A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing

Z.K. Peng\textsuperscript{a,*}, Peter W. Tse\textsuperscript{a}, F.L. Chu\textsuperscript{b}

\textsuperscript{a}Smart Asset Management Laboratory, City University of Hong Kong, Tat Chee Ave, Hong Kong, PR China
\textsuperscript{b}Department of Precision Instruments, Tsinghua University, Beijing 100084, PR China

Received 28 November 2003; received in revised form 6 January 2004; accepted 19 January 2004
Available online 14 March 2004

Abstract

For rolling bearing fault detection, it is expected that a desired time–frequency analysis method should have good computation efficiency, and have good resolution in both time domain and frequency domain. As the best available time–frequency method so far, the wavelet transform still cannot fulfill the rolling bearing fault detection task very well since it has some inevitable deficiencies. The recent popular time–frequency analysis method, Hilbert–Huang transform (HHT), has good computation efficiency and does not involve the concept of the frequency resolution and the time resolution. So the HHT seems to have potential to become a perfect tool for rolling bearing fault detection. However, in practical applications, the HHT also suffers from some unsolved deficiencies. To ameliorate these deficiencies, by seeking help from the wavelet packet transform (WPT) and a simple but effective method for intrinsic mode function (IMF) selection, an improved HHT is put forward in this studying. Several numerical study cases will be used to validate the capabilities of the improved HHT. Finally, the improved HHT’s performance in rolling bearing fault detection is compared with that of the wavelet based scalogram through experimental case studies. The comparison results have shown that (1) the improved HHT has better resolution both in time domain and in frequency domain than the scalogram; (2) the improved HHT has better computing efficiency than scalogram; (3) the HHT spectrum also has one unresolved and maybe inevitable deficiency—ripple phenomenon in its estimated frequency, which would mislead our analysis.

Keywords: Wavelet transform; Hilbert–Huang transform; Fault diagnosis; Vibration signal

\textsuperscript{*}Corresponding author.
\textit{E-mail address:} mepengzk@cityu.edu.hk (Z.K. Peng).

0888-3270/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
1. Introduction

Rolling bearings, as important components, are widely used in rotary machines; faults occurring in bearings must be detected as early as possible to avoid fatal breakdowns of machines that may lead to loss of production and human casualties. Faults that typically occur in rolling bearings are usually caused by localized defects in the outer-race, the inner-race, the rollers, or the cage. Such defects generate a series of impact vibrations every time a running roller passes over the surfaces of the defects, and as is well known, for different fault types, the impacts will appear with different frequency. To detect the faults in bearings, hitherto, many kinds of methods have been developed [1–5], and, without exception, their cores are various signal analysis techniques. In the early studies, Fourier analysis has been the dominating signal analysis tool for bearing fault detection. But, there are some crucial restrictions of the Fourier transform [6]: the signal to be analysed must be strictly periodic or stationary; otherwise, the resulting Fourier spectrum will make little physical sense. Unfortunately, the rolling bearing vibration signals are often non-stationary and represent non-linear processes, and their frequency components will change with time. Therefore, the Fourier transform often cannot fulfill the bearing fault diagnosis task pretty well. On the other hand, the time–frequency analysis methods can generate both time and frequency information of a signal simultaneously through mapping the one-dimensional signal to a two-dimensional time–frequency plane. Therefore, in the later studies, the time–frequency analysis methods are widely used to detect the faults in bearings since they can determine not only the time of the impact occurring but also the frequency ranges of the impact location, and hence can determine not only the existence of faults but also the causes of faults. Among all available time–frequency analysis methods, the wavelet transform maybe the best one and has been widely used for rolling bearing fault detection. However, the wavelet transform still has some inevitable deficiencies [7], including the interference terms, border distortion and energy leakage, all of which will generate a lot of small undesired spikes all over the frequency scales and make the results confusing and difficult to be interpreted. Additionally, for rolling bearing fault detection, the frequency ranges of the vibration signals that we need to analyse are often rather wide; and according to the Shannon sampling theorem, a high sampling speed is needed, and sequentially, large size samples are needed for the rolling bearing fault detection. Therefore, it is expected that the desired method should have good computing efficiency. Unfortunately, the computing of continuous wavelet transform (CWT) is somewhat time consuming and is not suitable for large size data analysis. Moreover, the occurrence of impacts in bearings is often rather frequent and the interval between two adjacent impacts is quite small. Hence, the desired time–frequency analysis methods for bearing vibration signal analysis should have fine resolutions both in time domain and in frequency domain. Due to the limitation of Heisenberg–Gabor inequality, the wavelet transform cannot achieve fine resolutions in both time domain and frequency domain simultaneously, therefore, although the wavelet transform has good time resolution in high-frequency region, it often cannot separate those impacts, for the time interval between them are often too small.

