Fault detection for planetary gearbox based on an enhanced average filter and modulation signal bispectrum analysis

Junchao Guo[1], Dong Zhen[1,*], Haiyang Li[2], Zhanqun Shi[1], Fengshou Gu[2], Andrew. D. Ball[2]

[1]School of Mechanical Engineering, Hebei University of Technology, Tianjin, 300401, China.

[2]Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK.

*Corresponding author: Dr. Dong Zhen. E-mail address: d.zhen@hebut.edu.cn

Abstract

Transient impulses are important information for machinery fault diagnosis. However, the transient features contained in the vibration signals generated by planetary gearboxes are usually immersed by a large amount of background noise and harmonic components. Even mathematical morphology (MM) is an excellent anti-noise signal processing method that can directly extract the geometry of impulse features in the time domain, but the four basic operators of MM can only extract one-way impulses while cannot extract the bidirectional impulses effectively at the same time. To accurately extract the impulse feature information, a novel method for fault detection of planetary gearbox based on an enhanced average (EAVG) filter and modulated signal bispectrum (MSB) is proposed. Firstly, the properties of the extracted impulses based on the four basic operators of MM will be divided into two categories of enhanced average operators. The four EAVG filters consist of the average weighted combination of enhanced average operators, and then the best EAVG filter is selected based on correlation coefficient to implement on the original vibration signal. It allows EAVG filter to extract positive and negative impulses of vibration signal, thereby improving the accuracy of planetary gearbox fault detection. Subsequently, the performance of the EAVG filter is influenced by the length of its structural element (SE), which is adaptively determined using an indicator based kurtosis. Then, the EAVG filter selects the optimal SE length to eliminate the interference of background noise and harmonic components to enhance the impulse components of the vibration signal. However, the nonlinear modulation components that are related to the fault types and severities are not extracted exactly and still remained in the filtered signal by EAVG. Finally, the MSB is utilized to the EAVG filtered signal to decompose the modulated components and extract the fault features. The advantages of EAVG over average (AVG) filter are clarified in the simulation study. In addition, the EAVG-MSB is validated by analyzing the vibration signals of planetary gearboxes with sun gear chipped tooth, sun gear misalignment and bearing inner race fault. The results indicate that the EAVG-MSB is effective and accurate in feature extraction compared with the combination morphological filter-hat transform (CMFH) and average combination difference morphological filter (ACDIF), and the feasibility of the EAVG-MSB are proved for planetary gearbox condition monitoring and fault diagnosis.

Keywords: enhanced average filter; modulation signal bispectrum; planetary gearbox; fault detection.

Nomenclature

SR	stochastic resonance
ELMD	ensemble local mean decomposition
FK	fast kurtogram
EWT	empirical wavelet transform
MCKD	maximum correlated kurtosis deconvolution
MM	mathematical morphology
MF	morphological filter
SE	structural element
MG	morphology gradient
DIF	difference filter
CMF	combination morphological filter
CMFH	combination morphological filter-hat transform
MTH	morphology filter-hat transformation
AVG	average of the closing and opening
CO	closing–opening
OC	opening–closing
ACDIF	average combination difference morphological filter
EAVG	enhanced average
DFT	discrete Fourier transform
CB	conventional bispectrum
MSB	modulated signal bispectrum
MSBc	modulation signal bicoherence
TEK	Teager energy kurtosis
FEF	feature energy factor
PSO	particle swarm optimization
QGA	quantum genetic algorithm

1. Introduction

Planetary gearboxes are the most widely used components in a variety of rotating machinery. Since planetary gearboxes often work under high-speed operation and long-term variable loads, it is easy to cause local damage and evolve into late failures, which will affect the operating efficiency and service life of the whole machine [1-3]. Therefore, fault detection of planetary gearboxes is critical to ensure mechanical reliability and avoid catastrophic accidents. In recent years, vibration signal analysis has become one of the most popular and common fault detection techniques due to the vibration signal with a large amount of information indicating the health of mechanical systems [4-6]. In addition, the vibration signal is mainly composed of impulse components, harmonic components and background noise, and its impulse components contain rich information of the state of the planetary gearbox. However, impulse components are often masked by a large amount of background noise and harmonic components from the mechanical parts and operating environments. Thus, the key issue in planetary gearbox fault detection is to highlight the impulse components through reducing background noise and harmonic components interference.

Nowadays, various typical signal processing methods have been used to extract impulse features from transient impulse components under background noise to detect planetary gearbox failures [7-13]. For instance, Lei et al. developed an adaptive stochastic resonance (SR) method to reduce the background noise of the planetary gearbox to highlight weak impulse features and fault detection accuracy [7]. Wang et al. applied the ensemble local mean decomposition (ELMD) to eliminate in-band noise in the fast kurtogram (FK) and improve the accuracy of fault diagnosis [8]. Miao et al. put forward an adaptive variational mode modal decomposition (VMD) and its performance was verified by rotating machinery compound signals [9]. Zhang et al. proposed the empirical wavelet transform (EWT) can be applied as a pre-filter to reduce background noise to enhance the impulse components of planetary gearbox [10]. Sun et al. explored a weighted sparse representation to extract the transient signal of the planetary gearbox under strong background noise [11]. Subsequently, Wang et al. proposed improved K-singular value decomposition (K-SVD) method to enhance the sparse representation of extracting transient impulse performance [12]. Li et al. studied an improved tunable Q-factor wavelet transform (TQWT) based on the characteristic frequency ratio to detect bearing transient signals [13]. Lyu et al. developed an improved maximum correlated kurtosis deconvolution (MCKD) to extract the impulse features of the planetary gearbox vibration signal and successfully verified its suitability for fault detection with heavy background noise [14]. Although the above approaches have been extensively studied and exploited for planetary gearbox fault detection, they focus on the denoising effects while ignoring the geometric characteristics of the signal [15]. Therefore, these methods will inevitably weaken useful information when eliminating background noise.

Mathematical morphology (MM) is an excellent anti-noise signal processing method, which is intended to amend the geometry characteristics of signals through structure element (SE) [16]. It was an image processing method initially introduced by Matheron [17]. Subsequently, Maragos [18] expanded MM to the field of signal processing. Currently, MM filters (commonly known as morphological filters (MFs)) have attracted widespread attention in the diagnosis of rotating machinery failures. In recent years, many scholars have done many research works on the fault diagnosis of mechanical systems or key components using morphological filter analysis method, and achieved some research achievements [19-21]. For instance, Li et al. put forward a novel differential filter ($G_{co\&oc}$) that consists the difference between closing–opening (CO) and opening–closing (OC) based on diagonal slice spectrum to extract

the fault features of rotating machinery [19]. Hu et al. proposed morphology gradient (MG) to extract impulse components from background noise and interference components [20]. Li et al. exploited the principle of max-relevance and min-redundancy to choose representatively fault feature scheme based on difference filter (DIF) for train bearing fault diagnosis [21]. Although MG, $G_{co\&oc}$, and DIF filters are widely used in the field of fault diagnosis, these filters will change the impulse component of the vibration signal due to the negative impulse being changed to the positive impulse [22]. To extract bidirectional impulse components, Dong et al. proposed the average (AVG) filter to obtain both positive and negative impulses in the signal, and successfully applied to fault diagnosis [23]. Subsequently, Yan et al. proposed the combination morphological filter-hat transform (CMFH) consisting of the combination morphological filter (CMF) and morphology filter-hat transformation (MTH), which can extract the impulse components from complicated vibration signal well [24]. In addition, it proves that it is superior to AVG and DIF. However, the amplitude of impulse components of AVG and CMFH is weakened [25]. In addition, there is a possibility that some of the effective impulse features cannot be obtained, resulting in a reduction in the accuracy of the fault detection. Thus, it is important to develop a new operator to extract the complete impulse components.

Inspired by Ref [26], the cascade of two basic operators with the same impulse feature properties can further enhance the impulse effect in the same direction. Considering the advantages of the novel cascade operators, a novel enhanced average (EAVG) filter is proposed in this paper. It not only extracts complete impulse components, but also effectively reduces background noise and harmonic interference. However, there are still some interference frequencies and nonlinear modulation components after EAVG filter analysis. The modulation signal bispectrum (MSB) is a typical demodulation approach, which has the advantage of demodulating the modulated components and suppressing some interference frequencies of signals [27-30]. Therefore, a new approach for fault diagnosis of planetary gearbox based on EAVG and MSB is proposed. The EAVG-MSB is used to detect planetary gearboxes with sun gear chipped tooth, sun gear misalignment and bearing inner race fault. A comprehensive analysis of the EAVG-MSB is performed and compared with the combination morphological filter-hat transform (CMFH) and average combination difference morphological filter (ACDIF) on the same experimental dataset. The results validate the effectiveness and superiority of the proposed EAVG-MSB.

The remainder of the paper is organized as follows. The EAVG is briefly introduced, and merits of EAVG over AVG are clarified based on the simulation study in Section 2. In Section 3, the comparison between MSB and conventional bispectrum (CB) is detailed and the theoretical basis of the MSB detector is summarized. In Section 4, the implementation process of the proposed EAVG-MSB is detailed. Vibration signals collected from planetary gearbox with sun gear misalignment, sun gear chipped tooth and bearing inner race fault are further applied to verify the EAVG-MSB in Section 5. Finally, the concluding remarks are provided in Section 6.

2. Impulsive components enhancement

2.1 EAVG

The mathematical morphological (MM) was first introduced in image processing methodology by Matheron and Serra [31]. Recently, MM has been widely used for rotating machine fault diagnosis based on impulsive components enhancement [32-34]. The basic morphological operations consist of dilation, erosion, closing and opening. Suppose that the definition domain of the input signal f (n) and selected SE g (m) are the discrete function over a domain $F = (0,1,...,N-1)$ and $G = (0,1,...,M-1)$ $(M \leq N)$, respectively. The above four basic morphological operators can be expressed as:

Dilation:

$$(f \oplus g)(n) = max\{f(n-m) + g(m)\}\{1 \leq n \leq N; 1 \leq m \leq M\} \tag{1}$$

Erosion:

$$(f \Theta g)(n) = min\{f(n+m) - g(m)\}\{1 \leq n \leq N; 1 \leq m \leq M\} \tag{2}$$

Opening:

$$(f \circ g)(n) = (f \Theta g \oplus g)(n) \tag{3}$$

Closing:

$$(f \bullet g)(n) = (f \oplus g \Theta g)(n) \tag{4}$$

where $\oplus$, Θ, o and $\bullet$ indicate the dilation, erosion, opening and closing operations, respectively. The performances of the four basic operators to extract impulse features are illustrated in Table 1.

Table 1. Performance of the four basic operators to extract impulse features.

