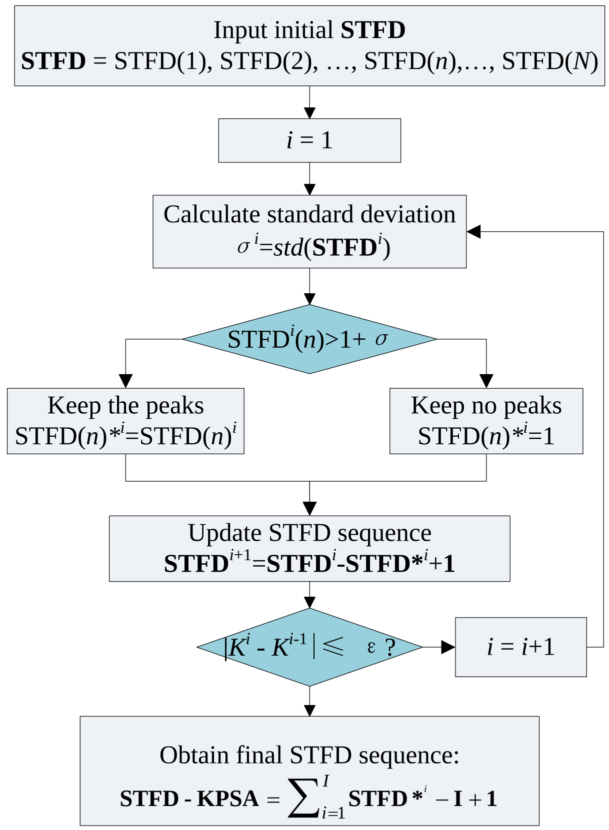


Fig. 3.6 Flowchart of the KPSA algorithm.

The detectability performance of a method can be expressed by the number of harmonics of the FCF revealed by the method using a map. The two axes are associated with two factors to be examined. In this case, the two factors are frequency and SIR, respectively. The levels of brightness represent the strength of energy: the brighter the plot, the higher the energy, as indicated in the Fig. 3.7 (This applies to all the detectability maps in Chapter 3 and Chapter 4). Fig. 3.7 shows such a detectability map for the proposed method. The simulated impulsive signal is defined by Eq. (3.7)with SNR of – 7 dB and the corresponding parameters of the simulated signal are the same as listed in Table 3.1.

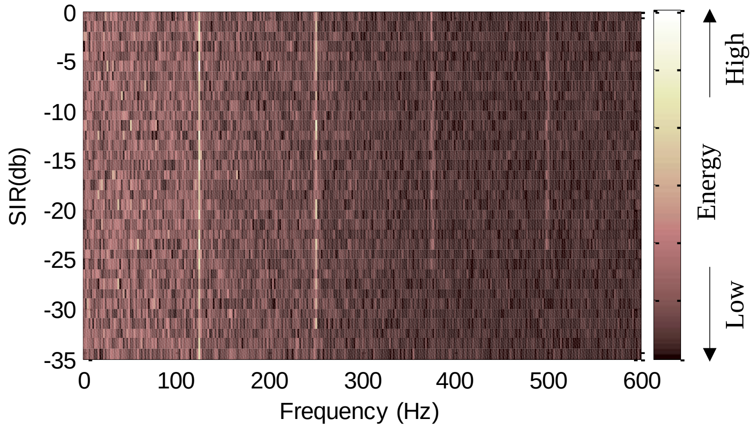


Fig. 3.7 Detectability map of the STFD-KPSA method

The obtained detectability map suggests that: a) the proposed method is capable of handling signal with SIR as low as – 33 dB, b) for signals with SIR ranging from 0 to – 25 dB, the FCF and three its harmonics can be detected by the proposed method, and c) the detectability then deteriorates with the decrease of SIR, as can be seen that only the FCF and one harmonic can be observed when SIR decreasing from – 25 dB to – 33 dB.

3.1.5 Computational complexity of the STFD-KPSA method

The computational efficiency is crucial to the success of an algorithm for on-line condition monitoring. As such, the computational complexity of the proposed method should be investigated. The Big O-notation is one of the most commonly used means of estimating computational complexity (Chang, 2003). The computational complexity of the STFD sequence calculation part is $O(Num_{win} \cdot N_{win})$, where Num_{win} is the number of sliding windows and N_{win} is the length of a window as defined before. This order of complexity can be calculated as follows

$$O(Num_{win} \cdot N_{win}) = O((N - N_{win} + 1) \cdot N_{win}) = O(N \cdot N_{win}) \tag{3.16}$$

where N is the signal length. The computational complexity of the KPSA can be estimated as $O(N \cdot I)$, where I is the number of iteration. As N_{win} and I are numbers for each case, the notations of $O(N \cdot N_{win})$ and $O(N \cdot I)$ indicates that the computational performance of STFD and KPSA algorithms changes linearly as the size of dataset. Therefore, the computational complexity of the whole algorithm can be approximated as $O(N)$. This suggests that the whole algorithm's performance is proportional to the size of the dataset being processed, leading to high computational efficiency in the general case. The graph of the running time of the proposed method is displayed in Fig. 3.8, where 'Time' represents the computing time in seconds. It shows that the running time almost linearly increases with the input signal size. For comparison, the time complexity graphs of the MMA and EO approaches are also plotted in Fig. 3.8. It can be seen that the MMA is the slowest. The EO is slightly more efficient than the proposed method, but both of them generally are highly computationally efficient. For instance, the running time of the proposed method is approximately 0.2 s for a signal of 20,000 sample points (on a desktop with Intel Core 2 CPU and 3.21 G memory), which is sufficient for online condition monitoring.

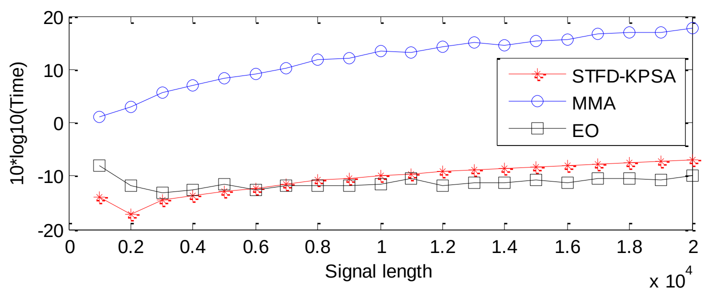


Fig. 3.8 Running time of the proposed STFD-KPSA, MMA and EO methods.

3.2 Simulated study

To validate the performance of the proposed signal, a composite signal is constructed by

$$x(t) = s(t) + i(t) + n(t) \tag{3.17}$$

where $s(t)$ is the simulated impulsive signal obtained according to Eq. (3.7), $i(t)$ stands for vibration interferences, and $n(t)$ is the white Gaussian noise. The multiple interferences generated by other electrical/mechanical components, like unbalanced shaft and gear meshing are simulated using sinusoidal functions $i(t) = \sum_{m=1}^{M} i_m(t) = \sum_{m=1}^{M} A_{im} \cos(2\pi f_{im} t)$. Eight interferences are added in this case, $f_{i1}\ldots f_{i7}$ =1000 Hz, 500 Hz, 300 Hz, 150 Hz, 80 Hz, 30 Hz, 15 Hz and 8 Hz respectively $A_{i1}\ldots A_{i7}$ = 0.5, 1, 1.5, 1, 1, 1, 1 and 1.5, making the SIR equal to – 21 dB. In addition, there is background noise in the real vibrating signal; hence white Gaussian noise $n(t)$ is added. The SNR relative to simulated impulse signal is -7 dB. The purpose of adding such interferences is to simulate real cases where the interferences as well as their harmonics would present. The mixed signal x of 16000 sampled data points is plotted in Fig. 3.9 (a), where the features of the impacts are severely obscured by interferences and noise. Fig. 3.9 (b) displays the

Hilbert enveloping spectrum of the simulated signal. As seen, the Hilbert enveloping spectrum is dominated by the frequency differences between interferences, and no frequency information associated with FCF can be found because the impulses are severely polluted by interferences and noise. Then the proposed method is applied and the results are shown in Fig. 3.9 (c) (output of the STFD method) and (d) (output of the STFD-KPSA method). Applying spectral analysis to STFD sequence and STFD-KPSA sequence respectively yields Fig. 3.9 (e) and (f). It can be seen that the effect of interferences is not completely eliminated using the STFD alone, as shown in Fig. 3.9 (e), where the second frequency line associated with interference differences has an amplitude almost comparative with that of the first FCF harmonic. This is due to the effect of multiple interferences that still exists in STFD representation. Fig. 3.9 (f) provides much cleaner result, where the FCF and three of its harmonics clearly identifiable because the combined effects of the interferences are further reduced using the proposed KSPA algorithm. The above analysis indicates better demodulation result of STFD-KPSA in comparison to the results obtained by the Hilbert transform and the STFD without KPSA as shown in Fig. 3.9 (b) and (e), respectively. It, hence, can be concluded that the proposed STFD-KPSA method is capable of extracting bearing fault feature and suppressing interferences and noise without using prefiltering.

For comparison, two of the state-of-the-art enveloping methods, i.e., energy operator (EO) and multi-scale morphological analysis (MMA), are applied to the same signal $x(t)$. The frequency spectra of resulting envelopes are presented in Fig. 3.10 (a) and (b), respectively. As observed, the results are still clouded by the interferences. In Fig. 3.10 (a), the combined effect of interferences (i.e., interference frequency differences) still dominates the spectrum and the FCF together with its harmonics are not easy to be identified. By means of EO, the FCF together with two harmonics can be seen in Fig. 3.10 (b). Nevertheless, the peaks associated with interferences, like the frequency differences between i_1 and i_2, i_2 and i_3, etc., still exist in the spectrum with quite high amplitudes. The detection results demonstrate that the proposed method outperforms EO and MMA when dealing with signals obscured by multiple interferences.

To further evaluate the capability of the proposed method in handling noise, it is employed to process the simulated signal defined by Eq. (3.17) with different levels of noise added. The results are plotted in Fig. 3.11, in which the FCF and at least two of its harmonics can be identified for SNRs from 0 to -8 dB. This level of noise handling capability should be sufficient for bearing fault detection in many applications.

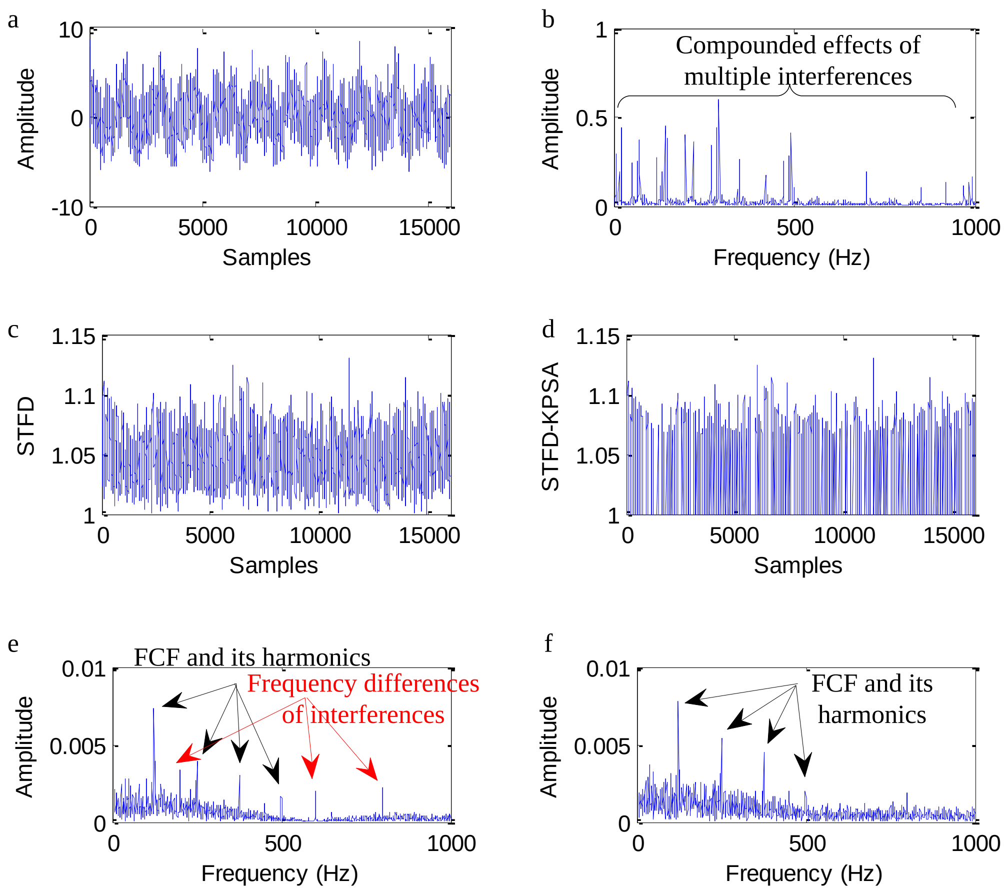


Fig. 3.9 Application of the proposed method to simulated signal *x* (SIR = - 21 dB; SNR = - 7 dB): (a) the composite signal, (b) the Hilbert envelop spectrum of (a), (c) the STFD sequence, (d) the STFD-KPSA result, (e) the frequency spectrum of (c), and (f) the frequency spectrum of (d).

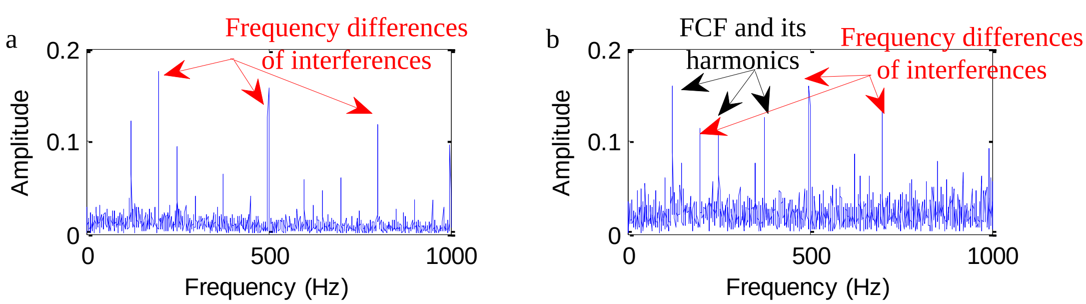


Fig. 3.10 Application of the MMA and EO methods to signal mixture *x*: (a) frequency spectrum of envelope signal obtained by MMA, and (b) frequency spectrum of envelope signal obtained by EO.

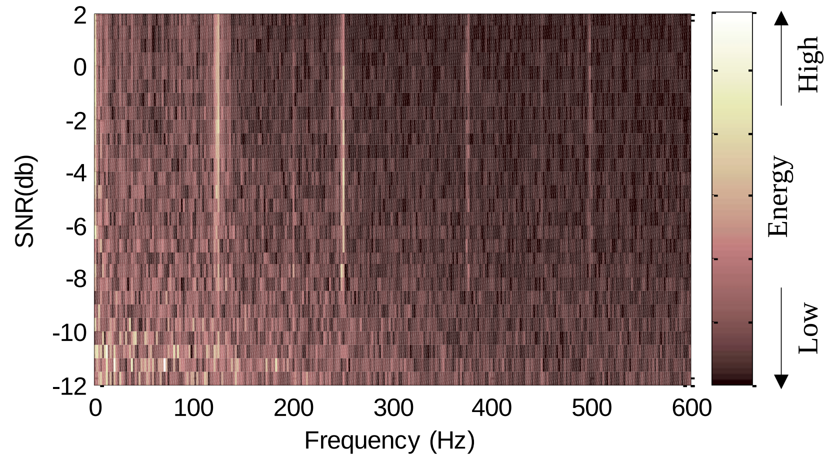


Fig. 3.11 The STFD-KPSA processing results of the signal mixture defined in Eq. (3.17) with different levels of noise.

3.3 Experimental study

The proposed method is further tested via experimental data collected in our lab using a SpectraQuest Machinery Fault Simulator (MFS-PK5M) test rig shown in Fig. 3.12. The vibration signals measured by an accelerometer (ICP® accelerometer, Model 623C01) are fed to an NI data acquisition board (NI_{TM} USB-6212 BNC) and recorded in a computer using the $LABVIEW_{TM}$ software. The sampling frequency is empirically set to be 20 KHz in the experiments, which has been validated by many experiments in our lab. This is mainly because the resonant frequencies are unknown. Even if accurate resonance frequencies could be calculated there are no practical means to determine which of the multiple possible resonances is likely to be dominant. Therefore, the frequency range of bearing vibration is often determined empirically. For example, a higher sampling frequency is used in the beginning to ensure that the full spectrum will be considered. After determining which range to study, a lower sampling frequency can be used, if desirable. To challenge the proposed method, the signal conditioner is removed during the experiments. Therefore the collected signal is not amplified and prefiltered, and hence is of low energy and contains interferences such as the power line frequency of 60 Hz, shaft rotational frequency, as well as background noise.

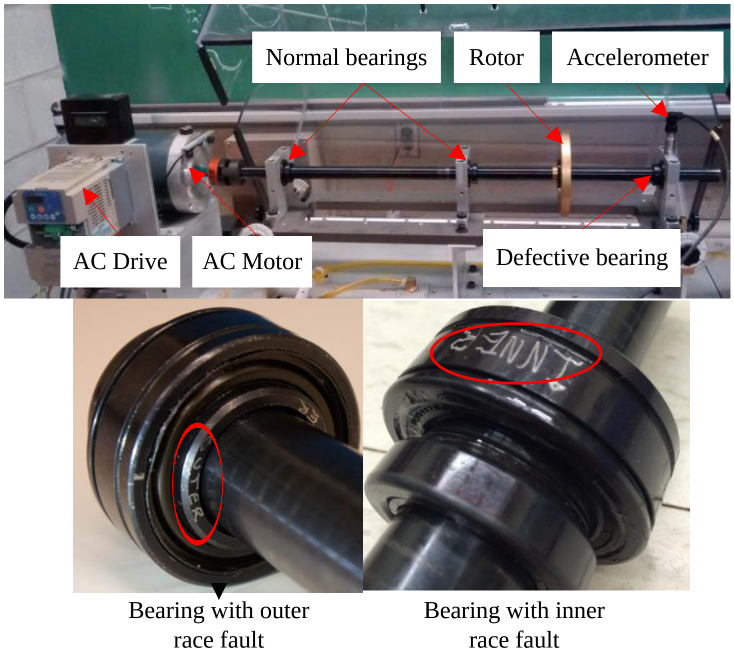


Fig. 3.12 Experimental setup and defective bearings.

The experiments are conducted for both inner race and outer race bearing faults. Two bearings of the same type are used, one with an inner race fault and the other one with an outer race defect, as marked on the surface of the bearings (see Fig. 3.12). The bearings used in the experiments in the thesis are all labelled by ‘outer’, ‘inner’, or ‘healthy’. Bearing parameters are listed in Table 3.2. The shaft rotation frequency f_r is 18 Hz (1080 rpm), and the FCF of outer race fault (f_{BPFO}) and inner race fault (f_{BPFI}) are 64.3 Hz and 97.74 Hz respectively.

Table 3.2 Parameters of faulty bearings used in the experiments.

Bearing type	Pitch Diameter	Ball Diameter	Number of balls	BPFO	BPFI
ER16K	38.52mm	7.94mm	9	$f_r \times 3.572$	$f_r \times 5.43$

Three locations of outer race fault are considered, i.e., top, middle and bottom, as shown in Fig. 3.13, where the marks on the bearings indicate the corresponding defect positions on the outer race inside the bearing. The only load in this test is the unbalanced rotor and the shaft. The contact between ball and the outer race is possible at any fault position due to the low load and the unbalanced rotor. In this case, the signal strength mainly depends on the distance between the sensor and the fault location. Therefore, the signal from the fault on the top is stronger as it is closer to the accelerometer; whereas the signal from bottom fault is weaker as it is far away from the accelerometer. The strength of the signal from the middle fault is between those obtained from the top and bottom fault locations.

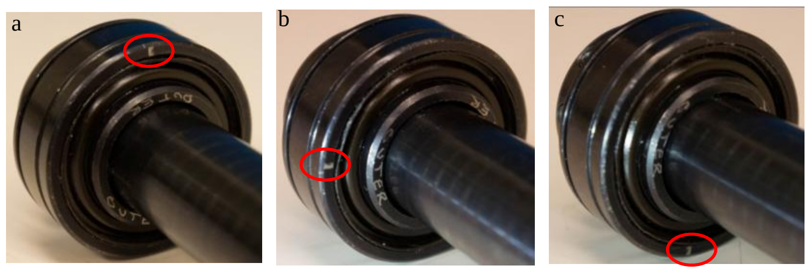


Fig. 3.13 Bearing outer race faults at three locations: (a) a fault on the top, (b) a fault in the middle (side), and (c) a fault at the bottom.

3.3.1 Outer race fault detection

Only the middle fault and bottom fault are considered in this subsection as they are more difficult to detect due to the relatively weaker impulses. The collected vibration signal of the middle fault and its Hilbert enveloping spectrum are plotted in Fig. 3.14 (a) and (b) respectively. In Fig. 3.14 (b), the shaft rotation frequency f_r (18 Hz) which is also denoted by f_{i1} and considered as one of the interferences and the power line interference frequency f_{i2} (60 Hz) as well as their harmonics dominate the landscape, but neither the outer race FCF f_{BPOF} (64.3 Hz) nor its harmonics can be observed. As the classical enveloping method fails to reveal the known outer race fault the proposed envelope demodulation is employed, yielding the envelope signal and its frequency spectrum shown in Fig. 3.14 (c) and (d) respectively. The impacts characterized by higher STFD values are more notable, whereas the interferences are removed by transferring them into much lower STFD peaks (Fig. 3.14 (c)). The f_{BPOF} and six of its harmonics can be easily identified on the spectrum of the STFD-KPSA result as shown in Fig. 3.14 (d).

Next, the signal from the outer race fault located at the bottom is analyzed. The raw signal is shown in Fig. 3.15 (a) and its classical enveloping spectrum is presented in Fig. 3.15 (b). It can be observed in Fig. 3.15 (b) that the fault-related frequency f_{BPOF} (64.3 Hz) and its harmonics are not discernible but the shaft rotation frequency (18 Hz) and power line frequency (60 Hz) as well as their harmonics clearly stands out the spectrum. The signal then is processed using the proposed method and the extracted envelope is illustrated in Fig. 3.15 (c). The Fourier transform of the extracted envelope leads to the result presented in Fig. 3.15 (d), from which the f_{BPOF} approximately 65.6 Hz and three of its harmonics can be clearly observed.

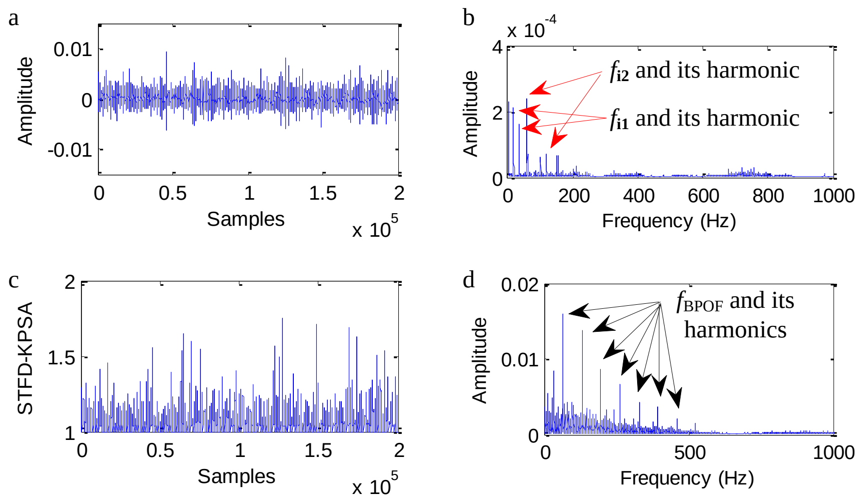


Fig. 3.14 Performance of the STFD-KPSA in detecting the side positioned outer race fault: (a) the raw vibration signal, (b) the Hilbert enveloping spectrum of the raw signal, (c) the STFD-KPSA result of the raw signal, and (d) the frequency spectrum of the STFD-KPSA result.

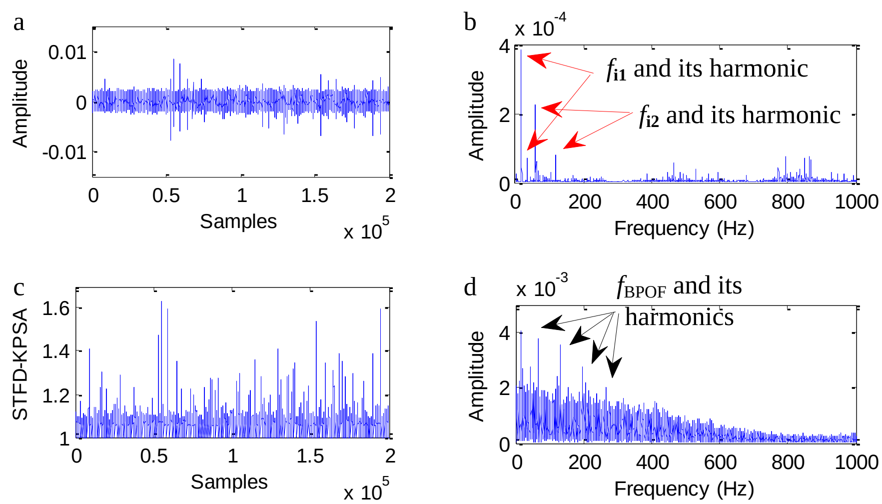


Fig. 3.15 Performance of the STFD-KPSA in detecting the bottom-positioned outer race fault: (a) the raw vibration signal, (b) the Hilbert enveloping spectrum of the raw signal, (c) the STFD-KPSA result of the raw signal, and (d) the frequency spectrum of the STFD-KPSA result.

According to the detection performance presented above for both cases, it can be concluded that the proposed method can diagnose faults from vibration data where the fault-induced impulses are severely masked by strong interferences. The major reason is that the proposed STFD-KPSA method can inherently suppress interferences and noise.

For comparison, the EO and MMA techniques are used for the signal generated by bottom-positioned outer race fault, with results illustrated in Fig. 3.16 (a) and (b), respectively. No information associated with fault can be revealed by the EO method. The MMA approach can merely reveal the fundamental FCF of outer race, as exhibited in Fig. 3.16 (b). The comparison demonstrates that the performance of the proposed method is superior to both the EO and MMA approaches.

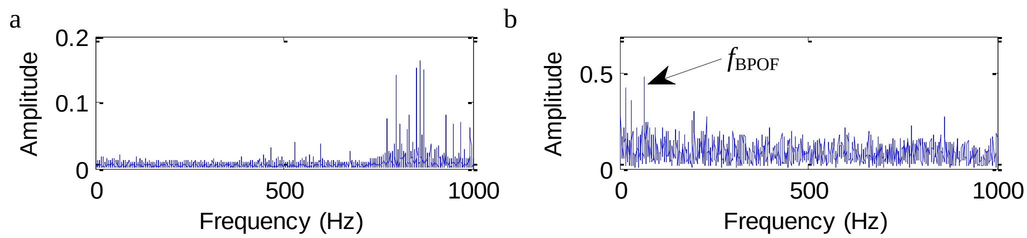


Fig. 3.16 Performance of the EO and MMA in detecting the bottom positioned outer race fault: (a) frequency spectrum of envelope signal obtained by EO, and (b) frequency spectrum of envelope signal obtained by MMA.

3.3.2 Inner race fault detection

To further examine the performance of the STFD-KPSA method, it is also applied to the inner race fault case. Like in the outer race case, the vibration signal of inner race fault is also processed using both the Hilbert transform and the proposed method with the results displayed in Fig. 3.17. The raw signal is plotted in Fig. 3.17 (a). Fig. 3.17 (b) presents its Hilbert envelope spectrum where only the shaft rotation frequency and its harmonics can be observed. Then, the proposed method is employed, yielding the envelope signal in Fig. 3.17 (c) and a much better frequency spectrum as shown in Fig. 3.17 (d). The FCF of inner race fault (f_{BPIF}) around 97.5 Hz and four of its harmonics are easily recognized.

The results obtained by the EO and MMA are illustrated Fig. 3.18 (a) and (b), respectively. Fig. 3.18 (a) reveals no useful frequency information, showing that the bearing fault fails to be detected. Fig. 3.18 (b) is obtained using the MMA, where the only peak that dominates the spectrum is the shaft rotational frequency f_{i1}. By contrast, the STFD-KPSA method provides a better detection outcome, as shown in Fig. 3.17 (d).

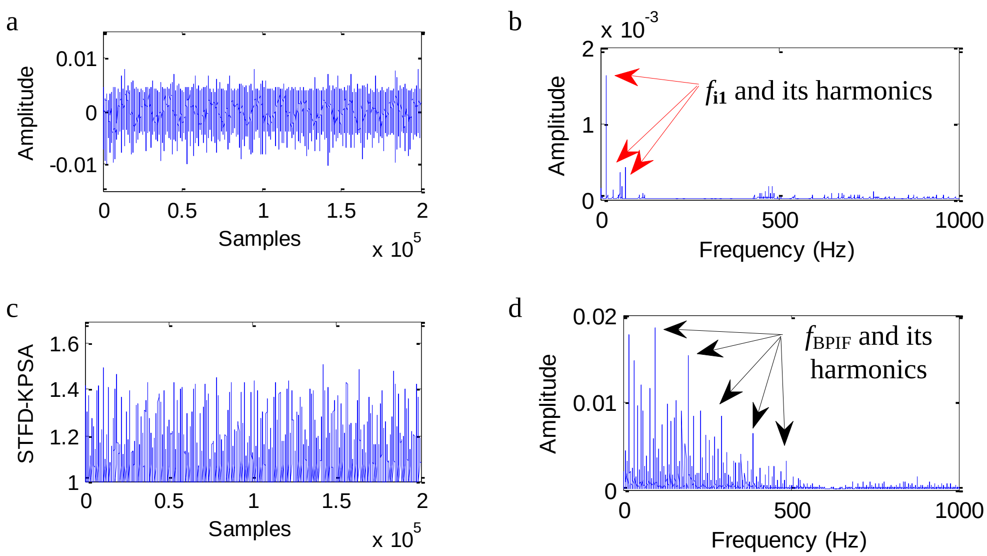


Fig. 3.17 Performance of the STFD-KPSA in detecting the inner race fault: (a) the raw vibration signal, (b) the Hilbert enveloping spectrum of the raw signal, (c) the STFD-KPSA result of the raw signal, and (d) the frequency spectrum of the STFD-KPSA result.

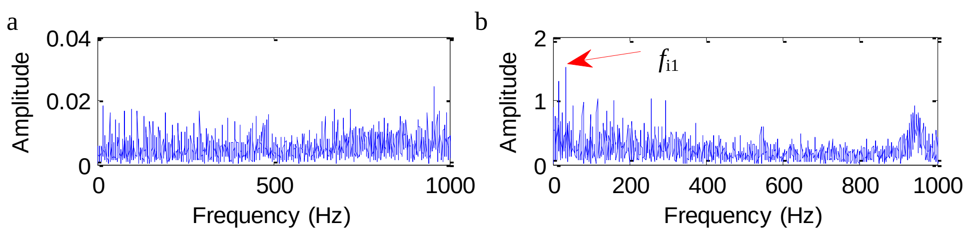


Fig. 3.18 Performance of the EO and MMA in detecting the inner race fault: (a) frequency spectrum of envelope signal obtained by EO, and (b) frequency spectrum of envelope signal obtained by MMA.

The experimental validation shows that the proposed method can detect both outer race and inner race flaws without any prefiltering. It is worth mentioning that there are still peaks of at non-fault frequencies in the spectra, as can be seen in Fig. 3.15 (d) and Fig. 3.17 (d). This is mainly due to that the signal conditioner in our experiments has been removed; the signals are thus not pre-denoising, pre-amplified and anti-aliasing. These frequencies could come from background noise, modulation and aliasing effects. However, the diagnosis decision is not influenced by these frequencies as they are suppressed and of low energy.

3.4 Conclusion

A new enveloping scheme has been proposed for bearing fault feature extraction from vibration signals without the involvement of prefiltering. The proposed method relates FD to the morphological appearance differences in signals. Based on this, the STFD transform is developed for the envelope demodulation of impulses. Problems related to STFD transform calculation using Kats FD are also discussed. It is observed that both fractal and non-fractal components have to be considered to obtain a smooth envelope signal. To achieve this, a weight coefficient is introduced to make the two components have almost the same order of magnitude when calculating the STFD sequence. In addition to impulse envelope extraction, the vibration interferences can be suppressed in the STFD representation with the aid of the proposed KPSA. Therefore, the proposed method can remove interferences without prefiltering process. The three main advantages of the proposed enveloping method are: (1) its ability of handling signal with very low SIR (as low as – 33 dB), (2) its filter-free nature, which mitigates the difficulties present in the selection of proper filter parameters in the classical prefiltering-based methods, and (3) its low computational complexity, making it suited for on-line bearing condition monitoring.

The performance of the proposed method has been validated using both simulated and experimental data. The results have shown that the proposed method is effective in extracting fault feature and suppressing interferences without resorting to any prefiltering step. The proposed method has also been compared favourably with the demodulation methods based on Hilbert transform, EO and MMA for interference suppression.

Chapter 4 Iterative Oscillatory Behavior based Signal Decomposition and Its Application to Bearing Fault Detection[2]

The technique based on STFD described in the previous chapter is effective for bearing fault signature extraction under constant speed via the direct demodulation without prefiltering. Even though this proposed enveloping method is capable of handling interferences, its effectiveness may be handicapped by harsh noise and high frequency interferences, like in-band interferences (i.e., those residing in the resonance frequency band). When there are multiple high frequency interferences and their energy is relatively high, the STFD-based method would become less effective (In Chapter 3, the frequencies of most considered interferences are much lower than that of the resonance). In reality, the data collected in an industrial setting may contain such signal components. This situation poses at least the following difficulties:

(1) The presence of in-band interferences can mislead the passband identification process of the popular high resonance frequency (HRF) technique as well as its variants such as the state-of-the-art spectral kurtosis (SK) based methods. The reason is that the in-band vibration interferences would negatively contribute to the used criterion such as kurtosis value and hence the true resonance band can be easily missed whereas a wrong band could be selected as the “resonance band”.

(2) Even if the in-band interferences do not mislead the pass-band identification process of the popular HRF technique and its variants, they will adversely affect the quality of the amplitude demodulation result after band-pass filtering.

(3) The out-band high frequency interferences can cause problems to the STFD methods because the interference-related effect on STFD representation becomes more pronounced with the increase of interference frequencies. As illustrated in the previous chapter, such an increase in the discrepancy between L and d leads to more and stronger interference-related ripples in the STFD plot which even make the proposed KPSA algorithm ineffective.

[2] Published in Expert Systems with Application, vol. 45, 2016, pp. 40-55. doi:10.1016/j.eswa.2015.09.039.

In the light of the above, this chapter presents an approach from the point of view of interference and noise removal, including multiple high frequency interferences resulted by other mechanical/electrical components and their harmonics. The proposed method is based on oscillatory behaviors, rather than frequency, of a signal to isolate bearing fault features from vibration interferences. Different from the conventional filter-based denoising techniques, this method is independent of frequency information and this feature can explain the reason why the proposed method is capable of removing in-band interferences. At the same time, it overcomes the shortcoming of wavelet transform (WT) decomposition methods that are ineffective in representing impulse-like signals and removing harmonic interferences.

The proposed method jointly applies morphological component analysis (MCA) and tunable Q-factor wavelet transform (TQWT) to decomposing a signal into two signal components, i.e., low- and high-oscillation components, according to whether they having sustained oscillations. This method is hereby called oscillatory behavior based signal decomposition (OBSD). The MCA, with the aid of the TQWT, is adopted to separate fault-induced transients (low-oscillation component) from other signal components like periodic interferences and noise (high-oscillation component and residual) because they oscillate differently. The TQWT is parameterized by a Q-factor and plays a role of distinguishing signal components that present different oscillatory behaviors. Thus a proper Q-factor specification is pivotal to the success of signal component separation. This chapter then develops the strategies to determiner Q-factor for the TQWT. Based on the MCA and TQWT, the objective function can be constructed. The low- and high-oscillation components can be obtained by solving the objective function using the split augmented Lagrangian shrinkage algorithm (SALSA); thus fault-induced transients is extracted. In practice, vibration data are often severely contaminated by strong noise and various harmonic interferences (and their harmonics). Therefore using OBSD once may fail to extract fault signatures from such vibration signals as a basis may be unable to represent all interferences at a time. To address this problem, the Iterative-OBSD (IOBSD) is first put forward in this chapter, which reinforces the capability of the proposed method in handling multiple interferences. It is worth mentioning that the proposed method extracts fault signatures on the basis of oscillatory behaviors, instead of frequency information, and it hence can deal with in-band interferences and noise, which outperforms other filter-based methods. This is verified by comparing the proposed method with one of the commonly used filtering techniques, namely, SK method. The proposed method has also been examined by both simulation and different experiments.

4.1 Signal oscillatory behavior based signal decomposition (OBSD)

The OBSD method expresses a signal as the sum of a 'high-oscillation' and 'low-oscillation' component. A high-oscillation component is defined as a time domain waveform that displays sustained oscillatory behaviors whereas a low-oscillation component is defined as a waveform that is consisted of non-oscillatory transients. Therefore, in this subsection, the oscillatory behavior of a signal is first explained, followed by the elaboration of signal decomposition based on oscillatory behaviors.

4.1.1 Oscillatory behavior

As described earlier, a high-oscillation component is consisted of sustained oscillations. Therefore, two signals displayed in Fig. 4.1 (a) and (b) respectively are classified as high-oscillation component, where each of them contains about six cycles of a sine wave multiplied by a bell-shaped function (Selesnick, 2011a); whereas, Fig. 4.1 (c) and (d) respectively present a low-oscillation signal that does not exhibit continuously oscillatory behaviors. As seen, each signal only contains a single cycle. It is reported in (Selesnick, 2011a) that a signal's oscillatory behavior can be quantified by Q-factor which is defined as the ratio of the center frequency of a signal to its bandwidth. The more oscillatory cycles comprising a signal, the higher its Q-factor is. The signals illustrated in Fig. 4.1 (a) and (b) are sustained, and accordingly they have higher Q-factors. In contrast, the signals in Fig. 4.1 (c) and (d) present no oscillations; hence they have lower Q-factors. However, Q-factors for Fig. 4.1 (a) and (b) (or for Fig. 4.1 (c) and (d)) are identical as they have the same degree of oscillation. Note that signals in Fig. 4.1 (a) and (b) are time-scaled version of one another. Likewise, the same can be said for signals in Fig. 4.1 (c) and (d). Therefore, observing Fig. 4.1 can conclude that: a) time-scaling a signal produces no effect on its oscillatory behaviors, and b) signals with the same Q-factors, i.e., the same oscillatory behaviors, may be either high frequency signals or low frequency signals, which means the Q-factor does not influenced by the frequency of signals. This feature is particularly important. For example, the traditional filter-based methods are non-effective in separating these signals, like signals in Fig. 4.1 (a) and (c) (or signals in Fig. 4.1 (b) and (d)), because they share the same frequency band. However, the OBSD provides an appealing means for addressing this issue.