In the recent years, another time–frequency analysis method named Hilbert–Huang transform (HHT) [8–10] has become more and more popular. The technique works through performing a time adaptive decomposition operation named empirical mode decomposition (EMD) on the signal; and then the signal will be decomposed into a set of complete and almost orthogonal
components named intrinsic mode function (IMF), which is almost monocomponent. Utilizing Hilbert transform on those obtained IMFs, we can get a full energy–frequency–time distribution of the signal, designated as the Hilbert–Huang spectrum. One of the advantages of HHT is that its most computation consuming step, EMD operation, does not involve the convolution and other time-consuming operations, therefore the HHT can deal with the signals of large size. Additionally, the Hilbert–Huang spectrum does not involve the concept of the frequency resolution and the time resolution but the instantaneous frequency. It seems that the HHT has the potential of becoming a perfect tool for rolling bearing fault detection. However, in practical applications, the HHT also has some unsolved problems, all of which will be caused by the EMD. First, the EMD will generate some undesired low amplitude IMFs at the low-frequency region and raise some undesired frequency components. Second, the first IMF may cover a wide frequency range at the high-frequency region and therefore cannot satisfy the monocomponent definition very well. Third, the EMD operation often cannot separate some low-energy components from the analysis signal, therefore those components may not be able to appear in the frequency–time plane. By seeking help from the wavelet packet transform (WPT) [11] and a simple but effective method for IMF selection, an improved HHT is put forward here which can alleviate those deficiencies of the HHT to certain extent. The detailed introduction about the improved HHT will be given in this paper, in the following, and comparison studies about its performances with that of the wavelet based scalogram in rolling bearing fault detection will be carried out; some numerical cases will also be studied.

The symptoms of rolling bearing fault and its detection methods together with their deficiencies are briefly introduced in this section. A potential tool for rolling bearing fault detection named HHT and its deficiencies are also briefly discussed in Section 1. Section 2 presents the theory of conventional HHT. Section 3 provides a detailed discussion on the deficiencies of the conventional HHT and to overcome those deficiencies, an improved HHT is put forward. In Section 4, the effectiveness of the improved HHT is tested by using simulated signals and experimental rolling bearing signals, and the results by wavelet scalograms are also given for comparison. Finally, Section 5 lays out the conclusive remarks.

2. Brief introduction of Hilbert–Huang transform

Hilbert transform [6], a well-known signal analysis method, is essentially defined as the convolution of signal \( x(t) \) with \( 1/t \) and can emphasise the local properties of \( x(t) \), as follows:

\[
y(t) = \frac{P}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} \, d\tau
\]

(1)

where \( P \) is the Cauchy principal value. Coupling the \( x(t) \) and \( y(t) \), we can have the analytic signal \( z(t) \) of \( x(t) \), as

\[
z(t) = x(t) + iy(t) = a(t)e^{i\varphi(t)},
\]

(2)

where

\[
a(t) = \left[ x^2(t) + y^2(t) \right]^{1/2}, \quad \varphi(t) = \arctan(y(t)/x(t)).
\]

(3)
\( a(t) \) is the instantaneous amplitude of \( x(t) \), which can reflect how the energy of the \( x(t) \) varies with time, and \( \phi(t) \) is the instantaneous phase of \( x(t) \).

One important property of the Hilbert transform is that if the signal \( x(t) \) is monocomponent, then the time derivative of instantaneous phase \( \phi(t) \) will be the physical meaning of instantaneous frequency \( \omega(t) \) of signal \( x(t) \), as the following:

\[
\omega(t) = \frac{d\phi(t)}{dt}.
\]

The pity is that, in almost all of practical applications, the signal to be treated hardly belongs to the monocomponent but to the multicomponent, and it prevents the important concept of the instantaneous frequency from being extensively applied. To make the instantaneous frequency applicable, Huang et al. [1] presented a signal decomposition method, referred to as the empirical mode decomposition (EMD), which is able to decompose a signal into some individual, nearly monocomponent signals with ‘Hilbert-friendly’ waveforms, named as intrinsic mode function (IMF) to which the instantaneous frequency defined by Eq. (4) can be applied. The EMD preprocessor-based Hilbert transform is named Hilbert–Huang Transform (HHT).

