A method of generating intrinsic mode functions through a filtering algorithm based on wavelet packet decomposition

YONG-SONG KIM*, TONG-CHOL CHOE, KUM-HYOK HWANG

[1]Department of Mechanics, Kim Il Sung University,

Pyongyang, Democratic People's Republic of KOREA

* Corresponding Author Email: mech3@ryongnamsan.edu.kp (YONG-SONG KIM)

Abstract: After a consideration of the previous works on the empirical mode decomposition (EMD) method, we proposed a method to generate intrinsic mode functions (IMFs) through a filtering algorithm based on the wavelet packet decomposition (GIFWPD), and on this basis, performed Hilbert spectrum analysis. The method generates IMFs of which each frequency band has the maximum of width but is not overlapped each other. We adopted the same conditions of IMF from the EMD method as the criterion of IMFs to be generated. But, in the proposed method, the IMFs are generated by a filtering algorithm based on wavelet packet decomposition instead of the sifting process in EMD. Some numerical simulations and comparisons are demonstrated in order to establish the applicability of the method.

Keywords: empirical mode decomposition, wavelet packet decomposition, Hilbert spectrum

1 Introduction

The empirical mode decomposition (EMD) method was first proposed by Huang et al. in 1998 [1]. The researchers introduced a novel conception of intrinsic mode function (IMF) and proposed the EMD method to decompose time series data into IMFs, and on this basis, developed a new kind of time-frequency spectrum analysis method for nonlinear and non-stationary processes, namely, Hilbert-Huang transform (HHT). After Huang et al.'s first study of the EMD method and the HHT, it has been widely applied to the field of stationary and non-stationary signal processing such as system identification, damage detection, structural health monitoring and earthquake engineering etc [2-5].

Although the HHT has a great ability to extract the properties of nonlinear and non-stationary signals, however, the EMD method has still a number of problems that need further attention. First, the EMD will generate undesirable IMFs at the low-frequency region that may cause misinterpretation to the result. Second, the first obtained IMF may cover too wide frequency range (mode mixing) that the property of monocomponent cannot be achieved. Third, the EMD method cannot separate signals that contain low-energy components [6-8].

In recent years, some efforts have been made to improve these problems. Mode mixing is often a consequence of signal intermittency. Wu and Huang [9] proposed a new noise-assisted data analysis, the Ensemble EMD (EEMD) method to overcome the scale separation problem without introducing a subjective intermittence test. The true IMF components are obtained as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. While the EEMD offers great improvement over the original EMD, it also has still some problems to be resolved. In EEMD, the noise amplitude and the number of ensemble have great influence on results and the EEMD produced IMFs do not satisfy the strict definition of IMF. In addition to that, in our trials, the EEMD turned out to be a time-wasting method, because, in general, an ensemble number of a few hundred will lead to a very good result.

Zhang [10] proposed the fast filtering to decompose signal into IMFs (FFDSI) method. In this method, the fast Fourier transform (FFT) filters out IMFs from the original signal. In our trials, this method is not very effective for non-stationary signals, for example, such ones as contain a strong trend. That is because the method is based on Fourier transform, of which application is limited only to linear and stationary signals. For non-stationary signals, the resulting Fourier transform will have little physical sense. Although FFDSI method has the drawbacks, yet we consider that its idea of generating IMFs by maximum frequency band-pass filtering is worthwhile to adoptable. If wavelet packet decomposition (WPD) is used instead of FFT, it will give better results.

In this paper, we propose a method to generate IMFs through a filtering algorithm based on the wavelet packet decomposition(GIFWPD), and on this basis, perform Hilbert spectrum analysis. The method generates IMFs of which each frequency band has the maximum of width but is not overlapped each other. We adopted the same conditions of IMF from the EMD method as the criterion of IMFs to be generated. But, in the proposed method, the IMFs are generated by a filtering algorithm based on wavelet packet decomposition instead of the sifting process in EMD. The IMFs generated by the proposed method are strictly orthogonal to each other.