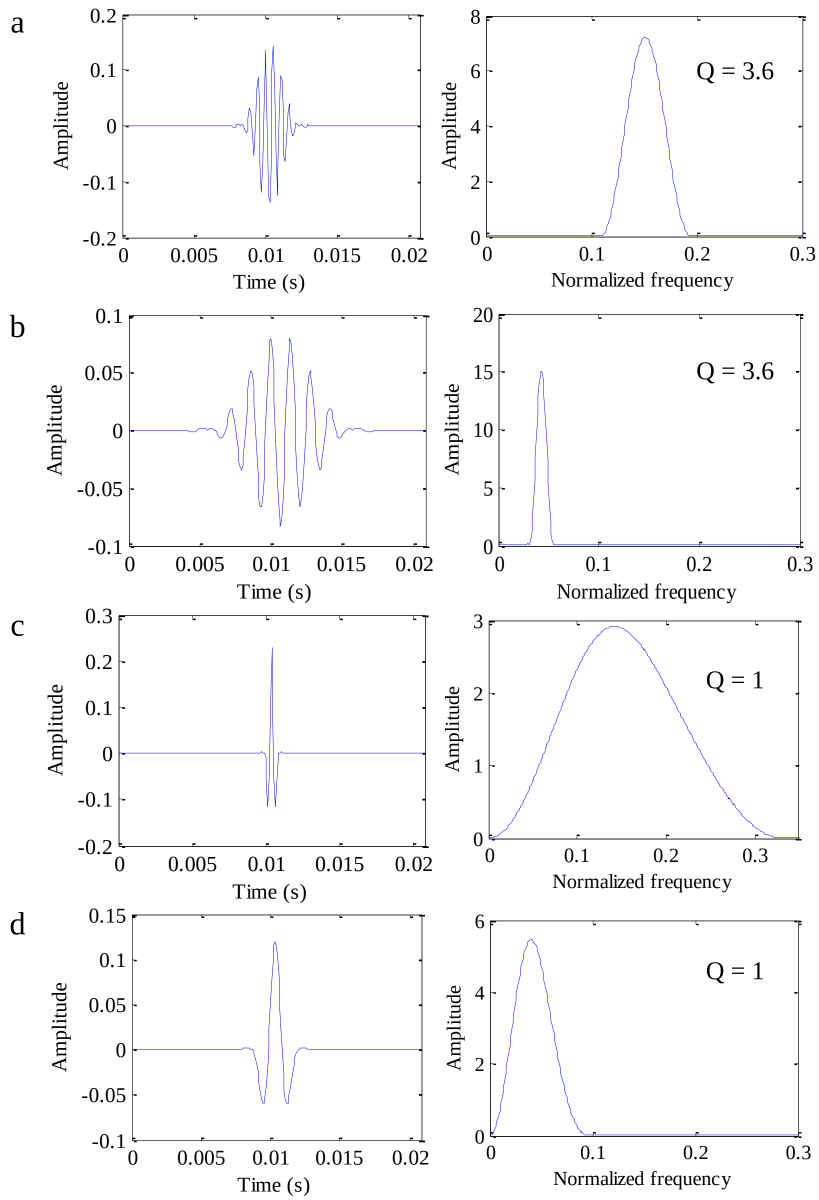


Fig. 4.1 Signals' oscillatory behaviors quantified by Q-factors: a high Q-factor characterized signal and its spectrum with normalized frequency of (a) 0.15 and (b) 0.04; a low Q-factor characterized signal and its spectrum with frequency (c) 0.15 and (d) 0.04.

The foregoing observation indicates that the Q-factors of signals are identical when the pulses exhibit the same oscillation behaviour, regardless of the frequency. Pulses with no or fewer oscillations have low Q-factors while pulses with more oscillations have high Q-factors.

4.1.2 Signal decomposition based on oscillatory behavior

In this subsection, signal decomposition based on the oscillatory behavior is given in detail. Signals measured from physical systems are often consisted of sustained oscillations and transients, i.e., high-oscillation component and low-oscillation component. Restated, the signal can be considered as a mixture of two separate signal components, each one having its own oscillatory behaviors. The transient is often a carrier of useful information for system health condition. Therefore, extracting the transient signal is a key step for monitoring system condition. To achieve this, the MCA which has been proved to be an effective tool in signal decomposition is employed (Bobin et al., 2007; Selesnick, 2011a). The detailed introduction of MCA is given below.

Suppose that a signal x, a combination of K independent components, i.e., $x=\sum_{k=1}^{K} x_k$ (x_k represents the kth component) is to be decomposed. All components' sparsest representations which can be sought over the dictionary including all bases are given by (Starck et al., 2005)

$$\{\mathbf{a}_1,...,\mathbf{a}_k\} = \underset{\{\mathbf{a}_1,...,\mathbf{a}_K\}}{\mathrm{Arg\,min}} \sum_{k=1}^{K} \| \mathbf{a}_k \|_1 \text{ s.t. } x=\sum_{k=1}^{K} \mathbf{\Phi}_k \mathbf{a}_k, \tag{4.1}$$

where $\mathbf{\Phi}_k$ is the kth basis, $\mathbf{a}_k$ is the coefficient vector of the kth component. Once $\mathbf{a}_k$ and $\mathbf{\Phi}_k$ are obtained, each component can be estimated as $x_k=\mathbf{\Phi}_k\mathbf{a}_k$. It is important to point out that the successful separation of the signal content can be reached relying on assumptions that $\mathbf{\Phi}_k$ is effective in representing one component and highly ineffective in representing the other components. If the signal under analysis is noisy, it cannot be exactly decomposed into K components. Suppose the given signal x is corrupted by noise. The objective is to use the given noisy signal x to produce an estimate $b=\sum_{k=1}^{K}\mathbf{\Phi}_k\mathbf{a}_k$ of the target signal. In order to accurately obtain the sparse representation of target signal in the noisy version, there are two goals that are described as follows:

(1) The estimate b is expected to be sparse; therefore it should involve as few atoms from $\mathbf{\Phi}$ as possible. This implies that $\sum_{k=1}^{K}\| \mathbf{a}_k \|_1$ should be minimized.

(2) The estimate b should be close to x in a least square sense to ensure that b is close to x. Mathematically, this aims at minimizing $\| \sum_{k=1}^{K} \mathbf{\Phi}_k \mathbf{a}_k - x \|_2^2$.

Considering the above two goals, it can transform this multi-objective optimization issue into a regularized problem, which can be expressed as

$$\{\mathbf{a}_1, ..., \mathbf{a}_k\} = \underset{\{\mathbf{a}_1, ..., \mathbf{a}_K\}}{\mathrm{Arg\,min}} \sum_{k=1}^{K} \gamma_k \| \mathbf{a}_k \|_1 + \| \sum_{k=1}^{K} \mathbf{\Phi}_k \mathbf{a}_k - x \|_2^2, \tag{4.2}$$

where γ_k is a regularized parameter. Then Eq. (4.2) is the noise-cognizant version of MCA method for noisy signals. In this way, the decomposition of signals is approximate and some residual that is represented by none of bases is produced. Thus, with respect to $\mathbf{a}_k$, the objective function of MCA can be written as

$$J(\mathbf{a}_1, ..., \mathbf{a}_k) = \sum_{k=1}^{K} (\| x - \mathbf{\Phi}_k \mathbf{a}_k \|_2^2 + \gamma_k \| \mathbf{a}_k \|_1). \tag{4.3}$$

As long as the basis for each signal component is created, the coefficients of K signal components can be obtained by minimizing Eq.(4.3), and accordingly each signal component can be estimated by $x_k = \mathbf{\Phi}_k \mathbf{a}_k$. In this study, we only have high-oscillation component and low-oscillation component, and so $K = 2$. Then objective function of Eq. (4.3) can be modified as

$$J(\mathbf{a}_1, \mathbf{a}_2) = \| x - \mathbf{\Phi}_1 \mathbf{a}_1 - \mathbf{\Phi}_2 \mathbf{a}_2 \|_2^2 + \gamma_1 \| \mathbf{a}_1 \|_1 + \gamma_2 \| \mathbf{a}_2 \|_1 \tag{4.4}$$

where subscript ‘1’ and ‘2’ represent low-oscillation component and high-oscillation component respectively.

4.2 TQWT and the development of IOBSD

As discussed above, in order to minimize the objective function and obtain the coefficients for each signal component, the bases $\mathbf{\Phi}_{1,2}$ that are used to represent each signal component have to be created. This is also important for the success of the MCA method as the bases $\mathbf{\Phi}_{1,2}$ play a role of distinguishing two components with different oscillatory behaviors. Therefore, in this subsection, the construction of bases is elaborated. To ensure that the two components i.e., high-oscillation component and low-oscillation component, are decomposed successfully, the high Q-factor basis is comprised of all high Q-factor wavelets, and similarly the low Q-factor basis is consisted of all low Q-factor wavelets. Then a high Q-factor basis efficiently represents the component with sustained oscillatory behavior and highly inefficiently representing the component with non-oscillatory behavior, vice versa. This can be done using a technique called tunable Q-factor Wavelet Transform (TQWT) (Selesnick, 2011b), where the Q-factor as discussed in subsection 4.2.1 representing oscillatory behaviors is easily specified. By means of specifying different Q-factors, bases with different oscillatory behaviors can be constructed. The details of TQWT and its application for MCA are described in the following.

4.2.1 TQWT

In this subsection, the TQWT that are parameterized by Q-factor is introduced. Like other kinds of wavelet transform, the TQWT is also based on the multi-rate filter banks, as shown in Fig. 4.2, where the TQWT is implemented through iteratively applying the two-channel filter bank on its low-pass channel. The wavelet coefficients of subband signals are denoted by $w^j(n)$ for j=1, 2,…,J+1. LPS α and HPS β in Fig. 4.2 represent low pass scaling and high pass scaling respectively, where α and β are the frequency domain scaling parameters. They either improve or decrease the sampling rate of the signal, i.e., the sampling rate of next stage is α (low pass channel) or β (high pass channel) times of that of the previous stage. To ensure that the wavelet transform will not be excessively redundant and in the meantime the perfect reconstruction of the filter banks is achievable, one has to have $0<\alpha$, $\beta<1$ and $\alpha+\beta>1$. Note that the TQWT is specified directly in the frequency domain and implemented using radix-2 FFTs, which can be more efficient.

For the filter bank implementation in Fig. 4.2, the frequency responses $H_0(\omega)$ and $H_1(\omega)$ are determined as follows to guarantee perfect reconstruction (Selesnick, 2011b)

$$H_0(\omega)=\begin{cases} 1 & |\omega|\leq(1-\beta)\pi \\ \theta(\dfrac{\omega+(\beta-1)\pi}{\alpha+\beta-1}) & (1-\beta)\pi\leq|\omega|\alpha\pi \\ 0 & \alpha\pi\leq|\omega|\leq\pi \end{cases} \tag{4.5}$$

$$H_1(\omega) = \begin{cases} 0 & |\omega| \leq (1-\beta)\pi \\ \theta(\dfrac{\alpha\pi - \omega}{\alpha + \beta - 1}) & (1-\beta)\pi \leq |\omega| \alpha\pi \\ 1 & \alpha\pi \leq |\omega| \leq \pi \end{cases} \tag{4.6}$$

where α and β are, as explained previously, the scaling parameters in the frequency domain and θ (•) is a function defined as $\theta(\omega) = 0.5(1 + \cos\omega)\sqrt{2 - \cos\omega}, |\omega| \leq \pi$. Among these parameters, β is directly related to Q-factor and the relationship between them is formulated as $Q = (2\text{-}\beta) / \beta$. Also it has previously made $\alpha + \beta > 1$ in order to get redundancy denoted by r which can be calculated by $r = \beta / (1\text{-}\alpha)$. The r represents a compromise between the desired redundancy and increased computational complexity. As the analyzed signal cannot be decomposed into unlimited number of stages, the maximum number of stages J_{max} has to be set according to the signal length as follows (Selesnick, 2011b)

$$J_{max} = \left\lfloor \frac{\log(\dfrac{L}{4(Q+1)})}{\log(\dfrac{r(Q+1)}{r(Q+1)-2})} \right\rfloor \tag{4.7}$$

where $\lfloor z \rfloor$ stands for the largest integer that is less than z, L is the signal length. The number of stages J should not be greater than J_{max}. Therefore, once r, Q-factor and J are specified, a specific wavelet transform is determined. For example, four kinds of eight-stage wavelet transform, as shown in Fig. 4.3, are generated by four different Q-factors respectively. Observing Fig. 4.3, it can be concluded that: a) with different Q-factors, the oscillatory behaviors of TQWT-generated bases are different, and more specifically, the larger the Q-factor, the more oscillation the basis presents, b) the wavelet of each stage is the time-scaling version of another but the Q-factor is not affected by the operation because the oscillation degrees of the wavelet and its time-scaled versions are unchanged, which ensures that a high-oscillation basis is comprised of all high-oscillation wavelets. A specific basis associated with a specific Q consists of J wavelets which are translated and/or scaled versions of each other but all have the same degree of oscillation, due to the same Q-factor. As such, all the bases can be created similarly in a systematic manner.

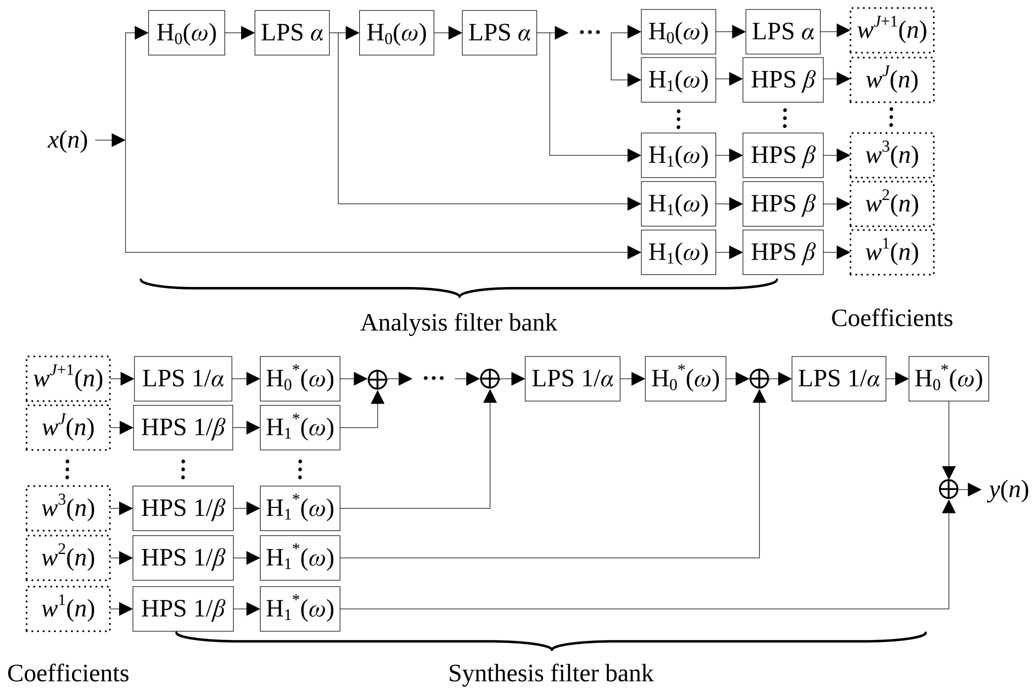


Fig. 4.2 Analysis and synthesis filter banks for the TQWT.

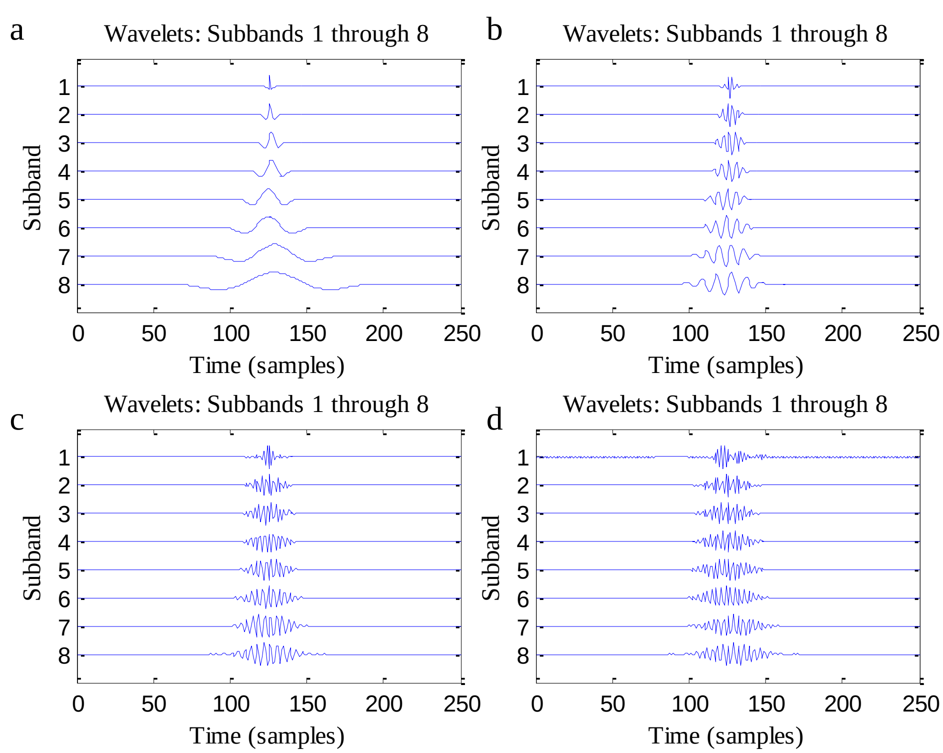


Fig. 4.3 Wavelet subbands of TQWT bases associated with four different Q-factors (r = 3 and J = 8): (a) Q=1, (b) Q=3, (c) Q=7, and (d) Q=10.

4.2.2 Application of TQWT to MCA

Recalling Eq. (4.4), by means of TQWT, the $\mathbf{\Phi}_{1(2)}$ is the matrix that are constituted by the synthesis functions of low(high) Q-factor TQWT and $\mathbf{a}_{1(2)}$ is the coefficients of low(high) Q-factor TQWT, i.e., $\mathbf{w}_{1(2)}$. Then Eq. (4.4) can be re-written as

$$\begin{aligned} J(\mathbf{w}_1,\mathbf{w}_2) &= \gamma_1 \|\mathbf{w}_1\|_1 + \gamma_2 \|\mathbf{w}_2\|_1 + \| x - \mathbf{\Phi}_1\mathbf{w}_1 - \mathbf{\Phi}_2\mathbf{w}_2 \|_2^2 \\ &= \gamma_1 \|\mathbf{w}_1\|_1 + \gamma_2 \|\mathbf{w}_2\|_1 + \| x - \mathrm{TQWT}_1^{-1}(\mathbf{w}_1) - \mathrm{TQWT}_2^{-1}(\mathbf{w}_2) \|_2^2 \end{aligned} \tag{4.8}$$

where TQWT^{-1} denotes the inverse TQWT, and γ_1 and γ_2 are subband-dependent regularization parameters. They are set to be proportional to the l_2-norm of the wavelet at every stage, which are computed as (Selesnick, 2011a)

$$\gamma_1^j = A_1 \| wlet_1^{\ j} \|_2,\ \gamma_2^j = A_2 \| wlet_2^{\ j} \|_2, \tag{4.9}$$

where $wlet_1^j$ and $wlet_2^j$ represent the wavelets of high Q-factor and low Q-factor at jth stage respectively, and A_1 and A_2 are parameters used to tune the weight between estimated signal components and estimation error. With the employment of TQWT, the next problem is to minimizing the objective function of Eq. (4.8) for the purpose of obtaining coefficients $\mathbf{w}_1$ and $\mathbf{w}_2$. The high-oscillation and low-oscillation components of the concerned signal can then be acquired as follows

$$x_k = \mathrm{TQWT}^{-1}(\mathbf{w}_k),\ k = 1, 2\,. \tag{4.10}$$

The minimization of objective function is discussed in the following subsection.

4.2.3 Split Augmented Lagrangian Shrinkage Algorithm (SALSA)

Although the objective function is convex, minimizing it may be difficult due to the non-differentiability of the l_1-norm and the large number of variables. Several algorithms, like Iterated Soft-threshold Algorithm (ISTA) (Daubechies et al., 2004), Fast Iterated Soft-threshold Algorithm (FISTA) (Beck and Teboulle, 2009) and SALSA (Afonso et al., 2010), have been developed for this kind of objective functions. It has been found that the SALSA is particularly effective for OBSD because it involves l_2-norm regularized problem that can be solved easily due to the tight frames of the TQWT (Afonso et al., 2010).

SALSA is based on the techniques called variable splitting to convert an unconstrained optimization problem to a constrained optimization problem. Specifically, the idea is to split the variable into a pair of variables, each to serve as the argument of the function. Next, it minimizes the sum of the functions under the constraint that the two variables have to be equal.

Considering the unconstrained optimization problem Eq. (4.8), the SALSA is based on casting the minimization problem

$$\min_{\mathbf{w}} \; f_1(\mathbf{w}) + f_2(\mathbf{w}), \tag{4.11}$$

where $\mathbf{w}$ are denoted as $\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}$, f_1 and f_2 are functions that are written as $f_1(\mathbf{w}) = \gamma_1 \| \mathbf{w}_1 \|_1 + \gamma_2 \| \mathbf{w}_2 \|_1$ and $f_2(\mathbf{w}) = \| x - \mathbf{\Phi w} \|_2^2$, with $\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \end{bmatrix}$.

In order to obtain constrained optimization problem, variable splitting technique is then used. A new variable, say $\mathbf{v}$, is created to serve as the argument of f_2, under the constraint of $\mathbf{v} = \mathbf{w}$. Then Eq. (4.11) can be rewritten as

$$\min_{\mathbf{w},\mathbf{v}} \; f_1(\mathbf{w}) + f_2(\mathbf{v}) \;\; \text{s.t.}\, \mathbf{v} = \mathbf{w}, \tag{4.12}$$

where $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}$ is used. Clearly, Eq. (4.12) is equivalent to Eq. (4.11); however, Eq. (4.12) is a constrained problem. The augmented Lagrangian function for this problem is defined as (Figueiredo et al., 2009)

$$\Im(\mathbf{w}, \mathbf{v}; \boldsymbol{\lambda}, \mu) = f_1(\mathbf{w}) + f_2(\mathbf{v}) + \boldsymbol{\lambda}(\mathbf{v} - \mathbf{w}) + \mu \| \mathbf{v} - \mathbf{w} \|_2^2 \tag{4.13}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrangian multipliers and μ denotes the penalty parameter. This method is called augmented Lagrangian method (Meireles et al.) (Meireles et al.), also known as the method of multipliers (MM), which lies in minimizing $\Im(\mathbf{w}, \mathbf{v}; \lambda, \mu)$ with respect to $\mathbf{w}$ and $\mathbf{v}$. The so-called ALM method involves an iterative process, updating μ in each iteration until the stop criterion is satisfied; see (Afonso et al., 2010) for details. The main steps of ALM for Eq. (4.13) can be written as follows:

$$(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}) \in \arg\min_{\mathbf{w},\mathbf{v}} \; f_1(\mathbf{w}) + f_2(\mathbf{v}) + \mu \| \mathbf{v} - \mathbf{w} - \mathbf{d}^k \|_2^2 \tag{4.14}$$

$$\mathbf{d}^{k+1} = \mathbf{d}^k - (\mathbf{v}^{k+1} - \mathbf{w}^{k+1}) \;. \tag{4.15}$$

The minimization of Eq. (4.14) is non-trivial as it contains non-separable quadratic and possibly non-smooth terms. In the research work of (Figueiredo et al., 2009; Afonso et al., 2010), it has been found that this minimizing problem can be solved using an algorithm called Alternating Direction Method of Multipliers (ADMM), in which Eq. (4.14) is solved by alternating minimization with respect to **w** and **v**, while keeping the other variables fixed. The resulting steps are in the following.

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} f_1(\mathbf{w}) + \mu \| \mathbf{v}^k - \mathbf{w} - \mathbf{d}^k \|_2^2 \tag{4.16}$$

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} f_2(\mathbf{v}) + \mu \| \mathbf{v} - \mathbf{w}^{k+1} - \mathbf{d}^k \|_2^2 \tag{4.17}$$

$$\mathbf{d}^{k+1} = \mathbf{d}^k - \mathbf{v}^{k+1} + \mathbf{w}^{k+1} \tag{4.18}$$

where k is the iteration index and μ is the penalty parameter specified by a user. Substituting $f_1(\mathbf{w}) = \gamma_1 \| \mathbf{w}_1 \|_1 + \gamma_2 \| \mathbf{w}_2 \|_1$ and $f_2(\mathbf{v}) = \| x - \mathbf{\Phi v} \|_2^2$ into Eq. (4.16) and (4.17) respectively, we can see that Eq. (4.16) is an l_1-norm de-noising problem and Eq. (4.17) is l_2-norm problem (convex quadratic function). The minimization of the Eq. (4.16) can be achieved by the following soft-thresholding equation (Selesnick, 2011a):

$$\mathbf{w}^{k+1} = \mathrm{soft}(\mathbf{v}^k - \mathbf{d}^k, \frac{\boldsymbol{\gamma}}{2\mu}) \tag{4.19}$$

where soft is the soft-thresholding rule with threshold $T = \boldsymbol{\gamma}/(2\mu)$, $\mathrm{soft}(x,T) = \mathrm{sign}(x)\max(0, 1 - T/|x|)$. Since Eq. (4.17) is a strictly convex quadratic problem, the solution of it is unique and can be obtained by derivation:

$$\mathbf{v}^{k+1} = (\mathbf{\Phi}^t\mathbf{\Phi} + \mu\mathbf{I})^{-1}(\mathbf{\Phi}^t x + \mu(\mathbf{w}^k + \mathbf{d}^k)) \tag{4.20}$$

where $\mathbf{\Phi}^t$ is the transposed matrix of $\mathbf{\Phi}$. Since the TQWT is a tight frame, the equations of $\mathbf{\Phi}_1\mathbf{\Phi}_1^t = \mathbf{I}$ and $\mathbf{\Phi}_2\mathbf{\Phi}_2^t = \mathbf{I}$ can be obtained. Using the matrix inverse lemma, it can write:

$$(\mathbf{\Phi}^t\mathbf{\Phi} + \mu\mathbf{I})^{-1} = \frac{1}{\mu}\mathbf{I} - \frac{1}{\mu(\mu+2)}\mathbf{\Phi}^t\mathbf{\Phi}. \tag{4.21}$$

Therefore, the iterative algorithm of SALSA can be summarized as follows:

$$\mathbf{w}^{k+1} = \mathrm{soft}(\mathbf{v}^k - \mathbf{d}^k, \frac{\boldsymbol{\gamma}}{2\mu}) \tag{4.22}$$

$$\mathbf{v}^{k+1} = \frac{1}{\mu}\mathbf{u}^k - \frac{1}{\mu(\mu+2)}\mathbf{\Phi}^t\mathbf{\Phi}\mathbf{u}^k \tag{4.23}$$

$$\mathbf{u}^k = \mathbf{\Phi}^t x + \mu(\mathbf{w}^k + \mathbf{d}^k) \tag{4.24}$$

$$\mathbf{d}^{k+1} = \mathbf{d}^k - \mathbf{v}^{k+1} + \mathbf{w}^{k+1} \ . \tag{4.25}$$

The minimization of Eq. (4.8) can be achieved by the iterative operation constructed by Eq. (4.22) -(4.25). Consequently, each signal component x_1 and x_2 can be obtained using Eq. (4.10).

In summary, the proposed OBSD method contains three key techniques, i.e., the MCA method, TQWT and SALSA. The MCA is used to decompose the concerned signal into signal components with different oscillatory behaviors. The crucial point of the success of MCA is to construct proper bases that are used to distinguish each component. This can be done by the technique called TQWT, where the Q-factor is tunable and easy to be specified. As a result, different bases with different oscillatory behaviors are constructed. The last SALSA is an iterative algorithm and applied to minimizing the objective function of the MCA, and thus the coefficients of each signal component are obtained. The low- and high-oscillation components are then constructed by the inverse TQWT.

4.2.4 The fault feature extraction of bearing and IOBSD

Vibration signals collected from bearings in a mechanical system often are consisted of periodic transient component, interfering signals from shaft misalignment, rotor unbalance and electrical components, etc., and strong background noise. As said previously, periodic transient component generated by bearing faults is often masked by noise and interferences, particularly at the early stage of faults. Therefore, the main challenge for bearing condition monitoring is to extract the periodic transient component from the raw signal polluted with interferences and noise.

Based on the definition of the oscillatory behavior described above, fault-generated transients are considered as the low-oscillation component because it presents non-oscillatory behaviors; in contrast, the cyclic interferences that apparently present sustained oscillation are classified as the high-oscillation component. The noise that cannot be represented by both high- and low-oscillation bases is classified as residual. Then by means of the proposed OBSD method, the transients considered as low-oscillation component can be separated from interferences and noise.

However, the vibration signals of defective bearings in a mechanical system often contain various vibration interferences. Even if such interferences present sustained oscillations and are classified as high-oscillation component, they may be different in terms of the level of oscillation. This is reflected by the fact that some of them exhibit more sustained oscillations and some show fewer sustained oscillations. In such a case, the separation of signal components may not be successful if using only a single Q_2 to represent all sustained oscillatory signal contents. Particularly, the sustained oscillatory behaviors of interferences differ widely. The consequence of using a single Q_2 is that some components with fewer sustained oscillations may be mistakenly classified as low-oscillation component. This will be seen in subsections 4.4 and 4.5.

To solve this problem, more than one Q_2 is needed for sustained oscillatory components. One straightforward method is to use multiple high Q-factors to represent various sustained oscillatory components at a time. However, as mentioned above, the successful decomposition of MCA requires that a basis efficiently represents a component but inefficiently represents others, which means that the bases should be of low coherence (Abrial et al., 2007; Selesnick, 2011a). The main issue caused by using multiple Q_2 at a time is that it may generally reduce the incoherence among bases associated with different Q_2 and accordingly decrease the likelihood of successful decomposition. Inspired by this, the IOBSD approach is proposed. Instead of using more than two bases at a time, the IOBSD iteratively 'peels' the interferences, gradually removing interferences that construct the high-oscillation component. The decomposed low- and high-oscillation components of the signal x are denoted by x_1^i and x_2^i at ith iteration, respectively. Then the IOBSD can be described in the following

$$x_1^1 = \mathrm{IOBSD}(x),\ i = 1 \tag{4.26}$$

$$x_1^i = \mathrm{IOBSD}(x_1^{i-1}),\ i = 2,\ldots. \tag{4.27}$$

The schematic representation of the proposed approach is shown in Fig. 4.4**.** It is noted that the number of iterations has to be properly controlled because excessively decomposing a signal may lead to an over smooth result and thus strong deviation from the target signal (i.e., the transients in this work) is produced. The kurtosis that is widely used as a measurement of signal impulsiveness is adopted to construct the stopping criterion. The underlying principle of applying the kurtosis to bearing fault feature extraction is that the signal component primarily consists of periodic interferences and noise produces a kurtosis value of less than or equal to 3. In contrast, the signal component including impulses results in a kurtosis value greater than 3. In each iteration, the kurtosis value of the high-oscillation component x_2^i is calculated. The reason for calculating kurtosis of high-oscillation component instead of low-oscillation component is that kurtosis values greater than 3 are not necessarily related to the fault-related impulses (Luo et al., 2013). Then comparing the calculated kurtosis value with 3, if it is less than 3 the iterative procedure continues; otherwise it stops and the final result is the low-oscillation component in *i*th iteration, i.e., x_1^i. The reasons are as following: a) a kurtosis value of a signal below 3 is an indicator of periodic interferences and Gaussian noise, and b) the high-oscillation component is mainly composed of periodic interferences and noise, and if a kurtosis value of a high-oscillation component (x_2^i) is greater than 3 it means that no more interferences are left in the low-oscillation component (x_1^i); hence no more iteration is required and the x_1^i is the final result. The stopping criterion is expressed below:

$$SC^i = \text{kurt}(x_2^i) - 3 \geq 0,\ i = 1, \ldots \tag{4.28}$$

where kurt(•) denotes the kurtosis value.

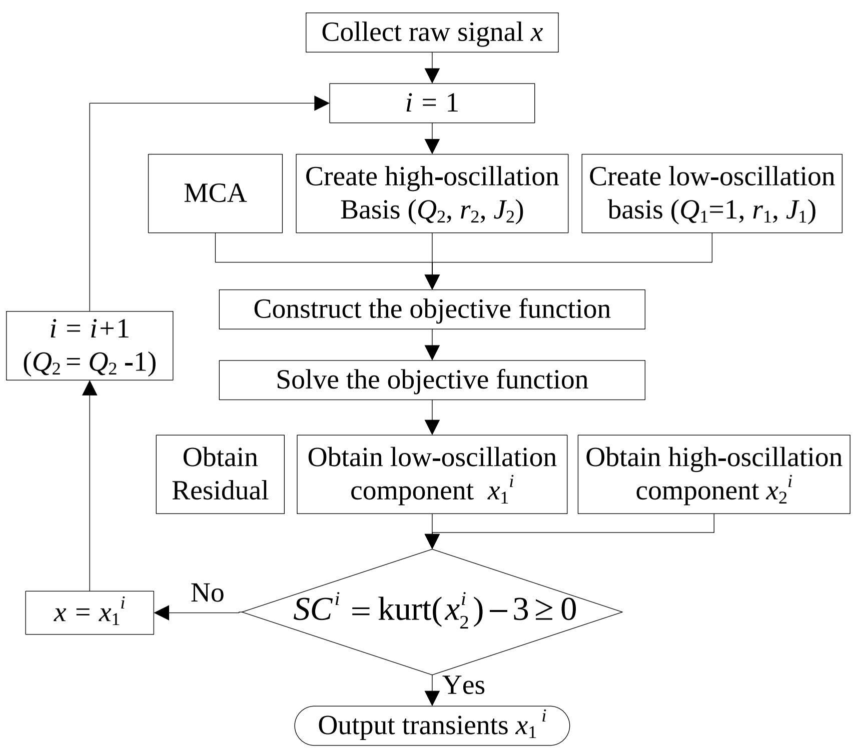


Fig. 4.4 Schematic representation of the proposed IOBSD.

4.3 The parameter determination for IOBSD

4.3.1 Parameters determination for MCA and SALSA

For MCA and SALSA, there are two parameters, i.e., regularization parameters and penalty parameter, to be selected. The former defined in Eq. (4.9) are determined by A_1 and A_2, which tune the relative energy between the two estimated signal components and also affect the smoothness of them. The larger A_1 and A_2, the smoother the resulting signals are. It has been proven by Cai et al. (2013) that A_1 and A_2 can provide an effective decomposition results except the case that the two values are far apart from each other. Therefore, A_1 and A_2 are selected to be very close. Note that A_1 and A_2 cannot be too large in order to avoid over-smooth results. Penalty parameter μ associated with SALSA has an impact on the convergence rate of the algorithm. Large μ leads to a slower convergence rate; however, it results in a small residual. To achieve a desirable balance between the convergence rate and residual, μ is set to 2 in this study.

4.3.2 Parameter determination for TQWT

For TQWT, as mentioned before, three parameters need to be determined: the Q-factor, the number of stage J and redundancy r. To obtain the overcomplete bases and make wavelets cover frequency band as much as possible, the maximum number of stages J_{max} (specified by Eq.(4.7)) is employed in this work at the expense of more computation. Fortunately, the TQWT can be implemented quickly using the radix-2 FFTs. The redundancy r is set to 5 by considering the trade-off between the desired redundancy and complex calculations. Thus, the only remaining issue is to specify the appropriate Q-factors to create proper bases for the low- and high-oscillation components.

4.3.2.1 Q-factor for low-oscillation component (Q_1)

When rolling elements interact with damaged area, a series of impulsive signal is generated as shown in Fig. 4.5 (a). Enlarging Fig. 4.5 (a) and only observing an impulse as displayed in Fig. 4.5 (b), the transient part of the impulse has a ‘similar’ curve shape with the wavelet of Q =1 which is presented in Fig. 4.5 (c). Hence, the Q-factor representing low-oscillation component is set to 1, i.e. $Q_1 = 1$.

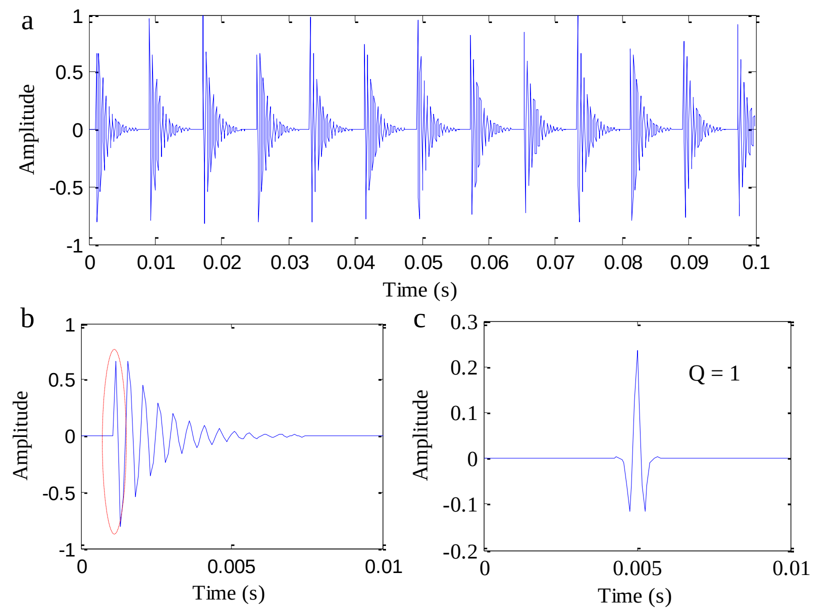


Fig. 4.5 (a) Simulated faulty bearing vibration signal, (b) an impulse from (a), and (c) wavelet with Q = 1.

4.3.2.2 Q-factor for high-oscillation component (Q_2)

As stated previously, to ensure the success of the MCA, bases of Q_1 and Q_2 should be of low mutual coherence. More specifically, a zero value of coherence between two bases indicates complete independence of the two. The mutual coherence between two bases are formulated as

$$C_{max} = \max_{\substack{n\in(1,\dots J_A)\\ l\in(1,\dots J_B)}} C_{(An,Bl)} = \sqrt{\frac{Q_1 + 1/2}{Q_2 + 1/2}}, \quad Q_2 > Q_1 \tag{4.29}$$

where A_n and B_l represent wavelets of A and B at decomposition levels n and l respectively. The right-hand side of the above equation is adopted from the ref. (Selesnick, 2011a). It can be seen that C_{max} depends on the Q-factors of the two bases. Ideally, C_{max} should be as close to zero as possible. For this purpose, Q_2 should be as high as possible. However, too high Q_2 incurs a substantial amount of computations. In the case of bearing fault detection, Q_1 = 1 has been selected. Then the high Q-factor has to be in a suitable range to avoid high computational requirements while in the meantime the coherence between high- and low-bases can be as small as possible.

The relationship between gains of C_{max} and Q_2 is plotted in Fig. 4.6. The gain of C_{max} denoted by ΔC_{max} is expressed by

$$\Delta C_{max}(Q_2) = C_{max}(Q_2 - 1) - C_{max}(Q_2), \quad for \;\; Q_2 \geq 2 \tag{4.30}$$

where $C_{max}(Q_2)$ is the maximal coherence value between the Q_1 basis and a given Q_2 basis. By observing Fig. 4.6, it can be seen that the gain of C_{max} drops monotonically with the increment of Q_2, and it almost keeps unchanged for Q_2 larger than 10 (less than 1.4%). This means that the contribution made by increasing Q_2 to the reduction of C_{max} is quite significant at the beginning; however, it drops monotonically with the increase of Q_2 and is negligible after Q_2 reaching 10. As seen in Fig. 4.6, most of changes of ΔC_{max} is achieved when $Q_2 \leq 10$. Quantitatively, the C_{max} is decreased by 62.2% (= $\sum_{Q_2=2}^{10} \Delta C_{max}(Q_2)$) when Q_2 reaches 10. Keeping increasing Q_2 from 11 to 20 can reduce the C_{max} by 10.6%; however it incurs complex computation. For example, for a signal with 32768 (=8192*4) data length, the J_{max} is 178 (calculated by Eq. (4.7)) for Q_2 equal to 10 while the J_{max} reaches 310 when increasing Q_2 to 20. Considering the negligible contribution and the substantial amount of computations caused by additional increment in Q_2, the maximal Q_2 is fixed at 10.