Morphological operator	Negative impulse	Positive impulse
Erosion	Smoothing	Reducing
Dilation	Reducing	Smoothing
Opening	Preserving	Reducing
Closing	Reducing	Preserving

From Table 1, it can be found that the erosion and opening operators are utilized to extract negative impulse. In contrast, the dilation and closing operators are capable of extracting positive impulse. Inspired by Ref [26], the properties of the extracted impulse features based on the four basic operators are divided into two categories of cascade operators. The cascade operator that extracts positive impulse consists of dilation and closing while the other cascades of erosion and opening has an opposite effect. Among them, when the dilation and closing operators are composed, there are two sequential composition modes and a positive impulse cascade operator can be obtained respectively. The two positive impulse cascade operators can be expressed as:

$$F_{DC}(f(n)) = (f \oplus g \bullet g)(n) \tag{5}$$
$$F_{CD}(f(n)) = (f \bullet g \oplus g)(n) \tag{6}$$

Similarly, when the erosion and opening operators are composed, there are also two sequential composition modes and a negative impulse cascade operator can be obtained respectively. The two negative impulse cascade operators can be expressed as:

$$F_{EO}(f(n)) = (f \Theta g \circ g)(n) \tag{7}$$
$$F_{OE}(f(n)) = (f \circ g \Theta g)(n) \tag{8}$$

To extract bidirectional impulse components, the average weighted combination of $F_{DC}(F_{CD})$ and $F_{EO}(F_{OE})$ is calculated as the enhanced average (EAVG) filter. Therefore, the four $EAVG$ filters are defined as follows:

$$EAVG_{DC-EO}(n) = \frac{F_{DC}(f(n)) + F_{EO}(f(n))}{2} \tag{9}$$

$$EAVG_{DC-OE}(n) = \frac{F_{DC}(f(n)) + F_{OE}(f(n))}{2} \tag{10}$$

$$EAVG_{CD-EO}(n) = \frac{F_{CD}(f(n)) + F_{EO}(f(n))}{2} \tag{11}$$

$$EAVG_{CD-OE}(n) = \frac{F_{CD}(f(n)) + F_{OE}(f(n))}{2} \tag{12}$$

To evaluate the impulses extraction performance of the four EAVG filters, a simulated signal $x_o = \cos(60\pi t) + 1.5\cos(100\pi t)$ is analyzed. The sampling frequency is 1024 Hz and signal length is 1024. The waveform of the simulated signal is depicted in Fig. 1. Then four basic morphological filter operators with flat SE length L= 7 are applied to process the simulated signal in Fig. 1, and the filtered signals using four basic morphological operators are shown in Fig. 2. It demonstrates by Fig. 2 that dilation (Fig. 2 (a)) reduces negative impulse and smooth positive impulse while erosion (Fig. 2 (b)) operator has an opposite result. The closing (Fig. 2 (c)) operator is able to reserve positive impulses and remove negative ones. On the contrary, the opening (Fig. 2 (d)) operator can reduce positive impulses and preserve negative ones.
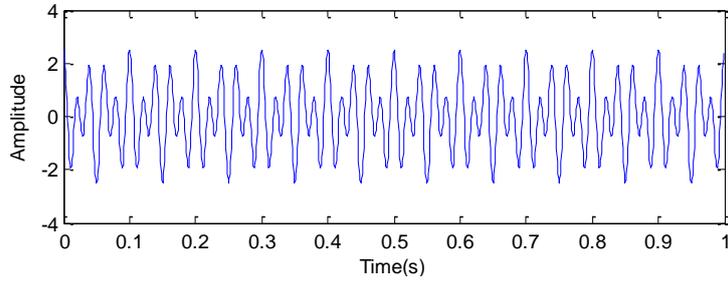
Fig. 1. Waveform of the simulated signal.
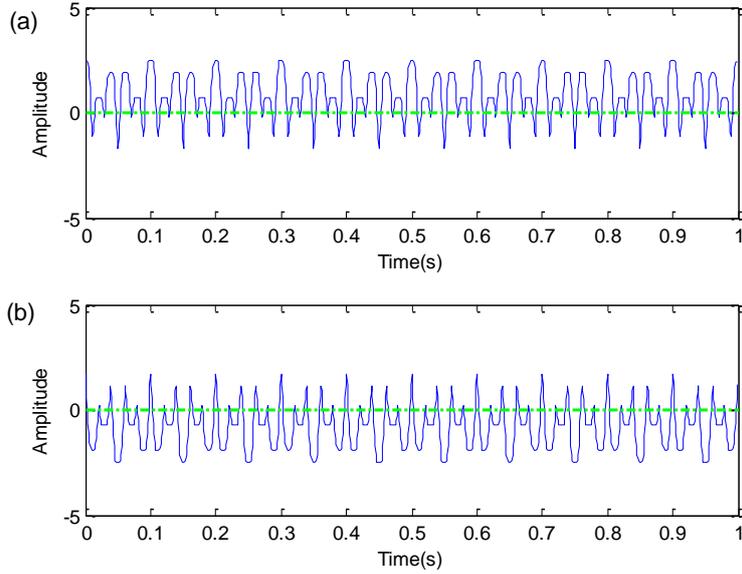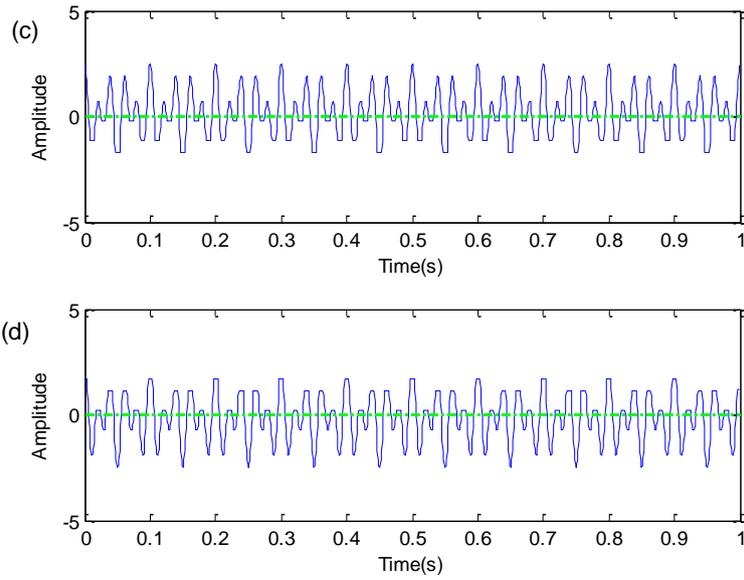
Fig. 2. Results of the four basic morphological operators: (a) dilation (b) erosion (c) closing (d) opening.

Considering the cascade of two basic operators with the same impulse feature properties, the impulse effect can be further enhanced in the same direction and verified as shown in Fig. 3. It can be seen that F_{DC} and F_{CD} enhance the effect of extracting positive impulse while F_{EO} and F_{OE} enhance the effect of extracting negative impulse compared with the dilation, closing, erosion and opening, respectively.
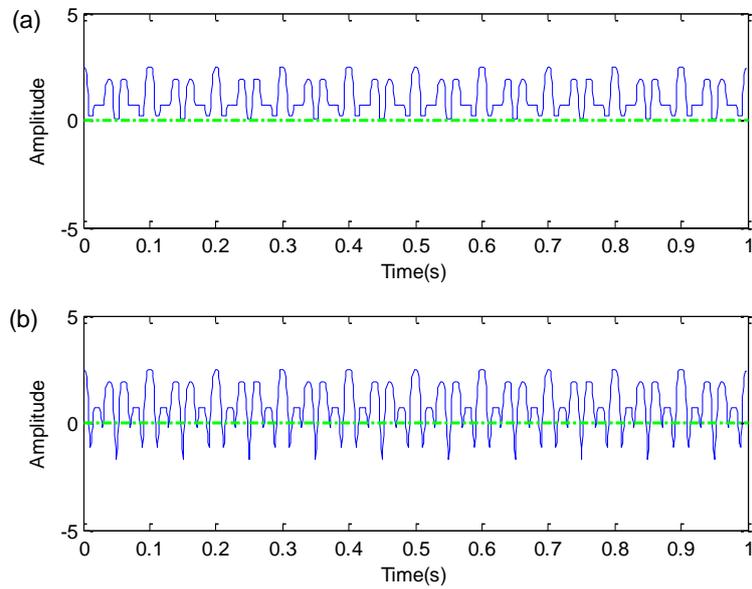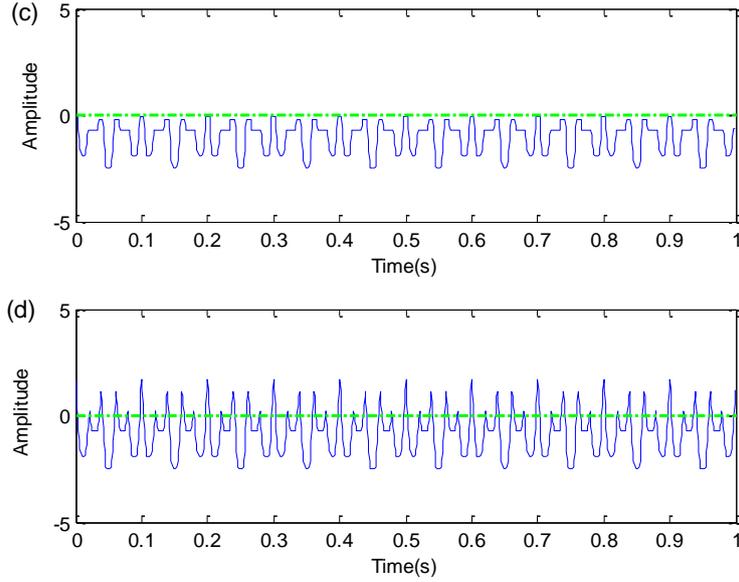
Fig. 3. Results of the four cascades operators (a) F_{DC} (b) F_{CD} (c) F_{EO} (d) F_{OE}.