An IMF is a function that satisfies the two following conditions: (a) the number of extrema and the number of zero crossings must either equal or differ at most by one in whole data set, and (b) the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at every point. An IMF represents simple oscillatory mode imbedded in the signal. Actually, the IMF does not always guarantee a perfect instantaneous frequency definition under all conditions and is only nearly monocomponent. Nevertheless, many applications had shown that, even under the worst conditions, the instantaneous frequency defined as Eq. (4) is still tenable for an IMF. In practice, at any given time, most of the signals may involve more than one oscillatory mode, that is, the signal has more than one instantaneous frequency at a time locally. With the simple assumption that any data consisting of different simple IMFs, the EMD is developed to decompose a signal into IMF components. The EMD process can be summarised in the following table (Table 1):

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>The EMD algorithm</td>
</tr>
</tbody>
</table>

1. Initialise: \( r_0 = x(t) \), and \( i = 1 \);
2. Extract the \( i \)th IMF
   1. Initialise: \( h_{i(k-1)} = r_i \), \( k = 1 \)
   2. Extract the local extrema and minima of \( h_{i(k-1)} \)
   3. Interpolate the local extrema and the minima by cubic spline lines to form upper and lower envelopes of \( h_{i(k-1)} \)
   4. Calculate the mean \( m_{i(k-1)} \) of the upper and lower envelopes of \( h_{i(k-1)} \)
   5. Let \( h_k = h_{i(k-1)} - m_{i(k-1)} \)
   6. If \( h_k \) is a IMF then set IMF \( i = h_k \), else go to step b) with \( k = k + 1 \)
3. Define \( r_{i+1} = r_i - IMF_i \)
4. If \( r_{i+1} \) still has least 2 extrema then go to step (2) else decomposition process is finished and \( r_{i+1} \) is the residue of the signal
At the end of the procedure we have a residue \( r_n \) and a collection of \( n \) IMFs \( c_i (i = 1, 2, \ldots, n) \). Summing up all IMFs and the final residue \( r_n \), we obtain

\[
x(t) = \sum_{i=1}^{n} c_i + r_n.
\] (5)

Having obtained the IMFs, we can utilise the Hilbert transform to each IMF, and compute the instantaneous frequency and amplitude according to Eqs. (4) and (3). After performing the Hilbert transform on each IMF component, it can express the signal in the following form:

\[
x(t) = \sum_{j=1}^{n} a_j(t) \exp \left( i \int \omega_j(t) \, dt \right).
\] (6)

The EMD preprocessor-based Hilbert transform is named Hilbert–Huang transform (HHT). Eq. (6) enables us to represent the instantaneous amplitude and frequency as functions of time in a three-dimensional plot or contour map. The frequency–time distribution of the amplitude is designated as the Hilbert–Huang spectrum, \( H(\omega, t) \).

3. Improved Hilbert–Huang transform

Ideally, the IMFs will be a kind of complete, adaptive and almost orthogonal representation for the analysed signal, and so the HHT would be the perfect tool for the non-stationary and non-linear signal analysis. But there are still problems in practice. The serious problem is that large swings will occur near the ends of signal in the spline fitting process. Furthermore, the end swings can eventually propagate inward and corrupt the whole signal span and cause some undesirable IMFs in the EMD process, especially in the low-frequency components. Fig. 1 shows a multicomponent signal that contains two frequency components. Fig. 2 gives this signal’s IMFs and the residue produced by the EMD. Obviously, only the first two IMFs are the real components of the signal and the others are the pseudocomponents that have low frequency and will be represented as low-frequency components in the Hilbert spectrum and mislead our analysis.

To eliminate those pseudocomponents, a simple but effective IMFs selection method is presented.