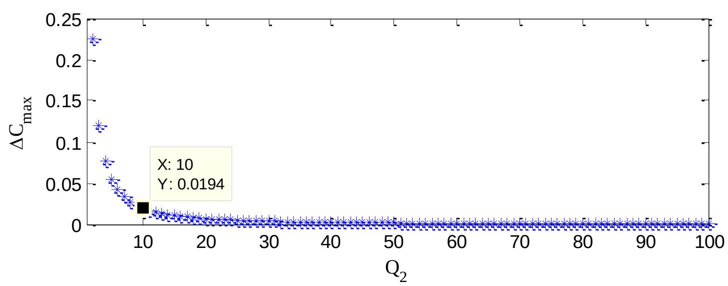


Fig. 4.6 The evolving history of ΔC_{max} versus Q_2

Based on the discussion above, the steps of high Q-factor determination for IOBSD are as follows. The interferences with more sustained oscillations are removed in the previous iteration and then interferences with fewer sustained oscillations are removed in the next iteration by reducing the Q_2. This can be done by setting the first Q_2 (in the first iteration) equal to the maximal Q_2 (=10) and then reducing Q_2 by 1 for next iteration. The iterative operation continues and Q_2 is kept reducing until the stop criterion is satisfied or $Q_2 < 2$. In this way, the Fig. 4.4 where the schematic representation of IOBSD is presented can be modified as shown in Fig. 4.7.

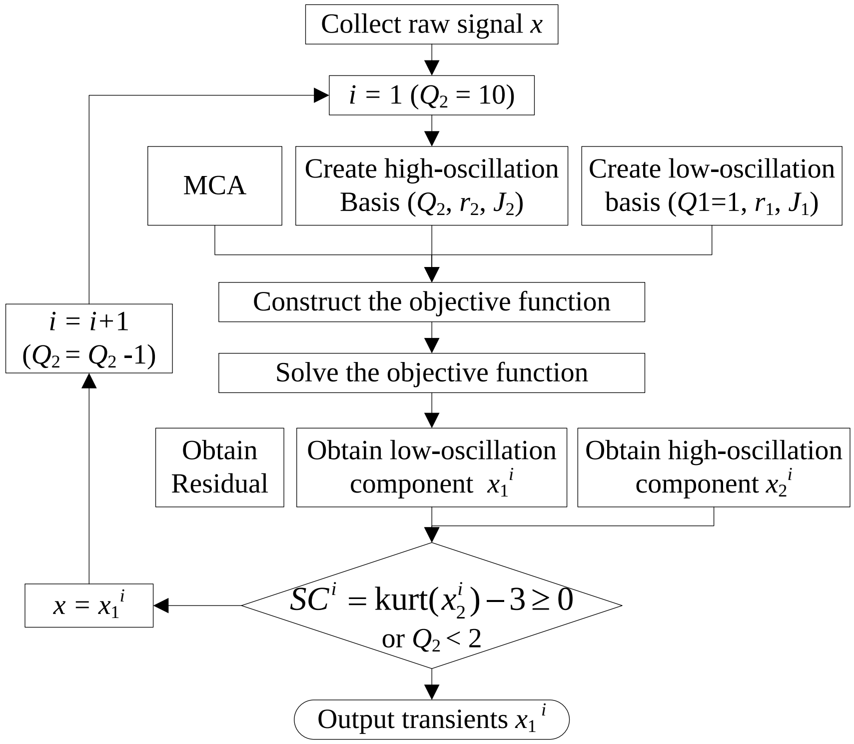


Fig. 4.7 Schematic representation of the proposed IOBSD method after Q_2 has been determined.

Note that the stopping criterion that is defined by Eq. (4.28) is often satisfied first. It is observed that eight iterations are sufficient for most applications for defective bearing vibration signal analysis.

4.4 Simulation assessment

To investigate the effectiveness of the proposed method, the simulated data of faulty bearings with noise and a variety of interfering signals are constructed, and then are processed by the proposed IOBSD method. The comparison between the IOBSD and the SK method is also made.

4.4.1 Faulty bearing simulated signal with slippage

The simulated signal of a faulty bearing with slippage can be modeled as (Liang and Bozchalooi, 2010):

$$s(t) = \sum_{m=-M}^{M} L_m e^{-\beta(t - mT_p - \sum_{i=-M}^{m} \tau_i)} \sin(\omega_r (t - mT_p - \sum_{i=-M}^{m} \tau_i)) \mathrm{u}(t - mT_p - \sum_{i=-M}^{m} \tau_i) \qquad (4.31)$$

where M is half of the number of the fault generated impulses, L_m is the amplitude of mth fault impulse, β the decay parameter, T_p is the time period corresponding to the fault characteristic frequency, τ_i presents the effect of random slippage of the rollers and is modeled as a uniformly distributed random variable with a zero mean and a standard deviation of $0.01T_p$, ω_r is the excited resonance frequency, and $u(\cdot)$ is the unit step function. The parameters of the simulated model mentioned above are given in Table 3.1. The sampling frequency is 12 KHz and the signal length is 8192*4=32768. The raw simulated signal of faulty bearing with slippage and its spectrum are plotted in Fig. 4.8 (a) and (b), respectively.

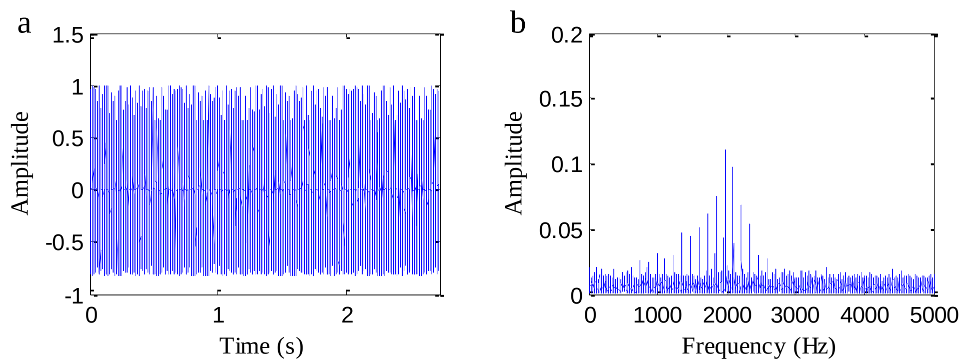


Fig. 4.8 Simulated signal of a defective bearing with slippage: (a) the simulated signal in time domain and (b) frequency representation of (a).

4.4.2 The application of the IOBSD to the simulated data

To imitate the real working condition of bearings, interferences originated from other machine elements and noise are added into the simulated signal,

$$x(t) = s(t) + n(t) + \mathrm{I}(t) \tag{4.32}$$

where $s(t)$ is the simulated faulty bearing signal as described in subsection 4.4.1, $n(t)$ is the additive white Gaussian noise and $\mathrm{I}(t)$ is the interfering signal. The main sources of interferences are vibrations of other mechanical components and its multiples, such as unbalanced rotor, misaligned shaft, gears as well as some signal contents from electrical devices. These interferences are discrete in frequency and hence they are also called discrete frequency noises, which can be expressed as

$$\mathrm{I}(t) = \sum_{m=1}^{M} \mathrm{I}_m(t) = \sum_{m=1}^{M} A_m \cos(\omega_m t - \varphi_m) \tag{4.33}$$

where A_m, ω_m and φ_m are the amplitude, frequency and initial phase respectively. In this simulated case, there are six interferences (I_1, I_2,…, I_6) added to the simulated impulsive signal s and their frequencies are 4.8 KHz, 2 KHz, 1.9 Hz, 1 KHz, 300 Hz and 150 Hz respectively. The frequency of I_3 is in the vicinity of the fault-induced resonance frequency and the frequency of I_2 is exactly identical to the resonance frequency, therefore they can be used to evaluate the in-band interferences removal ability of the proposed method. The amplitudes and initial phases of all interferences are set to 1 and 0 respectively. The SNR related to the simulated faulty bearing signal s is -7. The composite signal is plotted in Fig. 4.9 (a). Fig. 4.9 (b) presents the frequency spectrum of the raw signal, where all frequency contents of interferences can be clearly seen. The envelope spectrum of the raw signal is illustrated in Fig. 4.9 (c), showing that the interference frequencies and their differences stand out of the spectrum and hence the FCF and its harmonics cannot be observed. Therefore, the classical Fourier transform and Hilbert envelope demodulation are incapable of dealing with this signal which is severely tainted by noise and a wide range of interferences. Moreover, the frequencies of I_2 and I_3 reside in the frequency band covered by the resonance frequency. In this case, filters-based approaches may work ineffectively as interference frequencies and resonance frequency share the same frequency band. One of the state-of-the-art filtering methods, i.e., SK method, is employed to process the signal mixture. The kurtogram, the envelope of the filtered signal and its spectrum are shown in Fig. 4.10. It can be seen in Fig. 4.10 (a) that the center frequency of the SK-selected pass-band is 210 Hz; however, the resonance frequency of the simulated signal is 2 KHz. This indicates that the in-band interferences adversely affect the selection of proper pass-band as they have effects on kurtosis values. The envelope of the filtered result and its spectrum are exhibited in Fig. 4.10 (b) and (c), respectively. As seen in Fig. 4.10 (c), neither FCF nor its harmonics are recognizable, indicating that SK method is no longer effective in such case.

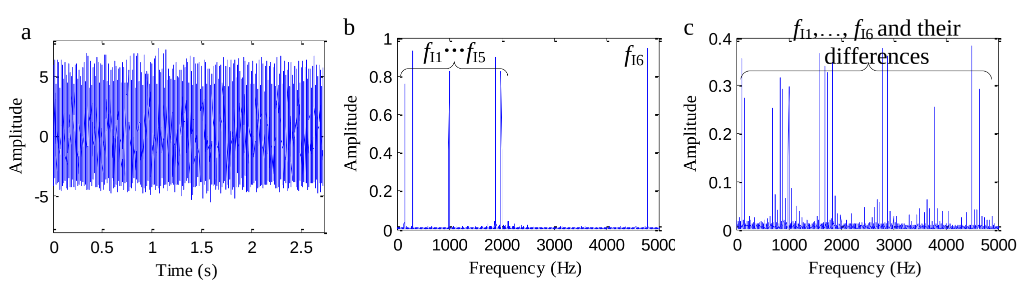


Fig. 4.9 (a) the simulated impulsive signal mixed with noise (SNR = -7 dB) and seven interferences (SIR = -18.8 dB), (b) the FFT spectrum, and (c) the envelope spectrum.

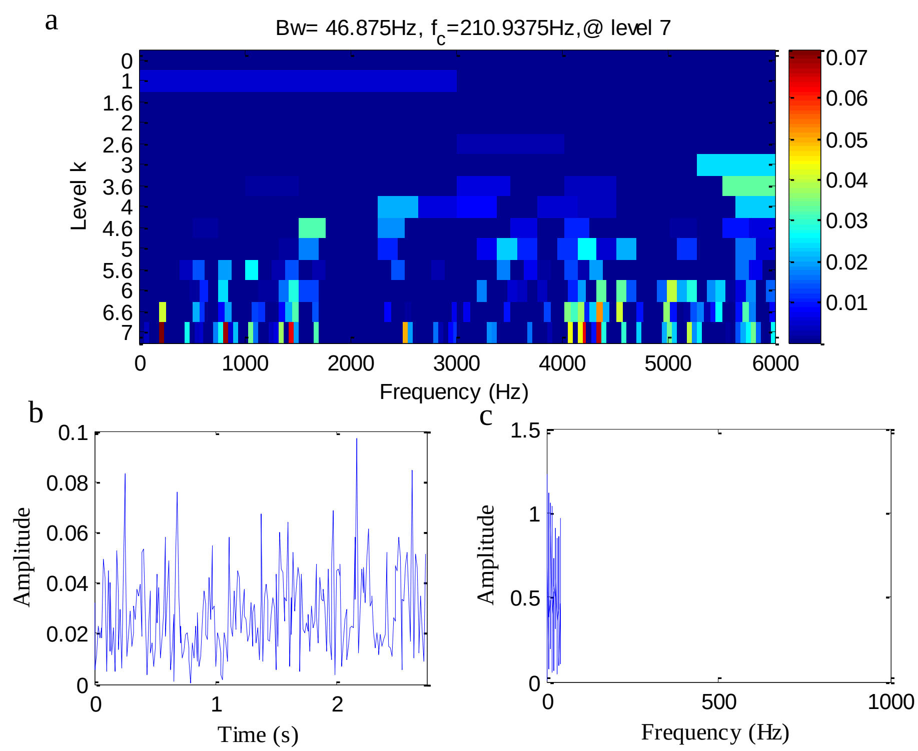


Fig. 4.10 Processing results using SK method: (a) the kurtogram obtained by SK method, (b) the envelope of the filtered signal, and (c) the spectrum of the envelope.

Next, the data is analyzed by the proposed approach. The selection of Q_1 and Q_2 has been discussed previously. The parameters A_1 and A_2 that are associated with the minimization of objective function should not be far away from each other and also not be very large; hence both of them are set 1. Two iterations for IOBSD are carried out for this case (kurtosis values of high oscillation components are 2.91 and 10.51, respectively), hence the outputs of the second iteration are the final results. The outputs of the first iteration are shown in Fig. 4.11. The minimization of the objective function is illustrated in Fig. 4.11 (a), showing that the objective function converges using the SALSA algorithm with the selected A_1 and A_2. Fig. 4.11 (b) displays the residual that is represented by neither the high-oscillation basis nor the low-oscillation basis. The obtained high-oscillation component, as shown in the left side of Fig. 4.11 (c), is mainly consisted of signal components with sustained oscillation, i.e., the interferences. This can be validated by its spectrum in middle of Fig. 4.11 (c), in which the dominating frequency contents are the frequencies of interferences. The envelope spectrum of the high-oscillation component is plotted in the right side of Fig. 4.11 (c). As observed, no frequency information related to transients is revealed in both the spectrum and envelope spectrum of high-oscillation component. The obtained low-oscillation component is plotted in the left side of Fig. 4.11 (d) and its corresponding FFT spectrum is illustrated in the middle of this row. Although the interference frequencies are still observable in the FFT spectrum of low-oscillation component, the amplitudes of them are much smaller. This indicates that the first iteration separates a large portion of interferences from the transients but not all of them. In this case, the envelope spectrum of low-oscillation component, as shown in the right side of Fig. 4.11 (d), still has the following two problems: a) in the low frequency band, the FCF and its harmonics are still clouded by interference frequencies and their differences, and b) there are four obvious frequency lines associated with interferences and their differences standing out the high frequency band. The two issues show that some interferences and/or their differences are also classified as low-oscillation component as their oscillatory behavior is far away from that of the current high-oscillation basis but close to that of the low-oscillation basis.

Thankfully, these problems mentioned above can be solved by the proposed IOBSD. With the previous low-oscillation component as the input signal of the next iteration, the iterative operation keeps being carried out until the stopping criterion is met. The outputs of the second iteration are shown in Fig. 4.12. The minimizing history of the objective function and the obtained residual are displayed in Fig. 4.12 (a) and (b), respectively. The components with sustained oscillatory behaviors are further classified as high-oscillation component, as shown in Fig. 4.12 (c) (left side), and the transients remain being considered as low-oscillation component illustrated in Fig. 4.12 (d) (left side). Their FFT spectra are respectively shown in the middle of Fig. 4.12 (c) and (d). As can be seen, the FFT spectrum of high-oscillation component is dominated by low frequency interferences which are not observable in the spectrum of low-oscillation component, showing that the second iteration further separates the remaining interferences from transients. The envelope spectrum of the low-oscillation component is presented in Fig. 4.12 (d) (right side), where the FCF at 125 Hz and four its harmonics can be easily recognized.

To verify the effects of multiple interferences including in-band ones on the SK method, the six interferences are removed from the simulated signal and the resulting signal is processed by the SK method. As shown in Fig. 4.13, the selected center frequency is 1875 Hz and bandwidth is 750 Hz, covering the frequency band that resonant frequency is located. Therefore, it is sufficient to say that the interferences have adverse effects on the filter parameter selection. The proposed method in this chapter is free from filters and hence can deal with this kind of signal based on the oscillatory behaviors. The above simulation illustrates that the interference handling capability is improved by proposing the IOBSD. The comparisons also show the superiority of proposed method to the widely used SK method in terms of processing signals with multiple interferences.

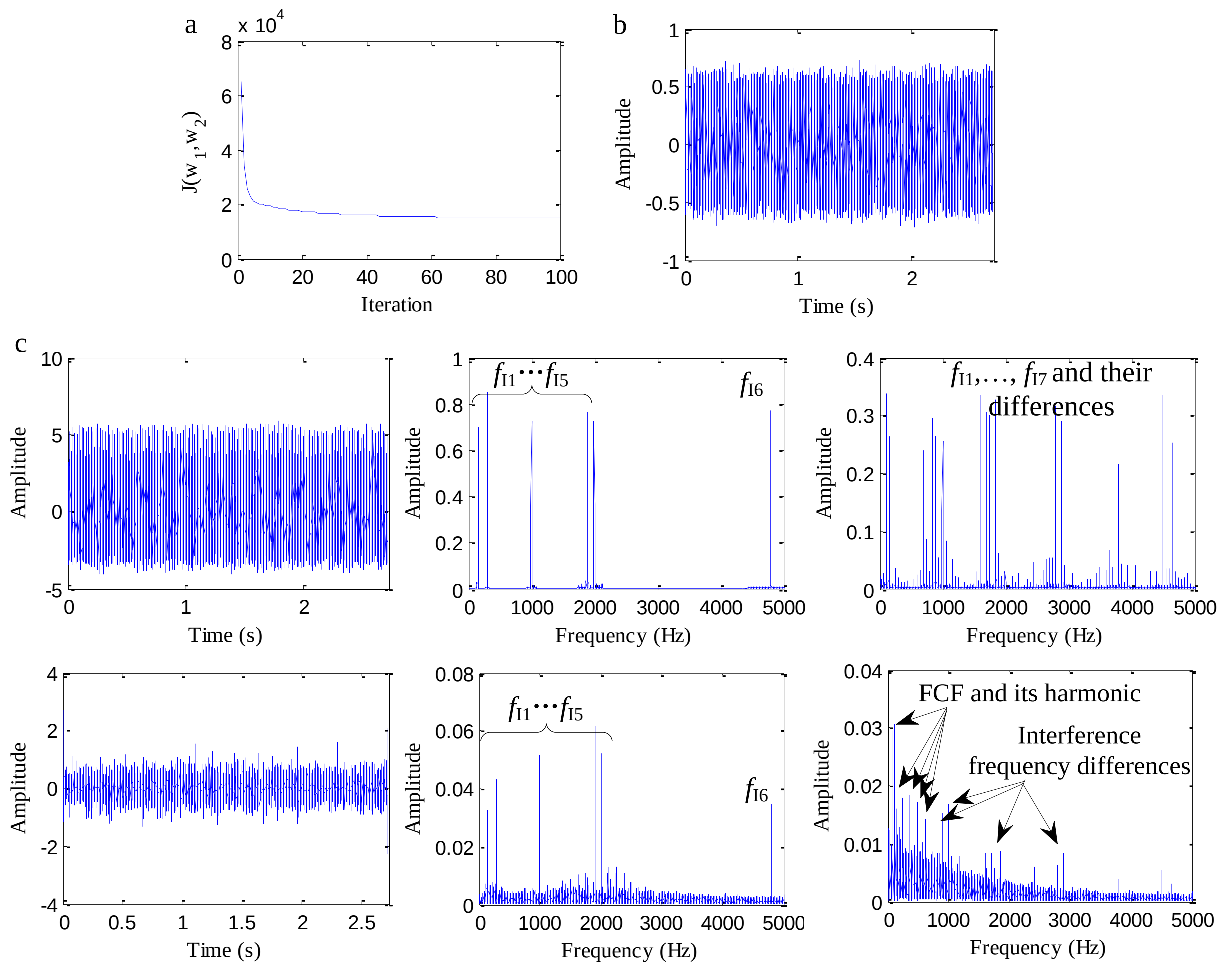


Fig. 4.11 The detecting result of the composite signal by the proposed method (in the first iteration): (a) the reduction of objective function, (b) the residual, (c) the high-oscillation component, its FFT spectrum and envelope spectrum, and (d) the low-oscillation component, it FFT spectrum and the envelope spectrum.

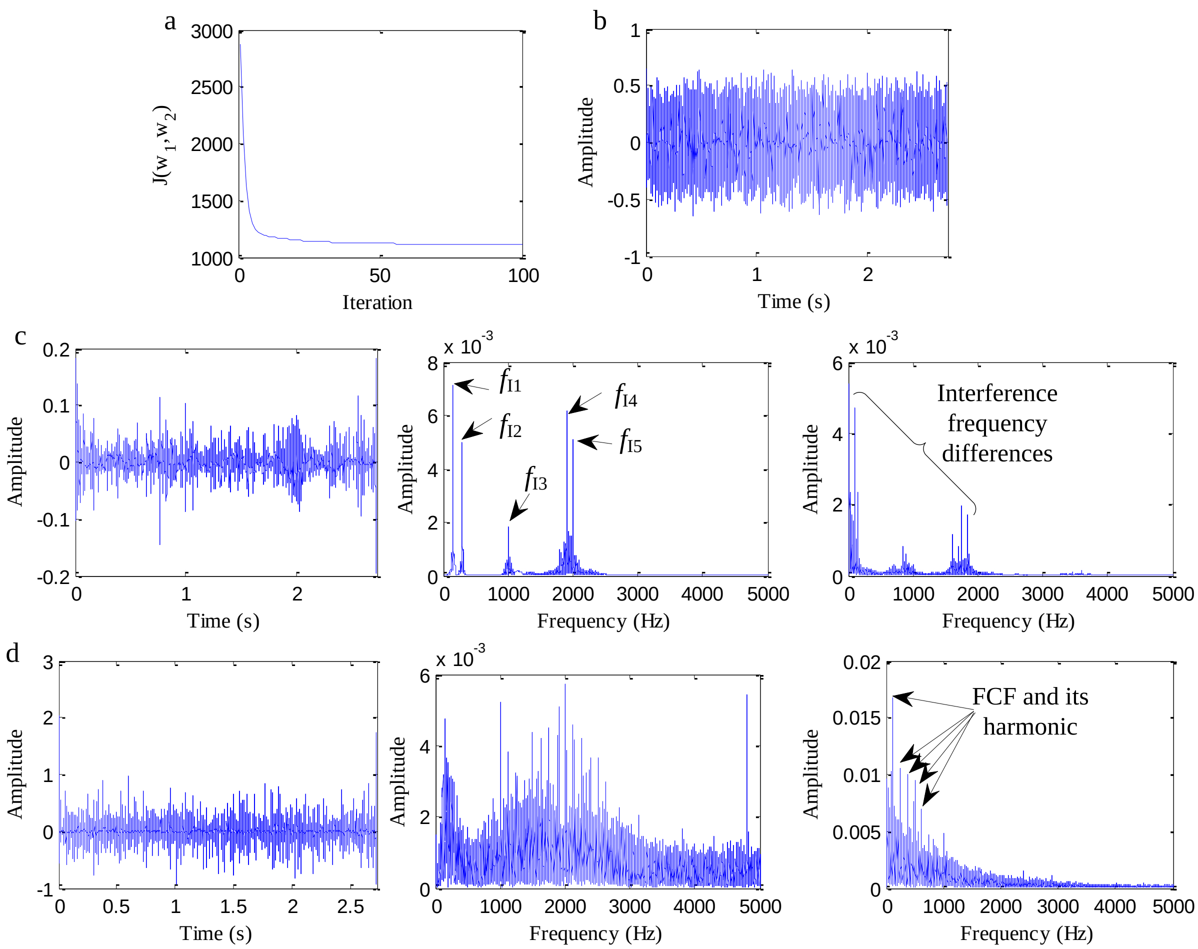


Fig. 4.12 The final detecting result of the composite signal by the proposed method (in the second iteration): (a) the reduction of objective function, (b) the residual, (c) the high-oscillation component, its FFT spectrum and envelope spectrum, and (d) the low-oscillation component, its FFT spectrum and envelope spectrum.

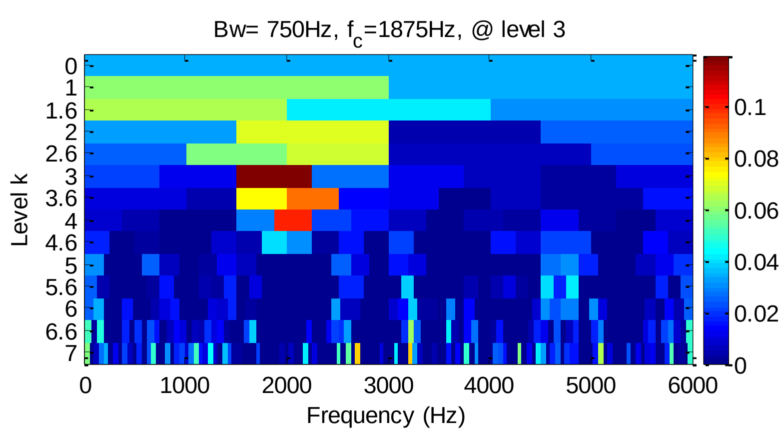


Fig. 4.13 The kurtogram obtained by SK for the above simulated signal without interferences.

4.4.3 Noise and interference handling abilities of the proposed OBSD method

As revealed above, the proposed technique can detect the faults from vibration signal contaminated by specific noise and interference level. In this subsection, the capabilities of dealing with different noise and interference levels of the method will be evaluated.

The simulated impulse-like signal generated by Eq. (4.31) is polluted by different noise levels (SNR ranging from 2 to -12) and then processed using the reported OBSD method. The processing result is displayed Fig. 4.14 (a), illustrating that the proposed method can deal with signal with SNR as low as - 10 dB. To test the interference handling capability, the simulated impulsive signal of SNR = - 7 is mixed by different levels of interferences. Fig. 4.14 (b) presents the detection performance of the OBSD, suggesting that the fault can be detected from signal with SIR ≥ - 25 dB. It is worth mentioning that no prefiltering processes are involved and only a single OBSD is used for both cases. If the IOBSD is applied, the detectable region in terms of both noise and interference handling capabilities should be broadened.

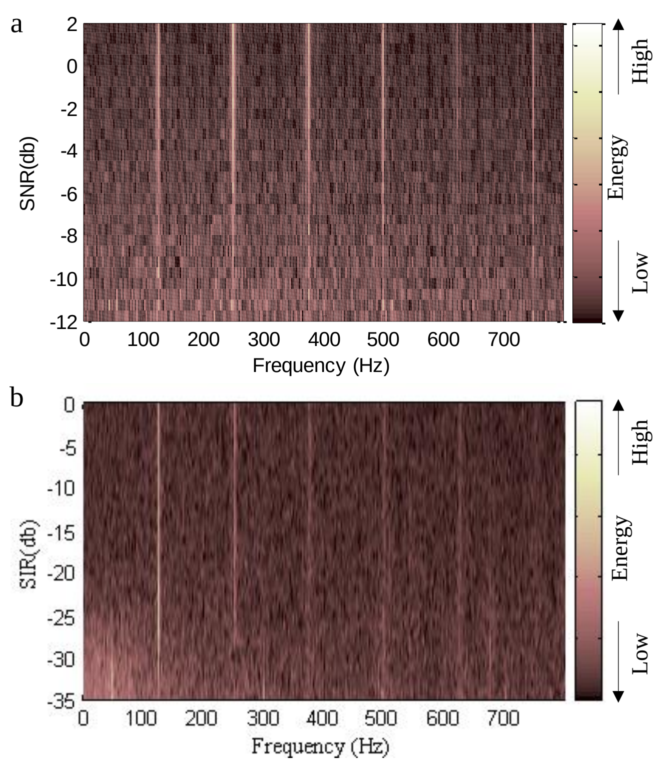


Fig. 4.14 Noise and interference handling abilities of the proposed OBSD method: (a) processing result for simulated impulsive signal with different noise levels, and (b) processing result for simulated noisy signal (SNR = - 7 dB) with different interference levels.

4.5 Experiment evaluation

To further validate the proposed method, the experimental data collected in Chapter 3 are utilized again. The data points used in this chapter are 2^{17}=131,072 and signal length is 6.55 sec (= 131,072/20,000). Except for the data used in Chapter 3, additional experiments are also carried out to further test the proposed method in this section.

4.5.1 Outer race fault detection

The experimental setup and related information have been presented in section 3.5. Again, the vibration signals of side (middle) and bottom-positioned outer race faults are to be processed, respectively. The raw vibration data for the side (middle)-positioned outer race fault and its envelope spectrum are plotted in Fig. 4.15 (a) and (b) respectively. As expected, no sign of a bearing fault could be observed in Fig. 4.15 (b) and the frequency spectrum is again dominated by the interferences, i.e., the shaft rotational frequency f_{I1} (18 Hz) and the power line frequency f_{I2} (60 Hz) as well as their harmonics.

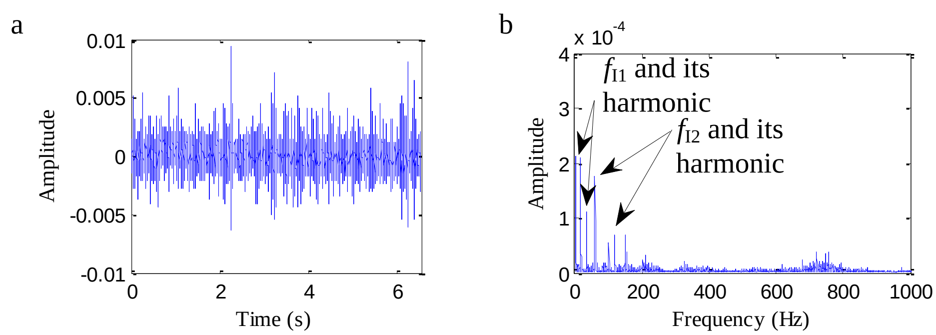


Fig. 4.15 Detecting result of outer race with middle fault by Hilbert envelope demodulation: (a) the raw vibration signal, and (b) the Hilbert envelope spectrum.

Then, the proposed method is then employed to process the vibration signal in Fig. 4.15 (a). The parameters A_1 and A_2 in this case are set to 1.35 and the algorithm iterates once in this case (kurtosis of high-oscillation component is 4.2392). The final results that are the outputs of the first iteration are shown in Fig. 4.16. Fig. 4.16 (a) illustrates that, with the selected parameters, the objective function converges after around ten iterations. The resulting high-resonance component that is mainly consisted of sustained oscillations, i.e., cyclic interferences caused by unbalanced rotor and the power line frequency, is exhibited in Fig. 4.16 (b). This can also be reflected by the spectrum of high-oscillation component in Fig. 4.16 (c), where the f_{I1} together with its harmonic and f_{I2} dominate the spectrum, indicating that high-oscillation component primarily contains harmonic interferences. The extracted transient signal and its envelope spectrum are shown in Fig. 4.16 (d) and (e) respectively. The impulsive features in Fig. 4.16 (d) are more evident, leading to a better detecting result in Fig. 4.16 (e) where the f_{BPOF} and several its harmonics can be clearly identified. The existence of the fault has been determined.

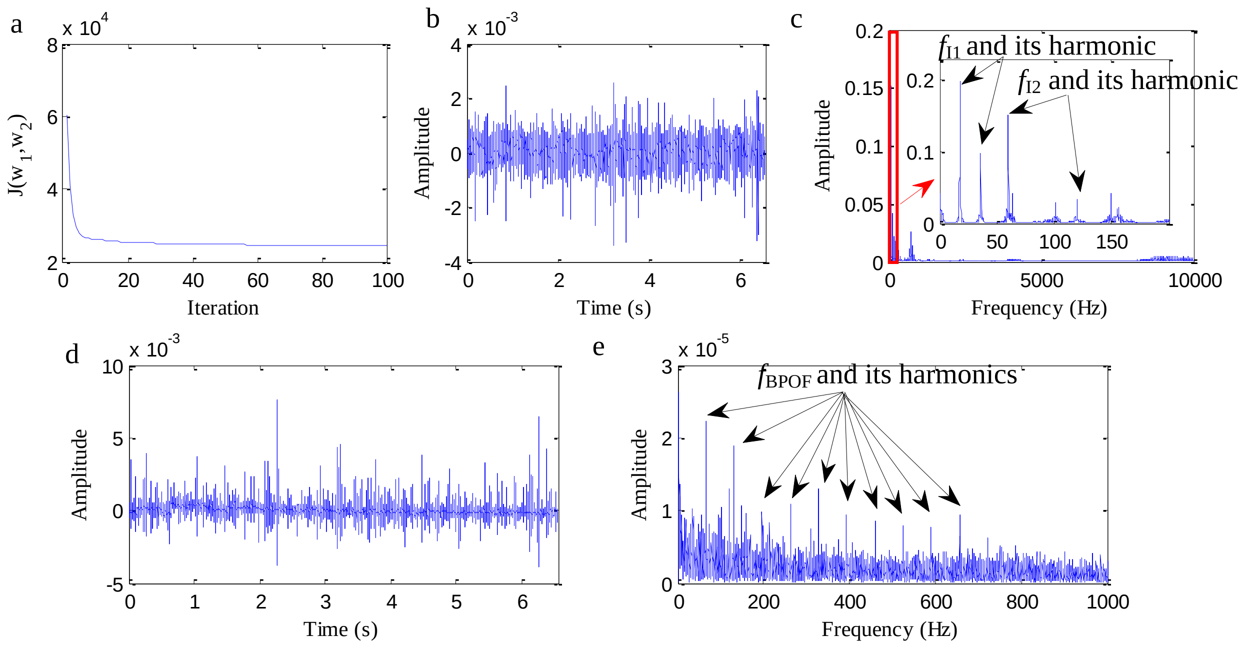


Fig. 4.16 The detecting result of outer race with middle fault by the proposed method: (a) the reduction of objective function, (b) the high-oscillation component, (c) frequency spectrum of (b), (d) the low-oscillation component, and (e) the envelope spectrum of (d).

The same data from middle fault is analyzed by the SK method. The resulting kurtogram and envelope of the filtered signal are shown in Fig. 4.17 (a) and (b), respectively. Fig. 4.17 (c) presents the frequency representation of the envelope signal, in which the f_{BPOF} and its harmonics are observable. However, compared with the proposed method, there are not as many harmonics as in the Fig. 4.16 (d), which indicates that the optimal center frequency and bandwidth are not selected by the SK method.

Next, the signal from bottom positioned outer race fault is analyzed following the same procedure as for side-positioned outer race fault. However, as the outer race fault located in the bottom of the fault is further away from the sensor (compared with the side-positioned outer race fault) the impulsive signal is much weaker. The raw signal is exhibited in Fig. 4.18 (a) and its

Hilbert envelope spectrum is plotted in Fig. 4.18 (b). Likewise, only the shaft rotational frequency f_{I1} (18 Hz) and the interference f_{I2} (60 Hz) together with their harmonics can be observed. No information associated with the fault is revealed by Hilbert demodulation.

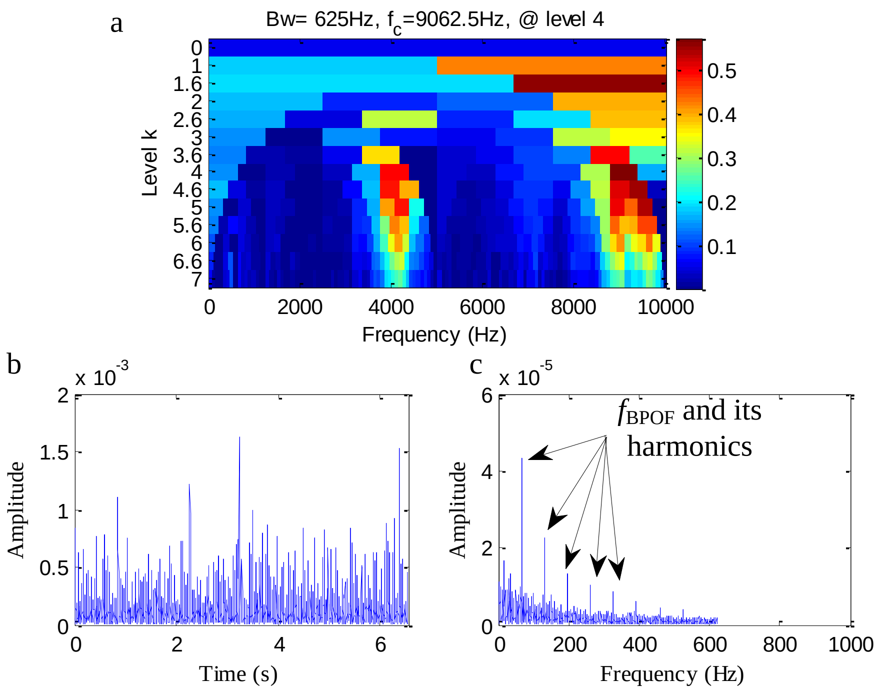


Fig. 4.17 The detecting results of outer race with middle fault using SK method: (a) the kurtogram, (b) the envelope of filtered signal, and (c) the spectrum of the envelope.

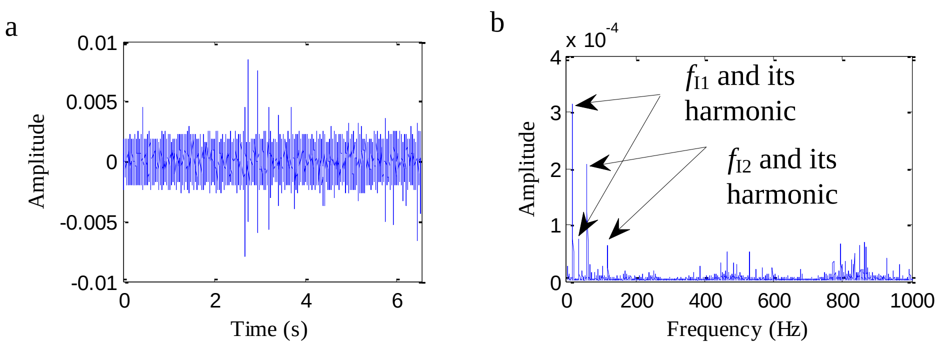


Fig. 4.18 The Detecting result of outer race with bottom fault by Hilbert envelope demodulation: (a) the raw vibration signal, and (b) the Hilbert envelope spectrum.

The proposed method is then employed to better the detecting result. The parameters A_1 and A_2 are selected to be 1.35. The algorithm iterates once (kurtosis value of the high-oscillation component is 3.1244.) and the final results are shown in Fig. 4.19. It can be seen from Fig. 4.19 (a) that the convergence of the objective function is achieved. As expected, the high-oscillation component in Fig. 4.19 (b) is constituted by the interferences, which is better reflected by its spectrum as presented in Fig. 4.19 (c) where the main frequency contents are f_{I1}, f_{I2} and their harmonics. The periodic transient signal, i.e., the low-oscillation component, is displayed in Fig. 4.19 (d) and the corresponding envelope spectrum is demonstrated in Fig. 4.19 (e). Comparing the raw signal in Fig. 4.18 (a) with the obtained low-oscillation component in Fig. 4.19 (e), it can be seen that the effect of interferences is reduced and hence the impulses are more pronounced. The more obvious improvement can be observed in the envelope spectrum of the low-oscillation component, as presented in Fig. 4.19 (e), where the f_{BPOF} at 65.9 Hz and three its harmonics are easily identified.