To extract bidirectional impulse components, the average weighted combination of $F_{DC}(F_{CD})$ and $F_{EO}(F_{OE})$ is calculated as the EAVG filters. The results of the four EAVG filters are shown in Fig. 4. It can be seen that EAVG_CD-OE extracts the positive and negative impulse of the signal better than other EAVG filters. In order to further compare the measured results of four EAVG filters, correlation coefficient is performed to measure the similarity between the original signal and $EAVG$ filtered signal to detect the impulse component. Given the original signal $x(t)$ and $EAVG$ filter signal $y(t)$, the correlation coefficient p_{xy} is expressed as follow [35]:

$$p_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} \tag{13}$$

$$C_{xy} = E\{[x(t) - u_x][y(t) - u_y]\} = \lim_{T \to \infty} \frac{1}{T} \int_0^T \{[x(t) - u_x][y(t) - u_y]\} dt \tag{14}$$

where C_{xy} present the covariance of $x(t)$ and $y(t)$. $u_x(u_y)$ and $\sigma_x(\sigma_y)$ indicate the mean values and standard variance of $x(t)$ and $y(t)$, respectively. The correlation coefficients of the four EAVG filters are depicted in Table 2. As can be observed from Table 2, the EAVG_CD-OE has maximum correlation coefficients. Therefore, EAVG_CD-OE is selected as the best EAVG filters.
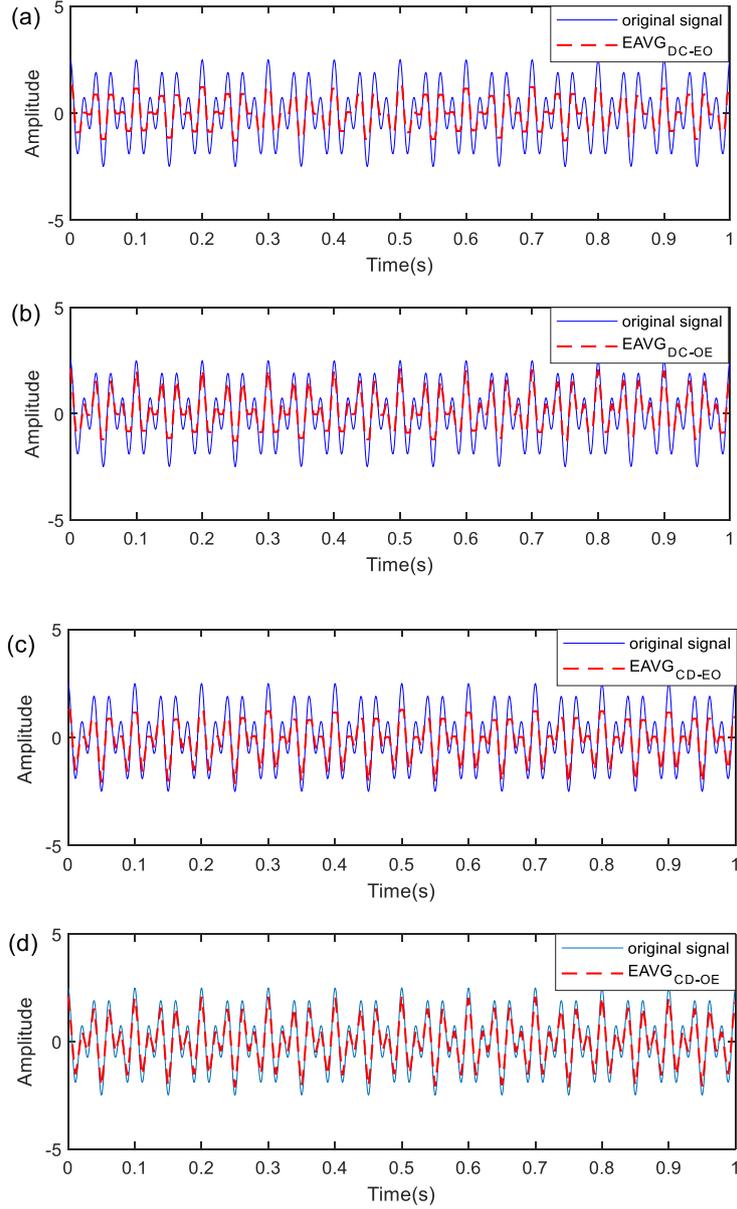
Fig. 4. Results of the four EAVG filters: (a) $EAVG_{DC-EO}$ (b) $EAVG_{DC-OE}$ (c) $EAVG_{CD-EO}$ (d) $EAVG_{CD-OE}$.

Table 2. The correlation coefficients of four EAVG filters.

Evaluation index	$EAVG_{DC-EO}$	$EAVG_{DC-OE}$	$EAVG_{CD-EO}$	$EAVG_{CD-OE}$
Correlation coefficients	0.9538	0.9707	0.9704	0.9883

To illustrate the advantage of EAVG, the representative morphological filter, i.e., AVG [23] is also taken into account for the simulated signal. The analysis results of AVG and $EAVG_{CD-OE}$ are displayed in Fig. 5. It can be observed that AVG extracts impulses inferior to the EAVG. The results show that EAVG filter is more suitable for extracting bidirectional impulse components.
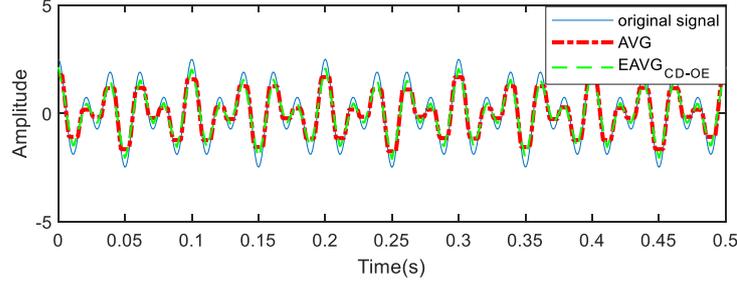
Fig. 5. Result of the comparison of AVG and EAVG filters.

2.2 Selection of structural element length

The structural element (SE) is very important for morphological filter (MF) to extract impulse components. SE is decided through three elements, i.e., shape, height and length. Studies have shown that the shape and height of SE have little effect on the results of morphological filters [36]. Therefore, to reduce the computational complexity of morphological operations, a flat SE with height $H = 0$ is selected in this paper. In addition, it is shown that the flat length L plays a decisive role in the MF analysis. Currently, the optimal length is usually chose by observation [37] or some measurement criteria (e.g., signal-to-noise ratio (SNR) [38], particle swarm optimization (PSO) [39], fault feature ratio (FFR) [27] and quantum genetic algorithm (QGA) [40] etc.). According to [41], the maximum kurtosis is proposed to optimize the SE length due to it measures the similarity between the stochastic process and Gaussian process to detect the presence of the impulse components, and has been proven to be an valid approach for selecting the optimal SE length [21, 42-43]. The definition of kurtosis is expressed as follow:

$$kurtosis = \frac{(1/N)\sum_{n=1}^{N}(y(n)-\bar{y})^4}{\left[(1/N)\sum_{n=1}^{N}(y(n)-\bar{y})^2\right]^2} \tag{15}$$

where $y(n)$ is a $EAVG$ filtered signal and $\bar{y}$ is the mean value of $y(n)$. Hence, in view of the merits of maximum kurtosis, an adaptive selection approach according to the kurtosis is utilized to obtain adaptively the optimal SE length of the enhanced average ($EAVG$) filter in this paper.

3. Feature extraction

3.1 Conventional bispectrum

Conventional bispectrum (CB) has nonlinear system detection, phase information reservation and Gaussian noise cancellation, and is usually applied to measure nonlinear quadratic coupling caused by two interacting waves [44-45]. For a discrete vibration signal $x(t)$, its discrete Fourier transform (DFT) is expressed as [46]:

$$X(f) = \sum_{t=-\infty}^{\infty} x(t)\,e^{-2j\pi t} \tag{16}$$

By DFT, the CB $B(f_c, f_x)$ is expressed as:

$$B(f_c, f_x) = E < X(f_c)X(f_x)X^*(f_c + f_x) > \tag{17}$$

where X^* indicates the complex conjugate of $X(f)$, $E <>$ stands the expectation operator, f_c, f_x and $f_c + f_x$ are three different frequencies. Unlike the power spectrum, the amplitude of the CB comprises of both the amplitude and phase of the original signal. If the three frequency parts f_c, f_x and $f_c + f_x$ are independent components, the phase of each frequency component conforms to the uniform random distribution law and is distributed on $(-\pi, \pi)$. Applying Eq. (17) to calculate the statistical expectation, the CB will be close to zero because of the random phase mixing effect, so applying this method can significantly suppress random noise. On the contrary, if the three frequency components: f_c,

f_x and $f_c + f_x$ are nonlinearly coupled with each other, their phases have a specific relationship. Even if the three frequencies are random, the phases are in accordance with the following equation:

$$\emptyset(f_c) + \emptyset(f_x) = \emptyset(f_c + f_x) \tag{18}$$

Therefore, the CB statistical mean will not tend to zero. This nonlinear coupling appears as a peak of the bifrequency $B(f_c, f_x)$ in the CB analysis.

3.2 Modulation signal bisepctrum

From spectrum features of points, Eq. (17) can only react to the nonlinear relationship between f_c, f_x and $f_c + f_x$ frequency components. It neglects the possibility that the existence of $f_c - f_x$ may also be because of the nonlinear interaction between f_c and f_x. Therefore, it is not sufficient to analyze the modulation components in the vibration signal. In order to improve the description of the characteristics of the vibration signal by the CB, the modulated signal bispectrum (MSB) can be expressed as [27]:

$$B_{MS}(f_c, f_x) = E < X(f_c + f_x)X(f_c - f_x)X^*(f_c)X^*(f_c) > \tag{19}$$

Unlike the CB in Eq. (17), this definition considers both $f_c + f_x$ and $f_c - f_x$ to characterize the nonlinear coupling in the modulated signal. It indicates that a bispectral peak is distinctly represented at bifrequency $B_{MS}(f_c, f_x)$.

The overall phase of the MSB is written as below:

$$\emptyset_{MS}(f_c, f_x) = \emptyset(f_c + f_x) + \emptyset(f_c - f_x) - \emptyset(f_c) - \emptyset(f_c) \tag{20}$$

when two components f_c and f_x are coupled, their phases are defined by the following relationships:

$$\begin{cases} \emptyset(f_c + f_x) = \emptyset(f_c) + \emptyset(f_x) \\ \emptyset(f_c - f_x) = \emptyset(f_c) - \emptyset(f_x) \end{cases} \tag{21}$$

By substituting Eq. (21) into Eq. (20) represents that the total phase of the MSB is zero and the amplitude of the MSB is computed by means of the product of the four magnitudes, indicating the maximum of the complex product. Consequently, a bispectral peak appears at bifrequency (f_c, f_x). The Eq. (20) consists of the $(f_c + f_x)$ and $(f_c - f_x)$ to detect the nonlinearity of the modulated components. If $(f_c + f_x)$ and $(f_c - f_x)$ are both because of nonlinear effect between f_c and f_x, a bispectral peak appears at bifrequency $B_{MS}(f_c, f_x)$. Thus, it is more efficient and distinct in meaning the modulation characteristics in a noise signal.

To quantify the sideband amplitude more exactly, MSB is modified by removing the high influence of f_c components via magnitude normalisation. The MSB sideband estimator (MSB-SE) is expressed as [27]:

$$B_{MS}^{SE}(f_c, f_x) = \frac{B_{MS}(f_c, f_x)}{\sqrt{|B_{MS}(f_c, 0)|}} \tag{22}$$

where $B_{MS}(f_c, 0)$ denotes the squared power spectrum estimation at $f_x = 0$.