![A simulated multicomponent signal](image)
Observing that the IMF is almost an orthogonal representation for the analysed signal, the real IMF components will have relative good correlation with the original signal. On the other hand, the pseudocomponents will only have poor correlation with the signal. Thus, we can use the correlation coefficients $m_i$ of IMFs and the signal as a criterion to decide which IMFs should be retained and which IMFs should be eliminated. To avoid eliminating some real IMFs with low amplitude, all IMFs and the signal will be normalised at first that the correlation coefficients are at most 1.0. After obtaining all correlation coefficients $m_i$ ($i = 1, \ldots, n$; $n$ is the number of IMFs), we can compare them with a hard threshold $\lambda$. The IMF selection criterion can be stated as the following (Table 2):

Often, the hard threshold $\lambda$ can be a ratio of the maximal $\mu_i$, that is

$$\lambda = \max(\mu_i)/\eta \quad (i = 1, \ldots, n),$$

where $\eta$ is a 1.0 bigger ratio factor. In this studying, $\eta = 10.0$ is used.

Now, let us use the IMF selection method on the example in Fig. 3. Table 3 lists the correlation coefficients of all the obtained IMFs and the signal.

It can be seen that only the first two IMFs have good correlation with the analysis signal and have large correlation coefficients. With IMF selection criterion, only the first two IMFs are retained and three others are eliminated and added to the residue $r_n$. The final residue is shown in Fig. 3. In the new residue, the end swing effect is shown obviously.
Besides the pseudo-IMF problem, there are still some cases where the first IMF will cover a wide frequency range at the high-frequency part and therefore cannot satisfy the monocomponent definition very well. Additionally, some low-energy components will be masked by the high-energy components during the EMD operation. Therefore those low-energy frequency components cannot be shown in the frequency–time plane of the signal. To illustrate this, a simulated signal generated through adding a relative weak sinusoidal signal to the signal in Fig. 1 is operated by EMD. This signal together with its IMFs and their respective FFT spectrums are given in Fig. 4. Obviously, the IMF1 contains many frequency components and, on the other hand, the other IMF is almost monocomponent. In the IMF1 case, the Hilbert transform will fail in getting the true frequency pattern. But, it is worthy to note that, if the Hilbert transform is used on this IMF1, the obtained instantaneous frequency will still be limited in its own frequency range since, after all, the IMF1 satisfies the aforementioned two conditions that an IMF should satisfy. Therefore, the calculated instantaneous frequency of IMF1 will be still somewhat meaningful.

As mentioned above, IMF represents simple oscillatory mode imbedded in the signal and, in each cycle, defined by the zero crossing, involves only one mode of oscillation without complex riding waves. In fact, with this definition, an IMF is not restricted to a narrow band signal, and it can be both amplitude and frequency modulated just as the IMF1 in Fig. 4. To solve the problem, a natural and intuitive ideal is to decompose the signal to some narrow band signals at first, and then use EMD operation on those narrow band signals, and thus the obtained IMFs will also have narrow frequency bands and their instantaneous frequencies calculated by Eq. (4) will be more close to the real frequency pattern of the IMF. The wavelet packet transform (WPT) maybe the best choice of the preprocessor for the HHT. It is well known that the WPT is orthogonal, complete, local and computing efficient. In the WPT, a signal is split into an approximation and detail through a couple of low band filter (LF) and high band filter (HF), respectively. The approximation and the detail are then themselves split into a second-level approximation and detail, and the process is repeated. For an n-level decomposition, the signal will be decomposed...
into 2\(^n\) narrow band signals. Additionally, the low-energy components will be decomposed into different bands. So the WPT is the best choice of the preprocessor for the HHT.

The HHT with the WPT as preprocessor plus the IMF selection method is here called the improved HHT. In the following sections, the performances of the improved HHT spectrum will be compared with the well-established wavelet base scalogram.

4. Comparative study of improved HHT and wavelet transform

Wavelet transform has become a well accepted time–frequency analysis tool and has been widely used in vibration signals analysis [12–14]. Since the wavelet transform is very well known, the details about it will be omitted. In fact, the HHT has also even been used in vibration signal analysis in different applications. Yang and Sun used the HHT to interpret the non-linear response of the crack-induced rotor [15]. Yang and Lei [16] proposed an HHT-based damage identification approach and applied it to the ASCE structural health monitoring benchmark structure. In this section, the performances of the improved HHT and the wavelet-based scalogram will be compared with each other through several numerical cases. Finally, both the improved HHT and the scalogram will be used to analyse the experimental signals of rolling bearings having different faults.
4.1. Numerical cases study