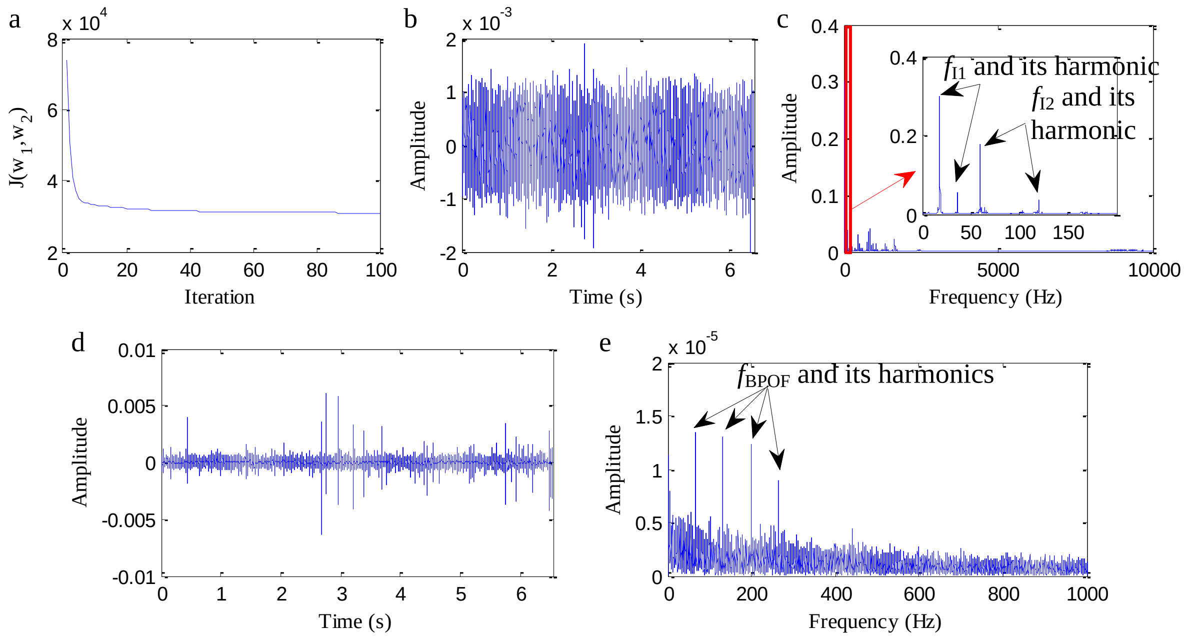


Fig. 4.19 The detecting result of outer race with bottom fault by the proposed method: (a) the reduction of objective function, (b) the high-oscillation component and its spectrum, (c) the low-oscillation component, and (d) the envelope spectrum of (c).

To compare the proposed method with the widely used SK method, the same signal from the bottom fault is also analyzed by the SK method. Fig. 4.20 (a) and (b) respectively display the kurtogram obtained by SK method and the envelope of filtered signal. The spectrum of the envelope in Fig. 4.20 (b) is illustrated in Fig. 4.20 (c), where the effect of interferences still exists and no information related to the bottom fault can be seen. In such case, the SK method fails to detect the fault. There are two possible reasons for this: a) due to high level of noise and interferences, the SK method is incapable of determining the right center frequency and bandwidth, and b) even though the right pass-band is properly determined, the fault-related information still cannot be seen because of strong in-band noise and interferences. The reason that the FCF and its harmonics of middle fault still can be detected even if the optimal pass-band is not selected is that the impulse-like signal is stronger than that of the case of bottom fault.

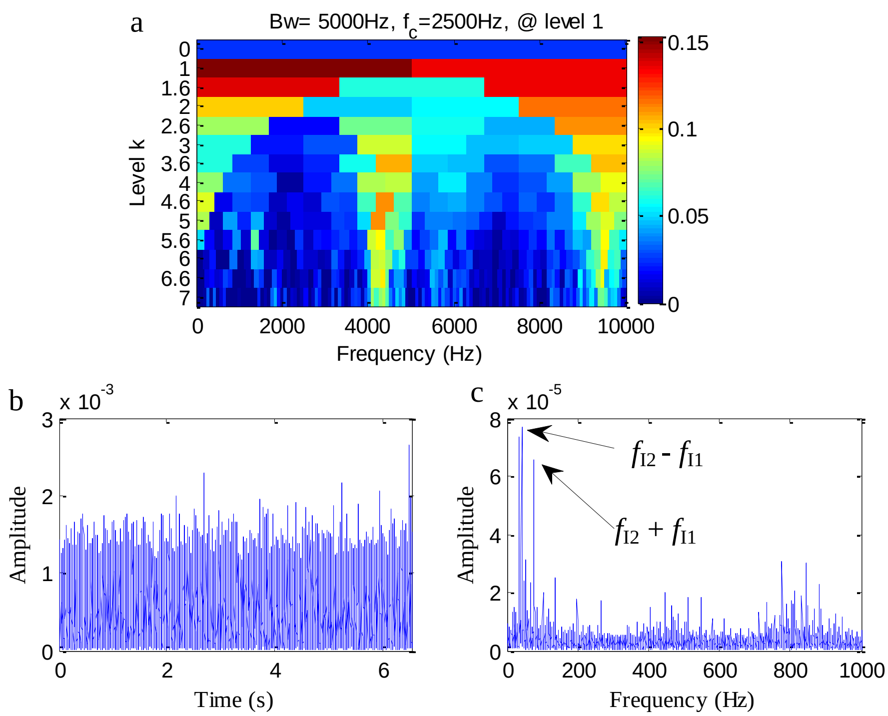


Fig. 4.20 The detecting results of outer race with bottom fault using SK method: (a) the kurtogram, (b) the envelope of filtered signal, and (c) the spectrum of the envelope.

4.5.2 *Inner race fault detection*

The bearing with inner race faults is of the same type as the one used in the outer race fault tests. The experimental setup and data are the same as those reported in Chapter 3. The shaft rotating frequency f_r is 18 Hz (1080 rpm) and vibration signals are sampled at a rate of 20KHz. The faulty characteristic frequency denoted by f_{BPIF}, according to parameters listed in , can be calculated as 97.66 Hz. The raw signal as shown in Fig. 4.21 (a) is first analyzed by Hilbert envelope demodulation. The result is presented in Fig. 4.21 (b), where the interference (f_{I1} = 18 Hz) caused by the unbalanced mass and its harmonics are identified. The frequency line (f_{I2} = 60 Hz) between the second and third harmonics is the power line frequency. However, no fault-related frequency peaks can be observed in Fig. 4.21 (b).

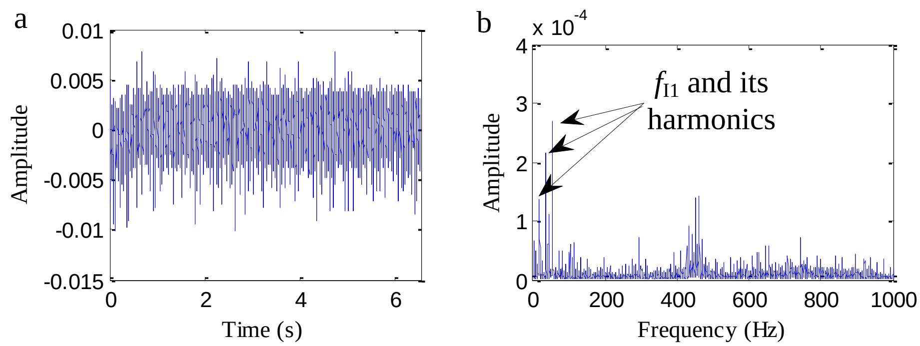


Fig. 4.21 Detection of inner race fault using Hilbert envelope demodulation: (a) the raw signal from the bearing with inner race fault, and (b) its Hilbert envelope spectrum.

The results obtained by the IOBSD method are displayed in Fig.4.22 and Fig. 4.23. The kurtosis values of each high-oscillation component are shown in Table 4.1. Since the kurtosis of the 4th high-oscillation component is larger than 3, four iterations for the proposed method are carried out in this case, with Q_2 being 10, 9, 8 and 7 respectively. The parameters of objective function A_1 and A_2 in this case are selected to be 0.1 and 0.15, respectively. Fig.4.22 presents the processing results of the 1st iteration. From Fig.4.22 (a), it can be seen that the minimization of the objective function is achieved. The residual which can be neither represented by low-oscillation basis nor high-oscillation basis is plotted in Fig.4.22 (b). The resulting high-oscillation component and its spectrum are displayed in Fig.4.22 (c) and (d). In the spectrum, the main frequency contents are the interference frequency and its harmonic, which demonstrates that the main signal component in the high-oscillation component is the interference. By looking at Fig. 4.21 (e) and (f) where the resulting low-oscillation component and its spectrum are presented respectively, the interference frequency f_{I1} and its harmonics still stand out the spectrum. Neither FCF of inner race (denoted by f_{BPIF}) nor its harmonics can be identified. Hence, one iteration is not sufficient to eliminate the effect of interferences in this case because the fault-induced impacts are too weak and severely corrupted by strong interferences.

Table 4.1 Kurtosis values of high-oscillation component at each iteration.

Iteration No.	1st	2nd	3rd	4th
Kurtosis values	2.3246	2.4646	2.1273	3.6491

Therefore, the iterative procedure continues until the stopping criterion is satisfied. The final results, i.e., the results of the fourth iteration, are presented in Fig. 4.23. As before, Fig. 4.23 (a) illustrates that the objective function converges fast and Fig. 4.23 (b) shows the residual of the decomposition. The high-oscillation component and its FFT spectrum of this iteration are exhibited in Fig. 4.23 (c) and (d), respectively. It can be seen that the main content of the high-oscillation component is the interference since its spectrum is dominated by the interference frequency. The extracted transient signal is shown in Fig. 4.23 (c), where the effect of interferences is weakened to a great extent and the impulsive characteristic is more pronounced. Therefore, the f_{BPIF} at 97.66 Hz and six its harmonics are recognizable in its envelope spectrum, as plotted in Fig. 4.23 (d).

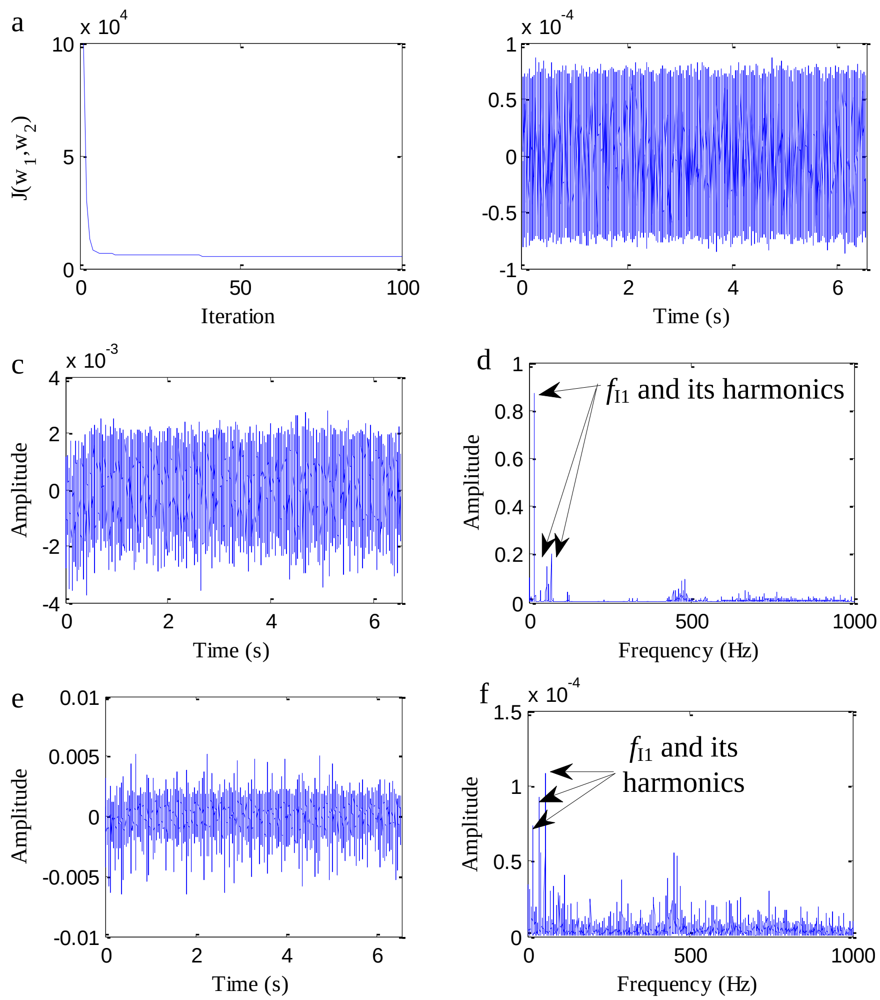


Fig.4.22 Detection of inner race fault using the IOBSD method (results of the first iteration): (a) the reduction of objective function, (b) the residual, (c) the high-oscillation component and (d) its FFT spectrum, (e) the low-oscillation component and (f) its envelope spectrum.

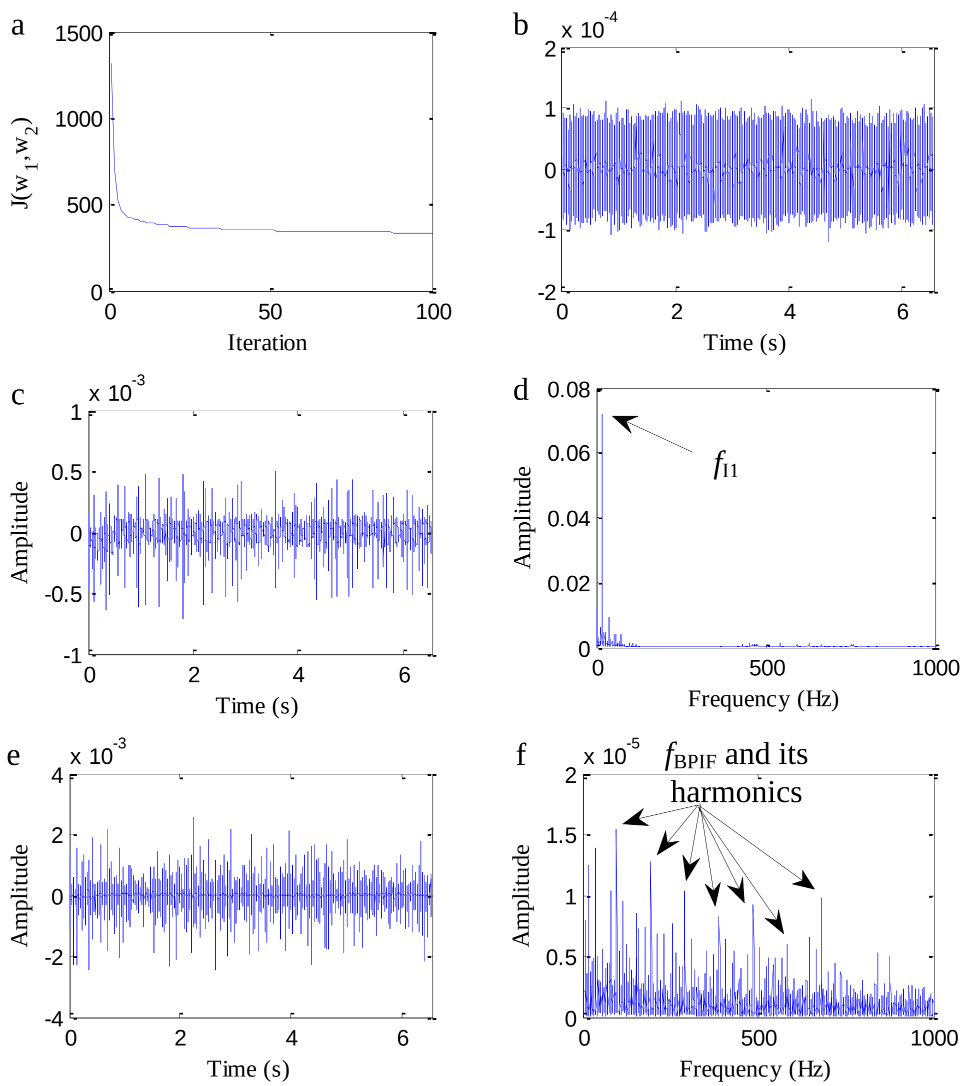


Fig. 4.23 Detection of inner race fault using the IOBSD method (results of the last iteration): (a) the reduction of objective function, (b) the residual, (c) the high-oscillation component and (d) its FFT spectrum, (e) the low-oscillation component and (f) its envelope spectrum.

Likewise, for comparison, the SK method is also applied to inner race fault detection, resulting figures shown in Fig. 4.24. According to the kurtogram as shown in Fig. 4.24 (a), the selected center frequency and bandwidth are 7187.5 Hz and 625 Hz, respectively. The envelope of the filtered signal is shown in Fig. 4.24 (b) and its frequency representation is presented in Fig. 4.24 (c). We can see that the f_{BPIF} and one of its harmonic are discernible; by comparison, more harmonics can be observed in Fig. 4.23 (d) obtained by the proposed method. The comparison indicates that the proposed method is more powerful in extracting impulse-like signals.

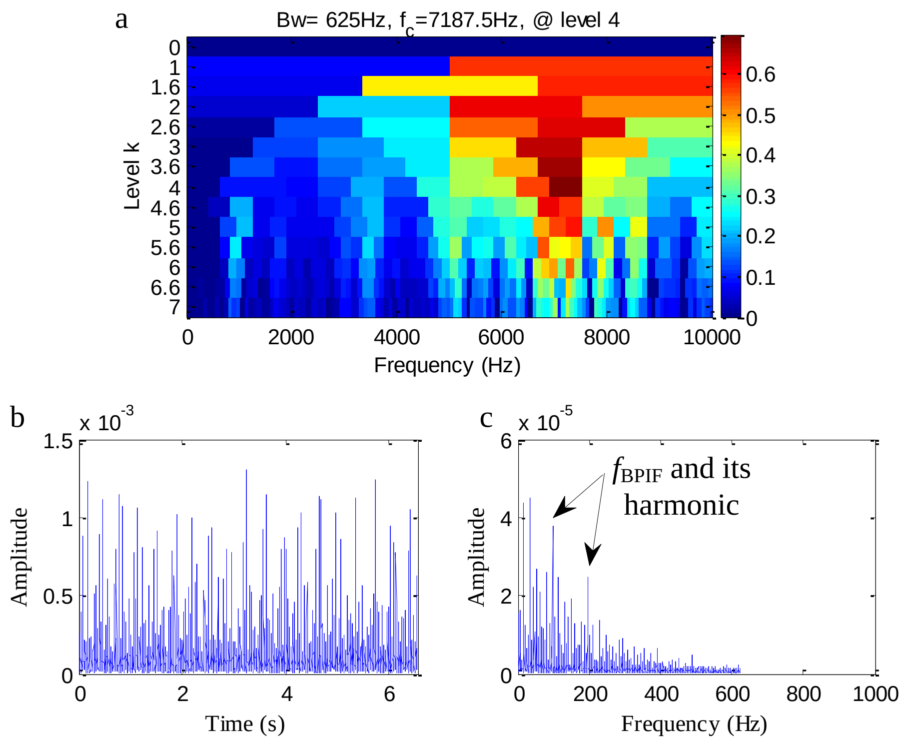


Fig. 4.24 The detecting result of inner race fault using SK method: (a) the kurtogram, (b) the envelope of filtered signal, and (c) the spectrum of the envelope.

4.5.3 Additional outer race fault test

4.5.3.1 Experimental setup #1

To further test the proposed method in handling relatively high frequency interferences and noise, the extra test is done on another SpectraQuest machinery fault simulator in our lab at the University of Ottawa (Fig.4.25). In this case, the shaft is connected with a gearbox by belts and two sheaves to generate interferences. The diameter ratio of the two sheaves is 1:2.6. The gear meshing frequency is calculated based on the motor shaft rotational frequency trend, the number of pinion (drive gear) teeth and the pulley belt speed reduction ratio (i.e., gear meshing frequency = motor shaft rotational speed × number of pinion teeth/speed reduction ratio). In this case, the number of pinion teeth is 18 and hence the gear meshing frequency = motor shaft frequency×18/2.6. An unbalanced flywheel (with two screws added to the flywheel to create mass imbalance and hence introduce additional interferences. A faulty bearing with an outer race fault is installed on the right side of the shaft. The left bearing supporting the shaft and those in the motor are healthy. The details of the faulty bearing are provided in Table 3.2 and its picture is presented in Fig. 3.12. The vibration data are collected using four identical accelerometers amounted at different locations but only the data from the sensor mounted on the steel plate that support the two bearing housings is used. As this sensor is far from the faulty bearing, its data is more difficult to process. The shaft rotation speed is 35 Hz (2100 rpm) and vibration signals are collected at a sampling rate of 12 KHz. The FCF for the outer race fault is 125 Hz (=3.572×35 Hz). The gear meshing frequency is 242.3 Hz (= 35Hz×18/2.6) and the mass unbalance frequency is the same as the shaft frequency, i.e., 35 Hz.

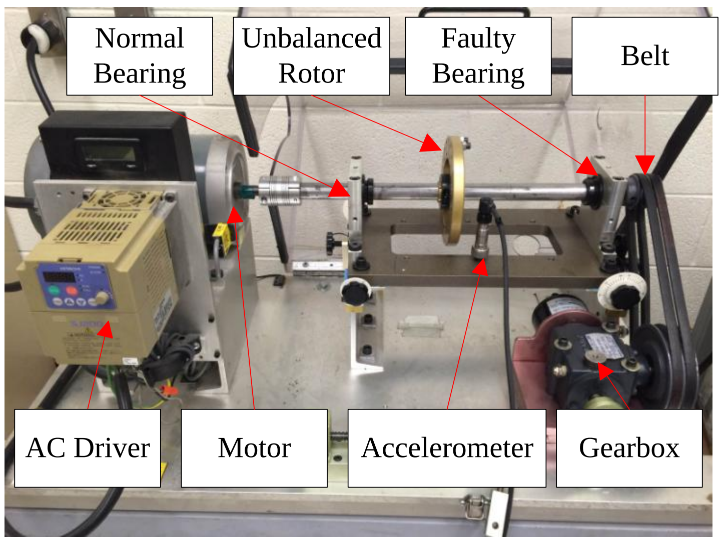


Fig.4.25 Experimental setup.

The original vibration signal and its envelope spectrum are shown in Fig. 4.26 (a) and (b) respectively. As can be seen, the envelope spectrum is dominated by the interference stemming from gear meshing. The FCF and a few of its harmonics cannot be seen on the spectrum. As such, the proposed method is then applied to analyze the collected vibration signal and the results are shown in Fig. 4.27. The minimization of the objective function is reached, as can be seen in Fig. 4.27 (a). The extracted high-oscillation component, Fig. 4.27 (c), is expected to mainly contain gear meshing vibration as it presents sustained oscillatory behavior. This can be reflected by its spectrum in Fig. 4.27 (d), where the gear meshing frequency and its harmonic stand out. Fig. 4.27 (e) presents the extracted fault-induced transients, i.e., low-oscillation component. Envelope demodulating the resulting low-oscillation component leads to the spectrum in Fig. 4.27 (f), in which the FCF (f_{BPOF}) at 125Hz and five its harmonics can be clearly identified. The parameters A_1 and A_2 in this case are 0.35.

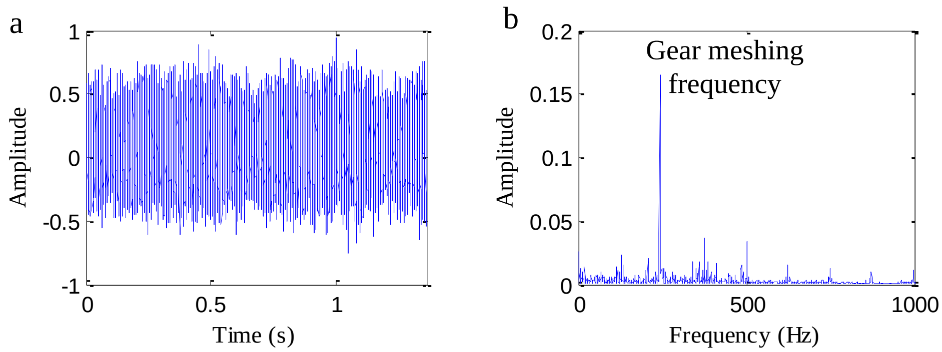


Fig. 4.26 Additional outer race fault test #1: (a) the raw vibration signal, and (b) the envelope spectrum of the raw signal.

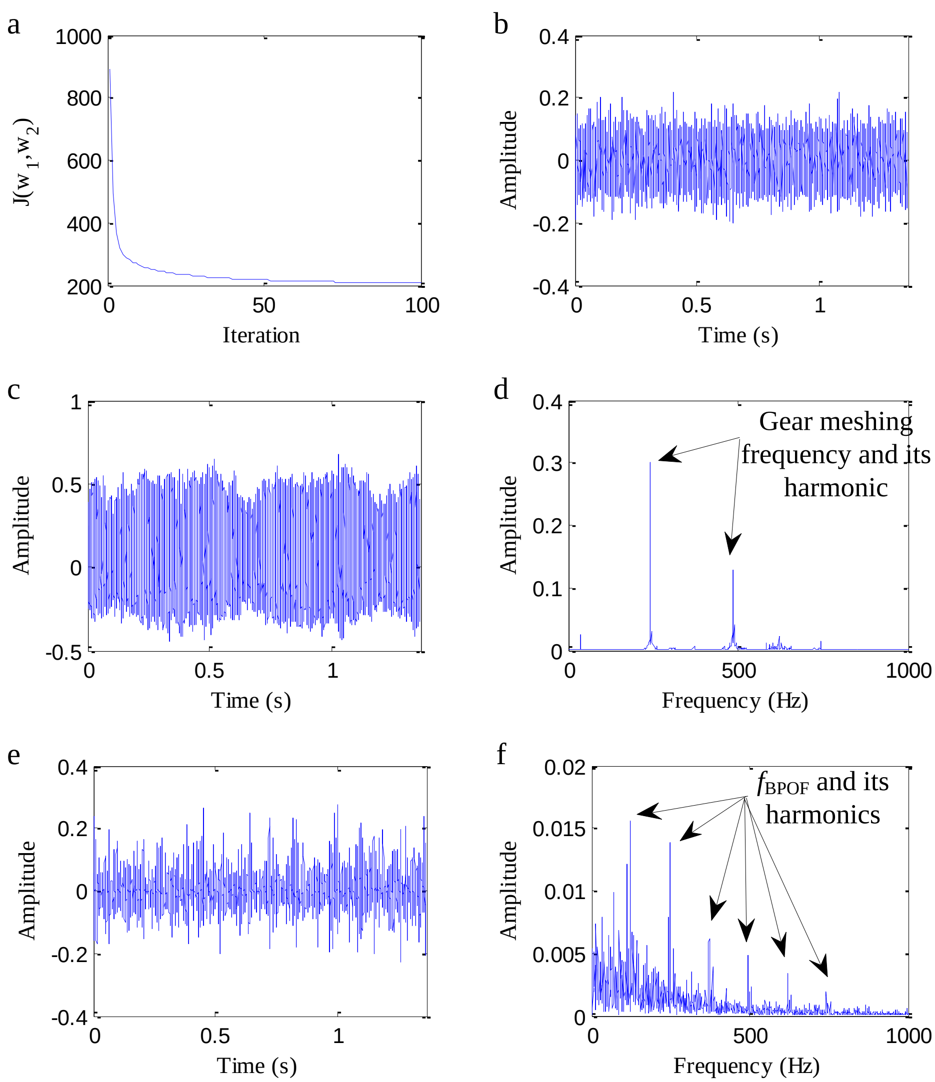


Fig. 4.27 Detection of outer race fault (experimental #1) using the IOBSD method (results of the last iteration): (a) the reduction of objective function, (b) the residual, (c) the high-oscillation component and (d) its FFT spectrum, (e) the low-oscillation component and (f) its envelope spectrum.

4.5.3.2 Experimental setup #2

In this test, the accelerometer is installed on the simulator base at a location close to the belt pulley and gearbox but far away from the faulty bearing (ER10K, bearing parameters are listed in Table 5.2 and its picture is shown in Fig. 5.6), as seen in Fig. 4.28. A imbalanced load rotor (11.1 lb) is mounted on the shaft. The gearbox is driven by the shaft via a belt connection. The shaft rotating speed f_r is 16.5 Hz (990 RPM) in this experiment and hence the FCF is 50.5 Hz (=3.052f_r). The meshing frequency f_{mesh} of gearbox is 114.2 Hz (= 16.5×18/2.6 Hz). The signals are sampled at 12 KHz.

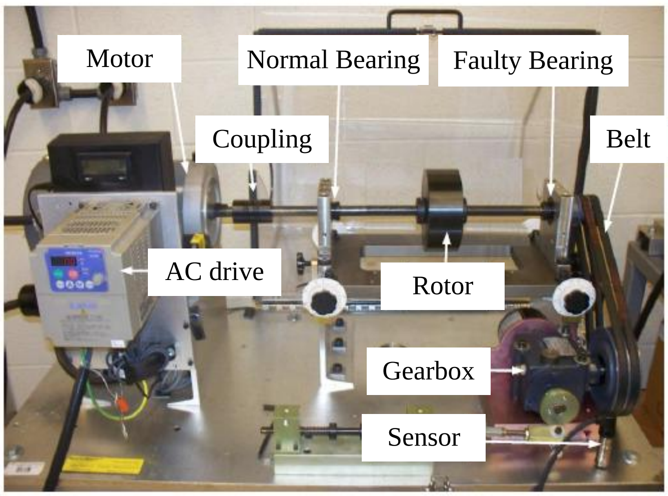


Fig. 4.28 Experimental setup #2.

The measured vibration data and its envelope spectrum are shown in Fig. 4.29 (a) and (b), respectively. Obviously, the landscape of Fig. 4.29 (b) is dominated by the gearbox meshing frequency and, as such, the FCF as well as its harmonics of outer race cannot be identified. The results of the proposed IOBSD method are presented in Fig. 4.30, from which it can be concluded that the primary components of high-oscillation component are interferences from gearbox (Fig. 4.30 (c) and (d)) and the low-oscillation component is mainly composed of fault-induced transients (Fig. 4.30 (e) and (f)). The f_{BPOF} and its harmonics can be recognized in the Fig. 4.30 (f).

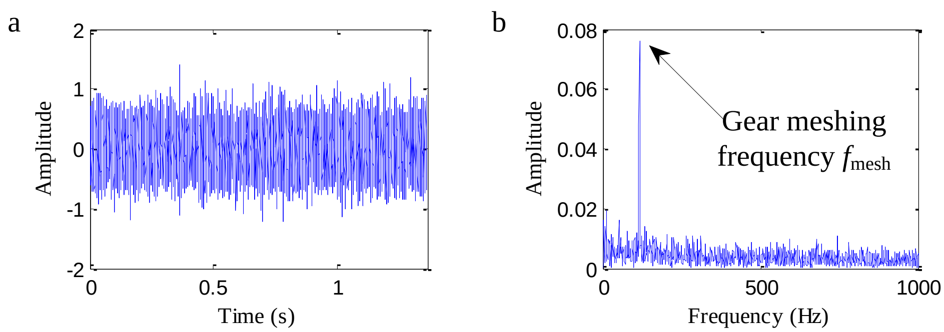


Fig. 4.29 Additional outer race fault test #2: (a) the raw vibration signal, and (b) the envelope spectrum of the raw signal.

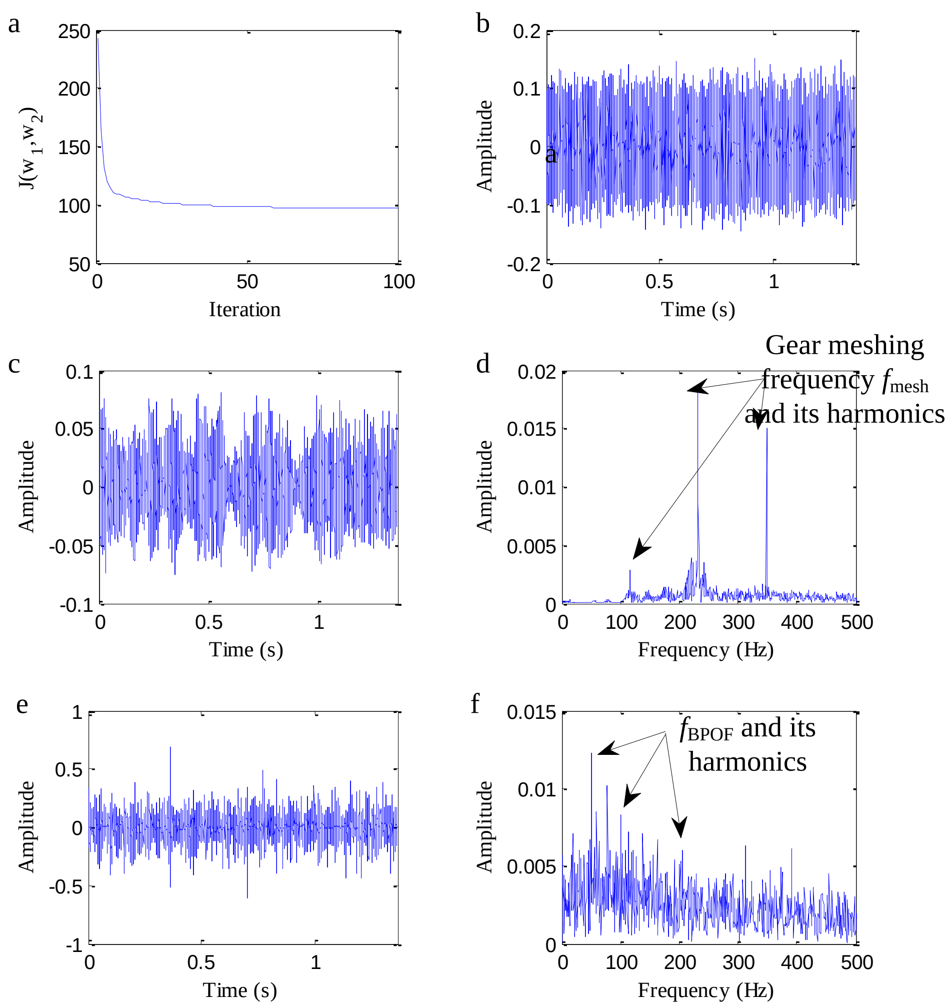


Fig. 4.30 Detection of outer race fault (experimental #2) using the IOBSD method (results of the last iteration): (a) the minimization process of the objective function, (b) the resulting residual, (c) the resulting high-oscillation component and (d) its FFT spectrum, (e) the resulting low-oscillation component and (f) its envelope spectrum.

For comparison, the processing result of the proposed STFD technique is displayed in Fig. 4.31. It can be seen that the spectrum of the envelope obtained by STFD method in Fig. 4.31 (b) is dominated by random frequencies at low frequency band (as shown in the zoomed-in figure) and vibration-interference-related frequencies, such as the harmonics of gear meshing frequency and the frequency difference between gear meshing and shaft rotating speed (labelled in Fig. 4.31 (b)). No fault-associated frequency information is revealed because the fault feature is severely obscured by noise, interferences and its harmonics as the sensor is placed close to gearbox but away from defective bearing. The comparison suggests that the IOBSD method performs better when dealing with signal contaminated by multiple interferences and strong noise.

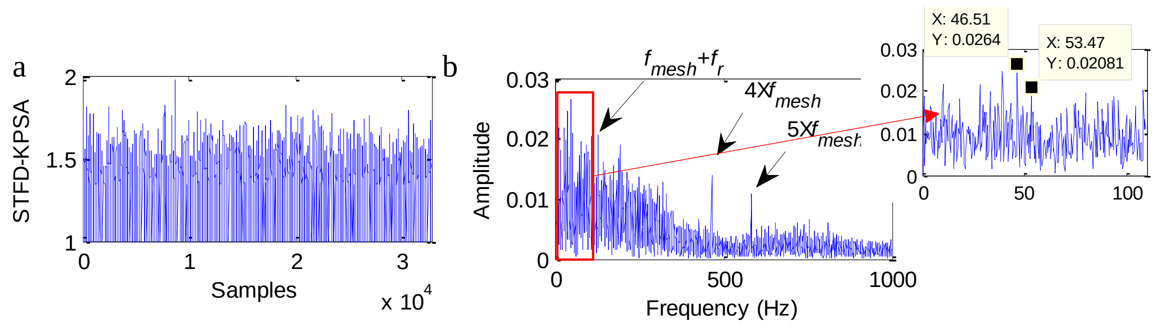


Fig. 4.31 Detection result of outer race fault (experimental #2) using the STFD method: (a) the resulting STFD-KPSA representation, and (b) the spectrum of (a).

4.6 Conclusion

An oscillatory behavior based signal decomposition (OBSD) has been reported in this chapter and its application to bearing fault detection has been explored. The rationale of the proposed method is that signal components presenting different oscillatory behaviors can be separated using the MCA technique as long as appropriate bases are generated for each signal component representation. The TQWT provides an appealing means of the proper basis creation by specifying Q-factors. The objective function can then be constructed via MCA technique and TQWT. Signal components with different oscillatory behaviors can be attained by solving the objective function.

The proposed method has three innovative features: a) the strategies of determining Q-factors for low- and high-oscillation components are developed in this chapter according to shape resemblance and mutual coherence between bases, b) an iterative OBSD method is proposed to tackle signals whose components presents a wide range of oscillatory degrees, and c) the oscillatory behavior based signal decomposition is the third feature that fundamentally distinguishes the proposed method from most, if not all, of existing frequency-based decomposition techniques.

The main advantages of this method are obvious: a) it can remove interferences with frequencies in the near vicinity of the resonant frequency and trim down in-band noise due to its frequency-information-dependent feature, b) no prior knowledge of the signals such as the fault-excited resonance is needed, and c) no additional noise reduction steps are involved. Finally, the method has been validated by both simulated signals and experimental data.

4.7 Summary for the two proposed methods for bearing fault feature extraction

The interference removal and noise reduction method proposed in this chapter extracts bearing fault feature from a perspective different from the method in the previous chapter. Both of them serve for the same purpose of fault feature extraction. However, each has its own advantages. The method in Chapter 3 is more powerful for low frequency interference removal and is more computationally efficient. It is a type of envelope demodulation technique; hence the Hilbert transform is not involved in this method. Nevertheless, its effectiveness deteriorates with the increase of the interference frequency complexity, particularly for multiple high frequency interferences. This issue can be solved by the oscillatory-behavior-based ‘filtering’ method presented in this chapter. This oscillatory behavior based method is capable of separating fault-induced transients from interferences as well as noise via generating proper bases for transients as well as periodic interferences and can more effectively handle multiple high frequency interferences.

Therefore, each of the two methods presented in Chapters 3 and 4 has its unique characteristics, and they are companion techniques in the application to bearing fault feature extraction. Generally, for the online application and the working condition without many high interferences (such as in the low speed condition), the STFD-KPSA is recommended. For the bearing working under environment with high frequency interferences, i.e., high speed condition, the IOBSD method is better suited.

Part II Bearing Fault Identification under Time-varying speed condition

Chapter 5 A Joint Application of Short Time Fractal Dimension Transform and Generalized Demodulation to Bearing Fault Identification under Time-varying Rotational Speed[3]

The bearing fault features can be captured via two methods proposed in the previous two chapters and then the repetition frequency of the fault related impulses can be obtained by spectrum analysis for fault detection and fault type diagnosis. However, this is the case for bearings running at time-invariant rotational speed. In reality, bearings in wind turbine, mining equipment and many other machines often work at a variable speed. Such speed fluctuations may cause “smearing” of the discrete frequencies in the frequency representation, making the fault related frequencies much less observable as discrete frequency lines in spectra (Urbanek et al., 2011). As a result, signal processing approaches that are developed for signal analysis at a constant speed may no longer be valid or become less effective for bearing condition monitoring under variable speed.

Bearing Fault detection under time-varying rotational speed has, therefore, attracted much attention in recent decades (McBain and Timusk, 2009; Zimroz et al., 2011; Urbanek et al., 2013). The conventional way for bearing fault diagnosis under variable rotational speed generally includes prefiltering, resampling based on shaft rotating IF and order spectrum analysis. However, its application is confined by three major obstacles: a) knowledge-demanding parameter determination required by prefiltering, b) unavailable shaft rotating frequency for resampling as it is coupled with IFCF by a FCC which cannot be decided without knowing what fault actually exists, and c) complicated and error-propagating resampling process. As such, we propose a new method to address these problems. The proposed method, which is free from prefiltering and resampling, mainly contains the following steps: a) extracting envelope via the previously developed STFD transform, requiring no prefiltering, b) with the extracted envelope signal, performing STFT to get a clear TFR, from which the IFCF and the basic demodulator for GD can be obtained, c) applying the generalized demodulation to the envelope signal with the current demodulator, converting the trajectory of the current time-frequency component into a

[3] Published in Mechanical Systems and Signal Processing, vol. 68-69, 2016, p. 15-33.

linear path parallel to the time axis, d) frequency analyzing the demodulated signal, followed by searching the amplitude of the constant frequency where the linear path is situated. Updating demodulator via multiplying the basic demodulator by different real numbers (i.e., coefficient λ) and repeating the steps c) - d), the resampling-free order spectrum is then obtained. Based on the resulting spectrum, the final diagnosis decision can be made. The proposed method for its implementation on the example of simulated data is presented. Finally, experimental data are employed to validate the effectiveness of the proposed technique.