3.3 Simulations and comparisons

To verify the identification performance of the MSB for the small modulation process, the MSB is compared with the CB by the numerical simulation signal in Eq. (23):

$$x(t) = cos(2\pi f_1 t) + 0.004cos[2\pi(f_1 - f_2)t] + 0.004\,cos[2\pi(f_1 + f_2)t] + 0.002cos[2\pi(f_1 - f_3)t] + 0.002cos[2\pi(f_1 + f_3)t] + 0.01cos(2\pi f_4 t) + 0.005N \tag{23}$$

where $f_1 = 50\ Hz$ is the carrier frequency, $f_2 = 24\ Hz$ is a large amplitude modulation component, $f_3=15$ Hz is a small amplitude modulation component and $f_4=10$ Hz is an interference frequency. The signal $x(t)$ is sampled with a sampling frequency of 512 Hz and 1024 samples. Fig. 6 shows the waveform and spectrum of the simulated signal, respectively.
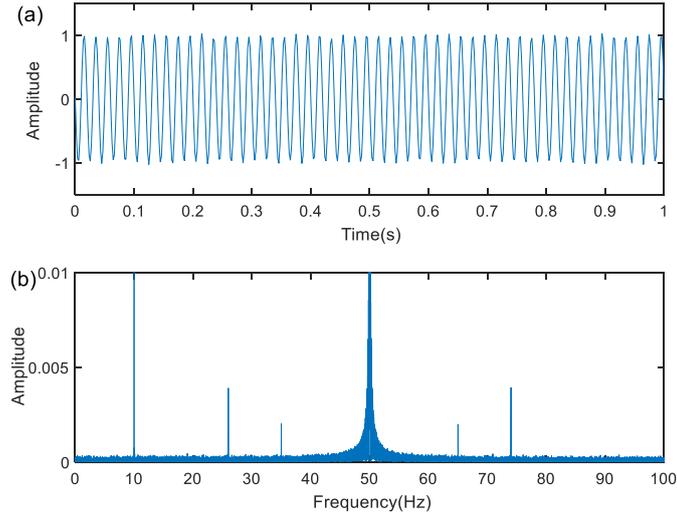
Fig. 6. Simulated signal: (a) waveform (b) spectrum.

Fig. 7 (a) and (b) illustrate the results of MSB and CB, respectively. The amplitudes of the MSB are clearly identified at the two peaks at (24, 50) Hz and (15, 50) Hz in Fig. 7 (a). However, in Fig. 7(b), CB shows a unique peak at (50, 50) Hz. In addition, there is no indication that the 24 Hz and 15 Hz modulation components in the simulation signal, indicating that the modulated signal cannot be detected. To further illustrate the effectiveness of MSB, the modulation signal bicoherence (MSBc) was calculated to detect the degree of coupling between three components, which is defined as follows:

$$b_{MS}^2(f_c, f_x) = \frac{|B_{MS}(f_c, f_x)|^2}{E < |X(f_c)X(f_c)X^*(f_c)X^*(f_c)|^2 > E < |X(f_c+f_x)X(f_c-f_x)|^2 >} \tag{24}$$

Fig. 8 (a) and (b) illustrate the results of MSBc and conventional bicoherence, respectively. The MSBc results illustrate unique peaks at (24, 50) Hz and (15, 50) Hz in Fig. 8(a). Clearly, MSBc can effectively indicate the nonlinear interaction between the carrier signal and modulated signal. However, the corresponding peaks cannot be observed in Fig. 8(b) to verify possible nonlinear interactions. Consequently, the simulation results demonstrate that the MSB detects modulation characteristics more accurately than CB. At the same time, it also proves that CB may generate misleading results and is not adequate for analyzing modulated signals.
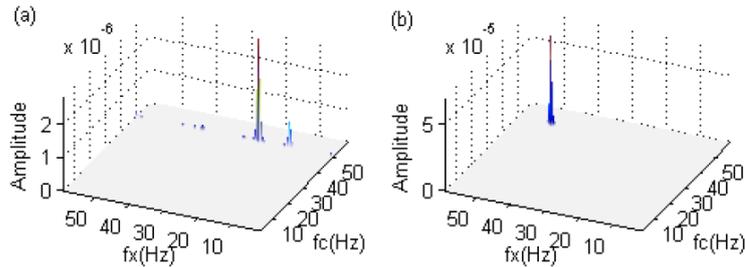

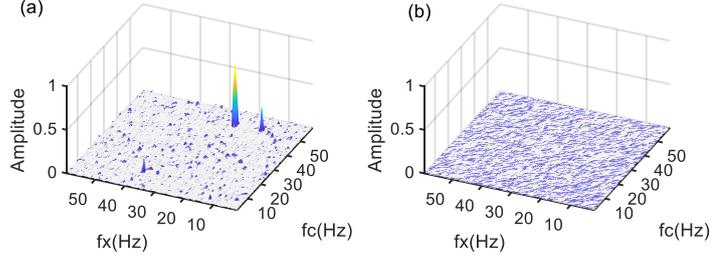**Fig. 7.** The result of the simulated signal: (a) MSB (b) CB.

Fig. 8. The result of the simulated signal: (a) MSBc (b) conventional bicoherence.

3.4 MSB detector

To obtain reliable suboptimal f_c slices, $B_{MS}^{SE}(f_c^n, f_x)$ is decided based on the compound MSB slice $B(f_c)$, which is computed by averaging the main MSB peaks in the incremental direction of the f_x [30]:

$$B(f_c) = \frac{1}{M-1}\sum_{m=2}^{M} B_{MS}^{SE}\left(f_c, m\Delta f\right) \tag{25}$$

where Δf presents the frequency resolution in the f_x orientation.

To get more robust results, the average of several suboptimal MSB slices is applied to improve the MSB detector:

$$B(f_x) = \frac{1}{N}\sum_{n=1}^{N} B_{MS}^{SE}(f_c^n, f_x) \quad , (f_x > 0) \tag{26}$$

where N denotes the total number of selected f_c suboptimal slices, depending on the main of the peaks themselves. In summary, the flowchart of the MSB detector is displayed in Fig. 9.
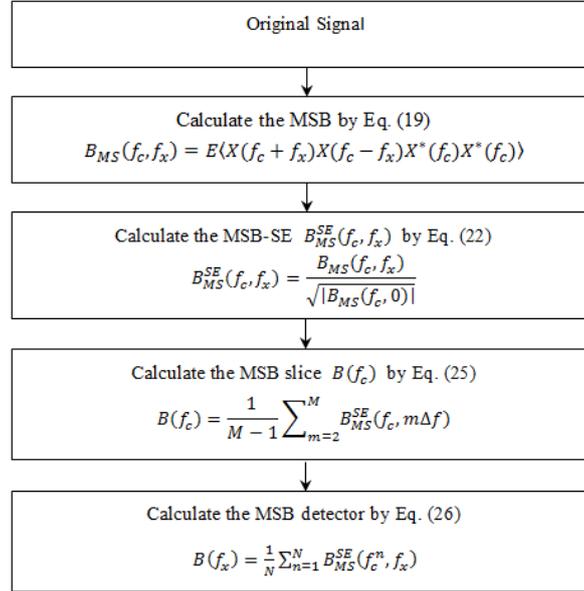
Fig. 9. The flowchart of the MSB detector.

4. The implementation process of the proposed method

In this subsection, a new analysis approach for fault detection of planetary gearbox based on EAVG and MSB is proposed. This method takes advantage of the transient signal extraction performance of EAVG and the modulation component demodulation and interference component suppression ability of the MSB, and achieves the purpose of fault feature enhancement and extraction. The specific implementation process in Fig. 10 and working procedures are summarized as follows:

Step 1: The properties of the extracted impulses based on the four basic operators will be divided into two categories of cascade operators.

Step 2: Calculate the EAVGs consist of the two categories of cascade operators.

Step 3: Select the optimal EAVG based on correlation coefficient to implement on the original vibration signal.

Step 4: Exploit the maximum kurtosis index to optimize the SE length of EAVG.

Step 5: Acquire filtered signal by EAVG using the optimal SE length.

Step 6: Calculate the MSB using Eq. (19) of the filtered signal.

Step 7: Calculate the MSB-sideband estimator $B_{MS}^{SE}(f_c, f_x)$ by Eq. (22).

Step 8: Calculate the compound MSB slice $B(f_c)$ by Eq. (25).

Step 9: Calculate the MSB detector $B(f_x)$ using Eq. (26) to extract the fault characteristic frequencies.
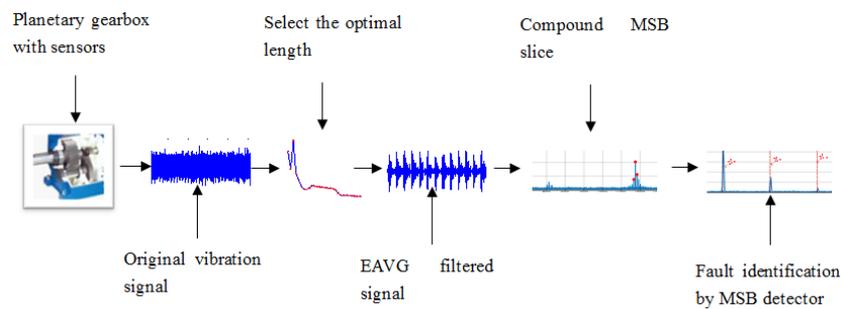
Fig. 10. Flowchart of the EAVG-MSB.

5. Experiments validation

5.1 Experimental system and data acquisition

The test platform of the planetary gearbox is presented in Fig. 11. The test platform is composed of an induction motor, two planetary gearboxes, a flexible type coupling, and a DC generator which acts as a mechanical load. The vibration signal of the planetary gearbox was collected by a vibration sensor mounted on top of the planetary gearbox with the sensitivity of $100mv/g$. The sampling frequency of the experiment data was 71428 Hz and signal length were 300000 data points. The fault modes consist of a sun gear chipped tooth and inner race fault, as presented in Figs. 12 (a) and (b), respectively. Main specifications and fault characteristic frequencies of the inner race and sun gear of the planetary gearbox are listed in Tables 3 and 4, respectively.
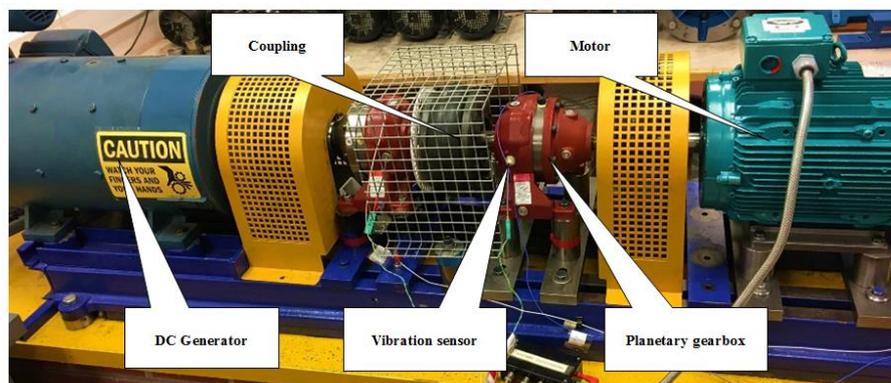

Fig. 11. The test platform of the planetary gearbox.

Fig. 12. Fault modes: (a) sun gear chipped tooth (b) inner race fault.