Case 1: A signal with two sinusoidal components. In this case, the signal shown in Fig. 2 will be analysed by both the improved HHT and the wavelet-based scalogram. The analysis results are shown as Fig. 5. It is worth noting here that the frequency is in a normalised frequency (Hz), while, the time-scale is in a number of sample points, and the rest of the figures (Figs. 6–9) will have the similar denotations. It can be seen that, although both of them have represented the true frequency patterns of this signal, the difference is clearly visible, that is, the scalogram has
different frequency resolutions at different frequency parts, the sinusoidal component with high frequency is shown as a wider frequency band than the one with low frequency in the time–frequency plane; on the other hand, the Hilbert–Huang spectrum has uniform resolution for all frequency part and all sinusoidal components have the same wide frequency bands. It is because different from the scalogram, the Hilbert–Huang spectrum does not involve the concept of the frequency resolution and time resolution but represents the instantaneous frequency. Therefore, once the EMD has decomposed a signal into real monocomponent IMFs successfully, the Hilbert–Huang spectrum will reflect the signal’s instantaneous frequency pattern, and not the time–frequency pattern with limit resolution that the scalogram can only give.

**Case 2: A signal with three sinusoidal components:** In this case, the signal shown in Fig. 4 will be analysed. Obviously, the three frequency components contained in the signal have been shown clearly by both the improved HHT spectrum and the wavelet-based scalogram. And similar to the results of Case 1 shown in Fig. 5, the scalogram has different frequency resolutions for the different sinusoidal components with different frequencies, and the improved HHT spectrum has uniform resolution for all sinusoidal components. A notable phenomenon is that, according to the HHT spectrum show, the sinusoidal component with highest frequency would have a frequency that changes periodically with time, but, in fact, its frequency is constant. This ripple phenomenon in the estimated frequency will often occur in cases when using the Hilbert transform on the signal that cannot satisfy the monocomponent conditions strictly, and will mislead our analysis thereby.
Additionally, it can be seen that the frequency with ripples oscillates around the real frequency value, and therefore we can still think that the estimated frequency can reflect the real frequency pattern of the analysed signal, but only in a mean sense. Furthermore, the variation range of the oscillating amplitude of the estimated frequency is narrower than the bandwidth of the scalogram, thus it could be said that the HHT spectrum still has better frequency resolution than scalogram even in a mean sense.

**Case 3: A signal with changed frequency.** In this case, a signal composed of components with changed frequencies will be analysed by both the improved HHT spectrum and the scalogram. Fig. 7 shows the temporal waveform of the simulated signal (the left bottom diagram), its FFT spectrum (the right bottom diagram), its result generated by improved HHT (the left top diagram), and the scalogram (the right top diagram). From the results shown in the HHT spectrum and the scalogram, one can find that the signal contains two sinusoidal components, among which the one with relative high frequency occurs during the first half lifespan of the signal and the one with low frequency occurs during the second half part. Once again, it can be seen that the concentration of high frequency is lower than that of the low frequency in scalogram, that is, the component with high frequency has wide bandwidth while the component with low frequency has narrow bandwidth. Additionally, the time resolution of the low frequency is lower than that of the high frequency in the scalogram. The low time resolution makes the component with low frequency, which should appear in the scalogram at the 500th sample point, has already appeared...
in the scalogram at about the 400th sample point. On the other hand, the HHT spectrum has captured the exact times of all components’ appearances and disappearances. Here, the ripple phenomenon occurs again in the HHT spectrum, especially visible for the component with high frequency.

With the comparison results, we can find that, in the above three numerical cases study, the improved HHT spectrum has shown better performances than the scalogram since the former has better resolution both in frequency domain and in time domain than the latter, and time resolution and frequency resolution are two important criteria in performance evaluation for a time–frequency analysis tool. But the HHT will suffer from the ripple phenomenon which often occurs when one obtained IMF cannot satisfy the monocomponent conditions strictly. Although the effects of the ripple phenomenon are not very serious in these cases, we should pay attention to it in other applications.

4.2. The experimental signal analysis for rolling bearing

In this section, both the improved HHT spectrum and the wavelet-based scalogram will be used to analyse two sets of experimental bearing signals and their performances will be compared.
All vibration signals were collected from an experimental testing machine with accelerometers at a sampling rate of 65.4 kHz. The geometric parameters of the rolling element bearing are as follows: the rolling element diameter \( d = 7.5 \text{ mm} \), pitch diameter \( D = 39.45 \text{ mm} \), contact angle \( \alpha = 5–20 \), and number of rolling elements \( n = 26(2 \times 13) \). The actual rotating speed monitored by the accelerometer was found to be \( f = 25 \text{ Hz} \). All signals analysed in the following are of large size and contain 8192 sample points.