The STFD transform, which is used for envelope extraction and interference suppression, has been detailed in the Chapter 3; hence its introduction in this chapter is omitted.

5.1 Resampling-free diagnosis strategy via STFT and GD

In this section, the GD is firstly introduced, followed by the development of a bearing fault detection strategy via joint application of GD and STFT based on the above described STFD transform. The envelope of the signal is extracted by the proposed STFD transform without filters. Based on the extracted envelope, a clearer TFR of the signal can be attained by STFT. The TFR obtained by conventional Hilbert transform is then compared with that of the proposed envelope method. With the improved TFR from the proposed STFD method, the IFCF which is used to calculate the demodulator of GD can be extracted. The resampling-free diagnosis strategy is then developed based on the extracted IFCF and GD.

5.1.1 Overview of GD

GD is a novel signal processing tool, which is particularly effective for the separation of multi-component non-stationary signals (Cheng et al., 2009; Feng et al., 2011). The key of the GD lies in the development of a signal transform, i.e., Generalized Fourier Transform (GFT), which is possible to transform the curved instantaneous frequency trajectory of an interested component into a linear path paralleling to the time axis. As a result, it avoids overlapping among signal components in the time-frequency domain. Considering a signal $x(t)$, the GFT is formulated by (Olhede and Walden, 2005)

$$X_G(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi[ft+d(t)]} dt \tag{5.1}$$

where $d(t)$ is a real-valued demodulator depending on the time only and $e^{-j2\pi d(t)}$ is the transform kernel. It must be noted that this is the same as applying the standard FT to $x(t)e^{-j2\pi d(t)}$. It is completely equivalent to the standard FT when $d(t) = 0$. Therefore, standard FT can be considered as a special case of GFT. Following the theory for standard FT, the inverse of GFT is defined as

$$x(t) = \int_{-\infty}^{\infty} X_G(f) e^{j2\pi[ft+d(t)]} df = e^{j2\pi d(t)} \int_{-\infty}^{\infty} X_G(f) e^{j2\pi ft} df \quad . \tag{5.2}$$

Observing the Eq.(5.2), it can be obtained that if $X_G(f) \equiv \delta(f - f_0)$ then $x(t)$ can be written as

$$x(t) = e^{j2\pi[f_0 t + d(t)]}, \tag{5.3}$$

which implies that a signal with instantaneous frequency $f(t) = f_0 + d'(t)$ is mapped to the frequency point $f = f_0$, where $d'(t)$ is the first derivative of $d(t)$. Hence if we wish to map a signal, for example, a curved path in the time-frequency plane specified by $f_0 + d'(t)$ into a linear path, we simply need to estimate the function $d(t)$ to demodulate the signal. In this way, the time-frequency distribution of the demodulated signal $x(t)e^{-j2\pi d(t)}$ can be mapped to the linear line paralleling the time axis with frequency $f = f_0$.

In the view of the above, two important properties of GD can be summarized as follows.

1) The curved path in time-frequency distribution can be transformed into the linear path, paralleling the time axis.
2) The energy of the signal $x(t) = e^{j2\pi[f_0 t + d(t)]}$ would be more concentrated on the constant frequency $f = f_0$ after the GD, compared with that of the signal without GD.

5.1.2 Development of resampling-free diagnosis strategy

Based on the proposed envelope method and properties of GD mentioned above, the resampling-free strategy of detecting bearing fault under variable speed is developed in this subsection.

To begin with, a simulated defective bearing signal of variable speed (Fig. 5.1 (a)) is used to illustrate the idea of the proposed method. The details of the simulated signal will be given in the following section. The raw signal is first STFD transformed to extract its envelope which is displayed in Fig. 5.1 (b). The vertical axis represents the amplitude of STFD transform, i.e., FD sequences. Applying the STFT to the obtained envelope signal, a clear TFR can be observed in Fig. 5.1 (c), where the IFCF and its harmonics can be easily recognized. The IFCF can then be successfully extracted based on the peak search algorithm and the function of the IFCF denoted by $f(t)$ can be approximated by least square fitting. The demodulator $d(t)$ of the IFCF-related signal component is then obtained by

$$d(t) = \int (f(t) - f_0)dt, \tag{5.4}$$

where f_0 is the constant frequency in which the linear line is situated.

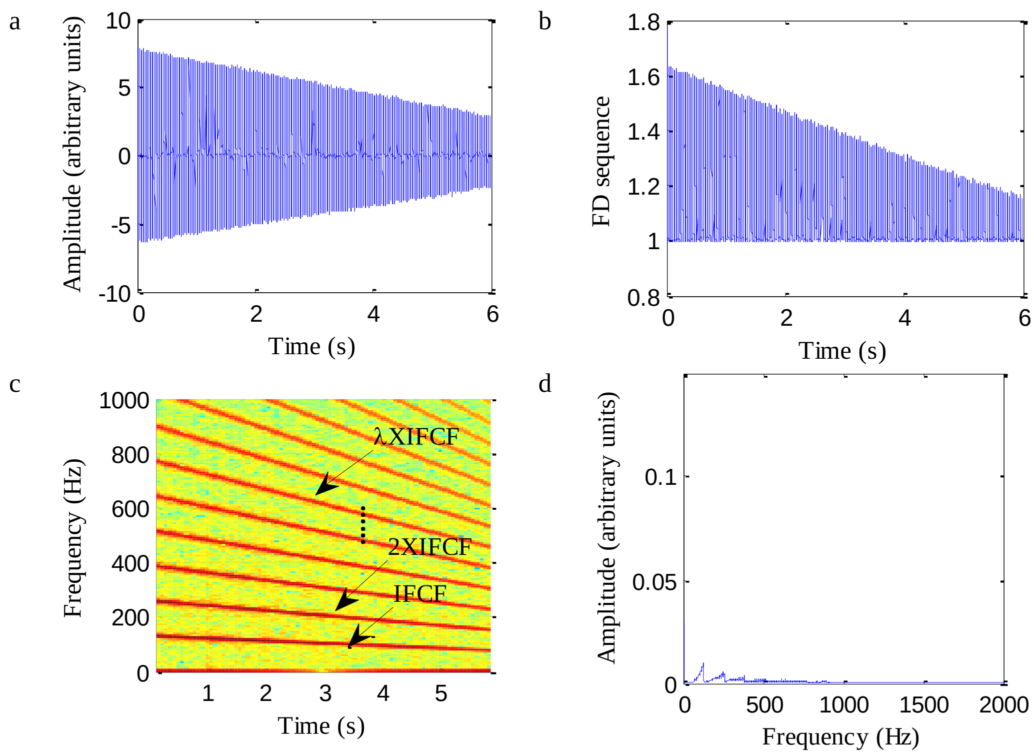


Fig. 5.1 (a) The raw signal, (b) the envelope signal obtained by the STFD transform, (c) the TFR of the envelope signal by STFT, and (d) frequency spectrum of the envelope signal.

With the demodulating function $d(t)$ for the time-frequency component IFCF, the GD is then applied to the envelope signal, converting the curved path of the IFCF into a horizontal line paralleling to the time axis in the time frequency plane, as shown in Fig. 5.2 (a). It can be observed that the curved trajectory of IFCF in Fig. 5.1 (c) is converted into the linear path located at the frequency f_0 (in this case, $f_0 = 129.4$ Hz) in Fig. 5.2 (a), as circled by a black ellipse.

Applying spectral analysis to the demodulated signal obtained by GD under the demodulator $d(t)$, it can be seen that the energy of the IFCF-related signal component is centralized at the frequency f_0, with a peak shown in the spectrum of Fig. 5.2 (b). This is due to the properties of GD, making the curved trajectory converting into a straight line paralleling to horizontal; thus, the energy which originally spread in a frequency bandwidth (due to the curved path) is concentrated at the frequency f_0. For comparison, the frequency spectrum of the FD sequence without GD analysis is presented in Fig. 5.1 (d), where no apparent peaks associated with IFCF, shaft IF and their harmonics can be seen.

However, faulty bearing vibration signal is multicomponent. Besides the IFCF-related component, there are components associated with harmonics of IFCF and shaft IF together with its harmonics. The harmonics of IFCF indicate the existence of fault and shaft IF can help determine the fault type. Therefore, analyses of these components are also vital for bearing fault detection, which are detailed in the following.

For the 1st harmonic of the IFCF, its function is $2f(t)$ and hence the corresponding demodulator is to double the basic demodulator $d(t)$ i.e., $2d(t)$. Performing generalized demodulation for the signal component related to the 1st harmonic of the IFCF, its linear path can be obtained in the time-frequency plane. The TFR and frequency representation of generalized demodulated signal are exhibited in Fig. 5.2 (c) and (d), respectively. As seen in Fig. 5.2 (c), the resulting horizontal line path circled by green ellipse is located in at the frequency $2f_0$ (=2*129.5). Accordingly, a peak at frequency $2f_0$ is recognizable at the frequency spectrum of the generalized demodulated signal, which is demonstrated in Fig. 5.2 (d).

Likewise, the horizontal lines of 2nd and 3rd harmonics of the IFCF can be obtained when the signal is generalized demodulated using demodulators $3d(t)$ and $4d(t)$, respectively. In general, a signal component related to an arbitrary harmonic of the IFCF, e.g., the (λ-1)th harmonic, can be demodulated using demodulator $\lambda d(t)$ for λ = 2, 3 … The corresponding peaks in their frequency spectra can be identified, locating at frequency $2f_0$, $3f_0$ ... λf_0 …, respectively.

Following the same logic, since IFCF is proportional to the instantaneous shaft rotational frequency, demodulating the raw signal with demodulator $\lambda d(t)$ for λ = 1/FCC, 2/FCC, …, and then applying frequency analysis to the demodulated signals, the peaks which are related to shaft rotational frequency and its harmonics can be revealed in the frequency spectra. Such peaks can be used, together with the IFCF, to identify fault types.

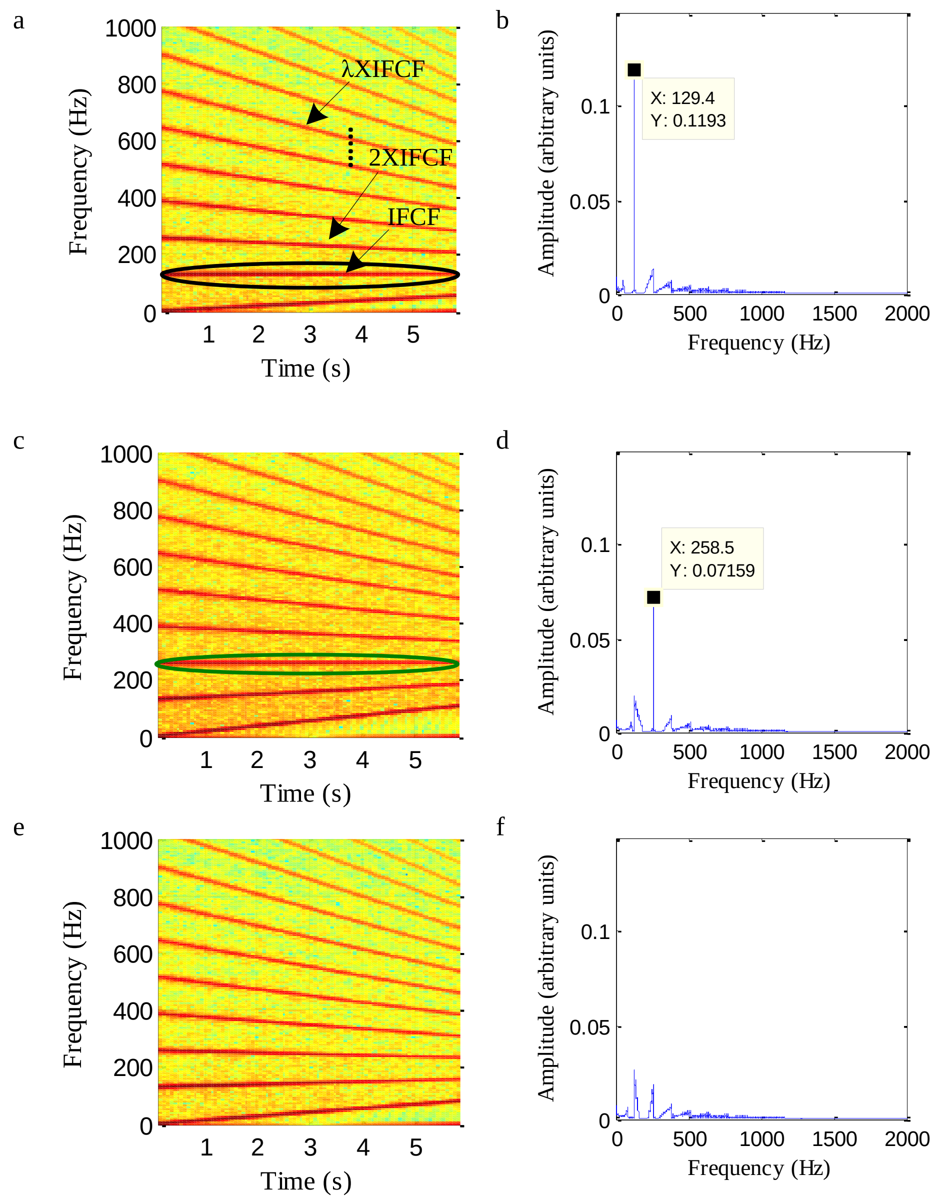


Fig. 5.2 The illustration of GD application: (a) TFR of GD-processed signal with basic demodulator $d(t)$, (b) frequency spectrum of GD-processed signal with $d(t)$. (c) and (d) are the counterparts of (a) and (b) with demodulator $2d(t)$, and (e) and (f) are the counterparts of (a) and (b) with demodulator $1.5d(t)$.

If the demodulator $\lambda d(t)$ is related neither to the IFCF and its harmonics nor to the shaft IF and its harmonics, no apparent time-frequency components can be observed on the time-frequency planes because no related signal components are included in the vibration signal. Therefore, no remarkable peaks can be identified on the frequency spectra of the demodulated signals. For example, as shown in Fig. 5.2 (e) where the TFR of the envelope signal demodulated by a phase function $1.5d(t)$ is shown, no corresponding horizontal line is discernible. As a result, the energy is still dispersive and no apparent peak can be identified at the frequency $1.5f_0$ (= 194.25).

Therefore, if the GD is iteratively applied to the raw signal with λ ranging from 0 to λ_{max} at a step size μ ($0 < \mu < 1$), followed by frequency analysis of every generalized demodulated signal with the demodulator denoted as $\lambda d(t)$ and search of the amplitude value corresponding to each frequency λf_0, the plot between the coefficient in front of $d(t)$ λ and the corresponding amplitude can be attained. We name the plot resampling-free order spectrum since no resampling process is involved. The horizontal axis is the coefficient λ with the resolution μ and the vertical axis is the amplitude corresponding to each constant frequency λf_0 where the straight frequency line is located. The detection strategy can hence be concluded as follows: if there are multiple peaks of significant amplitude located at integer λ values, the existence of a bearing fault can be confirmed. The peaks situated between 0 and 1 give the information of fault types.

Based on the above illustration, the procedure of bearing fault detection using STFD transform and GD can be summarized in the following:

1) Envelope demodulate the raw signal using the proposed STFD transform.
2) Apply STFT to the envelope signal in order to obtain the basic demodulator $d(t)$.
3) Multiply basic demodulator $d(t)$ by λ_i, $\lambda_i \in (\mu, 2\mu, ..., \lambda_{\max})$.
4) For $\lambda_i d(t)$, implement the GD for the envelope signal consisting of three steps:
    - Hilbert transform the raw signal $s(t)$ to create an analytic signal $s_a(t) = s(t) + jH[s(t)]$, where $jH[s(t)]$ is the Hilbert transform of the original signal.
    - Apply the GD to the analytic signal and obtain a new signal $y(t) = s_a(t)e^{-j2\pi\lambda_i d(t)}$.
    - Hilbert transform the new signal $y(t)$ to avoid negative frequencies and generate a new analytic signal $z(t) = y(t) + jH[y(t)]$.
5) Perform frequency analysis on the generalized demodulated signal.
6) Search the amplitude value corresponding to the frequency $\lambda_i f_0$ in the frequency spectrum.
7) Repeat step 3) – 6) until λ_i equals to λ_{max} and plot coefficients λ_i versus amplitudes.

The flowchart of the proposed method for bearing fault detection under time-varying speed is presented in Fig. 5.3. The process starts with enveloping the vibration signal by STFD transform, followed by the TFR acquisition via STFT. Then the function of IFCF (or basic demodulator) can be obtained by least square curve fitting based on the TFR. Next, the GD converts the curved trajectories of IFCF and shaft IF together with their harmonics into linear paths paralleling to the horizontal axis. This is followed by Fourier transform of generalized demodulated signals with demodulator $\lambda_i d(t)$, $\lambda_i \in (0, \mu, 2\mu, ..., \lambda_{\max})$. In this way, the resampling-free order spectrum is obtained by searching the amplitude of the frequency ($\lambda_i f_0$) where each linear path is located. Consequently, bearing fault detection and diagnosis can be made based on the resampling-free order spectrum. The resampling step has been completely eliminated from the proposed approach.

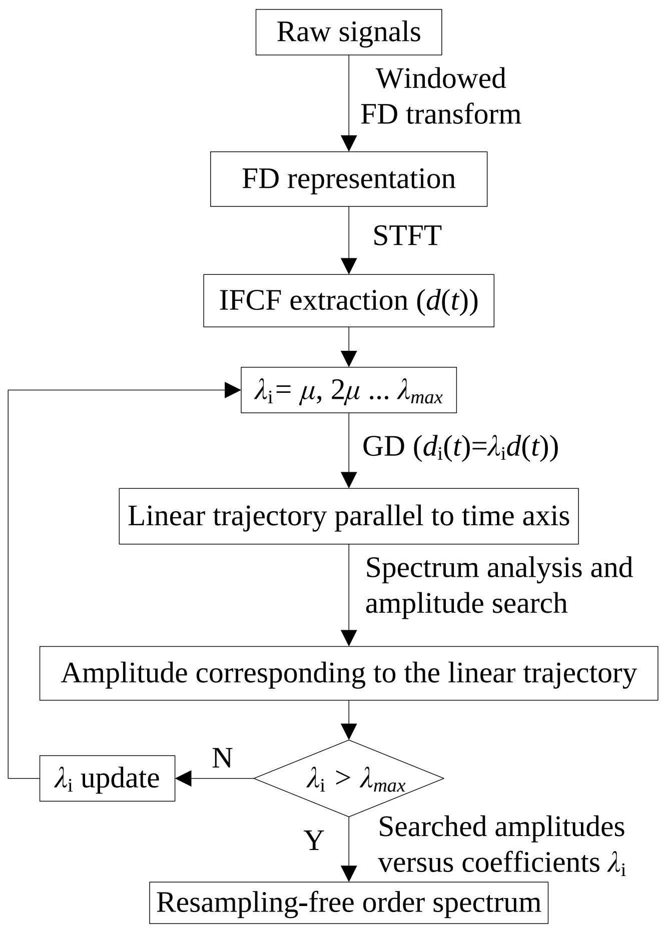


Fig. 5.3 Flowchart of the proposed method for bearing fault detection under variable speed condition without resampling algorithm and a tachometer.

5.2 Simulated case

A simulated signal of a faulty bearing under ramp-down speed is constructed in this section. The performance of the proposed method is then examined by the simulated signal. The model of faulty bearing vibration is expressed as (Liang and Bozchalooi, 2010)

$$s(t)=\sum_{m=-M}^{M} L_m e^{-\beta(t-mT_p-\sum_{i=-M}^{m}\tau_i)}\sin(\omega_r(t-mT_p-\sum_{i=-M}^{m}\tau_i))\,\mathrm{u}(t-mT_p-\sum_{i=-M}^{m}\tau_i) \tag{5.5}$$

where M is half of the number of the fault generated impulses, L_m is the amplitude of mth fault impulse, β the decay parameter, T_p is the time period corresponding to the fault characteristic frequency, τ_i presents the effect of random slippage of the rollers and is modeled as a uniformly distributed random variable with a zero mean and a standard deviation of βT_p, ω_r is the excited resonance frequency, and u(t) is a unit step function.

The above model is a good example for vibration of faulty rolling bearing under constant speed. In order to imitate the time-varying rotational speed, the model has to be trimmed. In this chapter, we generate a signal of faulty bearing under ramp-down speed. The modified model for variable speed bearing is written as

$$s(t)=\sum_{m=1}^{N} A_m e^{-\beta(t-t_m)}\sin(\omega_r(t-t_m))\mathrm{u}(t-t_m)+n(t) \tag{5.6}$$

where N is the signal length, $A_m = A(N/f_s - \eta t_m)$ is amplitude which is inversely proportional to time instant t_m since this is a speed-down case, A and η are constants, $n(t)$ is random noise, and t_m is the occurrence time of the mth impulse (m = 1, 2, ...). According to Eq. (5.6), the critical step for generating simulated data under variable speed is to confirm the parameter t_m. The IFCF function is denoted by $f(t)$ and then $1/f(t)$ means the time interval of between two adjacent fault impulses. In this way, the occurrence time of the first impulse t_1, as shown in Fig. 5.4, can be computed by $1/f(0)$, where $f(0)$ cannot be equal to 0. Taking the slippage into consideration, the express of t_1 is given as follows:

$$t_1=\beta[1/f(0)]+1/f(0)=(\beta+1)[1/f(0)] \tag{5.7}$$

where β is used to count slippage effect and varies from 0.01 to 0.02. Based on t_1, the instantaneous frequency between t_1 and t_2 can be obtained and the occurrence time of the second impulse can be computed as:

$$t_2 = t_1 + (\beta+1)[1/f(t_1)] = (\beta+1)[1/f(0)+1/f(t_1)]. \tag{5.8}$$

According to the above analysis, the occurrence time of each fault impulse can be calculated using a recursive scheme, which is expressed below:

$$t_m = (\beta+1)[1/f(0)+1/f(t_1)+...+1/f(t_{m-1})],\ m=1,2,... \tag{5.9}$$

The illustration of the above analysis is presented in Fig. 5.4.

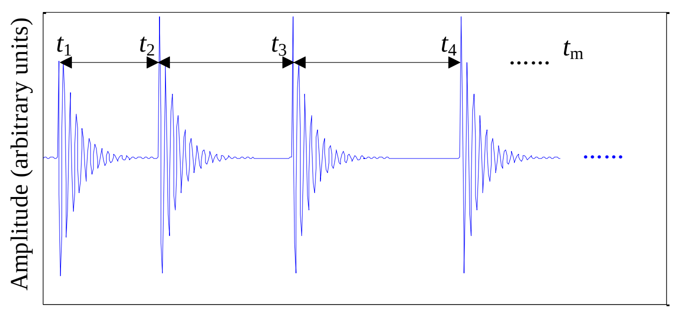


Fig. 5.4 The demonstration of generating simulated signals under time-varying speed.

Table 5.1 Parameters of the simulation model.

Parameters	N	f_s	ω_r	β	η	A	FCC	μ
Value	65536	20 KHz	2KHz×2π	800 Hz	0.7	4/3	3.7	0.01 ~ 0.02

The parameters of the model are listed in Table 5.1. We set the IFCF and order to $f(t)$ = - 2.5t + 35 and 3.7 respectively in this study. The simulated impulse-like signal is shown in Fig. 5.1 (a). In addition to background noise, vibration signals of faulty bearing are often polluted by cyclic interferences resulted by mechanical or/and electrical components, thereby two cyclic interferences with frequency 60 Hz and 200 Hz respectively and amplitude 1 for both being added into the simulated signal. To reflect the phenomena of the resonance being demodulated by shaft rotating frequency, the signal s is multiplied by $(1+0.1\cos(2\pi f_r(t)t))$, where $f_r(t)$ is the shaft rotating frequency calculated by $f(t)/3.7$. The composite signal is plotted in Fig. 5.5 (a) and the TFR of its envelope signal by Hilbert transform is presented in Fig. 5.5 (b). It can be seen that the IFCF and its harmonics are clouded by the frequencies of interferences, which would have adverse effects on the IFCF extraction based on peak search algorithm. As shown in Fig. 5.5 (e), the extracted IFCF deviates from the true IFCF that is plotted in red solid line during the interval of 4 s to the end. Then the composite signal is processed by proposed STFD transform and the resulting envelope is shown in Fig. 5.5 (c). The improvement of using the STFD transform can be observed in Fig. 5.5 (d), where the TFR of STFD transformed signal is shown using STFT. As seen, the effect of interferences is completely eliminated and thus the estimated IFCF which can be extracted via peak search algorithm is shown in Fig. 5.5 (f). For comparison, the true IF (= $f(t)$/3.7) and estimated IF ((= $f_e(t)$/3.7) are also exhibited in Fig. 5.5 (f), respectively. Visually, the estimated IF fits the true IF very well. To quantity the accuracy of estimation, the following error analysis is made. The largest absolute error is 0.071, which is located at true IF equal to 35, hence leading the relative error to 0.2% at this point. The Mean Square Error (MSE), which is usually used to measure the difference between the estimator and the real value, is 2.059×10^{-6}. The MSE is computed by $\sum_{i=1}^{n}(\mathrm{IF}_{e,i}-\mathrm{IF}_{t,i})^2 \big/ n$, where $\mathrm{IF}_{e,i}$ and $\mathrm{IF}_{t,i}$ is the estimated IF and true IF at the ith sampling point respectively. All error analysis demonstrates that the extracted IF highly matches the real IF, indicating that the proposed envelope method together with the STFT is effective in extraction of IFCF. Therefore, the estimated IFCF f_e (t) can be obtained by fitting the data in a least-square sense, which is formulated as

$$f_e(t) = -9.21t + 129.24 \ . \tag{5.10}$$

Integrating the term with variable t in Eq. (5.10), the demodulator of GD for the simulated case is obtained as $d_s(t) = -4.605t^2$, where d_s(t) denotes the basic demodulator of the simulated signal. Following the procedures given in section 3.2, the relationship between amplitudes and coefficient of the basic demodulator, i.e., the resampling-free order spectrum, is plotted in Fig. 5.5 (g). There are five peaks associated with IFCF and its harmonics, respectively. Specifically, λ = 1 corresponding to IFCF and λ = 2 corresponding to the first harmonic of IFCF. The rest can be done in the same manner. It has to be noted that, below the IFCF, three peaks related to shaft IF and its harmonics are observable. The corresponding values are 0.27, 0.54 and 0.81, equal to the reciprocal of the order, its double and triple respectively, which is used to distinguish fault types in real cases.

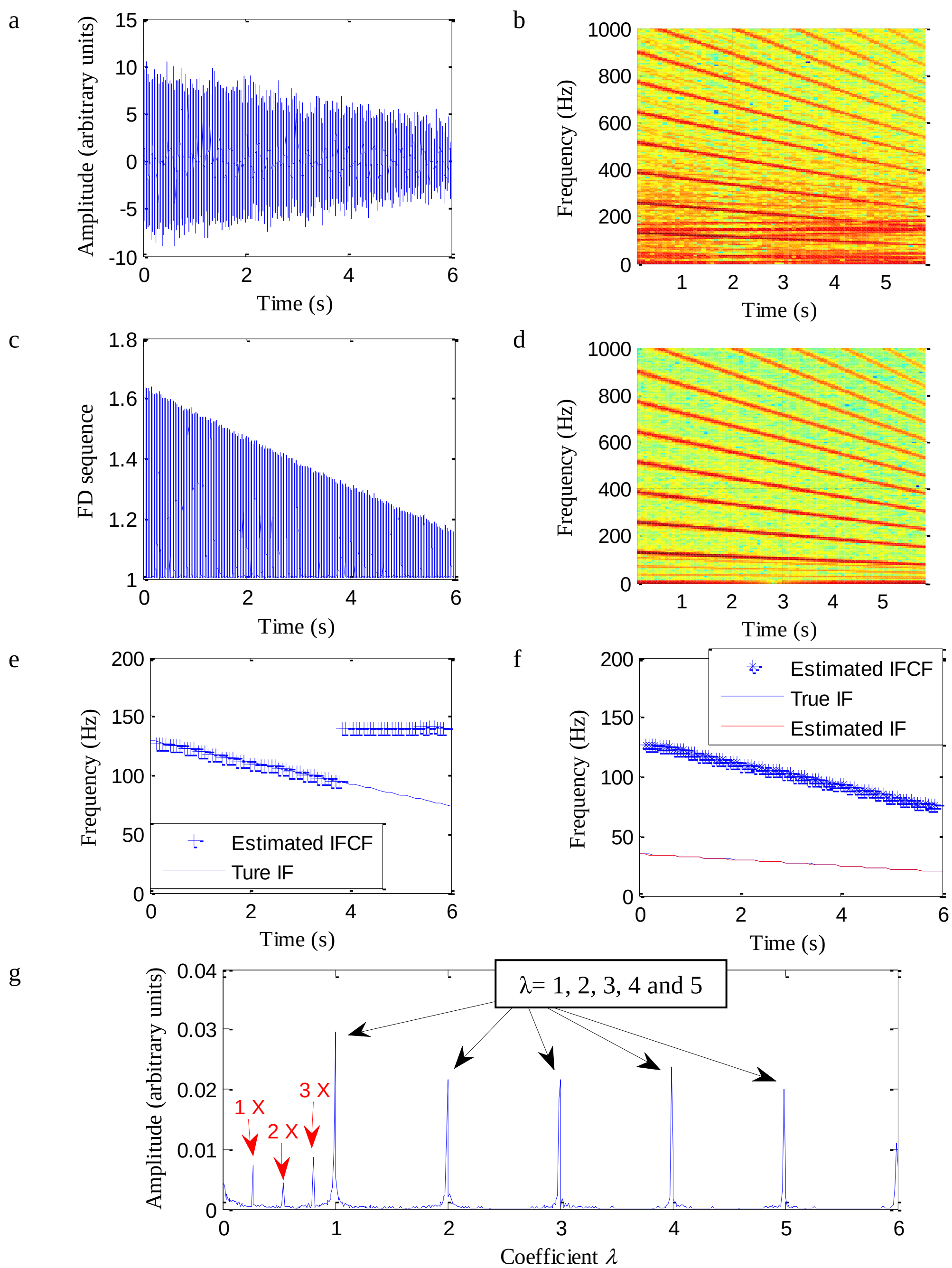
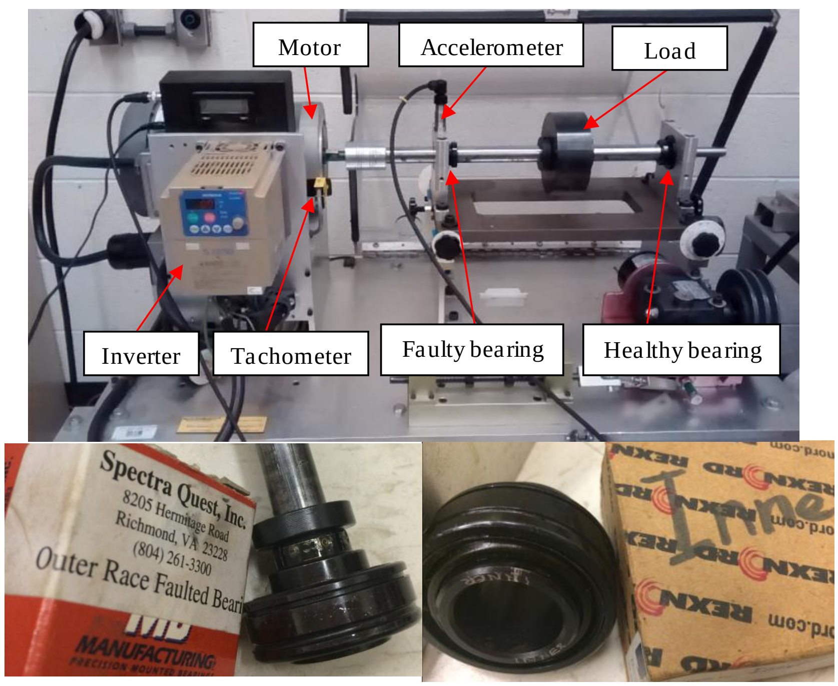


Fig. 5.5 Processing results of the proposed method for simulated data: (a) the composite signal, (b) TFR of the envelope obtained by Hilbert transform, (c) the envelope signal obtained by STFD transform, (d) TFR of (c), (e) estimated IFCF based on (b), (f) estimated IFCF based on (d), corresponding estimated IF and true IF, and (g) the resampling-free order spectrum.

The results above show that the proposed technique is effective in bearing fault detection under variable speed. By the STFD transform, the envelope of the impulsive signal is successfully extracted without prefiltering, thereby facilitating the acquisition of a clear TFR and the estimation of IFCF. Based on the IFCF, basic demodulator is then obtained by integration with respect to variable *t*, which paves the way for GD operation. By virtue of the developed strategy, the finial detection result can be accordingly made based on the resampling-free order spectrum. Next, the effectiveness of the proposed method is evaluated by experimental data.

5.3 Experimental evaluation

In this section, diagnosis performances of the proposed method for real cases are tested. The experiments are carried out on a SpectraQuest Machinery fault simulator (MKF-PK5M) at our lab. The experiment setup is shown in Fig. 5.6. The shaft is driven by a motor controlled by an AC inverter. There are two bearing mounted to support the shaft and the load. The unhealthy one is positioned at the left side. Tachometer and accelerometer are mounted on the test rig to collect the shaft speed and vibration signal, respectively. After processed by a signal conditioner (PCB PIEZOTRONIC 482C15), the data is fed to an NI data Acquisition Module (NI USB-6212 BNC) and is recorded by the computer with Labview software. Two experiments are carried out, for outer race and inner race respectively.



Bearing (ER10K) with outer race fault Bearing (ER16K) with inner race fault

Fig. 5.6 Experimental setup for bearing fault detection under time-varying speed and the defective bearings.

5.3.1 *Outer race fault detection*

Two ER10K bearings and a load mass of 5.03 kg is mounted on a steel shaft with diameter 15.875 mm (5/8 inch). The left bearing has a fault on outer race. The parameters of the bearing are listed in Table 5.2. The test lasts for 5.46 s and the sampling frequency is 24 KHz. During this period, the shaft rotational speed increases from 18.2 Hz to 35 Hz following a nearly linear pattern.

Table 5.2 Parameters of the bearings for the outer race fault test.

Parameter	Bearing type	Fault type	Pitch diameter(mm)	Ball diameter(mm)	Number of balls	IFCF(Hz)
Value	ER10K	Outer race fault	33.50	7.94	8	$3.052f_r$

The raw vibration signal and the TFR of its envelope obtained by Hilbert transform are shown in Fig. 5.7 (a) and (b) respectively. As seen, Fig. 5.7 (b) does not present a clear TFR pattern relative to the IFCF, from which the extracted IFCF shown in Fig. 5.7 (e) is quite messy and does not present the pre-set linear pattern at all. Then the raw signal is processed by STFD transform and the resulting envelope signal is presented in Fig. 5.7 (c). Applying STFT to the obtained envelope leads to the TFR shown in Fig. 5.7 (d), where a clear IFCF pattern and its harmonics are recognized. From the TFR of the envelope signal, the IFCF is extracted by peak search algorithm, which is illustrated in Fig. 5.7 (e). The estimated IF of the shaft that is obtained by dividing IFCF by the order (=3.052) together with the true IF are demonstrated in Fig. 5.7 (e), for comparison. To evaluate the accuracy of the IFCF estimation, the maximal absolute error, relative error at point of maximal absolute error and MSE are calculated. The maximum absolute error is 2.13 Hz, which happens at t equalling to 3.883 s. The true frequency at this time point is 31.87 Hz, thus the relative error at this point is 6.68%. The MSE, calculated as the same manner mentioned above, is 8.7×10^{-3}. The quantitative error analysis shows that the estimating result is reliable. Then the function of IFCF can be approximated by fitting the extracted data best in a least-square sense, as shown below

$$f_o(t) = 9.067t + 55.65 \ . \tag{5.11}$$

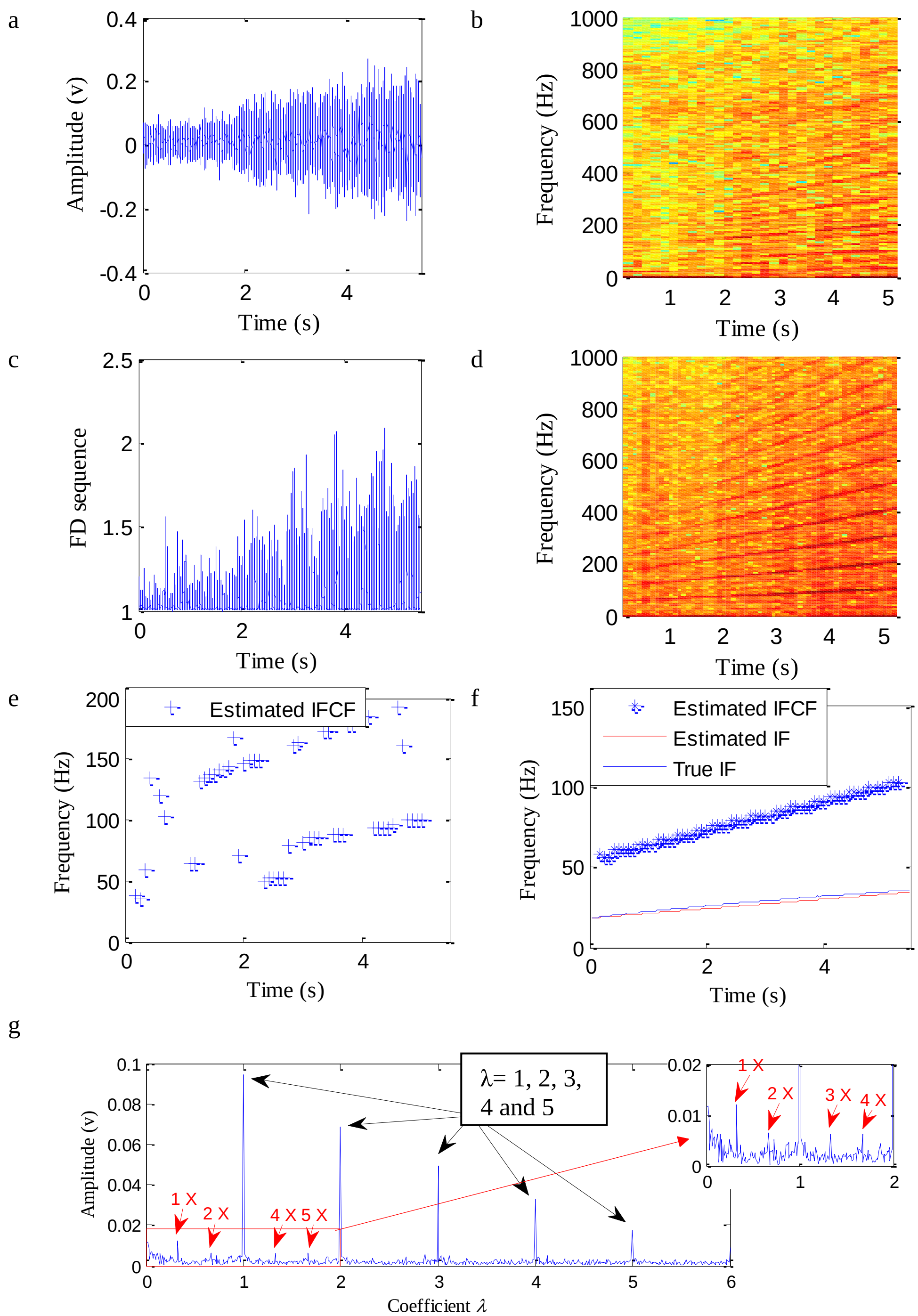


Fig. 5.7 Detecting performance of the proposed method for outer race fault: (a) the raw signal, (b) TFR of the envelope obtained by Hilbert transform, (c) the envelope obtained by STFD transform, (d) TFR of (c), (e) estimated IFCF based on (b), (f) estimated IFCF based on (d), corresponding estimated IF and true IF, and (g) the resampling-free order spectrum.