Table 3. Main specification and fault characteristic frequencies of the planetary gearbox bearing.

Bearing designation	Ball numbers d (mm)	Pitch Diameter D_m (mm)	Ball Number z	Contact Angle β
	12	54	7.9	$0°$
6008	f_o	f_i	f_b	f_c
	49.25	65.17	33.60	4.10

Table 4. Main specification and fault characteristic frequencies of the planetary gearbox.

Gear	Number of teeth	Characteristic frequency (Hz)
Sun gear f_{sf}	10	24.10
Planet gears(number) f_{pf}	26(3)	9.77
Ring gear f_{rf}	62	3.89

5. 2 Planetary gearbox with sun gear chipped tooth fault diagnosis

The waveform and spectrum of the vibration signal of the test planetary gearbox with sun gear chipped tooth fault is illustrated in Fig. 13. It can be seen that the waveform of the measured vibration signals contains various impulsive components which are related to the sun gear faults as shown in Fig. 13 (a). However, it is difficult to identify the fault characteristic frequencies of the sun gear chipped tooth based on the spectrum as shown in Fig. 13 (b).
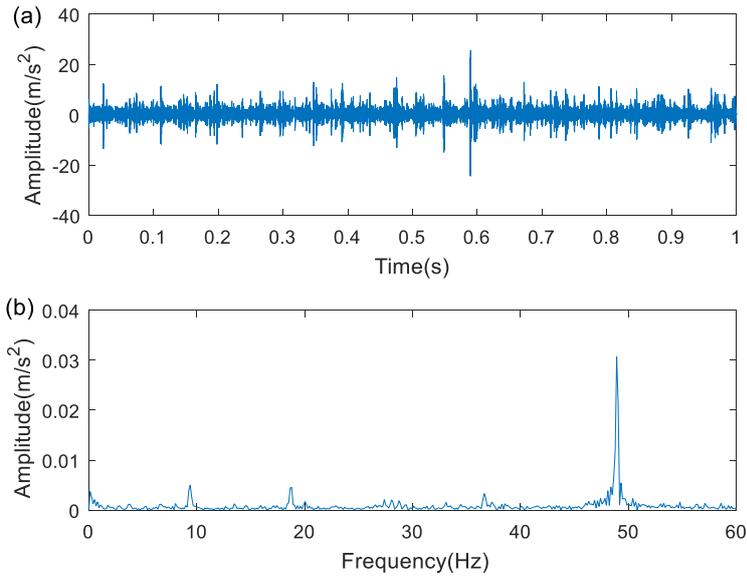
Fig. 13. Vibration signal for planetary gearbox with sun gear chipped tooth: (a) waveform (b) spectrum.

The EAVG-MSB is then applied to analyze the measured vibration signal that is depicted in Fig. 13 (a). The EAVG with different length of flat SEs are then performed in the raw data, the kurtosis values of the filtered signals by EAVG under different lengths of SEs are shown in Fig. 14, and the optimal length of SE occurs at 4 with maximum kurtosis value. The result of the EAVG filtered signal using the optimized length of SE is shown in Fig. 15. Only the sun gear rotational frequency f_{rs}, sun gear characteristic frequency $2f_{sf}$ can be identified, but sun gear characteristic frequency f_{sf} and modulation frequency $f_{sf} \pm f_{rs}$ and its harmonics are not recognized in Fig. 15 (b). In addition, there are still exists some interference frequencies and nonlinear modulation components after EAVG analysis. Then, the EAVG filtered signal is further analyzed by the MSB. Fig. 16 illustrates the results of the EAVG-MSB, and it can be clearly seen that some peaks correspond accurately and straightly to the sun gear rotational frequency f_{rs}, sun gear characteristic frequency f_{sf}, and their combinations $f_{sf} \pm f_{rs}$ and its harmonics. Therefore, it can be reasonably inferred that the EAVG-MSB is an effective fault detection method for the measured signal of the sun gear chipped tooth based on the fault related impulsive components enhancement by EAVG and fault characteristic frequency demodulated and extracted accurately by MSB.
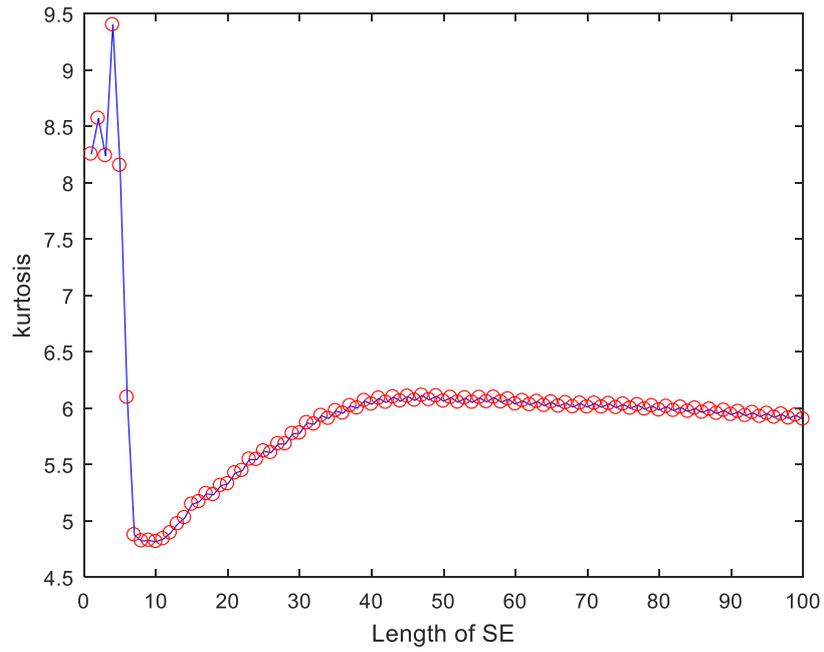
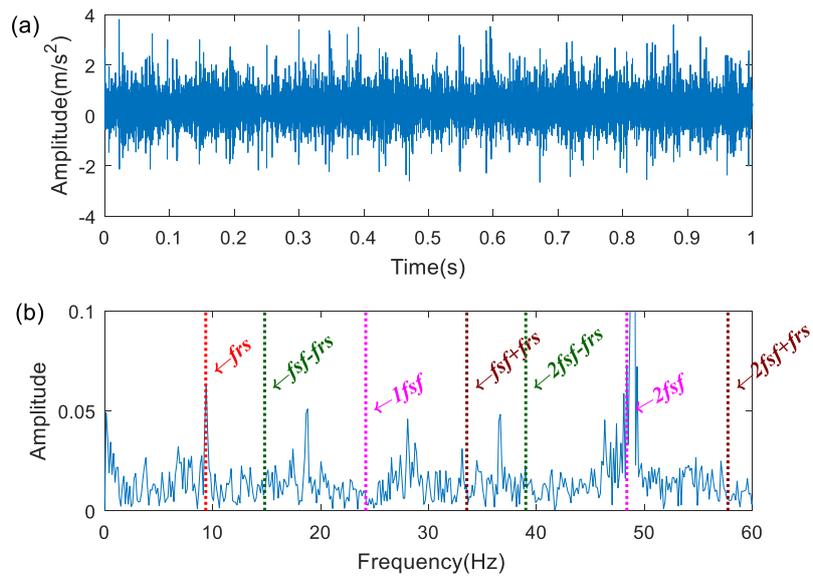**Fig. 14.** The kurtosis value of EAVG filtered signal under different lengths of SE.

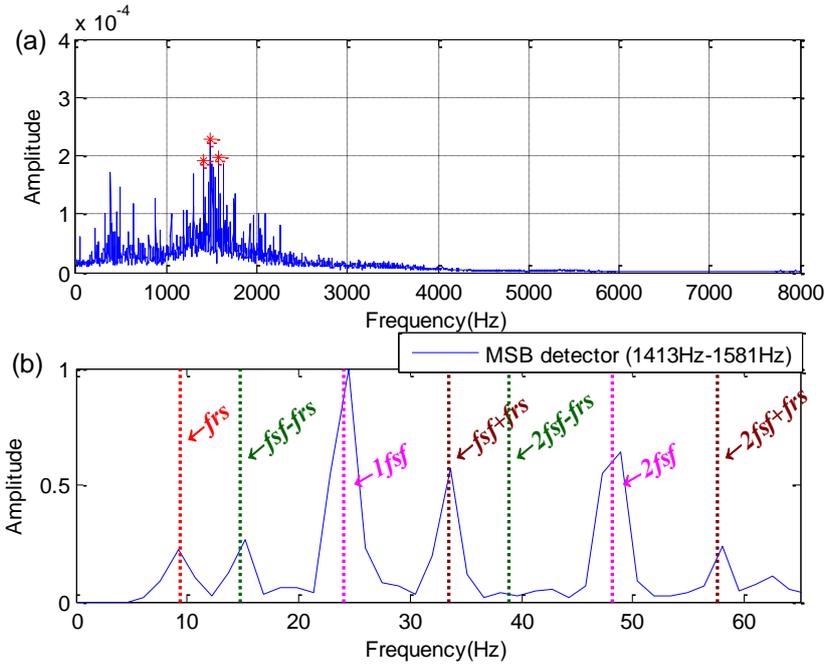**Fig. 15.** The analysis result of the EAVG: (a) waveform (b) spectrum.

Fig. 16. The analysis results of the EAVG-MSB: (a) MSB slice (b) MSB detector.

For comparison, the same waveform shown in Fig. 13 (a) is processed by the average combination difference morphological filter (ACDIF) [26] and the combination morphological filter-hat transform (CMFH) filter [24]. According to the literature [26], ACDIF is an effective analysis method for extracting fault-related frequency and its harmonics from background noise and interference components of vibration signal. In addition, it has been demonstrated that ACDIF is superior to other morphological filters (i.e. CMF [32] AVG [23] DIF [21]). The ACDIF with a flat SE is applied, and the optimal SE length of the ACDIF is decided by the Teager energy kurtosis (TEK). The TEK values of the filtered signals under different lengths of SE as showed in Fig. 17. The analysis result of the ACDIF with the length L= 57 is shown in Fig. 18. It can be seen that only the fault characteristic frequencies of the sun gear can be extracted. According to literature [24], the combination morphological filter-hat transform (CMFH) with a flat SE is introduced to detect rolling element bearing fault. The feature energy factor (FEF) values of the filtered signals under different lengths of SE are showed in Fig. 19. The CMFH with the optimal length L=5 are applied to analyze the measured vibration signal displayed in Fig. 13 (a) and the analysis result is shown in Fig. 20. There are a lot of heavy noise and some interference frequency components in Fig. 20, and it is impossible to extract the fault-related frequency and its harmonics. In order to further illustrate the superiority of the proposed EAVG-MSB method, the CPU running time of EAVG-MSB, ACDIF and CMFH approaches is tested on an Intel Core i5 4210U 2.40 GHz CPU with 8.00 GB RAM. The comparison results by the CPU running time are displayed in Table 5. As can be observed from Table 5, the proposed EAVG-MSB method has less CPU runtime than ACDIF, but more than CMFH. The reason why the calculation efficiency of EAVG-MSB is lower than CMFH mainly includes two aspects. On the one hand, the morphological operator of EAVG is more complicated than CMFH. On the other hand, MSB increases the CPU running time when calculating MSB slices. Therefore, it demonstrates that the proposed EAVG-MSB method is superior to the ACDIF and CMFH for planetary gearbox with sun gear misalignment fault diagnosis.