For the inner defect, when occurring, the impact has to pass more transfer segments to the outer racer surface, hence the impact components are usually rather weak in the vibration signal. Thus detecting the inner race defect is somewhat difficult. The characteristic frequency of the inner race defect can be calculated by the following formula:

\[
f_i = \frac{n}{2} \left( 1 + \frac{d}{D} \cos \alpha \right) f
\]

with the geometric parameters of the rolling element bearing given above, the theoretical characteristic frequencies \( f_i \) can be calculated as 383–387 Hz.

Fig. 8 gives a set of experimental signal with inner race defect and its FFT spectrum, improved HHT spectrum and scalogram. Both the improved HHT spectrum and the scalogram have shown the impacts in the bearing excitation range, about 0.08–0.18. Obviously, due to the better time resolution, the improved HHT spectrum has shown the signatures of the impacts clearer than the scalogram. Therefore, with the improved HHT spectrum, we can mark out almost all the impacts without any trouble; on the other hand, due to the limit time resolution, the interference terms [17] have occurred in the scalogram, hence some relatively slight impacts have been masked by other heavy impacts in the scalogram, and only those heavy impacts can be found with the scalogram. Here, we have selected five pairs of distinct impacts marked with legends. The rough time intervals, which can be conversed from the number of the sample points, between two consecutive impacts are about 2.5–3.0 ms, correspondingly, the estimated characteristic frequencies are about 333.3–400 Hz. The theoretical characteristic frequencies \( f_i \) are even located among the estimated characteristic frequencies. With those analysis results, we should take care whether there are defects on the inner race.

Compared to the inner race defects, the impacts caused by the outer race defects will be stronger in the vibration signals. Therefore detecting the outer race defect is relatively easy. The characteristic frequency of the outer race defect can be calculated by the following formula:

\[
f_o = \frac{n}{2} \left( 1 - \frac{d}{D} \cos \alpha \right) f
\]

with the geometric parameters of the rolling element bearing given above, the theoretical characteristic frequencies \( f_o \) can be calculated as 263–267 Hz.

Fig. 9 gives a set of experimental signals with outer race defect and its FFT spectrum, improved HHT spectrum and scalogram. Obviously, both the improved HHT spectrum and the scalogram have shown those impacts and have denoted their occurring time clearly and accurately in the bearing excitation range, about 0.08–0.18. Just as in the inner race defect case, the improved HHT spectrum has shown those impacts more clearly. Similarly, here, we have selected five pairs of distinct impacts marked with legends whose rough time intervals can also be known, about 3.5–4.0 ms. Correspondingly, the estimated characteristic frequencies are about 250.0–285.7 Hz.
The theoretical characteristic frequencies $f_0$ are just located among the estimated characteristic frequencies. With those analysis results, we should take care whether there are defects on the outer race.

In the two cases, it has taken more than 8 min to calculate the scalogram, but no more than 2 min for the improved HHT spectrum calculation. Therefore, it can be seen that the improved HHT spectrum has better computing efficiency than the scalogram. When the signal to be analysed has too many sample points, the improved HHT can be a better choice.

5. Conclusions

In this study, the deficiencies of Hilbert–Huang transform were presented, and to overcome those deficiencies, an improved HHT was proposed. In the improved HHT, the wavelet packet transform is used as a preprocessor to decompose the original signal into a set of narrow band signals at first, and then the EMD will be utilised on those obtained narrow band signals and some IMFs will be generated, finally a simple but effective IMF selection method is used to select the useful IMFs and eliminate the undesired IMFs. Several numerical study cases were used to validate the capabilities of the improved HHT. Finally, the improved HHT’s performance in rolling bearing fault detection was compared with that of the wavelet-based scalogram through experimental case studies. The comparison results have shown that

1. The improved HHT has better resolution both in time domain and in frequency domain than the scalogram. This makes the HHT more powerful for detecting the impacts in vibration signals, especially in the dense impact conditions where the scalogram cannot usually separate those impacts successfully because of its limit time resolution.
2. The improved HHT has better computing efficiency than the scalogram, which means that the improved HHT is more suitable for large size signal analysis.
3. The HHT spectrum often has ripple phenomenon in its estimated frequency. This may mislead our analysis.

Acknowledgement

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 1100/02E).

References