The basic demodulator of outer race fault is then obtained by integration, which is $d_o(t) = 4.53t^2$. The same procedures proposed above are employed to demodulate the raw signal.

Processing result is shown in Fig. 5.7 (g), where five peaks associated with IFCF and its harmonics can be observed, indicating the existence of the fault. Several peaks located at both sides of the IFCF can also be recognized in Fig. 5.7 (g). As marked, these peaks correspond to the reciprocal of the order (1/3.052 = 0.33) and its multiples, thereby revealing the fault type. By means of the proposed technique, the final detection decision can be made that there is a fault on the outer race of the bearing, which agrees the reality.

5.3.2 Inner race fault detection

In this subsection, the proposed approach is investigated by data from the bearing with inner race fault. The experiment is operated on the same test rig as outer race test. However, we replace some mechanical components. A larger shaft, 25.4 mm (1 inch) in diameter, is mounted on the test rig to support two ER16K bearings and the load which is the same as the last experiment. The detailed parameters of ER16K bearing are list in Table 5.3. To make the experiment more general, the shaft speed changing pattern is different from the previous test. It first grows from 17 Hz to 33.3 Hz and then drops down to 23.8 Hz. The sampling frequency is 12 KHz in this case and signal of 17.9 s is collected.

Table 5.3 Parameters of the bearings for inner race fault test.

Parameter	Bearing type	Fault type	Pitch diameter(mm)	Ball diameter(mm)	Number of balls	IFCF(Hz)
Value	ER16K	Inner race fault	38.52	7.94	9	$5.43f_r$

In Fig. 5.8 (a), the original signal of inner race fault is presented. Fig. 5.8 (b)-(g) are the counterparts of Fig. 5.7 (b)-(g). Comparing Fig. 5.8 (d) with (b), the trend of IFCF and its harmonics are more visible after the application of STFD transform. The more apparent improvement obtained using the proposed STFD transform can be observed by examining Fig. 5.8 (e) and (f). The extracted IFCF in Fig. 5.8 (e) is disorganized. Fig. 5.8 (f) presents the extracted IFCF based on the envelope signal obtained by STFD transform, where the true IF of the shaft and the predicted IF are also plotted. The maximal absolute error, appearing at 11.67 s, is 2.72 Hz in this case. The true IF at this time point is 33.43 Hz, leading the relative error at this location to 8.1%. The MSE is 9.3×10^{-3} for the prediction. Applying the least-square fitting, the IFCF function is acquired as follows

$$f_I(t) = -0.0564t^3 + 0.8582t^2 + 3.7042t + 92.88\,. \tag{5.12}$$

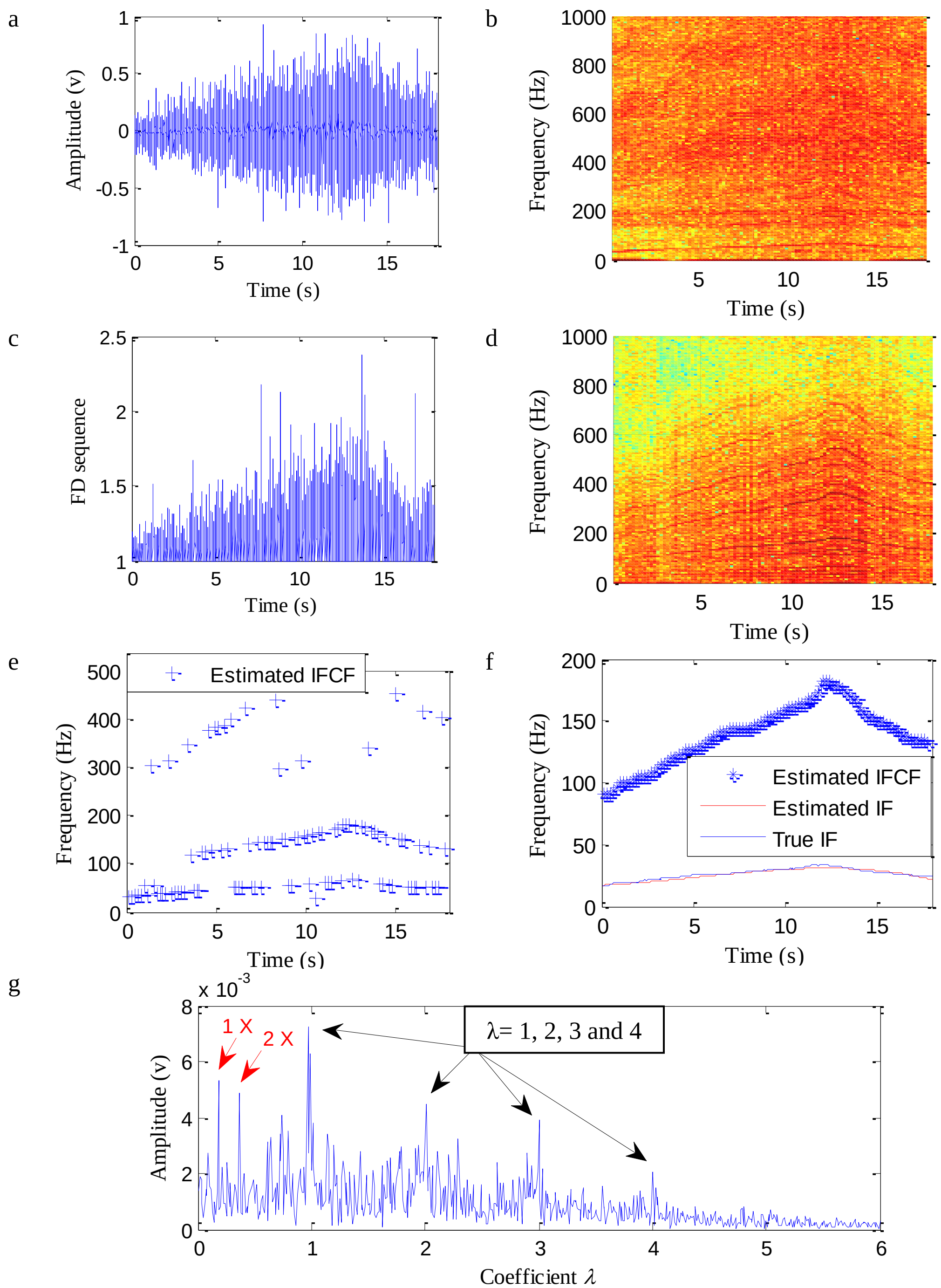
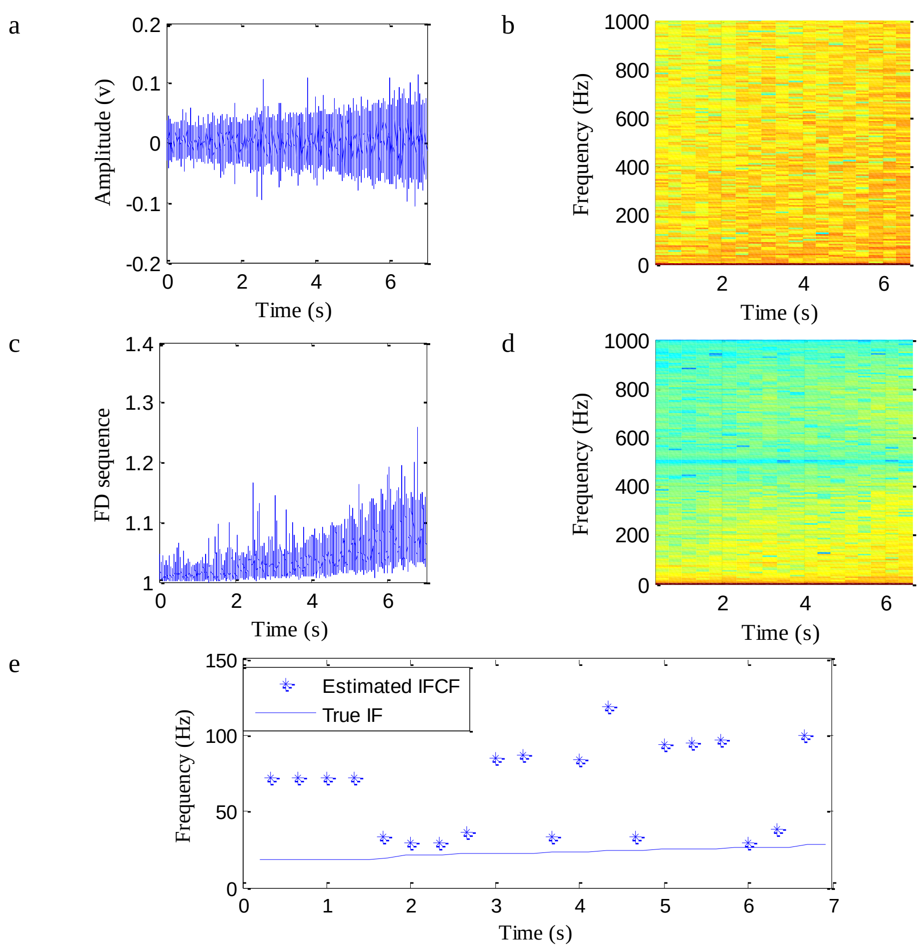


Fig. 5.8 Detecting performance of the proposed method for inner race fault: (a) the raw signal, (b) TFR of the envelope obtained by Hilbert transform, (c) the envelope obtained by STFD transform, (d) TFR of (c), (e) estimated IFCF based on the TFR in (b), (f) estimated IFCF based on the TFR in (d), corresponding estimated IF and true IF, and (g) the resampling-free order spectrum.

By integrating the first three terms of Eq. (5.12), the basic demodulator of inner race fault is expressed as $d_i(t)$ = - $0.0141t^4 + 0.30t^3 + 1.852t^2$. Following the strategy in subsection 3.2 gives the resampling-free order spectrum illustrated in Fig. 5.8 (g). Peaks locating at λ = 1, 2, 3 and 4 can be identified; thus the existence of the fault is confirmed. The reciprocal of the order (1/5.43=0.184) and its multiple located below the IFCF are also observable, which relates the fault type to inner race. Therefore, the effectiveness of the proposed approach is validated by inner race data.

5.3.3 Healthy bearing test

For comparison, the vibration signal from healthy bearing is measured in this subsection to examine the proposed approach. In this test, the experimental setup is the same as the one described in the section 5.3.2 except that the bearing with inner race fault is replaced by a healthy one. The shaft rotational speed in this case increases from 18.75 Hz to 28.2 Hz during the time duration of 7s.



f

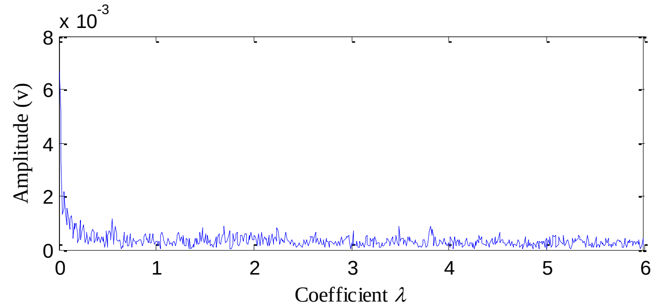


Fig. 5.9 Detection performance of the proposed method for healthy bearing: (a) the raw signal, (b) TFR of the envelope obtained by Hilbert transform (c) the envelope obtained by STFD transform, (d) TFR of (c), (e) estimated IFCF based on the TFR in (d), corresponding estimated IF and true IF, and (f) the resampling-free order spectrum.

The raw signal is shown in Fig. 5.9 (a) and the TFR of its envelope signal obtained by Hilbert transform is demonstrated in Fig. 5.9 (b). The envelope that is extracted by STFD transform and the corresponding time-frequency distribution are presented in Fig. 5.9 (c) and (d), respectively. As seen in Fig. 5.9 (d), no obvious IFCF trend is exhibited as the healthy bearing results no fault-related impacts, which is as expected. Therefore, the shaft rotational speed cannot be reflected by the extracted IFCF, as observed in Fig. 5.9 (e). Fig. 5.9 (f) shows the resampling-free order spectrum based on the demodulator obtained by integrating the extracted IFCF vector, where no peaks situated at integral λ values can be seen. Consequently, the final detection decision that the bearing is in a healthy condition can be made.

5.4 Conclusion

A technique without the use of prefiltering and resampling algorithm is proposed for bearing fault diagnosis under time-varying rotational speed. Two main parts constituting the proposed method are the STFD transform used for envelope extraction without prefiltering and the STFT-based GD for obtaining the resampling-free order spectrum. The envelope extraction via the STFD transform is a pre-process for STFT and GD, helping obtain the demodulator for GD via STFT. Fourier transforming the GD-processed signal, the amplitude of the constant frequency resulted by GD can stand up and be searched on the spectrum. Vibration signals of a faulty bearing are multicomponent; hence, iteratively employing generalized demodulation, frequency analysis and amplitude search for the signal under analysis, the resampling-free order spectrum can then be obtained. The diagnosis decision can then be made.

The major advantages of the proposed method are (a) the envelope of impulses can be effectively extracted in virtue of the STFD transform without filtration, thus relieving the worry of filter bandwidth and filter center selection; (b) by means of GD collaborating with STFT, an innovative diagnosis strategy is developed with no need of resampling algorithm, which simplifies the diagnosis technique, improves the computational efficiency and also distinguishes this method from most of current methods for bearing fault detection under variable speed; and (c) since the proposed method is resampling-free there is no need to worry about shaft rotational frequency. The effectiveness of the proposed technique has been tested by both simulated and experimental data.

It is worth mentioning that, as long as the IF can be accurately estimated based on the vibration signal, the proposed method can be applied to different conditions, regardless of the changing pattern of the rotational speed. As illustrated by the experimental examinations, both the outer race and inner race faults can be diagnosed no matter the speed changes linearly or nonlinearly.

Chapter 6 Dual Demodulation Transform and Synchrosqueezing for Time-frequency Analysis and Bearing Fault Diagnosis[4]

The technique developed in Chapter 5 for bearing fault identification under time-varying speed conditions has solved the problems caused by prefiltering and resampling. However, to obtain a clear TFR and to accurately estimate the IFCF, it heavily relies on an effective envelope demodulation method. TFA has proven to be a powerful tool in nonstationary signal processing as it can reveal insights about both time and frequency. However, its application to bearing condition monitoring has often been undermined due to the blurry and ambiguous representations, which is particularly true under variable speed conditions. Targeting these issues as well as the problems to be addressed for existing TFA methods (mentioned in Section 2.4), this chapter proposes a novel TFA method suited to bearing fault detection under variable speed.

The energy concentration level is an important indicator for time-frequency analysis (TFA). The primary issue for most of the existing TFA methods is the TFR diffusion caused by weak energy concentration which would lead to ambiguous results and thus misleading signal analysis results, particularly for nonstationary multicomponent signals. To improve the energy concentration level, we propose a dual demodulation transform (DDT). The rationale of the DDT is that 1) a forward demodulation can improve the energy concentration level of a signal by transforming its IF trajectory into a line of a constant frequency, and 2) focusing on a short window around the time instant of interest, a backward demodulation can recover the original frequency at the time instant without affecting the improved energy concentration level. The improved TFR of the entire signal can be attained via repeating the backward demodulation at every time instant of interest. Based on the DDT, an iterative DDT (IDDT) is developed to analyze multicomponent signals that are subjected to different modulating sources for their constituent components. The IDDT iteratively demodulates each constituent component to attain its TFR and the TFR of the whole signal is obtained by superposing all the TFRs of constituent components. The cross-term free and more energy concentrated TFR is then generated by the proposed IDDT. The effectiveness of the proposed method in TFA is tested by both simulated

[4] Submitted to Journal of Sound and Vibration, under revision.

monocomponent and multicomponent signals. The application of the proposed method to bearing fault detection is exploited. Bearing condition and fault pattern can be revealed by the proposed method resulting TFR.

6.1 The presentation of the proposed method

6.1.1 Dual demodulation transform (*DDT*)

6.1.1.1 Generalized demodulation

The overview of GD has been introduced in subsection 5.1.1. Next, we will present a brief proof that the GD can map a signal to an arbitrary constant frequency. This feature is important to analyze nonstationary multicomponent signal whose IFs of constituting components are closely-positioned.

Considering a signal $x(t) = e^{i\pi\psi_x(t)}$ ($\psi_x(t) = f_0 t + \psi(t)$). Demodulating the signal $x(t)$ by the function $e^{-i2\pi(\psi_x(t)-f_c t)}$ leads to

$$X_G(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi[ft+\psi_x(t)-f_c t]} dt. \tag{6.1}$$

where f_c represents an arbitrary constant frequency.

Its inverse GFT is then derived

$$x(t) = \int_{-\infty}^{\infty} X_G(f) e^{i2\pi[ft+\psi_x(t)-f_c t]} df = e^{i2\pi[\psi_x(t)-f_c t]} \int_{-\infty}^{\infty} X_G(f) e^{i2\pi ft} df \;. \tag{6.2}$$

If $X_G(f) \equiv \delta(f - f_c)$, $x(t)$ can then be written as

$$x(t) = e^{i2\pi\psi_x(t)}, \tag{6.3}$$

which is identical to the previously defined $s(t)$, indicating that the signal can be mapped to any constant frequency f_c. The feature of mapping the signal with curved trajectory into a linear path parallel to the time axis at an arbitrary constant frequency can avoid the frequency overlap, thereby facilitating the filtering operation and signal component separation.

Two important GD properties are re-summarized as follows:

1) The curved path in TFR can be transformed into the linear path, parallel to the time axis.

2) After the demodulation, the energy of the signal $x(t)=e^{i2\pi[f_0t+\psi(t)]}$ would be concentrated on the constant frequency $f = f_c$, as can be seen in Fig. 6.1.

6.1.1.2 DDT Proposal

The GD and its properties described above form the foundation for the proposed DDT method which is detailed in this subsection. To illustrate the idea of DDT, we start with a monocomponent AM-FM signal $x(t)$, given by

$$x(t)=A(t)e^{i\phi(t)}=A(t)e^{i2\pi[f_0t+\psi(t)]}\ . \tag{6.4}$$

where $A(t)$ is the instantaneous amplitude (IA), $\phi(t)$ is the instantaneous phase, f_0 is the constant frequency and $\psi(t)$ is the modulating source. The IF of the signal is accordingly defined as

$$f_x(t)=\frac{d\phi(t)}{2\pi dt}=f_0+\psi'(t)\ . \tag{6.5}$$

Suppose that the modulating source $\psi(t)$ of the signal is known. The original signal is firstly generalized demodulated by $d^f(t)=e^{-i2\pi\psi(t)}$, which can be formulated as

$$x_{d^f}(t)=x(t)d^f(t)=A(t)e^{i2\pi f_0t}\ . \tag{6.6}$$

Eq. (6.6) derived by the above mentioned GD is the forward demodulation and the $d^f(t)$ is the forward demodulator. The TFR of the original signal and the forward demodulated signal obtained by the STFT is plotted in Fig. 6.1 (a). As mentioned previously, we focus on a signal segment within a short window $[u - \Delta t,\ u + \Delta t]$ with $\Delta t \approx 2\pi/\phi'(u)$ (Daubechies et al., 2011), centered at time instant u. The resulting TFR within the short window is plotted in Fig. 6.1 (b). By looking at Fig. 6.1 (a) and (b), the following observations can be obtained. First it can be seen that, in Fig. 6.1 (a), the curved TF trajectory, $f_0 + \psi'(t)$, is converted into a straight line, f_0, parallel to time axis with more concentrated TFR ridges (the level of concentration is quantified by the darkness). Secondly, for an arbitrarily selected small time window $[u-\Delta t, u+\Delta t]$ with $\Delta t \approx 2\pi/\phi'(u)$ as plotted in Fig. 6.1 (b), the spectrum of the original signal segment is widely spread around the frequency $f_x(u) = f_0 + \psi'(u)$; whereas the spectral energy of the forward demodulated signal segment, i.e., the line segment at f_0, is much more concentrated. The improved level of concentration can be quantified by the spectrum on the right side of Fig. 6.1 (b). The above has echoed the two GD properties.

Although the energy concentration degree of the demodulated signal segment spectrum is improved, as shown in Fig. 1 (b), the spectrum does not reflect the actual frequency of the signal segment at the time instant u. Therefore, further processing is required. Shifting the TFR of the demodulated signal by $|\psi'(u)|$, the frequency of the TFR then becomes $f_0 + \psi'(u)$ ($\psi'(u) > 0$ means upwards shifting; $\psi'(u) < 0$ leads to downwards shifting), identical to the actual frequency at the current time instant u. The frequency shifting can be achieved via simply multiplying the generalized demodulated signal by the backward demodulator $d^b(t,u) = e^{i2\pi\psi'(u)t}$. This is based on one of the Fourier transform properties, i.e., shifting a function in the frequency domain corresponds to multiplication by a complex exponential function in the time domain. The further demodulated signal is then given by

$$x_d(t,u) = x_{d^f} \cdot d^b = x_{d^f} \cdot e^{i2\pi\psi'(u)t} = A(t)e^{i2\pi[f_0 t + \psi'(u)t]}. \tag{6.7}$$

It should be noted that the modulus of x_d is equal to that of x_{d^f}. This is the reason why the proposed DDT method can improve the energy concentration level. The above illustrated forward and backward demodulations can be done by a composite demodulator defined as

$$d(t,u) = d^f(t) \cdot d^b(t,u) = e^{-i2\pi[\psi(t) - \psi'(u)t]}. \tag{6.8}$$

Mathematically, Eq. (6.7) means that a signal segment is demodulated by the source which is a bivariate function of time variable t and time-shift variable u (as suggested by Eq. (6.8)). The original one-dimensional signal is transformed into a two-dimensional signal (bivariate signal), as seen in Eq. (6.7). The IF of the bivariate signal is

$$\begin{aligned} f_{x_d}(t) &= \frac{\partial}{2\pi\partial t}\{2\pi[f_0 t + \psi'(u)t]\} \\ &= f_0 + \psi'(u) \end{aligned}. \tag{6.9}$$

For a sufficiently small time window around u, $t \in [u\text{-}\Delta t, u+\Delta t]$, we can obtain that

$$f_{x_d}(t) = f_x(t)|_{[u-\Delta t, u+\Delta t]}. \tag{6.10}$$

Eq. (6.10) indicates that the original frequency at time instant u is restored via such demodulations. More importantly, the concentration level of the spectrum in the TF plane is preserved, as seen in Fig. 6.1 (c).

The STFT of the demodulated signal $x_d(t, u)$ is given by

$$X_{x_d}(u,\xi)=\int_{-\infty}^{\infty}x_d(t,u)g(t-u)e^{-i\xi(t-u)}dt \tag{6.11}$$

where $g(\cdot)$ is the Gaussian window function. As Gaussian window function has the minimal area of the Heisenberg box, it is usually applied to STFT (Wang et al., 2014). Then the TFR of the original signal $x(t)$ (for $t\in[u\text{-}\Delta t,\ u+\Delta t]$) can be approximated by $X_x(u,\xi)=X_{x_d}(u,\xi)$. The signal $x(t)$ can be recovered by the inverse DDT of $X_{x_d}(u,\xi)$ defined as

$$x(t)=\frac{1}{2\pi\,\|g(t)\|^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}X_{x_d}(u,\xi)g(t-u)e^{i\xi(t-u)}dud\xi\,. \tag{6.12}$$

If we apply the procedures above to the signal at every time instant, the signal is demodulated step by step. Connecting the small pieces of all time instants, the full TFR of the signal is then attained. With the DDT, not only can the actual frequency be retrieved but the energy concentration of TFR can also be improved.

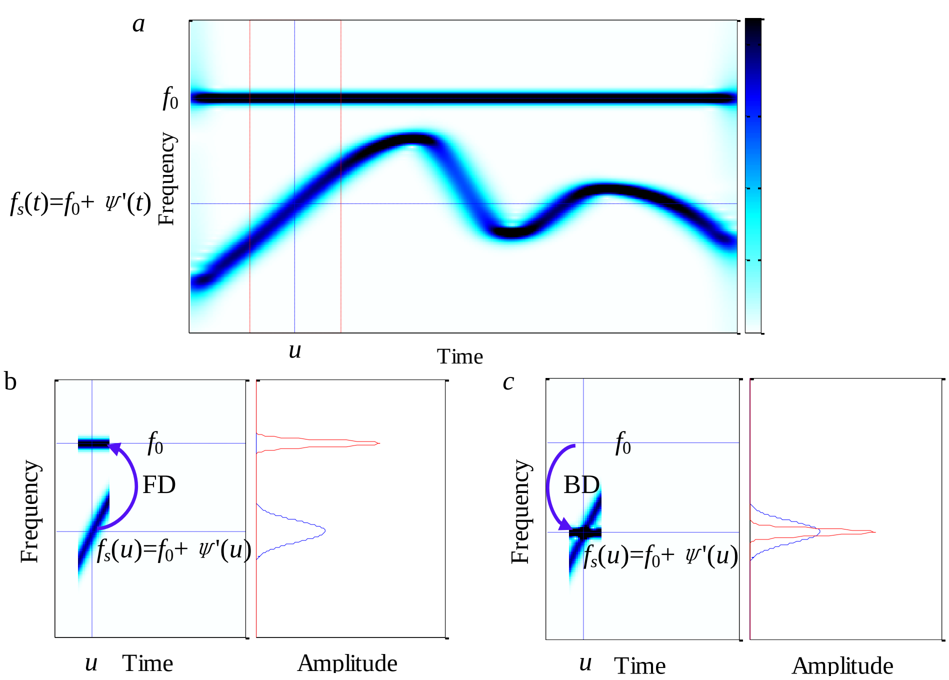


Fig. 6.1 The demonstration of the proposed DDT: (a) TFR of the original signal and generalized demodulated signal, (b) TFR of a signal segment and (c) TFR after frequency shifting (FD: forward demodulation, BD: backward demodulation), (the level of darkness reflects the level of energy) .

To better evaluate the performance of the proposed method, a comparison and quantitative analysis are given below. The difference before and after applying DDT to a signal is illustrated in Fig. 6.2. By comparing the TFRs respectively obtained by STFT and DDT, as shown in Fig 2 (a) and (b), it can be seen that the energy concentration level is strengthened and the TFR of the signal is sharpened by DDT. To provide quantitative analysis, the modulus maxima along the TFR ridge are searched. As shown in Fig. 6.2 (c), the modulus of the TFR obtained by DDT is much higher than that obtained by STFT because the energy is more concentrated. Quantitatively, the modulus maxima of the TF ridge generated by DDT is averagely improved by 122.44% (denoted by η), compared with that of the TFR attained by STFT. The number η is computed by

$$\eta = \frac{1}{I}\sum_{i=1}^{I}\frac{Mod(\mathrm{DDT}(i)) - Mod(\mathrm{STFT}(i))}{Mod(\mathrm{STFT}(i))} \tag{6.13}$$

where I is the length of DDT or STFT and $Mod(\cdot)$ denotes the modulus maxima.

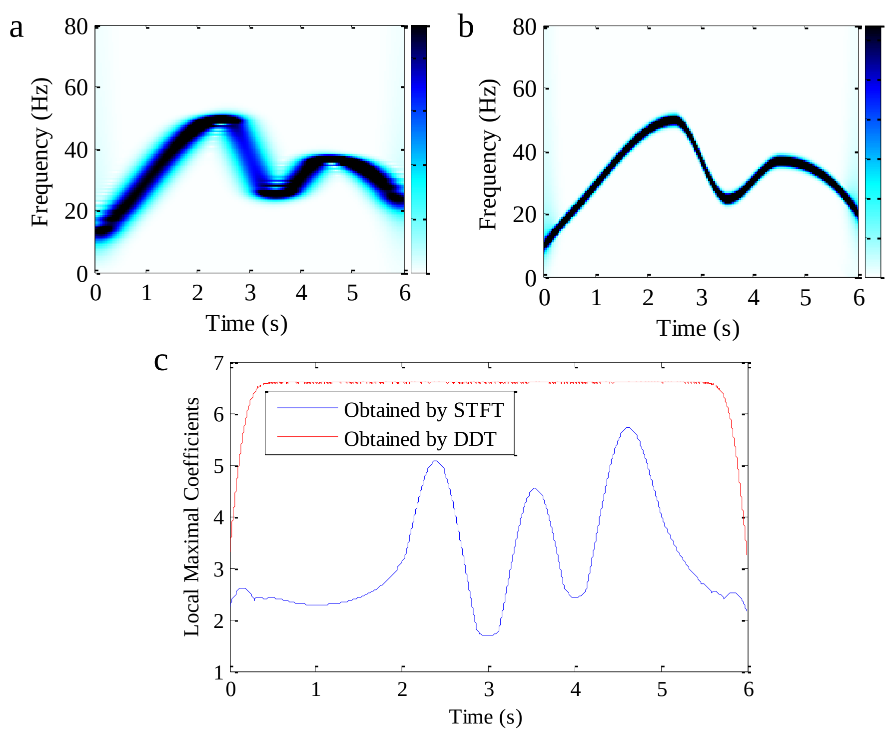


Fig. 6.2 Demonstration of the proposed DDT performance (without noise): (a) TFR obtained by STFT, (b) TFR obtained by DDT and (c) modulus maxima of the IF trajectories using STFT and DDT, respectively.

Next, the proposed DDT is applied to noisy signal (SNR = -8 dB) which is generated by adding white Gaussian noise to the signal used for Fig. 6.2. The TFR generated by STFT and the one obtained by DDT are presented in Fig. 6.3 (a) and (b), respectively. As can be seen, the TFR is sharpened by the DDT. The improvement can be clearly observed in Fig. 6.3 (c), where the modulus maxima of TF ridges using the DDT are much larger than that using the STFT. On average, the modulus maxima of TF trajectory are improved by 90.1% via the DDT in this case.

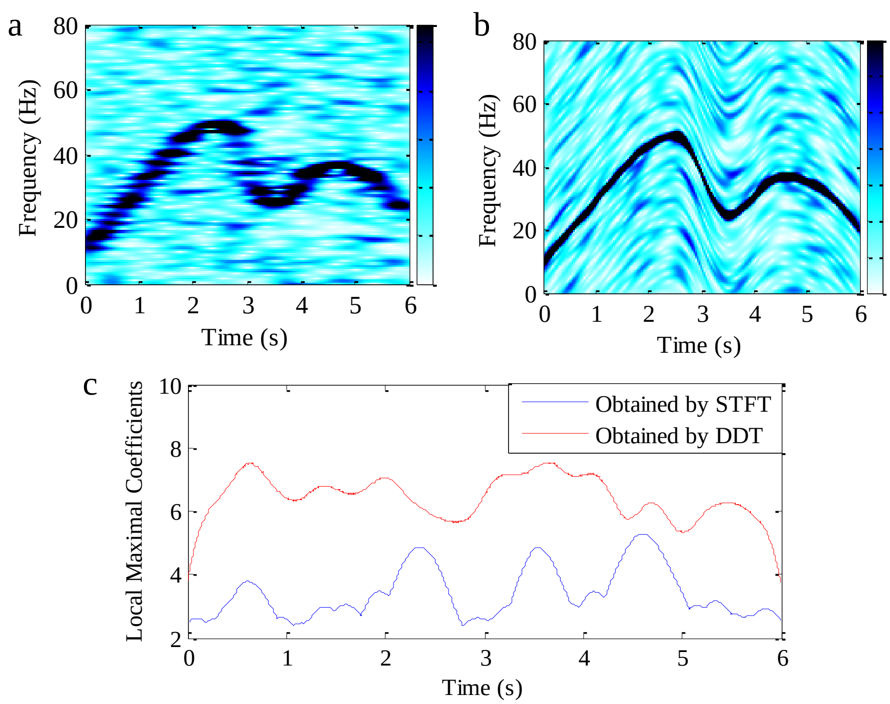


Fig. 6.3 Demonstration of the proposed DDT performance (with noise): (a) TFR obtained by STFT, (b) TFR obtained by DDT and (c) modulus maxima of the IF trajectories using STFT and DDT, respectively.

6.1.2 Iterative DDT (IDDT) for multicomponent signals

The proposed DDT for monocomponent signal TF analysis is detailed in the last subsection. However, in real applications, a signal usually contains multiple components and all components are frequency demodulated by different modulating sources. Single DDT is therefore not sufficient to demodulate all constituent components. As such, we propose the IDDT to address this issue. Compared with the DDT, the IDDT, as suggested by the name, involves an iterative procedure. Specifically, the multicomponent signal subjected to different modulating sources is iteratively demodulated by different demodulators, including a filtering operation between the forward demodulation and backward demodulation to separate each monocomponent signal from the other signal components. The TFR of the entire signal can be attained by superposing the

TFRs of all the signal components. Consider a multicomponent signal consisting of K components

$$x(t) = \sum_{k=1}^{K} x_k(t) = \sum_{k=1}^{K} A_k(t) e^{i2\pi[f_{0,k}t + \psi_k(t)]} . \tag{6.14}$$

The IF of the kth component can be

$$f_{s_k}(t) = f_{0,k} + \psi_k{}'(t) . \tag{6.15}$$

Based on the demodulator defined by Eq. (6.8), the kth demodulator can be expressed as

$$d_k(t,u) = e^{-i2\pi[\psi_k(t) - \psi_k'(u)t]} . \tag{6.16}$$

With the demodulator, the kth component can then be demodulated as

$$x_{d,k}(t,u) = x_k(t) \cdot d_k(t,u) = A_k(t) e^{i2\pi[f_{0,k}t + \psi_k'(u)t]} . \tag{6.17}$$

Except for the kth component, the rest of the signal can also be transformed into

$$x_d(t,u) = x_{d,k}(t,u) + \sum_{i=1, i\neq k}^{K} x_i(t) \cdot d_k(t,u) = x_{d,k}(t,u) + \sum_{i=1, i\neq k}^{K} x_{d,i,k}(t,u) \tag{6.18}$$

where the $x_{d,i,k}(t, u)$ is the demodulated item of the ith component by the kth demodulator, i.e.,

$$x_{d,i,k}(t,u) = x_i(t) \cdot d_k(t,u) = A_i(t) e^{i2\pi[f_{0,i}t + \psi_i(t) - \psi_k(t) + \psi_k'(u)t]} . \tag{6.19}$$

The IF of the ith component after being demodulated by kth demodulator is

$$f_i(t) = f_{0,i} + \psi_i'(t) + \psi_k'(u) - \psi_k'(t) \;\; for \, i = 1, ..., K, i \neq k . \tag{6.20}$$

By examining Eq. (6.20) and considering a small window around time instant u, one has $\psi_k'(u) - \psi_k'(t) \approx 0$. Hence the IF of the ith component can be approximated as $f_i(t)|_{[u-\Delta t, u+\Delta t]} = f_{0,i} + \psi_i'(t)$, which has seemingly no effect on the IF of kth component. However, when the kth and ith components are very close to each other, we may also have $\psi_i'(t) - \psi_k'(t) \approx 0$. Then the IF of ith component is $f_i(t)|_{[u-\Delta t, u+\Delta t]} = f_{0,i} + \psi_k'(u)$. In this case, the distorted IF of the ith component may have adverse effects on the closely located kth component, causing blurred TF or even misleading estimate for the IF of kth component. The above analysis shows that demodulating the entire signal by the demodulator of the target component is not favourable when handling signals with very closely located TF components.

Therefore, we propose to filter the target component after the use of the GD (i.e., the forward demodulation) via a band-pass filter whose center frequency is equal to the constant frequency $f_{0,k}$, and bandwidth equal to the minimum frequency difference between the instantaneous frequency of the kth component and the lower/upper ones. It is important to point out that there is no frequency overlap between the kth component and other components as the forward demodulation can transform its IF into an arbitrary constant frequency sharing completely different frequency band with others. With this operation, the adverse effects of other components on the kth component are eliminated. By iteratively applying the DDT plus forward demodulation-based filtering between the forward demodulation and backward demodulation, the TFR of each component can then be obtained. Accordingly the complete TFR of the original signal can be constructed by superposing the TFR of all the constituent components. The above procedures of IDDT for nonstationary multicomponent signal processing can be summarized in the flow chart shown in Fig. 6.4.

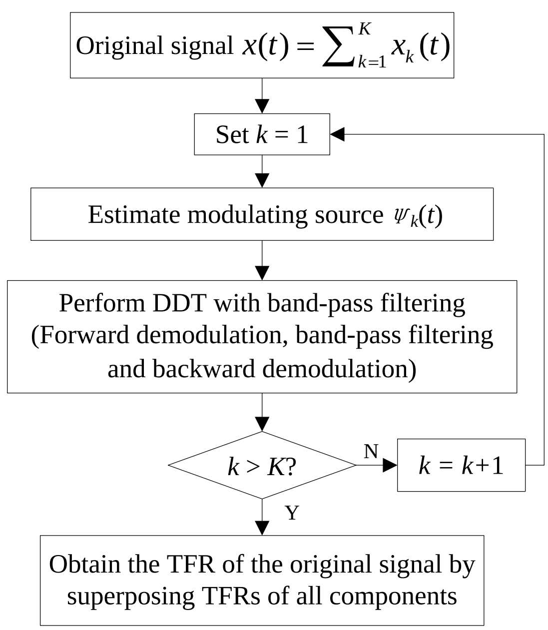


Fig. 6.4 The flow chart of the IDDT for nonstationary multicomponent signals.

The IDDT for nonstationary multicomponent signals is an improved version of the original DDT. With this enhancement, a substantially sharpened and cross-term free TFR can be achieved for a nonstationary multicomponent signal. The cross-term interferences are naturally avoided because the final TFR is a linear superposition of the TFRs of all individual components. This advantage is particularly important to obtain an unequivocal TFR from a nonstationary multicomponent signal.

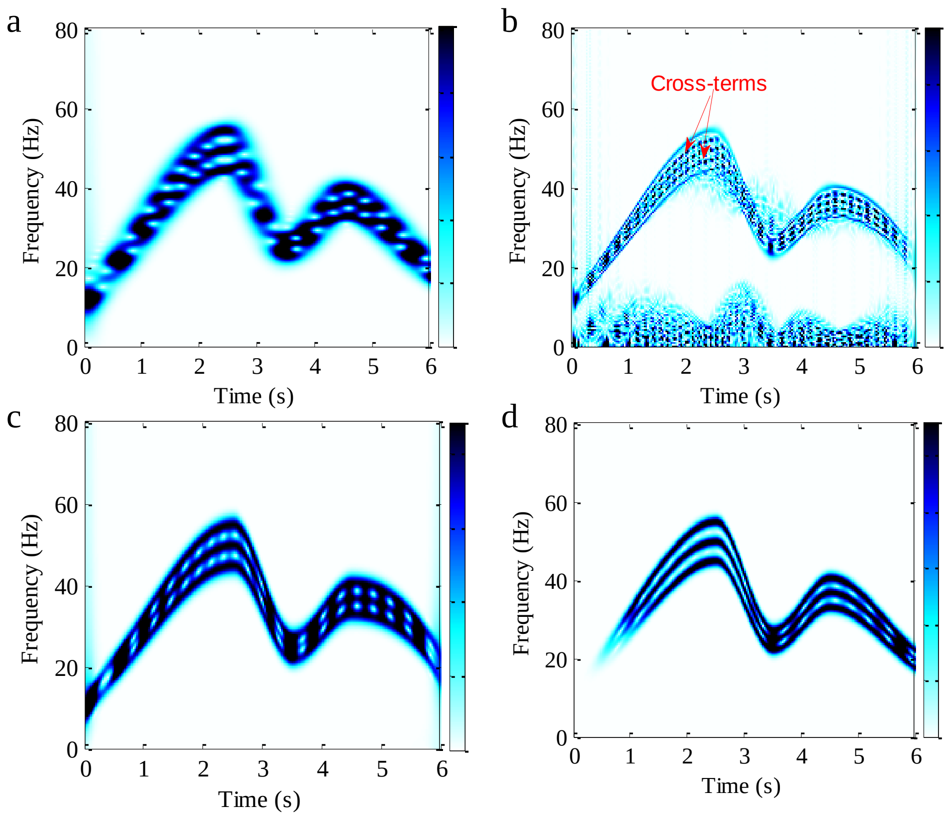


Fig. 6.5 TFRs of the nonstationary multicomponent signal obtained by: (a) STFT, (b) PWVD, (c) IDDT excluding filtering and (d) IDDT.