Table 5. The CPU running time of the EAVG-MSB, ACDIF and CMFH methods.

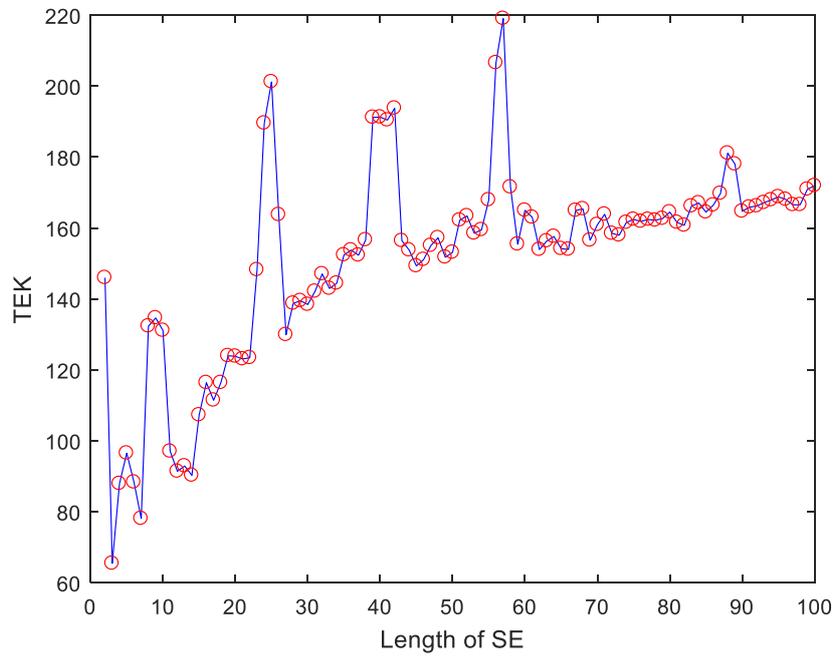| Methods	EAVG-MSB	ACDIF	CMFH
CPU running time	323.61	346.93	168.70

Fig. 17. The TEK value of ACDIF filtered signals under different lengths of SE.
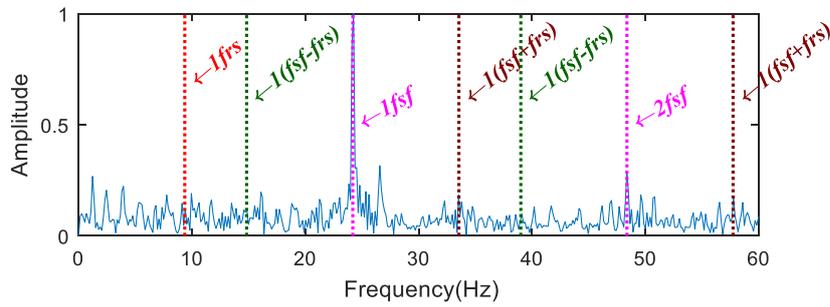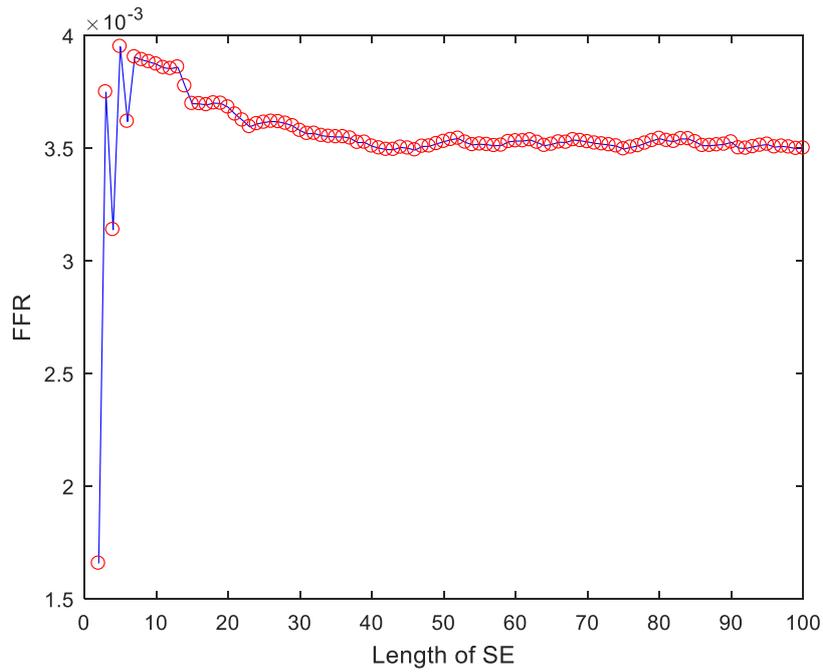

Fig. 18. Analysis results of the ACDIF.

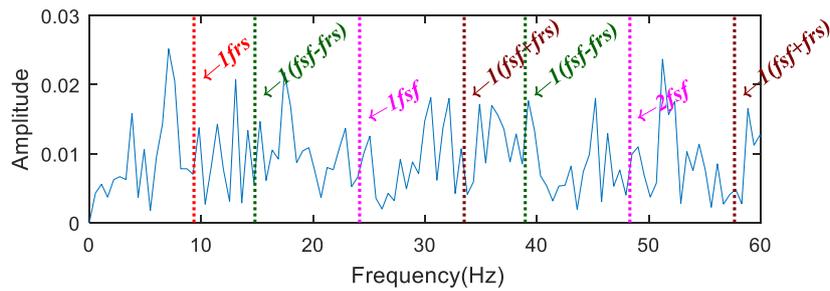**Fig. 19.** The FFR value of CMFH filtered signal under different lengths of SE.

Fig. 20. Analysis results of the CMFH.

5. 3 Planetary gearbox with sun gear misalignment fault diagnosis

Fig. 21 presents the waveform and spectrum of the test planetary gearbox with sun gear misalignment vibration signal. As seen in Fig. 21 (a), the waveform of the measured vibration signal involves various impulse components associated with the sun gear failure. And it is difficult to identify the fault characteristic frequencies of the sun gear misalignment fault from the spectrum as displayed in Fig. 21 (b).
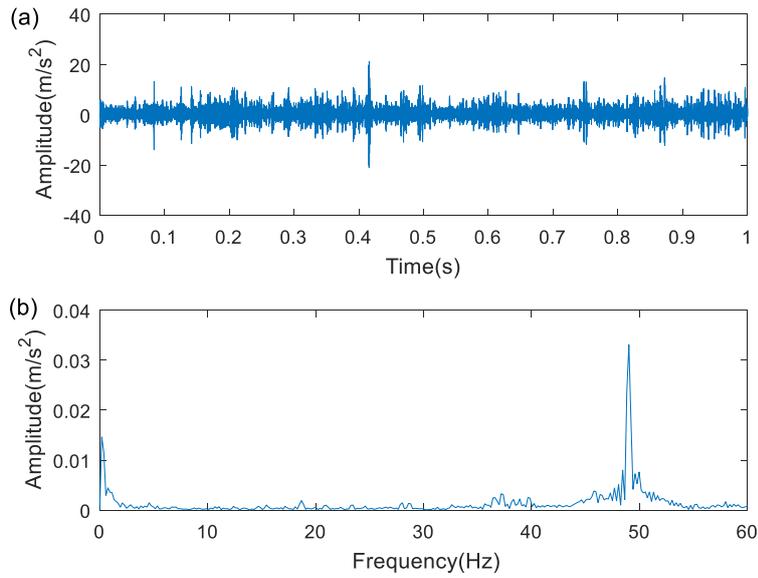
Fig. 21. Vibration signal for planetary gearbox with sun gear misalignment: (a) waveform (b) spectrum.
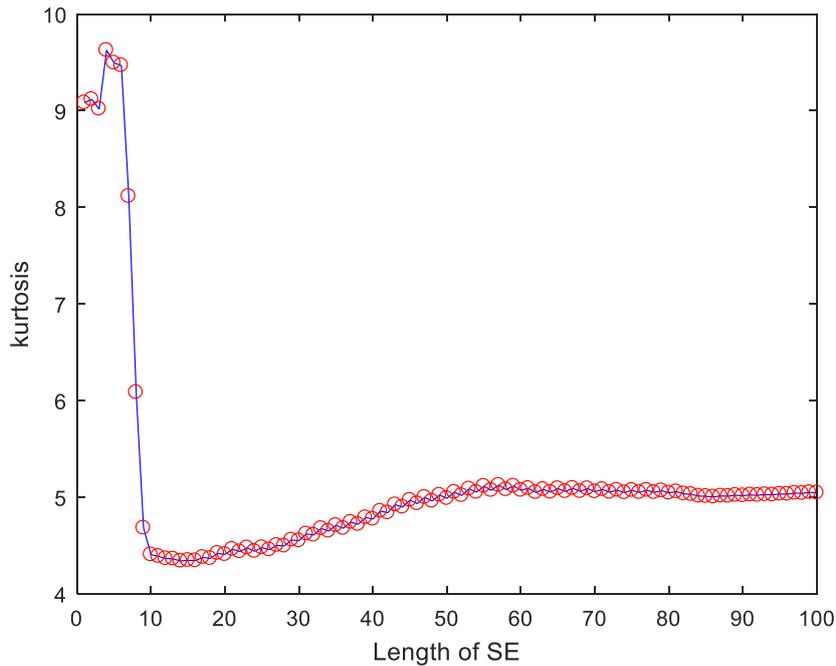

Fig. 22. The kurtosis value of EAVG filtered signal under different lengths of SE.

The EAVG-MSB is applied to process the measured vibration signal of the sun gear misalignment. The EAVG eliminates heavy noise and enhances the impulse component in the vibration signal using the SE length L = 4 determined by the criterion of maximum kurtosis as presented in Fig. 22. Then, the EAVG filter signal is further analyzed using the MSB. It can be clearly seen that the EAVG-MSB has distinctive peaks correspond accurately and straightly to the sun gear rotational frequency f_{rs}, sun gear characteristic frequency f_{sf}, and their combinations $f_{sf} \pm f_{rs}$ and its harmonic as shown in Fig. 23. For comparison, ACDIF and CMFH are applied to process the signal as depicted in Fig. 21 (a). The analysis result of the ACDIF with the length L= 37 is plotted in Fig. 24. It can be observed that the extraction impulse feature results using the ACDIF are not sufficiently obvious compared with the EAVG-MSB.

The analysis result of the CMFH with the length L= 6 is shown in Fig. 25. It can be seen from Fig. 25 that there is background noise and some interference frequency components, and it is impossible to extract the fault-related characteristic frequencies. For further comparison, the CPU running time is applied to evaluate the performance of EAVG-MSB, ACDIF and CMFH. The comparison results using the CPU running time are displayed in Table 6. Although the CPU running time of the CMFH is less than the proposed EAVG-MSB method, the CMFH cannot accurately extract the fault defect frequencies. Therefore, it proves that the proposed EAVG-MSB method outperforms the ACDIF and CMFH for planetary gearbox with sun gear misalignment fault diagnosis.

Table 6. The CPU running time of the EAVG-MSB, ACDIF and CMFH methods.

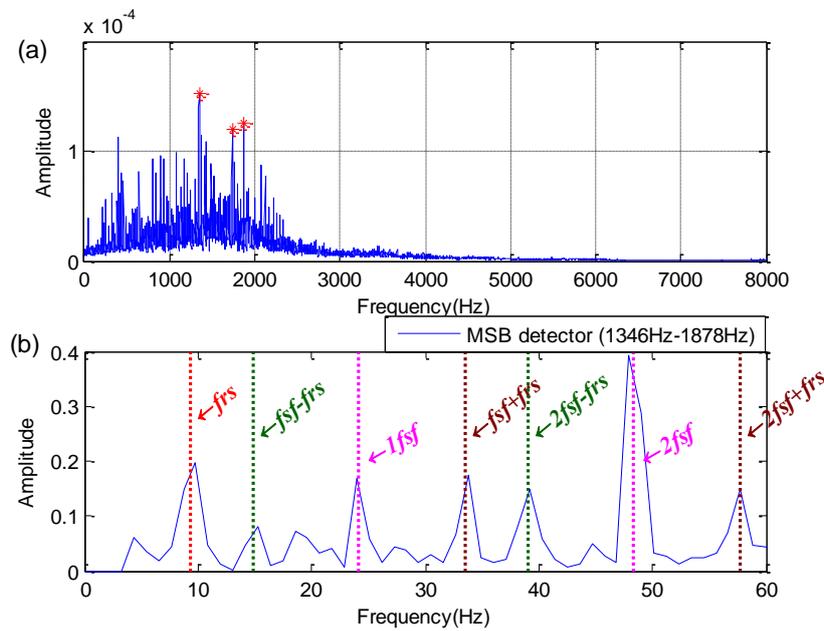| Methods	EAVG-MSB	ACDIF	CMFH
CPU running time	319.67	347.99	243.16

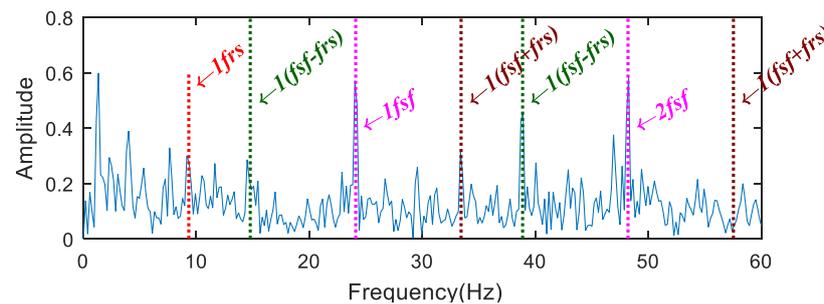**Fig. 23.** Analysis results of the EAVG-MSB: (a) MSB slice (b) MSB detector.

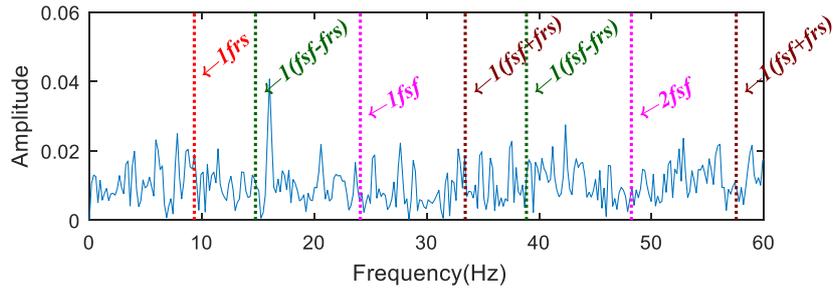
Fig. 24. Analysis results of the ACDIF.

Fig. 25. Analysis results of the CMFH.

5. 4 Planetary gearbox bearing with inner race fault diagnosis

Fig. 26 illustrates the waveform and spectrum of the test planetary gearbox bearing with inner race fault vibration signal. The signal contains various impulsive components which are related to the bearing inner race fault as shown in Fig. 26 (a). And it is difficult to recognize the fault characteristic frequencies of the inner race fault from the spectrum as depicted in Fig. 26 (b).
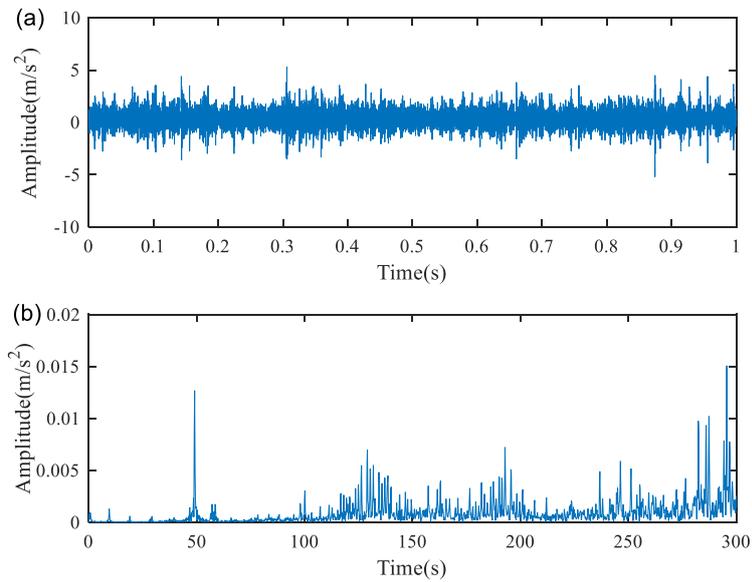

Fig. 26. Vibration signal for planetary gearbox bearing with inner race fault: (a) waveform (b) spectrum.
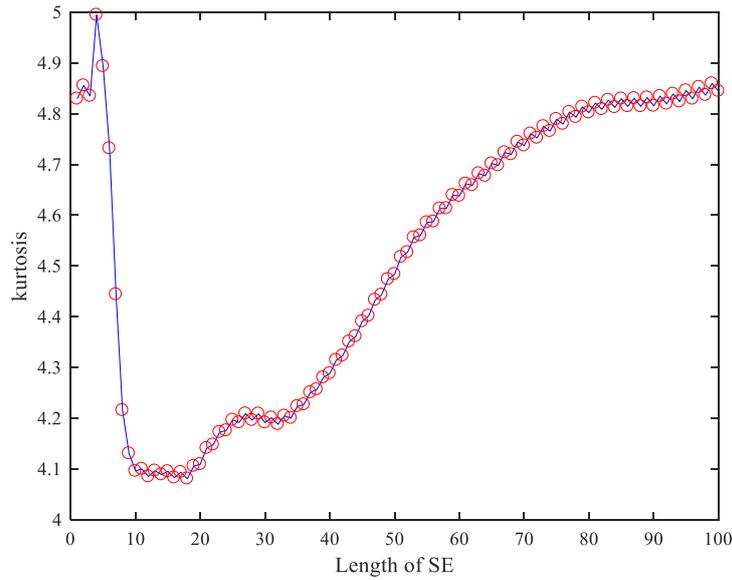
Fig. 27. The kurtosis value of EAVG filtered signal under different lengths of SE.

The EAVG-MSB is then used to analyze the measured vibration signal for fault characteristic frequency detection of the bearing inner race fault. The EAVG removes background noise and enhances the impulse component using the SE length L = 4 determined by the criterion of maximum kurtosis as illustrated in Fig. 27. The MSB is then further implemented into the signal filtered by the EAVG. It can be clearly seen that the EAVG-MSB has distinctive peaks correspond exactly and directly to the bearing inner race fault characteristic frequency f_i and its harmonics as presented in Fig. 28. For comparison, ACDIF and CMFH are employed to analyze the Fig. 26 (a). The analysis result of the ACDIF with the length L= 12 is displayed in Fig. 29. It can be observed that the extraction impulse feature results using the ACDIF are not sufficiently obvious compared with the EAVG-MSB. The analysis result of the CMFH with the length L= 4 is shown in Fig. 30. It can be seen from Fig. 30 that there is background noise and some interference frequency components, and it is impossible to extract the fault-related frequencies. The CPU running time of the EAVG-MSB, ACDIF and CMFH methods is computed in Table 7. From Table 7, it is obvious the CPU running time of the proposed EAVG-MSB method is more than CMFH, but EAVG-MSB can accurately extract the fault defect frequencies. Therefore, it certificates that the proposed EAVG-MSB method has better than ACDIF and CMFH for planetary gearbox bearing with inner race fault diagnosis.

Table 7. The CPU running time of the EAVG-MSB, ACDIF and CMFH methods.

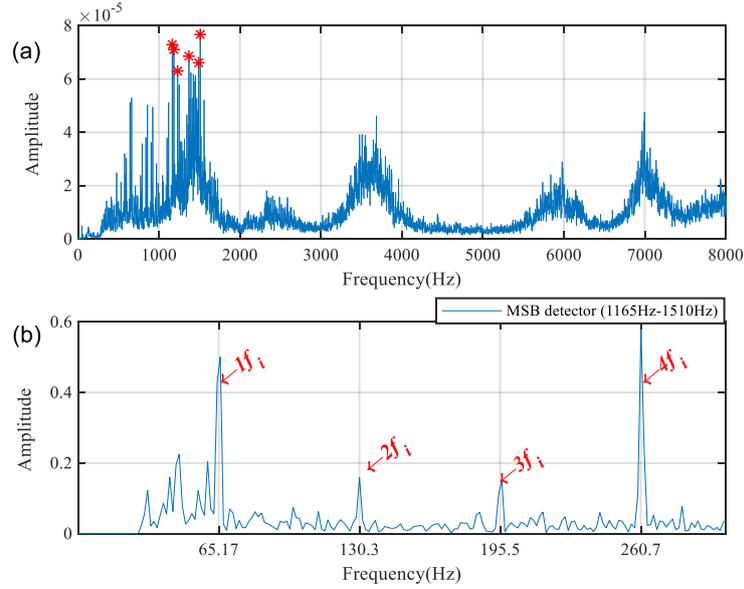| Methods	EAVG-MSB	ACDIF	CMFH
CPU running time	315.89	369.94	263.13

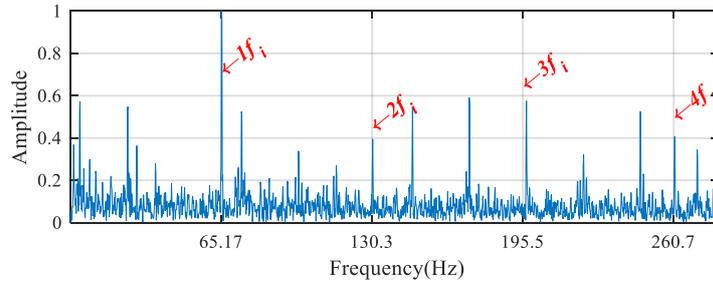**Fig. 28.** Analysis results of the EAVG-MSB: (a) MSB slice (b) MSB detector.