To illustrate the advantage, Fig. 6.5 is plotted, where the TFRs respectively obtained by STFT, Pseudo Wigner-Ville Distribution (PWVD), IDDT without filtering procedure and IDDT are shown. The STFT, Fig. 6.5 (a), can roughly show the trend of the signal frequency, but cannot distinguish the IF components at all, due to the limited TF resolution. The TFR obtained by PWVD, as shown in Fig. 6.5 (b), is undermined by cross terms. Particularly, when the IF components are closely located, the real IF components cannot be identified as they are clouded by the cross terms. The IDDT without involving filtering each signal component, Fig. 6.5 (c), improves the TF concentration and seizes the main features of signal components around the duration from 1.5 s to 3 s. However, there are blurred problems in the durations 3 s to 6 s. The result of IDDT, as presented in Fig. 6.5 (d), is almost free from blur and cross-terms, and hence clearly reveals each IF component of the signal. This is mainly due to the forward demodulation-based filtering processing which ensures the monocomponent requirement of DDT.

6.1.3 DDT-based synchrosqueezing transform

Although the blurriness and cross term issues have been alleviated by the IDDT, the TFR information during the first one second is missing as shown in Fig. 6.5 (d). To solve this problem, the synchrosqueezing transform (ST) which is originally proposed by Daubechies (Daubechies et al., 2011) to sharpen a TFR by reallocating the scalogram obtained from wavelet transform or STFT is employed to reassign the scalogram derived from IDDT. In this subsection, we will first introduce the STFT-based ST (Thakur and Wu, 2011), from which the DDT-based ST is detailed. For illustration, consider a signal having constant IF and IA, i.e., $x(t) = Ae^{i2\pi f_0 t}$. According to Eq. (6.11), the STFT of the signal can be expressed as

$$\begin{aligned} X_x(u,\xi) &= \int_{-\infty}^{+\infty} x(t) g(t-u) e^{-i\xi(t-u)} dt \\ &= A\int_{-\infty}^{+\infty} e^{i2\pi f_0 t} g(t-u) e^{-i\xi(t-u)} dt = Ae^{i\xi u} \int_{-\infty}^{+\infty} g(t-u) e^{-i(\xi-2\pi f_0)t} dt \\ &= AG(\xi - 2\pi f_0) e^{i2\pi f_0 u} \end{aligned} \tag{6.21}$$

where $G(\cdot)$ stands for the Fourier transform of the window function g. Eq. (6.21) indicates that the energy of signal will concentrate around the frequency line $\xi = 2\pi f_0$ in the TF plane. However in reality, the energy will distribute around the horizontal line, which raises the blurriness. The IF of the original signal can be obtained by performing the first-order derivative of X_x with respect to the time-shifting variable u, i.e.,

$$f_x(u,\xi) = \frac{-i}{2\pi X_x(u,\xi)} \frac{\partial X_x(u,\xi)}{\partial u} \quad . \tag{6.22}$$

To solve the blurriness, the TF distribution of STFT is reallocated by transforming the (u, ξ) plane to the $(u, f_x(u, \xi))$ plane via

$$T_x(u, f_l) = \sum_{\xi_k} X_x(u,\xi_k)(\Delta\xi)_k \tag{6.23}$$

where ξ_k, the kth discrete angular frequency, varies in the range specified by $| f_s(u,\xi_k) - f_l | \le \Delta f / 2$, f_l is the lth discrete frequency, $\Delta f = f_l - f_{l-1}$ and $\Delta\xi_k = \xi_k - \xi_{k-1}$. $X_x(u, \xi_k)$ is calculated only at the discrete value ξ_k and the synchrosqueezing result is determined at the center f_l of the interval $[f_l - (1/2)\Delta f, f_l + (1/2)\Delta f]$. In this way, the TFR is improved by squeezing the blurred ridges resulted from the STFT along the frequency axis into more concentrated ones.

The DDT-based ST is motivated from the STFT-based ST described above and is elaborated in the following. Replacing the original signal $x(t)$ and the original TFR $X_x(u, \xi)$ in Eq. (6.22) by the demodulated signal $x_d(t)$ and its TFR $X_{xd}(u, \xi)$ leads to the IF of the demodulated signal, i.e.,

$$f_{x_d}(u,\xi) = \frac{-i}{2\pi X_{x_d}(u,\xi)} \frac{\partial X_{x_d}(u,\xi)}{\partial u} \ . \tag{6.24}$$

Similar to those in the STFT-based ST, the u, ξ and f_{xd} are all discrete and $\Delta\xi_k$ is defined as $\Delta\xi_k = \xi_k - \xi_{k-1}$. When mapping from the (u, ξ) plane to the actual TF plane $(u,\ f_{x_d}(u, \xi))$ plane, the squeezing step for the DDT-based ST performing at the center f_l can be expressed as

$$T_x(u,\xi) = \sum_{\xi_k} X_{x_d}(u,\xi_k)(\Delta\xi)_k \tag{6.25}$$

where ξ_k, the kth discrete angular frequency, varies in the range determined by $| f_{x_d}(u,\xi_k) - f_l | \le \Delta f / 2$, f_l and Δf are defined as the same as those for the STFT-based ST.

Via the DDT-based ST, the DDT results can be further condensed by the synchrosqueezing and the TF readability is hence further improved. The IDDT-based ST is based on the DDT-based ST and involves synchrosqueezing for each signal component. After processing by the IDDT-based ST, the TFRs of the same multicomponent signal previously illustrated in Fig. 6.5 are presented in Fig. 6.6, where Fig. 6.6 (a) displays the result of the IDDT-based ST without filtering and Fig. 6.6 (b) shows the result of IDDT-based ST. It can be seen that the energy concentration of both TFRs is further enhanced, compared with Fig. 6.5. However, Fig. 6.6 (b) shows a better outcome as each IF component is much more discernible. Moreover, it can be seen in Fig. 6.6 (b) that the missing information in the first one second, as shown in Fig. 6.5 (d), is restored by ST. The comparison between Fig. 6.6 (a) and (b) also shows the importance of the pre-process of ST and further demonstrates the significance of the filtering operation after forward demodulation for multicomponent signals.

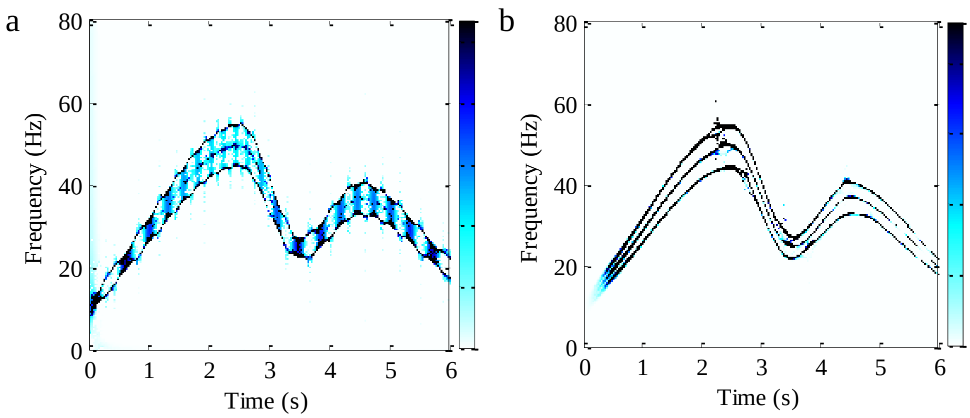


Fig. 6.6 TFRs of the nonstationary multicomponent signal obtained by: (a) IDDT-based ST excluding filtering and (b) IDDT-based ST.

6.1.4 How to implement the proposed DDT/IDDT in practice

In the previous subsections, it is assumed that the modulating source of the signal is known in advance, and thus the demodulator can be easily derived. Nevertheless, in most of real applications, the modulating source of the signal is often unknown, which makes it impossible for direct demodulation. The study in (Feng et al., 2015) estimates the IF using the conventional TFA methods such as STFT and WT. However, the inherent issue of the approach is that estimating IF once may not guarantee the required accuracy of the IF, hence leading to the smearing problem in the TF plane. Therefore, the proposed method in this study involves an iterative procedure to repeatedly improve IF prediction until the IF precision is acceptable. The initial IF is estimated from the original signal via the conventional TFA methods and used to roughly demodulate the signal. The more accurate IF can then be obtained from the enhanced TFR and used to further demodulate the signal to further enhance the TFR. The iteration continues until one of the stopping criteria is satisfied. The flow chart for the iteration is shown in Fig. 6.7, where δ represents MSE difference between two successive IFs, M stands for the maximum number of iterations. The IF $f_{x(1)}$ is initially estimated using STFT, from which the TFR $X_{xd(1)}$ can be calculated using DDT. Estimating IF $f_{x(2)}$ based on $X_{xd(1)}$, one can calculate the TFR $X_{xd(2)}$. The iteration continues until the one of the stopping criteria is satisfied.

By means of the IF refinement procedure, the estimated IF gradually approaches the real IF. The parameters M and δ are user-specified. In this study, they are set 20 and 10^{-4}, respectively. To illustrate the effectiveness of the IF refinement, the signal presented in subsection 2.2 is used again here. The simulated signal is a sinusoidal function with time-varying frequency and is contaminated by white Gaussian noise with a SNR of -8 dB. The IF is created by setting a few values at corresponding time instants and then applying interpolation (using *interp1* function in MATLAB$_{TM}$). Fig. 6.8 shows the improving process of the MSE between the estimated IF and actual IF. As can be seen, the MSE has been reduced from 1.4475 to 7.497×10^{-4} in only six iterations. No significant improvement is observed after seven iterations. This may suggest that we have practically obtained the true IF after six iterations.

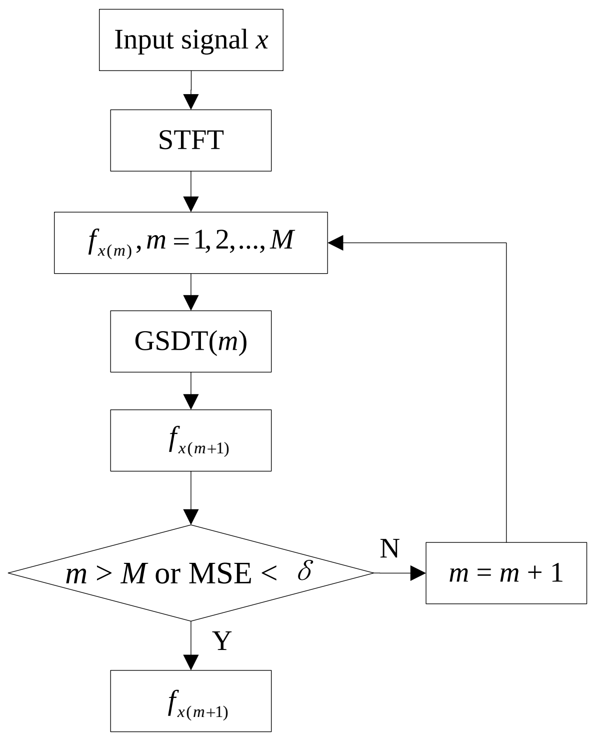


Fig. 6.7 The flowchart of IF refinement (MSE = mean squared error).

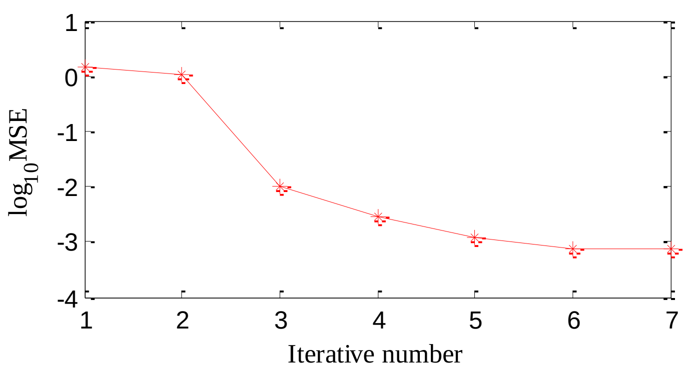


Fig. 6.8 The evolving history of the MSE between the estimated IF and the actual IF.

6.2 Simulated case study

In this section, simulation studies are presented to validate the capability of the proposed method in TFR enhancement for both monocomponent and multicomponent signals polluted by additive noise.

6.2.1 Monocomponent signal

Consider the monocomponent signal used in subsection 2.2. This signal is generated by a frequency modulated sinusoidal function. The IF is created by interpolation (using *interp1* function in MATLAB). The simulated signal $x(t)$ can be expressed as

$$x(t) = \cos(2\pi\psi(t)) + n(t) = \cos(2\pi\int_0^t f(t)dt) + n(t) \tag{6.26}$$

where the function of the IF obtained by curve fitting is

$$f(t) = 40\sin(0.36t + 0.6) + 10\sin(2t - 2.5) + 5.4\sin(3.6t - 7.6) \tag{6.27}$$

and $n(t)$ is the white Gaussian noise. In the practical implementation, the $\psi(t)$ is produced by computing the integral of the IF vector. Adding the noise to the monocomponent signal generates a synthetic signal with a SNR of -8 dB, as plotted in Fig. 6.9 (a). The actual IF of the synthetic signal is shown in Fig. 6.9 (b). Fig. 6.9 (e)-(f) present the TFRs obtained using the CWT, PWVD, STFT and STFT-based ST, respectively. It can be observed that the ineffective energy concentration capability of such methods lead to blurred TFRs. Therefore, the TFR ridges cannot accurately reflect the true IF.

The proposed DDT method is then applied to the monocomponent signal. The initial IF $f_{x(1)}(t)$ of the signal is estimated via the traditional STFT. According to the analysis in subsection 2.1, we choose f_c = 100 Hz, and hence the signal is mapped to the target signal of constant frequency f_c = 100 Hz. The composite demodulator defined by Eq. (6.8) can then be calculated for the first iteration, where the $\psi_{(1)}(t)$ and $\psi'_{(1)}(t)$ are respectively obtained by $\psi_{(1)}(t) = \int_0^t (f_{x(1)}(t) - f_c)\,\mathrm{d}t$ and $\psi'_{(1)}(t) = f_{x(1)}(t) - f_c$. With the demodulator, the DDT can be implemented and the TFR $X_{xd(1)}$ is obtained. The IF continues to be refined via the procedure given in Fig. 6.7 until one of the stopping criteria is reached. According to Fig. 6.8, the TFR of the 7th iteration is the final result, as shown in Fig. 6.9 (g). The TFR of the corresponding DDT-based ST is given in Fig. 6.9 (h).

Comparing Fig. 6.9 (c)-(f) with Fig. 6.9 (g) - (h) reveals that the proposed DDT method outperforms the existing TFR tools.

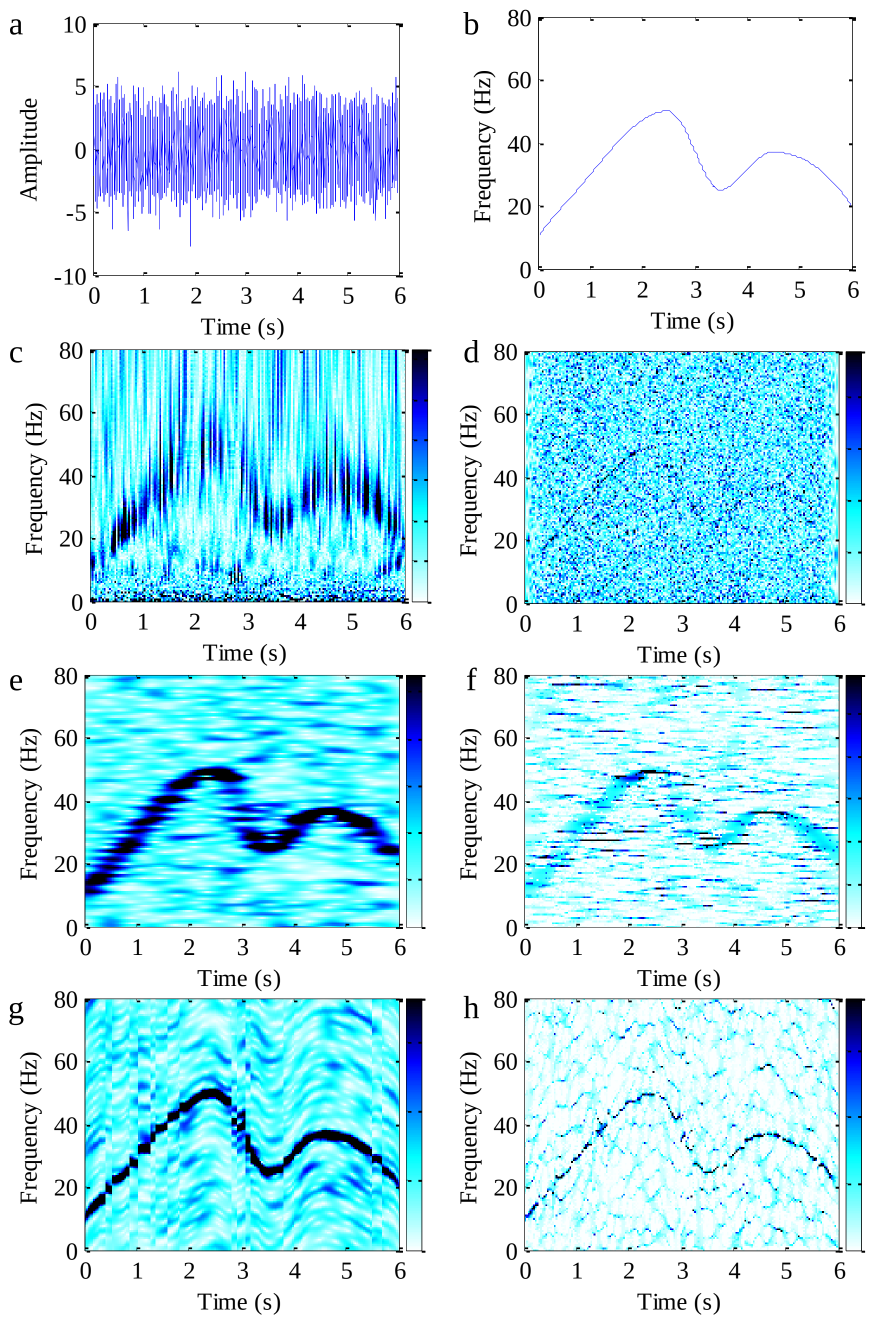


Fig. 6.9 The simulated monocomponent signal, real IF and TFRs of the signal generated by various TFA methods: (a) the time domain waveform, (b) the real IF, (c) – (h) TFRs of the signal obtained respectively by CWT, PWVD, STFT, STFT-based ST, DDT and DDT-based ST.

To illustrate the effectiveness of the IF refinement procedure, the TFRs obtained by the DDT and the DDT-based ST using the initial IF are given in Fig. 6.10 (a) and (b) respectively. It can be seen that, owning to the difference between real IF and the initial IF, the resulting TFR based on the initial IF in Fig. 6.10 (a) is not as clear as the final result in Fig. 6.9 (g). The significant difference can be observed by comparing Fig. 6.10 (b) with Fig. 6.9 (h). Fig. 6.10 (c) and (d) present the TFRs obtained by the DDT and the DDT-based ST using the real IF. Apparently, utilizing the real IF gives the best TFR result, which in turn shows that the accuracy of estimated IF has a significant impact on the TFR. The comparison demonstrates that the TFR of the signal can be enhanced by the IF refinement process.

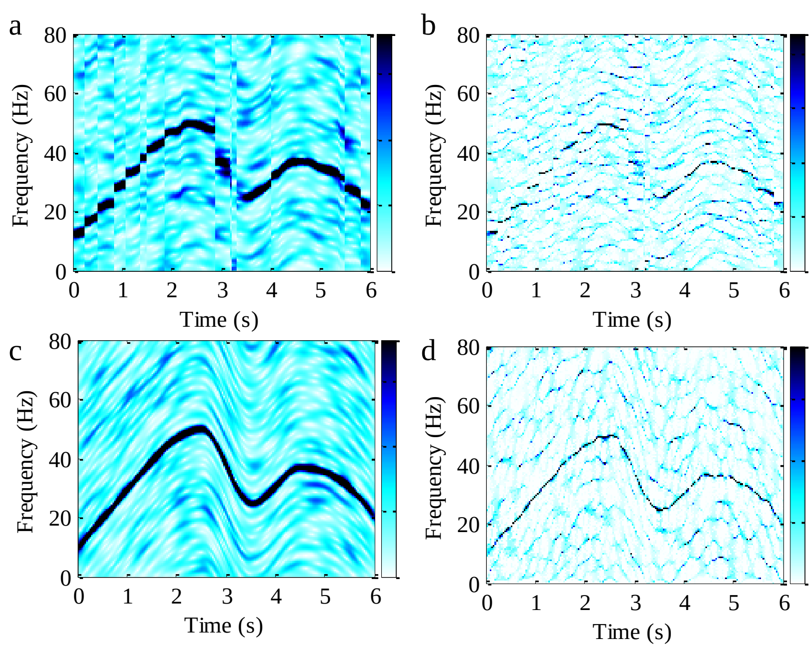


Fig. 6.10 The illustration of the effectiveness of the IF refinement procedure: TFRs obtained by DDT using (a) the initial IF and (c) the real IF respectively, and TFRs obtained by DDT-based ST using (b) the initial IF and (d) the real IF, respectively.

6.2.2 Multicomponent signal

In this subsection, we apply the proposed IDDT method to a nonstationary multicomponent signal with five closely located IF components whose modulating sources are different from each other. The multicomponent signal $s(t)$ is defined as

$$x(t) = \sum_{k=1}^{5} \cos(2\pi \int_0^t f_k(t)dt) + n(t) \tag{6.28}$$

where $f_k(t) = C_k f(t)$, $C_k = 0.5, 0.9, 1, 1.1, 1.5$, $f(t)$ is defined in the last subsection by Eq. (6.27) and $n(t)$ is the additive white Gaussian noise. The IF components are set proportional to each other because, in the application of bearing/gear fault diagnosis, the shaft frequency, IFCF and their harmonics are always proportional. The composite signal with a SNR of -8 dB has 6000 samples in the time interval [0, 6] and is plotted in Fig. 6.11 (a) as shown in the following.

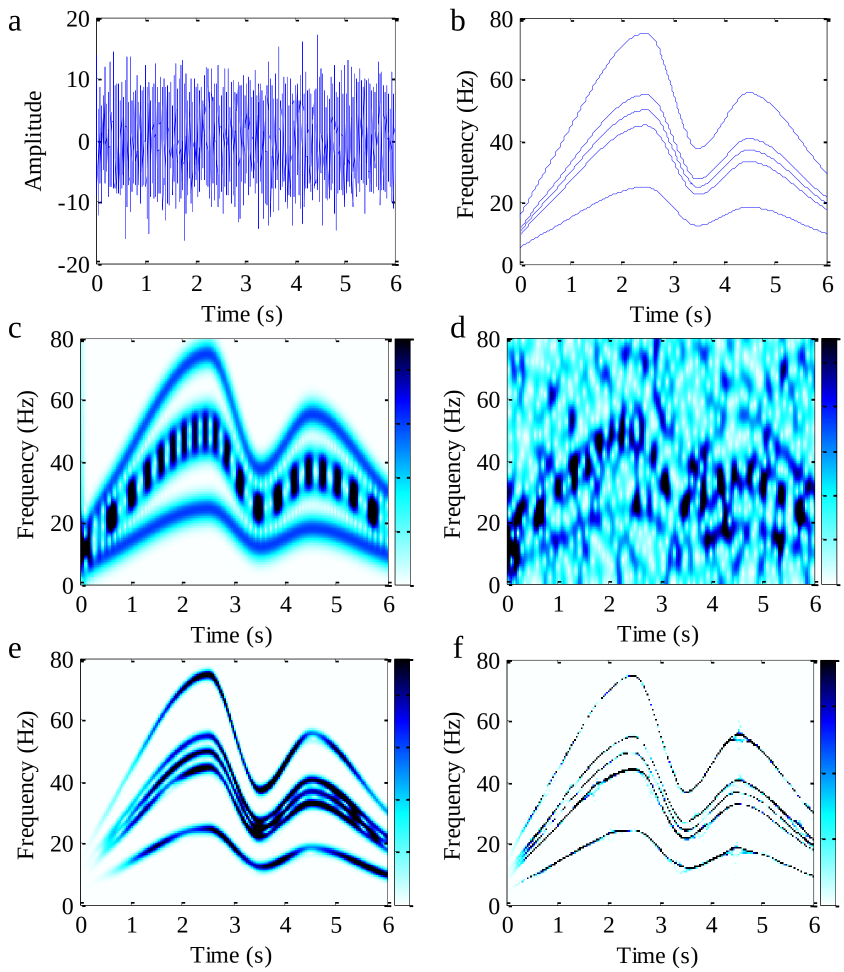


Fig. 6.11 TFRs of the multicomponent signal with a SNR of -8 dB: (a) the composite signal, (b) the real IFs of all components, (c)-(d) the TFRs obtained by STFT without and with noise, (e) TFR of the noisy signal using IDDT and (f) TFR of the noisy signal using IDDT-based ST.

The real IFs of all the signal components are shown in Fig. 6.11 (b). The STFT resulting TFRs of the multicomponent signal without and with noise are given in Fig. 6.11 (c) and (d), respectively. It can be observed that even though the signal is clean the IF components via STFT are not identified in the TF plane due to the spread energy distribution. The performance of

STFT becomes worse when the signal is mixed with noise (SNR = -8 dB), as presented in Fig. 6.11 (d). The proposed IDDT method is then applied to the composite signal and the resulting TFR is displayed in Fig. 6.11 (e). We can see that the energy for each IF component is more condensed; therefore, the TF components can be clearly recognized in the TF plane, particularly for the three very closely positioned IF components. The result is further improved by the synchrosqueezing, as seen in Fig. 6.11 (f) where the energy concentration of IF components is further condensed and all IF components are easily distinguished.

6.3 The application of the proposed method for bearing fault detection

The performance of a method for real applications is an important indicator to evaluate the method; therefore, the proposed method is applied to real cases in this section.

6.3.1 Inner race fault identification

The experiment is carried out on a SpectraQuest Machinery fault simulator (MKF-PK5M) in our lab at the University of Ottawa and its setup is shown in Fig. 6.12. The shaft is powered by a motor controlled by an AC inverter. There are two ER16K bearings mounted to support the shaft and the load. The pictures of the defective bearings are shown in Fig. 5.6. The one with inner race fault is positioned at the left side. The parameters of the bearings are given in Table 5.3. Tachometer and accelerometer are mounted on the test rig to collect the shaft speed and vibration signal, respectively. A load mass of 5.03 kg is mounted on a 25.4mm (1 inch) steel shaft. The shaft rotational speed f_r increases from 28.1 Hz to 33.7 Hz and then drops down to 28.1 Hz again. The sampling rate is 12 KHz and a signal of 6 s is collected. The data is fed to an NI data Acquisition Module (NI USB-6212 BNC) and is recorded by the computer with Labview software.

Since the signal from a faulty bearing is amplitude modulated and the signal component carrying the faulty information is often the envelope of the original signal, we propose to Hilbert demodulate the raw signal and then downsample the envelope signal. Via this operation, the computational efficiency is improved. The maxima of IFCF in our experiment is around 183 Hz (= 33.7 × 5.43). Taking three times of IFCF into consideration (two harmonics of IFCF is sufficient to show the existence of a fault), the new sampling frequency should be set greater than 1098 Hz according to Nyquist theorem. We select 1500 Hz in this case. It has to be mentioned that there might be aliasing resulted by high order harmonics after downsampling; however, these high order harmonics have relatively low energy and hence are of little effects on results. The original signal and the downsampled envelope signal are shown in Fig. 6.13 (a) and (b), respectively. Fig. 6.13 (c) is the STFT resulting TFR of the downsampled signal. We can roughly observe the IFCF and its two harmonics in Fig. 6.13 (c); the fault type of the bearing is, however, not revealed yet as the IF of the shaft is still ambiguous in the TF plane.

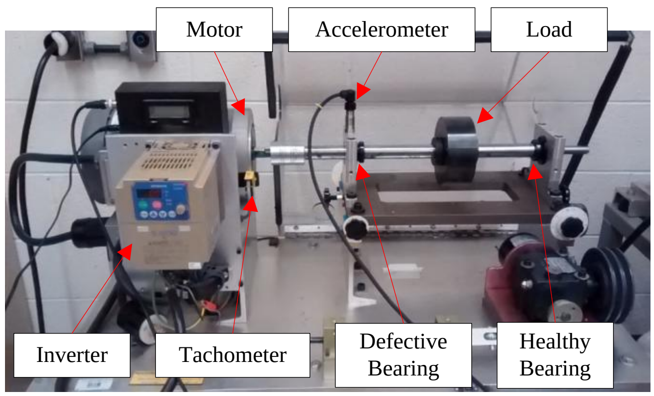


Fig. 6.12 Experimental setup.

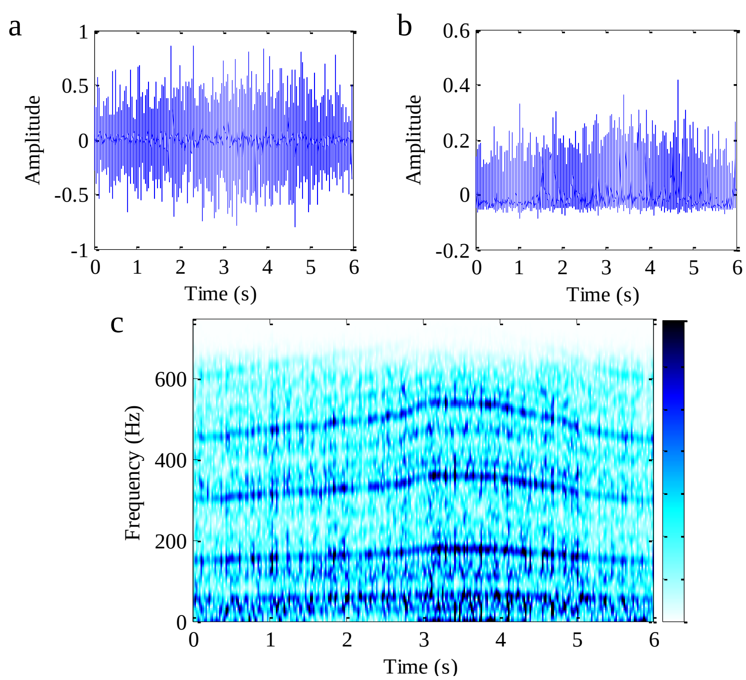


Fig. 6.13 The original signal of inner race fault and TFR obtained by STFT: (a) the original signal, (b) the downsampled signal and (c) TFR of the downsampled signal using STFT.

The proposed method is hence applied. To begin with, the IFCF with low-accuracy is first estimated based on the TFR obtained by STFT. Using the IF refinement procedure which is elaborated in subsection 6.1.5, the more accurate IFCF is obtained to further demodulate the signal. The final estimated IFCF is attained once the stopping criterion is satisfied ($M > 20$ or $\delta < 10^{-4}$). With the IFCF, the demodulator of the IFCF-related signal component can then be obtained. As the other components' IF of bearing vibration signal is proportional to the IFCF, the IF refinement procedure does not need to apply to every IF component. The demodulating sources for the 2×IFCF and 3×IFCF components can be directly calculated based on the demodulator of IFCF-related component. However, the demodulators for the shaft-related IF components are still unavailable because the fault type is unknown.

To solve this issue, the resampling-free order spectrum for shaft-related IF components is developed in this study. The horizontal axis of this spectrum represents the IFCF's coefficient λ (λ is a coefficient in front of the IFCF) which is multiplied by the IFCF to obtain different modulating sources (λ·IFCF), and the vertical axis is presented as amplitude. Since there is no resampling processing involved it is termed resampling-free order spectrum. The shaft IF f_r and its harmonics can be expressed as λ·IFCF, where λ can be calculated by i/K_{fc}, i = 1, 2… (K_{fc} is determined by the fault type and hence called the fault characteristc coefficient). For bearing fault diagnosis, setting the resolution of λ (denoted by $\Delta\lambda$) to be 0.001 is sufficient to distinguish each shaft-related IF (in this case, the resolution of shaft-related IFs is 0.184 (=1/5.43, as listed in Table 5.3). For the coefficient λ in the range of [0, λ_{max}], there are $\lambda_{max}/\Delta\lambda+1$ demodulators generated to demodulate the signal. Each demodulator can map an IF trajectory of a signal component at a constant frequency λf_c (the IFCF component is mapped to the constant frequency of f_c after GD). Iteratively performing frequency analysis on each generalized demodulated signal and searching the amplitude value corresponding to the frequency λf_c in the frequency spectrum, the resampling-free order spectrum is obtained, as shown in Fig. 6.14. In this case, we only consider the coefficient λ in the range of [0, 1] as this range contains the shaft IF and several its harmonics. As can be observed, the peaks on the spectrum represent the energy corresponding to these constant frequencies in which the shaft IF and its harmonics are situated after GD. According to the values of the peaks in Fig. 6.14, the demodulators of the shaft-related IF components can be determined. Only the shaft IF and its first harmonic are analyzed in this study, i.e., two demodulators are defined as λ_1·IFCF and λ_2·IFCF respectively (the values of λ_1 and λ_2 are as shown in Fig. 6.14). With the demodulators, the proposed IDDT can be implemented and the TFRs of the shaft-related IF components are obtained. Superposing the TFRs of IFCF and its two harmonics, as well as the shaft IF and its first harmonic leads to the TFR in Fig. 6.15 (a). The corresponding synchrosqueezing result is shown in Fig. 6.15 (b). As seen, both the IDDT and IDDT-based ST can provide TFR with satisfactory energy concentration and all IF components can be clearly identified. The IFCF and its harmonics demonstrate the existence of the fault, and the fault type is determined by the ratio between the shaft IF and the IFCF.

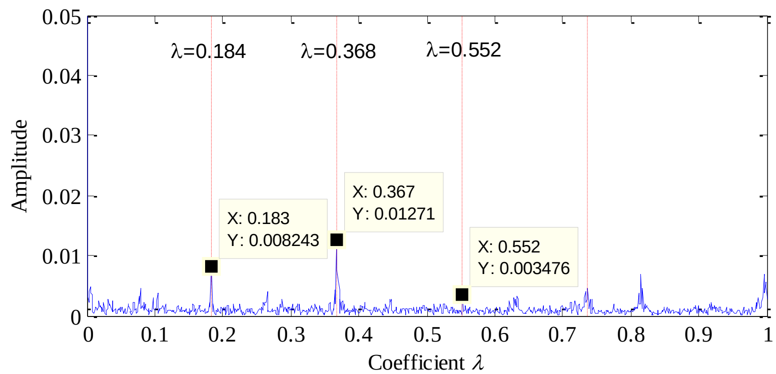


Fig. 6.14 The resampling-free order spectrum for $\lambda \in [0, 1]$ for inner race fault.

The comparison between Fig. 6.13 (c) and Fig. 6.15 reveals that the proposed method yields a better TFR than that resulting from the STFT, which also indicates that the proposed IDDT method can condense the energy and solve the blurred problems caused by multicomponents and time-variant instantaneous frequency. The above analyses show that the proposed method can effectively detect the fault for a real case.

6.3.2 Outer race fault identification

The bearings in Fig. 6.12 are replaced by two ER10K bearings and the faulty one is placed on the left side. A load mass of 5.03 kg is mounted on a steel shaft with diameter 15.875 mm (5/8 inch). The parameters of the bearing are listed in Table 5.2. The shaft rotational speed f_r increases from 18.2 Hz to 36 Hz following a nearly linear pattern. The sampling rate is 24 KHz and the data of the signal is 6 s long. The data is fed to an NI data Acquisition Module (NI USB-6212 BNC) and is recorded by a computer with Labview software.

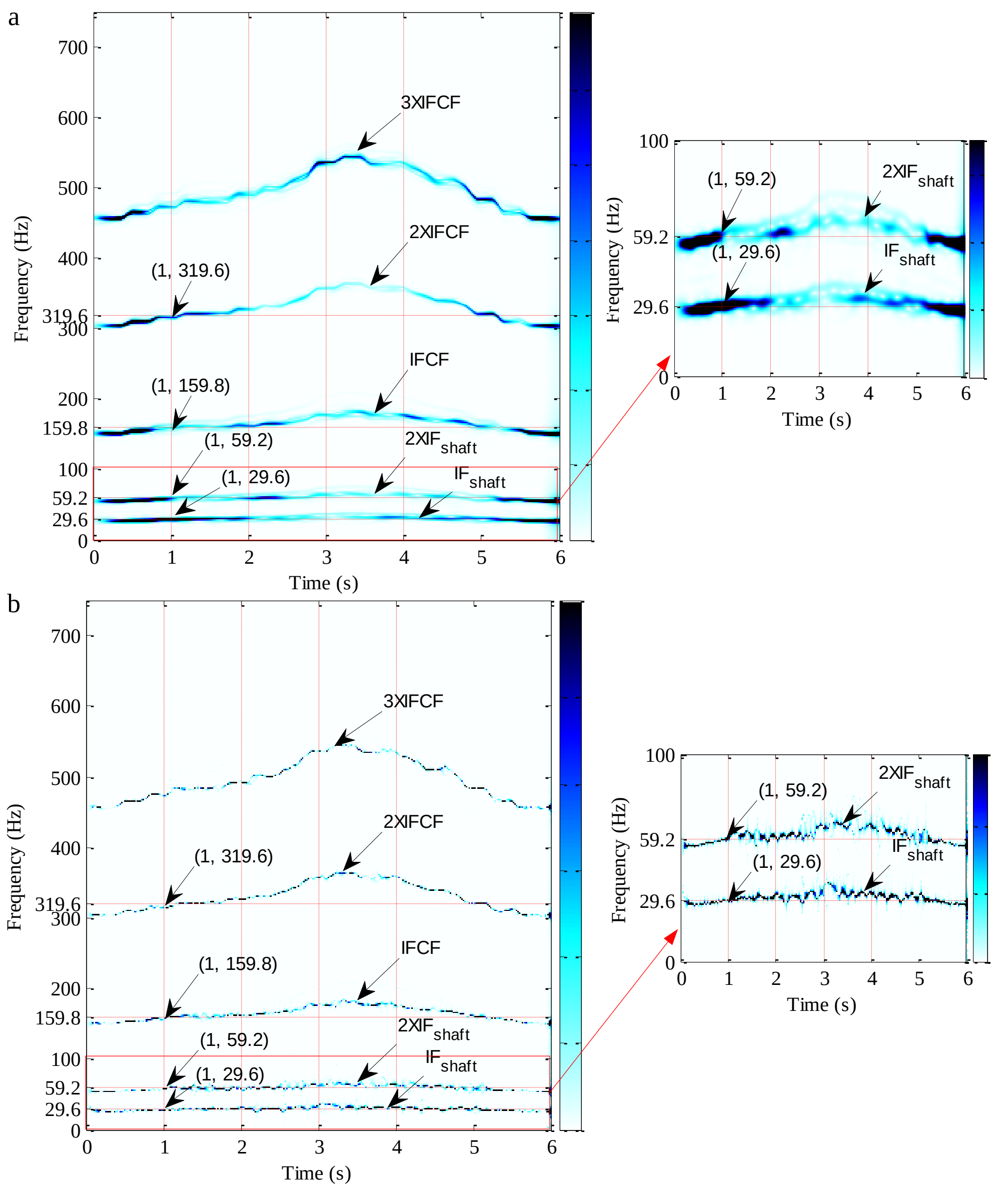


Fig. 6.15 TFRs of the downsampled signal for inner race fault obtained respectively by: (a) the IDDT and (b) the IDDT-based ST.

Likewise, the collected vibration signal in Fig. 6.16 (a) is first amplitude demodulated and then the extracted envelope signal is downsampled. Again, the IFCF and three of its harmonics are taken into consideration. The down sampling rate is then determined to be 1500 Hz. The downsampled envelope signal, displayed in Fig. 6.16 (b), is then processed using the classical STFT, yielding the TFR presented in Fig. 6.16 (c). As observed, the IFCF and its harmonics cannot be clearly identified, nor can the shaft IF and harmonics.