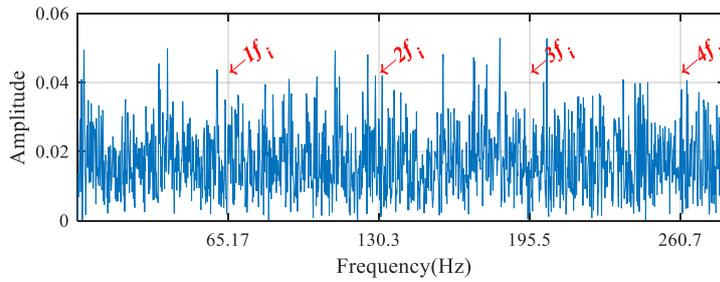
Fig. 29. Analysis results of the ACDIF.

Fig. 30. Analysis results of the CMFH.

6. Conclusions

This paper presented a novel algorithm based on the enhanced average (EAVG) filter and modulated signal bispectrum (MSB) for planetary gearbox fault detection. This method takes advantage of the fault-related impulse components enhancement performance of EAVG and the demodulation modulation component and the suppression interference frequency capability of MSB to accurately detect the fault of the planetary gearbox. The main conclusions of this paper are drawn as follows:

(1) In comparison with the AVG, the proposed EAVG can effectively extract the positive and negative impulses and suppress background noise and harmonic components interference.

(2) A criterion called kurtosis index is applied to adaptively optimize the SE length of EAVG, which

improves the extraction transient impulses performance of EAVG.

(3) MSB produces more accurate detection of the AM feature than CB, and is used to process filtered signal by EAVG to accurately extract fault features.

(4) The experimental analysis results illustrate that the proposed EAVG-MSB outperforms the ACDIF and CMFH and is more accurate and effective in enhancing the fault characteristic frequencies for planetary gearbox fault diagnosis.

In the future works, more applications need to be carried out to evaluate the performance and efficiency of the proposed EAVG-MSB method in the relevant industries. And the influence of the structural element length on the performance of EAVG needs to be analysis in detail to optimize the proposed EAVG-MSB method.

Acknowledgement

References

[1] Feng Z, Zhou Y, Zuo M, Chu F, Chen X. Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: A review with examples. Measurement 2017; 103: 106-32.

[2] Zhang M, Wang KS, Wei D, Zuo M. Amplitudes of characteristic frequencies for fault diagnosis of planetary gearbox. J Sound Vib 2018; 432 (13): 119-32.

[3] He Q, Song H, Ding X. Sparse signal reconstruction based on Time–Frequency manifold for rolling element bearing fault signature enhancement. IEEE T Instrum Meas 2016; 65 (2): 482-91.

[4] Feng Z, Chen X, Wang T. Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions. J Sound Vib 2017; 400: 71-85.

[5] Liu Z, Zuo M, Jin Y, Pan D, Qin Y. Improved local mean decomposition for modulation information mining and its application to machinery fault diagnosis. J Sound Vib 2017; 397 (9): 266-81.

[6] Meng Z, Zhan X, Li J, Pan Z. An enhancement denoising autoencoder for rolling bearing fault diagnosis. Measurement 2018; 130: 448-54.

[7] Lei Y, Han D, Lin J, He Z. Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. Mech Syst Sig Process 2013; 38 (1): 113-24.

[8] Wang L, Liu Z, Miao Q, Zhang X. Time-frequency analysis based on ensemble local mean decomposition and fast kurtogram for rotating machinery fault diagnosis. Mech Syst Sig Process 2018; 103: 60-75.

[9] Miao Y, Zhao M, Lin J. Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition. ISA Trans 2019; 84: 82-95.

[10] Zhang X, Wang J, Liu Z, Wang J. Weak feature enhancement in machinery fault diagnosis using empirical wavelet transform and an improved adaptive bistable stochastic resonance. ISA Trans 2019; 84: 283-95.

[11] Sun RB, Yang ZB, Luo W, Qiao BJ, Chen XF. Weighted sparse representation based on failure dynamics simulation for planetary gearbox fault diagnosis. Meas Sci Technol 2019; 30(4): 045008.

[12] Wang HQ, Ren BY, Song LY, Cui LL. A novel weighted sparse representation classification strategy based on dictionary learning for rotating machinery. IEEE T Instrum Meas 2019; DOI:10.1109/TIM.2019.2906334.

[13] Li Y, Liang X, Xu M, Huang W. Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform. Mech Syst Signal Process 2017; 86: 204-223.

[14] Lyu X, Hu Z, Zhou H, Wang Q. Application of improved MCKD method based on QGA in planetary gear compound fault diagnosis. Measurement 2019; 139: 236-48.

[15] Li B, Zhang PL, Wang ZJ, Mi SS, Liu DS. A weighted multi-scale morphological gradient filter for rolling element bearing fault detection. ISA Trans 2011; 50 (4): 599-608.

[16] Meng L, Xiang J, Wang Y, Jiang Y, Gao H. A hybrid fault diagnosis method using morphological filter–translation invariant wavelet and improved ensemble empirical mode decomposition. Mech Syst Signal Process 2015; 50-51: 101-15.

[17] J. Serra, L. Vincent, An overview of morphological filtering. Circ. Syst. Signal Process. 1992; 11 (1): 47-108.

[18] Maragos P, Schafer RW. Morphological filters-Part I: their set theoretic analysis and relations to linear shift-invariant filters, IEEE Trans. Acoust. Speech Signal Process. 1987; 35 (8): 1153-69.

[19] Li Y, Liang X, Zuo MJ. Diagonal slice spectrum assisted optimal scale morphological filter for rolling element bearing fault diagnosis. Mech Syst Sig Process 2017; 85: 146-61.

[20] Hu Z, Wang C, Zhu J, Liu X, Kong F. Bearing fault diagnosis based on improved morphological filter. Measurement 2016; 80: 163–78.

[21] Li Y, Zuo MJ, Chen Y, Feng K. An enhanced morphology gradient product filter for bearing fault detection. Mech Syst Sig Process 2018; 109: 166-84.

[22] Hu AJ, Sun JJ, Xiang L. Analysis of morphological filter's frequency response characteristics in vibration signal processing. Chin J Mech Eng 2012; 48 (1): 98-103.

[23] Dong Y, Liao M, Zhang X, Wang F. Faults diagnosis of rolling element bearings based on modified morphological method. Mech Syst Signal Process 2011; 25 (3): 1276-86.

[24] Yan X, Jia M, Zhang W, Zhu L. Fault diagnosis of rolling element bearing using a new optimal scale morphology analysis method. ISA Trans 2018; 73: 165-80.

[25] Hu A, Xiang L. Selection principle of mathematical morphological operators in vibration signal processing. J Vib Control 2014; 43(2): 420-25.

[26] Lv JX, Yu JB. Average combination difference morphological filters for fault feature extraction of bearing. Mech Syst Signal Process 2018; 100: 827-45.

[27] Gu F, Wang T, Alwodai A, Tian X, Shao Y, Ball AD. A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals. Mech Syst Sig Process 2015; 50-51: 400-13.

[28] Tian X, Gu JX, Rehab I, Abdalla GM, Gu F, Ball AD. A robust detector for rolling element bearing condition monitoring based on the modulation signal bispectrum and its performance evaluation against the Kurtogram. Mech Syst Sig Process 2018; 100: 167-87.

[29] Guo J, Shi Z, Li H, Zhen D, Gu F, Ball AD. Early fault diagnosis for planetary gearbox based wavelet packet energy and modulation signal bispectrum analysis. Sensors 2018; 18: 2908.

[30] Guo J, Zhen D, Li H, Shi Z, Gu F, Ball AD. Fault feature extraction for rolling element bearing diagnosis based on a multi-stage noise reduction method. Measurement 2019; 139: 226-35.

[31] Serra J. Image analysis and mathematical morphology. Academic Press, New York, USA, 1982.

[32] Wang J, Xu G, Zhang Q, Liang L. Application of improved morphological filter to the extraction of impulsive attenuation signals. Mech Syst Signal Process 2009; 23 (1): 236-45.

[33] He W, Jiang ZN, Qin Q. A joint adaptive wavelet filter and morphological signal processing method for weak mechanical impulse extraction. J Mech Sci Technol 2010; 8: 1709-16.

[34] Hu Z, Wang C, Zhu J, Liu X, Kong F. Bearing fault diagnosis based on improved morphological filter. Measurement 2016; 80: 163-78.

[35] Chen J, Zhou D, Lyu C, Lu C. An integrated method based on CEEMD-SampEn and the correlation analysis algorithm for the fault diagnosis of a gearbox under different working conditions. Mech Syst Signal Process 2017; 113: 102-11.

[36] Li B, Zhang P, Wang Z, Mi S, Zhang Y. Gear fault detection using multi-scale morphological filters. Measurement 2011; 44: 2078-89.

[37] Hu AJ, Tang GJ, An LS. New method of removing pulsed noises in vibration data. J Vib Shock 2006; 25 (1): 126-9.

[38] Zhang L, Xu J, Yang J. Yang D. Wang D. Multiscale morphology analysis and its application to fault diagnosis. Mech Syst Signal Process 2008; 22 (3): 597-610.

[39] Dong S, Tang B, Chen F. De-nosing method based on multiscale morphological filter optimized by particle swarm optimization algorithm. J Chongqing Univ 2012; 35: 7-12.

[40] Zhang X, Gao H, Huang H. Optimization design of mathematical morphology filter based on quantum genetic algorithm. J Southwest Jiao Tong Univ 2014; 49: 462-9.

[41] Raj AS, Murali N. Early classification of bearing faults using morphological operators and fuzzy inference. IEEE Trans Industr Electron 2013; 60 (2): 567-74.

[42] Li C, Liang M. Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis. J Sound Vib 2012; 331 (26): 5864-79.

[43] Yan X, Jia M. Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings. Mech Syst Signal Process 2019; 122: 56-86.

[44] Kovach CK, Oya H, Kawasaki H. The bispectrum and its relationship to phase-amplitude coupling. Neurolmage 2018; 173: 518-39.

[45] Dong G, Chen J, Zhao F. A frequency-shifted bispectrum for rolling element bearing diagnosis. J Sound Vib 2015; 339: 396-418.

[46] Orosco E, Diez P, Laciar E, Mut V, Soria C, Sciascio FD. On the use of high-order cumulant and bispectrum for muscular-activity detection. Biomed Signal Process 2015; 18: 325-33.