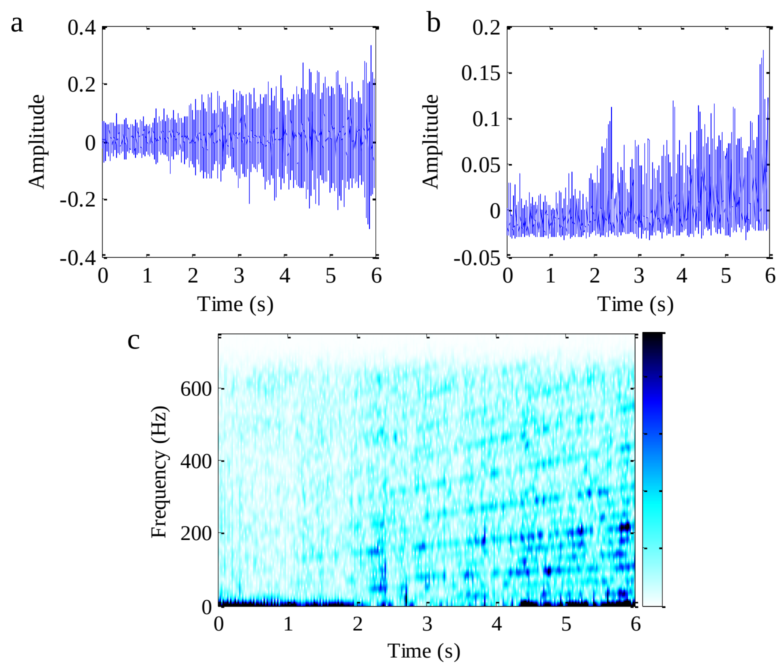


Fig. 6.16 The original signal of a bearing with a outer race fault and the TFR of the signal obtained by STFT: (a) the original signal, (b) the downsampled signal, and (c) TFR of the downsampled signal using STFT.

Next, the effectiveness of the proposed TFR method is examined in processing the same vibration data. The procedure is similar to that for the inner race fault case. The demodulator for the IFCF-related signal component can be obtained via the IF refinement procedure in Fig. 6.7. Demodulators of 2×IFCF- and 3×IFCF-related signal components are easily obtained by multiplying integers, i.e., λ·IFCF (λ = 2, 3). To calculate the demodulators for the shaft IF related signal components, the resampling-free order spectrum for the outer race fault case is obtained in the same manner as for inner race fault, as shown in Fig. 6.17. As can be seen in Fig. 6.17, the demodulators for shaft IF and its first harmonic can be attained, i.e., λ·IFCF (λ = 0.328, 0.567).

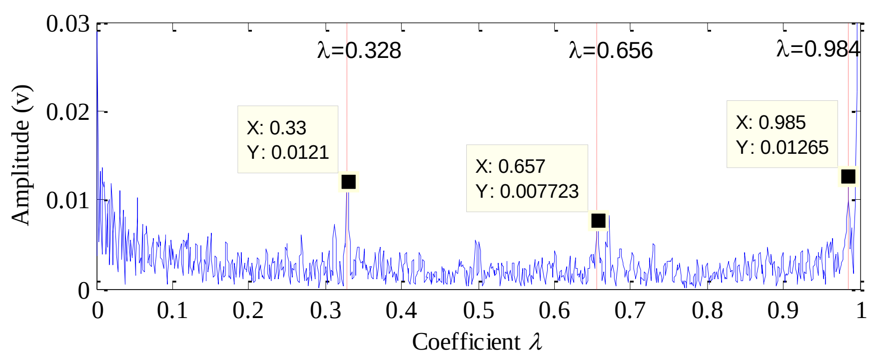


Fig. 6.17 The resampling-free order spectrum for $\lambda \in$ [0, 1] for outer race fault.

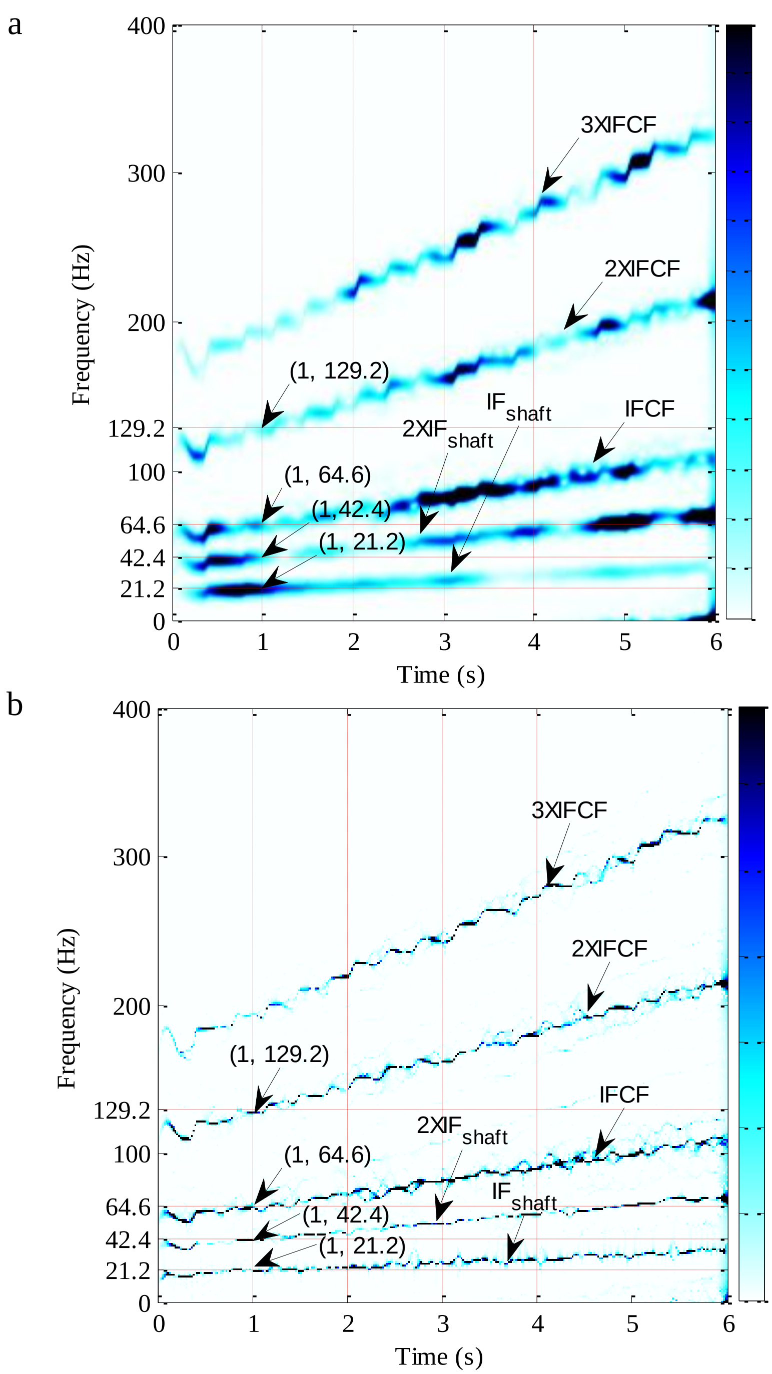


Fig. 6.18 TFRs of the downsampled signal for outer race fault obtained respectively by: (a) the IDDT, and (b) the IDDT-based ST.

With the obtained demodulators, the energy concentration level of each signal component can be improved and thus TFR of each signal component can be sharpened. Superposing the TFRs of all signal components of interest yields the TFR in Fig. 6.18 (a), from which the not only can IFCF and two of its harmonics be recognizable but also the shaft IF and its first harmonic are clearly observed. Fig. 6.18 (b) presents the synchrosqueezing result. It can be seen that the TFR of each component is further sharpened. Based on any of the two, it can be concluded that the bearing is defective and the location of the flaw is on the outer race. The outer race fault experiment hence further validates the effectiveness of the proposed method in bearing fault diagnosis.

6.4 Conclusion

This chapter has developed the dual demodulation transform (DDT) by utilizing the properties of the generalized demodulation and Fourier transform. The proposed DDT can effectively enhance the TF readability by improving the energy concentration level in the TF plane via the forward and backward demodulations. To better handle the nonstationary multicomponent signals with closely located IF components, the IDDT is explored. The IDDT can decompose nonstationary multicomponent signal into monocomponents with constant frequencies and enhance energy concentration of each component's TFR, thereby suppressing the bi-dimensional diffusion and naturally solving the blurred issue resulted by multicomponents and variable frequency. The proposed method also involves an IF refinement procedure to continuously improve the accuracy of estimated IF, which significantly alleviates the results' dependence of initially estimated IF.

The effectiveness of the proposed method in terms of energy concentration and nonstationary multicomponent signal TFA has been validated by both simulated and experimental signals. Compared with traditional TFA methods, the TF readability and estimated IF accuracy have been improved via the proposed method. The modulus maxima of TFRs for the simulated signal with (SNR = -8 dB) and without noise has been averagely improved by 90.1% and 122.44%, respectively, in comparison with the results using STFT. The MSE of the estimated IF has been reduced from 1.4475 to 7.497×10^{-4} via the IF refinement procedure for the simulated signal. By exploiting a resampling-free order spectrum, the proposed method can also be applied to bearing fault diagnosis under variable condition. The bearing condition can be precisely diagnosed and the fault type can also be identified. The analyses have demonstrated the effectiveness of the proposed method in TFA and bearing fault detection.

Chapter 7 Conclusions, Limitations and Future Research

7.1 Conclusions

7.1.1 For constant speed condition

Each time a rolling element moves over a damaged area under load, an impulse is generated and the resonance of the system is also excited. The main goal of bearing fault diagnosis is to extract the impulses and reveal the frequency information associated with the impulse repetition, i.e., FCF. However, the frequency of the impulse repetition modulates the resonance frequencies. The first challenge of bearing condition monitoring is, therefore, the modulation effect. If the signal is clear enough, the FCF can be revealed via the envelope demodulation techniques. Nevertheless, a measured vibration signal is often severely tainted by a variety of interfering interferences produced by other mechanical components and background noises, particularly at the early stage of a bearing failure. Based on the foregoing description, this issue can be addressed from two different perspectives: 1) developing an effective envelope demodulation method which can suppress interferences and noise, or 2) separating the impulsive signal from vibration interference and noise components. Most of the existing enveloping methods are susceptible to interferences and noise. The widely used methods for interference and noise removal are usually dependent on prefiltering. The difficulty of filter parameter specification largely confines the application scope of a filter-based method. Moreover most of the existing methods cannot remove in-band interferences. As such, this thesis is directed towards the development of two methodologies pertinent to the two perspectives mentioned above. The two methodologies are as follows.

(1) A FD-based Envelope Demodulation Method for Fast Bearing Fault Feature Extraction without Prefiltering

This method relates the FD to signal morphological complexity. By the proposed STFD transform, the envelope of impulses can be extracted. The calculation of STFD transform using Katz FD is first discussed. Based on the discussion, it has been found that both amplitude (fractal) component and time (non-fractal) component should be involved to obtain a satisfactory envelope signal that is neither too oscillatory nor over-smooth.

Due to the different signal complexity between fault-induced impacts and interferences, the proposed STFD transform can suppress interferences in the STFD representation and is particularly effective in the suppression of low frequency interferences. When the interference frequencies become large, the interferences would adversely affect the STFD representation. Therefore, the KPSA algorithm is proposed to overcome the interference-related effect on STFD representation. The proposed method, as described in Chapter 3, can handle signals with SIR as low as – 33 dB.

The proposed method has been validated by simulation and experiments. The comparison between the proposed method and other three widely used envelope techniques has shown that the proposed method can more effectively handle signals tainted by multiple interferences without resorting to prefiltering.

(2) Oscillatory behavior based signal decomposition (OBSD) for separating fault-induced transients from interferences and noise

It has been demonstrated in Chapter 3 that the STFD method can deal with signals for most real applications where the frequencies of interferences are much lower than the frequency of resonance. Even though the proposed KPSA can help extend the frequency range of the interferences to be handled, the effectiveness of the method inevitably deteriorates with the increased interference frequency complexity. Thus, the oscillatory behavior based signal decomposition method is proposed in Chapter 4.

The rationale of the method is that the fault-generated transients present different oscillatory behaviors from those of interferences and noise. Hence this method does not rely on frequency information and is effective in removing in-band interferences. As long as proper bases can be generated to represent signal components with different oscillatory behaviors, each signal component can be obtained using the MCA technique. The basis generation can be done via the TQWT which is parameterized by Q-factor. The method for specifying Q-factor for each signal component of interest is also studied in this thesis. To make the method capable of handling signals with a variety of interferences, the IOBSD is developed.

The proposed method has been evaluated by simulated and experimental data. The additional experiments where high frequency interferences are generated by gear meshing and its harmonics have also been carried out to test the proposed method’s performance of handling signals contaminated by strong noise and high frequency interferences. Detection results show that the proposed method is capable of isolating the fault-induced transients from gear-related signal components and noise. The technique is also favorably compared with one of the sate-of-the-art filtering methods, i.e., the SK technique.

7.1.2 For time-varying speed condition

In additional to noise and vibration interferences, frequency smearing on spectrum is another challenge in bearing fault diagnosis under variable speed. The traditional spectrum analysis techniques are no longer effective. Resampling the signal to convert it into a stationary signal in angular domain is one of the most frequently used methods to overcome the problem resulted by speed variations. However, the applications of resampling for bearings have been hindered by the two facts: 1) the shaft rotating IF is unavailable because it is coupled with IFCF by a FCC which cannot be determined without knowing the fault type, and 2) resampling operation can easily propagate errors and complicate the algorithm substantially. Therefore, a resampling-free method is highly desirable. Additionally, though the TFA is a powerful tool in theory for analyzing nonstationary signals, its actual applications to bearing fault diagnosis under time-varying speed is very limited because of the blurred and confusing representation. Motivated by the foregoing observations, a resampling-free technique for bearing condition monitoring under variable speed is proposed. In the meantime, to take the advantage of TFA method in processing nonstationary signal, a new TFA method has been developed and exploited to identify bearing faults. The two methods proposed for handling variable speed conditions are as follows.

(1) A resampling-free strategy based on GD for bearing fault diagnosis under variable speed

To start with, the proposed STFD is first employed to extract the envelope of the vibration signal. As revealed earlier, the STFD method is effective in fault feature extraction and interference and noise suppression. Hence, an envelope signal with high SNR and SIR can be obtained, from which the IFCF can be accurately estimated.

Targeting the issue caused by speed variations, the GD employed to map curved or oblique TF trajectory into a straight line parallel to time axis. By doing so, the energy of the signal component will concentrate around the frequency where the straight line is located. Performing frequency analysis to the demodulated signal, a peak will then stand out the spectrum. As the vibration signal from defective bearing is multicomponent in nature, an iterative process is used in this method. The resampling-free order spectrum can be obtained by searching amplitudes of the frequencies to which IF trajectories are respectively mapped. The detection decision can then be made according to the resampling-free order spectrum.

The simulation and experiments in Chapter 5 indicate that the proposed method can effectively detect the fault as well as reveal the fault locations without resampling.

(2) Dual demodulation transform and synchrosqueezing for TFA and its application to bearing fault identification

The TFR ambiguousness caused by weak energy concentration level is the major obstacle for its application. As such, a novel TFA technique termed DDT has been developed to improve the energy concentration level, thereby enhancing the readability of TFR. Since the vibration signal of a faulty bearing consists of multiple components, the IDDT is proposed to sharpen the TF trajectory of each component. A cross-term free and more energy-concentrated TFR is then generated.

The main task of TFA application to bearing fault detection is to reveal the shaft rotating related TF information on the TF plane. For this purpose, the resampling-free order spectrum is employed, from which the demodulators for shaft-related signal components can be determined. Thus, the shaft-related TF ridges can also be enhanced via the proposed IDDT.

Accurately predicting IFCF is pivotal to demodulation. Therefore, an IF refinement procedure is proposed to repeatedly improve the accuracy of the estimated IFCF in the Chapter 6.

The proposed method has been tested using simulated data with monocomponent and multicomponent signal contents. Experiments have also been done to validate its application to bearing fault identification, showing that the proposed method can effectively expose the fault-related as well as shaft-associated signal components of interest on the TF plane.

7.2 Limitations

This study has the following limitations:

(1) The capability of interference removal and denoising of the first two methods is not yet studied.

(2) The tuning of high Q-factor in Chapter 4 is not adaptive, which is computationally costly.

(3) The bearing fault diagnosis under both constant speed and variable speed cases have been studied. However, the effects of load variations are not taken into account in the experiments.

7.3 Future research

Partially motivated by limitations list above, further studies in the following directions are important:

(1) Obtain a comprehensive detectability map for STFD-KPSA and OBSD methods

In Chapter 3 and Chapter 4, the interference handling ability of both methods have been investigated under a certain SNR level (= - 7 dB). To better evaluate a new fault detection method, different noise levels should also be taken into consideration. Therefore, a three-dimensional (SIR, SNR and frequency) detectability map should be developed.

(2) Enhance the OBSD method by adaptively tuning Q-factor

The Q-factor selection has effects on the basis generation to represent signal components. Adaptively selecting Q-factor could further improve the effectiveness of the method but the challenge is how to guide the selection process and reduce the computation burden related to the adaptive selection process.

(3) Make the resampling-free diagnosis strategy independent of the pre-process

The resampling-free diagnosis strategy presented in Chapter 5 is based on a pre-process to obtain a clear envelope signal, from which the IFCF can be accurately approximated and a much clearer resampling-free order spectrum can be attained. The pre-process, however, inevitably complicates the method. Therefore, a technique that can extract the IFCF directly from the noisy vibration signal is preferred. In addition, improving the noise and interferences handling capability of the resampling-free strategy will also be an interesting research topic.

(4) Develop a mathematical model to quantitatively measure the energy concentration level improvement for the DDT technique

It has been demonstrated in Chapter 6 that the energy concentration level can be improved by the proposed DDT technique. However, a mathematical proof of the improvement scope has not been done and can be explored in future to better quantify the effectiveness of the DDT technique.

(5) Studying signals under varying load condition is also one of our recommendations for the future work.

In addition, only single fault case is considered in this thesis; there, however, might be multiple bearings with multiple faults in the industry. Hence, multiple fault diagnosis is also worth being explored in the near future.

References

Abrial, P., Moudden, Y., Starck, J.-L., Afeyan, B., Bobin, J., Fadili, J., and Nguyen, M. 2007. Morphological component analysis and inpainting on the sphere: Application in physics and astrophysics. Journal of Fourier Analysis and Applications **13**(6): 729-748.

Afonso, M.V., Bioucas-Dias, J.M., and Figueiredo, M.A. 2010. Fast image recovery using variable splitting and constrained optimization. IEEE Transactions on Image Processing **19**(9): 2345-2356.

Afshari, N., and Loparo, K.A. 1998. A model-based technique for the fault detection of rolling element bearings using detection filter design and sliding mode technique. Proceedings of the 37th Ieee Conference on Decision and Control, Vols 1-4: 2593-2598.

Altmann, J., and Mathew, J. 2001. Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis. Mechanical Systems and Signal Processing **15**(5): 963-977.

Amirat, Y., Choqueuse, V., Benbouzid, M., and Turri, S. 2011. Hilbert transform-based bearing failure detection in DFIG-based wind turbines. International Review of Electrical Engineering **6**(3): 1249-1256.

Antoni, J. 2006. The spectral kurtosis: a useful tool for characterising non-stationary signals. Mechanical Systems and Signal Processing **20**(2): 282-307.

Antoni, J. 2007. Fast computation of the kurtogram for the detection of transient faults. Mechanical Systems and Signal Processing **21**(1): 108-124.

Antoni, J., and Randall, R. 2002. Differential diagnosis of gear and bearing faults. Journal of Vibration and Acoustics **124**(2): 165-171.

Antoni, J., and Randall, R.B. 2004. Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms. Mechanical Systems and Signal Processing **18**(1): 89-101.

Antoni, J., and Randall, R.B. 2006. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mechanical Systems and Signal Processing **20**(2): 308-331.

Beck, A., and Teboulle, M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences **2**(1): 183-202.

Benbouzid, M.E.H., and Kliman, G.B. 2003. What stator current processing-based technique to use for induction motor rotor faults diagnosis? Ieee Transactions on Energy Conversion **18**(2): 238-244.

Bertrand, J., and Bertrand, P. 1992. A class of affine Wigner functions with extended covariance properties. Journal Of Mathematical Physics **33**(7): 2515-2527.

Bobin, J., Starck, J.L., Fadili, J.M., Moudden, Y., and Donoho, D.L. 2007. Morphological component analysis: An adaptive thresholding strategy. IEEE Transactions on Image Processing **16**(11): 2675-2681.

Borghesani, P., Ricci, R., Chatterton, S., and Pennacchi, P. 2013. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. Mechanical Systems and Signal Processing **38**(1): 23-35.

Borghesani, P., Ricci, R., Chatterton, S., and Pennacchi, P. (2014). Diagnostic of rolling element bearings with envelope analysis in non-stationary conditions. Advances in Condition Monitoring of Machinery in Non-Stationary Operations, Springer**:** 127-135.

Bozchalooi, I.S., and Liang, M. 2008. A joint resonance frequency estimation and in-band noise reduction method for enhancing the detectability of bearing fault signals. Mechanical Systems and Signal Processing **22**(4): 915-933.

Bozchalooi, I.S., and Liang, M. 2009. Parameter-free bearing fault detection based on maximum likelihood estimation and differentiation. Measurement Science and Technology **20**(6): 065102.

Cai, G., Chen, X., and He, Z. 2013. Sparsity-enabled signal decomposition using tunable Q-factor wavelet transform for fault feature extraction of gearbox. Mechanical Systems and Signal Processing **41**(1): 34-53.

Castiglioni, P. 2010. What is wrong in Katz's method? Comments on:"A note on fractal dimensions of biomedical waveforms". Computers in biology and medicine **40**(11): 950-952.

Chang, S.G., Yu, B., and Vetterli, M. 2000. Adaptive wavelet thresholding for image denoising and compression. IEEE Transactions on Image Processing **9**(9): 1532-1546.

Chang, S.K. (2003). Data structures and algorithms, World Scientific.

Chen, B., Wang, X., Yang, S., and McGreavy, C. 1999. Application of wavelets and neural networks to diagnostic system development, 1, feature extraction. Computers & Chemical Engineering **23**(7): 899-906.

Chen, Q., Chen, Z., Sun, W., Yang, G., Palazoglu, A., and Ren, Z. 2013. A new structuring element for multi-scale morphology analysis and its application in rolling element bearing fault diagnosis. Journal of Vibration and Control: 1077546313486163.

Cheng, J.S., Yang, Y., and Yu, D.J. 2009. Application of the improved generalized demodulation time–frequency analysis method to multi-component signal decomposition. Signal Processing **89**(6): 1205-1215.

Cheng, J.S., Yu, D.J., and Yang, Y. 2007. The application of energy operator demodulation approach based on EMD in machinery fault diagnosis. Mechanical Systems and Signal Processing **21**(2): 668-677.

Cocconcelli, M., Zimroz, R., Rubini, R., and Bartelmus, W. 2012. STFT Based Approach for Ball Bearing Fault Detection in a Varying Speed Motor. Condition Monitoring of Machinery in Non-Stationary Operations: 41-50.

Combet, F., and Gelman, L. 2009. Optimal filtering of gear signals for early damage detection based on the spectral kurtosis. Mechanical Systems and Signal Processing **23**(3): 652-668.

Combet, F., and Zimroz, R. 2009. A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform. Mechanical Systems and Signal Processing **23**(4): 1382-1397.

Daubechies, I., Defrise, M., and De Mol, C. 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on pure and applied mathematics **57**(11): 1413-1457.

Daubechies, I., Lu, J., and Wu, H.-T. 2011. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. Applied and Computational Harmonic Analysis **30**(2): 243-261.

Daubechies, I., and Maes, S. 1996. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. Wavelets in medicine and biology: 527-546.

Dempsey, P.J., Kreider, G., and Fichter, T. 2006. Investigation of tapered roller bearing damage detection using oil debris analysis. 2006 IEEE Aerospace Conference.

Donoho, D.L. 1995. De-noising by soft-thresholding. IEEE Transactions on Information Theory **41**(3): 613-627.

Donoho, D.L., and Johnstone, I.M. 1995. Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association **90**(432): 1200-1224.

Falconer, K. 2013. Fractal geometry: mathematical foundations and applications, John Wiley & Sons.

Feng, Z., Chu, F., and Zuo, M.J. 2011. Time–frequency analysis of time-varying modulated signals based on improved energy separation by iterative generalized demodulation. Journal Of Sound and Vibration **330**(6): 1225-1243.

Feng, Z.P., Chen, X.W., and Liang, M. 2015. Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. Mechanical Systems and Signal Processing **52**: 360-375.

Figueiredo, M.A., Bioucas-Dias, J.M., and Afonso, M.V. (2009). Fast frame-based image deconvolution using variable splitting and constrained optimization. IEEE/SP 15th Workshop on Statistical Signal Processing**:** 109-112.

Fyfe, K., and Munck, E. 1997. Analysis of computed order tracking. Mechanical Systems and Signal Processing **11**(2): 187-205.

Fyfe, K.R., and Munck, E.D.S. 1997. Analysis of computed order tracking. Mechanical Systems and Signal Processing **11**(2): 187-205.

Garner, D.R., and Leopard, A. (1985). Temperature measurements in fluid film bearings. Proceedings of the 13th Turbomachinery Symposium, College Station, Texas.

Georgoulas, G., Loutas, T., Stylios, C.D., and Kostopoulos, V. 2013. Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. Mechanical Systems and Signal Processing **41**(1–2): 510-525.

Gonçalves, P., and Baraniuk, R.G. 1998. Pseudo affine Wigner distributions: Definition and kernel formulation. IEEE Transactions on Signal Processing **46**(6): 1505-1516.

Gong, X., and Qiao, W. 2013. Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals. IEEE Transactions on Industrial Electronics **60**(8): 3419-3428.

Grossmann, A., and Morlet, J. 1984. Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM journal on mathematical analysis **15**(4): 723-736.

Hadjileontiadis, L.J. 2005. Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding-Part I: Methodology. IEEE Transactions on Biomedical Engineering, **52**(6): 1143-1148.

Hadjileontiadis, L.J., and Rekanos, I.T. 2003. Detection of explosive lung and bowel sounds by means of fractal dimension. IEEE Signal Processing Letters, **10**(10): 311-314.

Hao, R.J., Peng, Z.K., Feng, Z.P., and Chu, F.L. 2011. Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings. Measurement Science and Technology **22**(4): 045708.

He, Y.Y., Zhang, X.M., and Friswell, M.I. 2009. Defect Diagnosis for Rolling Element Bearings Using Acoustic Emission. Journal of Vibration and Acoustics-Transactions of the Asme **131**(6).

Higuchi, T. 1988. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena **31**(2): 277-283.

Hlawatsch, F., and Boudreaux-Bartels, G.F. 1992. Linear and quadratic time-frequency signal representations. IEEE Signal Processing Magazine, **9**(2): 21-67.

Ho, D., and Randall, R.B. 2000. Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signals. Mechanical Systems and Signal Processing **14**(5): 763-788.

Hong, H., and Liang, M. 2007. K-hybrid: a kurtosis-based hybrid thresholding method for mechanical signal denoising. Journal of vibration and acoustics **129**(4): 458-470.

Howard, I. 1994. A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis, DTIC Document.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., and Liu, H.H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**(1971): 903-995.

Katz, M.J. 1988. Fractals and the analysis of waveforms. Computers in biology and medicine **18**(3): 145-156.

Li, C., and Liang, M. 2012a. A generalized synchrosqueezing transform for enhancing signal time–frequency representation. Signal Processing **92**(9): 2264-2274.

Li, C., and Liang, M. 2012b. Time–frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform. Mechanical Systems and Signal Processing **26**: 205-217.

Li, H., Zhang, Y.P., and Zheng, H.Q. 2009. Gear fault detection and diagnosis under speed-up condition based on order cepstrum and radial basis function neural network. Journal of Mechanical Science and Technology **23**(10): 2780-2789.

Li, H., Zhang, Y.P., and Zheng, H.Q. 2010. Bearing fault detection and diagnosis based on order tracking and Teager-Huang transform. Journal of Mechanical Science and Technology **24**(3): 811-822.

Li, H., Zheng, H., and Tang, L. 2006. Wigner-Ville distribution based on EMD for faults diagnosis of bearing. Fuzzy Systems and Knowledge Discovery, Springer**:** 803-812.

Li, H., Zheng, H., and Tang, L. 2009. Bearing fault detection and diagnosis based on Teager-Huang Trasform. International Journal of Wavelets, Multiresolution and Information Processing **7**(05): 643-663.

Li, H., and Zheng, H.Q. 2008. Bearing fault detection using envelope spectrum based on EMD and TKEO. FSKD'08. IEEE Fifth International Conference on Fuzzy Systems and Knowledge Discovery,.

Li, Z., He, Z., Zi, Y., and Wang, Y. 2008. Customized wavelet denoising using intra- and inter-scale dependency for bearing fault detection. Journal of sound and vibration **313**(1–2): 342-359.

Liang, M., and Bozchalooi, I.S. 2010. An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection. Mechanical Systems and Signal Processing **24**(5): 1473-1494.

Lin, J., and Qu, L. 2000. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. Journal of sound and vibration **234**(1): 135-148.

Lin, J., and Zuo, M.J. 2003. Gearbox fault diagnosis using adaptive wavelet filter. Mechanical Systems and Signal Processing **17**(6): 1259-1269.

Lin, J., Zuo, M.J., and Fyfe, K.R. 2004. Mechanical fault detection based on the wavelet de-noising technique. Journal of vibration and acoustics **126**(1): 9-16.

Liu, H., Wang, J., and Lu, C. 2013. Rolling bearing fault detection based on the teager energy operator and elman neural network. Mathematical Problems in Engineering.

Luo, J.S., Yu, D.J., and Liang, M. 2012. Application of multi-scale chirplet path pursuit and fractional Fourier transform for gear fault detection in speed up and speed-down processes. Journal of sound and vibration **331**(22): 4971-4986.

Luo, J.S., Yu, D.J., and Liang, M. 2013. A kurtosis-guided adaptive demodulation technique for bearing fault detection based on tunable-Q wavelet transform. Measurement Science & Technology **24**(5).

Mallat, S.G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(7): 674-693.

Mba, D. 2003. Acoustic emissions and monitoring bearing health. Tribology Transactions **46**(3): 447-451.

McBain, J., and Timusk, M. 2009. Fault detection in variable speed machinery: Statistical parameterization. Journal of Sound and Vibration **327**(3–5): 623-646.

McFadden, P.D., and Smith, J.D. 1984. Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review. Tribology International **17**(1): 3-10.

Meireles, M.R.G., Almeida, P.E.M., and Simoes, M.G. 2003. A comprehensive review for industrial applicability of artificial neural networks. IEEE Transactions on Industrial Electronics **50**(3): 585-601.

Moon, F.C. 1992. Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers. New York, John Wiley.

Mortazavizadeh, S.A., and Mousavi, S.M.G. 2014. A Review on Condition Monitoring and Diagnostic Techniques of Rotating Electrical Machines. Physical Science International Journal **4**(3): 310-338.

Nikolaou, N.G., and Antoniadis, I.A. 2002a. Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted Morlet wavelets. Mechanical Systems and Signal Processing **16**(4): 677-694.

Nikolaou, N.G., and Antoniadis, I.A. 2002b. Rolling element bearing fault diagnosis using wavelet packets. NDT & E International **35**(3): 197-205.

Nikolaou, N.G., and Antoniadis, I.A. 2003. Application of morphological operators as envelope extractors for impulsive-type periodic signals. Mechanical Systems and Signal Processing **17**(6): 1147-1162.

Olhede, S., and Walden, A. 2005. A generalized demodulation approach to time-frequency projections for multicomponent signals. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **461**(2059): 2159-2179.

Patil, M.S., Mathew, J., and RajendraKumar, P.K. 2008. Bearing signature analysis as a medium for fault detection: A review. Journal of Tribology-Transactions of the ASME **130**(1).

Peng, Z.K., and Chu, F.L. 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. Mechanical Systems and Signal Processing **18**(2): 199-221.

Peng, Z.K., Tse, P.W., and Chu, F.L. 2005. A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. Mechanical Systems and Signal Processing **19**(5): 974-988.

Peter, W.T., Peng, Y., and Yam, R. 2001. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities. Journal of vibration and acoustics **123**(3): 303-310.

Qian, S., and Chen, D. 1999. Joint time-frequency analysis. Signal Processing Magazine, IEEE **16**(2): 52-67.

Qiu, H., Lee, J., Lin, J., and Yu, G. 2003. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. Advanced Engineering Informatics **17**(3–4): 127-140.

Qiu, H., Lee, J., Lin, J., and Yu, G. 2006. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal Of Sound and Vibration **289**(4): 1066-1090.

Raghavendra, B., and Dutt, N.D. 2010. Computing fractal dimension of signals using multiresolution box-counting method. International Journal of Information and Mathematical Sciences **6**(1): 50-65.

Rai, V.K., and Mohanty, A.R. 2007. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. Mechanical Systems and Signal Processing **21**(6): 2607-2615.

Randall, R.B. 2004. Detection and diagnosis of incipient bearing failure in helicopter gearboxes. Engineering Failure Analysis **11**(2): 177-190.

Renaudin, L., Bonnardot, F., Musy, O., Doray, J.B., and Remond, D. 2010. Natural roller bearing fault detection by angular measurement of true instantaneous angular speed. Mechanical Systems and Signal Processing **24**(7): 1998-2011.

Russell, P., Cosgrave, J., Tomtsis, D., Vourdas, A., Stergioulas, L., and Jones, G. 1998. Extraction of information from acoustic vibration signals using Gabor transform type devices. Measurement Science and Technology **9**(8): 1282.

Satish, L. 1998. Short-time Fourier and wavelet transforms for fault detection in power transformers during impulse tests. Science, Measurement and Technology, IEE Proceedings-, IET.

Sawalhi, N., Randall, R.B., and Endo, H. 2007. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mechanical Systems and Signal Processing **21**(6): 2616-2633.

Sejdić, E., Djurović, I., and Jiang, J. 2009. Time--frequency feature representation using energy concentration: An overview of recent advances. Digital Signal Processing **19**(1): 153-183.

Selesnick, I.W. 2011a. Resonance-based signal decomposition: A new sparsity-enabled signal analysis method. Signal Processing **91**(12): 2793-2809.

Selesnick, I.W. 2011b. Wavelet transform with tunable Q-factor. IEEE Transactions on Signal Processing **59**(8): 3560-3575.

Sevcik, C. 2010. A procedure to estimate the fractal dimension of waveforms. arXiv preprint arXiv:1003.5266.

Starck, J.L., Moudden, Y., Bobin, J., Elad, M., and Donoho, D. (2005). Morphological component analysis. Optics & Photonics International Society for Optics and Photonics**:** 59140Q-59140Q-59115.

Tandon, N., and Choudhury, A. 1999. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology International **32**(8): 469-480.

Tang, C., Miao, Q., and Pecht, M. 2011. Rolling element bearing fault detection: Combining energy operator demodulation and wavelet packet transform. IEEE Prognostics and System Health Management Conference (PHM-Shenzhen).

Teotrakool, K., Devaney, M.J., and Eren, L. 2009. Adjustable-Speed Drive Bearing-Fault Detection Via Wavelet Packet Decomposition. Instrumentation and Measurement, IEEE Transactions on **58**(8): 2747-2754.

Thakur, G., and Wu, H.-T. 2011. Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. SIAM Journal on Mathematical Analysis **43**(5): 2078-2095.

Tricot, C. (1995). Curves and fractal dimension, Springer.

Urbanek, J., Barszcz, T., and Antoni, J. 2013. A two-step procedure for estimation of instantaneous rotational speed with large fluctuations. Mechanical Systems and Signal Processing **38**(1): 96-102.

Urbanek, J., Barszcz, T., Sawalhi, N., and Randall, R. 2011. Comparison of amplitude-based and phase-based methods for speed tracking in application to wind turbines. Metrology and Measurement Systems **18**(2): 295-304.

Wald, R., Khoshgoftaar, T.M., Beaujean, P.P., and Sloan, J.C. 2010. Combining Wavelet and Fourier Transforms in Reliability Analysis of Ocean Systems. 16th ISSAT International Conference on Reliability and Quality in Design**:** 303-307.

Wang, D., Tse, P.W., and Tsui, K.L. 2013. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. Mechanical Systems and Signal Processing **35**(1): 176-199.

Wang, D., W Tse, P., and L Tse, Y. 2012. A morphogram with the optimal selection of parameters used in morphological analysis for enhancing the ability in bearing fault diagnosis. Measurement Science and Technology **23**(6): 065001.

Wang, J., Xu, G.H., Zhang, Q., and Liang, L. 2009. Application of improved morphological filter to the extraction of impulsive attenuation signals. Mechanical Systems and Signal Processing **23**(1): 236-245.

Wang, S.B., Chen, X.F., Cai, G.G., Chen, B.Q., Li, X., and He, Z.J. 2014. Matching demodulation transform and synchrosqueezing in time-frequency analysis. IEEE Transactions on Signal Processing **62**(1): 69-84.

Wang, T.Y., Liang, M., Li, J.Y., and Cheng, W.D. 2014. Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. Mechanical Systems and Signal Processing **45**(1): 139-153.

Wang, X., Chen, B., Yang, S., and McGreavy, C. 1999. Application of wavelets and neural networks to diagnostic system development, 2, an integrated framework and its application. Computers & chemical engineering **23**(7): 945-954.

Wang, Y.X., and Liang, M. 2011. An adaptive SK technique and its application for fault detection of rolling element bearings. Mechanical Systems and Signal Processing **25**(5): 1750-1764.

Yang, H.T., and Liao, C.C. 2001. A de-noising scheme for enhancing wavelet-based power quality monitoring system. IEEE Transactions on Power Delivery **16**(3): 353-360.

Yang, Y., Wang, H.h., Cheng, J.s., and Zhang, K. 2013. A fault diagnosis approach for roller bearing based on VPMCD under variable speed condition. Measurement **46**(8): 2306-2312.

Yang, Y., Yu, D., and Cheng, J. 2007. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement **40**(9): 943-950.

Yen, G.G., and Lin, K.C. 1999. Conditional health monitoring using vibration signatures. Proceedings of the 38th IEEE Conference on Decision and Control. **5:** 4493-4498.

Yu, D.J., Cheng, J.S., and Yang, Y. 2005. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. Mechanical Systems and Signal Processing **19**(2): 259-270.

Zhang, L.J., Xu, J.W., Yang, J.H., Yang, D.B., and Wang, D.D. 2008. Multiscale morphology analysis and its application to fault diagnosis. Mechanical Systems and Signal Processing **22**(3): 597-610.

Zhang, P.J., Du, Y., Habetler, T.G., and Lu, B. 2011. A Survey of Condition Monitoring and Protection Methods for Medium-Voltage Induction Motors. Ieee Transactions on Industry Applications **47**(1): 34-46.

Zhang, Y., and Ai, S. 2008. EMD based envelope analysis for bearing faults detection. Intelligent Control and Automation, 2008 IEEE 7th World Congress on WCICA.

Zhang, Y., Liang, M., Li, C., and Hou, S.M. 2013. Joint Kurtosis-Based Adaptive Bandstop Filtering and Iterative Autocorrelation Approach to Bearing Fault Detection. Journal of Vibration and Acoustics-Transactions of the Asme **135**(5).

Zhang, Y., and Randall, R.B. 2009. Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram. Mechanical Systems and Signal Processing **23**(5): 1509-1517.

Zhao, M., Lin, J., Xu, X., and Lei, Y. 2013. Tacholess Envelope Order Analysis and Its Application to Fault Detection of Rolling Element Bearings with Varying Speeds. Sensors **13**(8): 10856-10875.

Zimroz, R., Urbanek, J., Barszcz, T., Bartelmus, W., Millioz, F., and Martin, N. 2011. Measurement of instantaneous shaft speed by advanced vibration signal processing-application to wind turbine gearbox. Metrology and Measurement Systems **18**(4): 701-712.