This paper is organized as follows: In section 2, the method to generate IMFs through a filtering algorithm based on wavelet packet decomposition(GIFWPD) was proposed and some numerical simulations are demonstrated in order to establish the applicability of the method. In section 3, comparative studies of time-frequency analysis by the short time Fourier transform (STFT), the Wigner-Ville distribution (WVD) and the proposed method are presented for a real voice signal, a typical nonstationary signal of which frequencies vary with time. Finally, we offer the conclusion in section 4. The simulation results show that the proposed method has a good ability to decompose a signal data into IMFs and can also be perfectly used for the Hilbert spectrum analysis of non-stationary signals.

2 Generating IMFs through a filtering algorithm based on wavelet packet decomposition

The wavelet packet analysis is a generalization of wavelet decomposition.

The finally obtained approximation or detail portion of wavelet packet decomposition tree is called “terminal node”.

The naturally ordered terminal node does not match exactly the order defined by the number of oscillations (i.e. the order defined by the frequency) in wavelet packet analysis. So, for a frequency ordered analysis, it is convenient to define the frequency order obtained from the natural one recursively[11,12].

At an enough decomposition level, each reconstructed terminal node of the wavelet packet decomposition tree becomes a narrow-band filtered sub-signal. The reconstructed terminal nodes corresponding to relatively higher frequency region will satisfy the IMF conditions well. If each of frequency bands of these terminal nodes is enough narrow, the sum of these two or more reconstructed nodes will also be narrow-banded sub-signal, and should still satisfy the IMF conditions. That is, it can be said that the sum of a certain number of reconstructed nodes corresponding to relatively higher frequency region will also generate an IMF. According to ways of summing up the reconstructed nodes, a variety of decomposition into IMFs can be obtained. An approach of summing up the terminal nodes to guarantee the uniqueness of the decomposition into IMFs can be established by making the frequency band of each IMF as wide as possible but non-overlapped with each other.

The below describes how to realize the above approach. First, the last terminal node (the rightmost node corresponding to the highest frequency band) is fixed as a reference node. The reconstructions of all terminal nodes starting from the first node (the leftmost node corresponding to the lowest frequency band) to the last node (i.e. reference node) are made in the time domain and all these reconstructions are summed up. Then we check if the sum satisfies the IMF conditions or not. Of course, it will not, because the firstly obtained sum is just the same as the original signal. So, if not, the sum of nodes starting from the second node (next to the first node) to the reference node is again checked for the IMF conditions. If again not, the above procedure is repeated until the sum of nodes starting from a certain node to the reference node satisfies the IMF conditions. According as the starting node from which the sum is computed moves towards the right hand, that is, towards the reference node, the corresponding frequency band of the sum of reconstructed nodes will decrease in width. If the decrease of width of frequency band reaches to a certain width, the corresponding summed signal (filtered signal) should be an IMF, because, at least, the last one (i.e. reference node) is an IMF. The above procedure is repeated to filter the second IMF out of the remaining frequency band except the higher frequency band of the first obtained IMF. And, in this way, the procedure is repeated until the non-oscillating signal appears. After finishing the whole procedure, in results, we can obtain the IMFs of which each frequency band has the maximum of width but is not overlapped each other.

The selections of decomposition level and the base function of wavelet packet decomposition are two optional points. In theory, the wavelet decomposition can be continued indefinitely. In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. In practice, the suitable number of levels is selected based on the nature of the signal, or on a suitable criterion such as entropy [11]. The selection of base function of wavelet packet decomposition has an influence on the performance of the above procedure. Based on our experience, the Daubechies wavelet, db 43 was the best selection in the implementation with MATLAB. And, for the frequency ordered analysis, we must define the frequency order because the naturally ordered terminal node does not match exactly the order defined by the frequency in wavelet packet analysis.

The whole procedure to generate IMFs by filtering based on wavelet packet decomposition can be established as the following steps:

(1) Select an enough decomposition level l and decompose the signal $x(t)$ into wavelet packet decomposition tree. If necessary, find the best tree with the criterion of Shannon entropy. By rearranging the naturally ordered terminal nodes in order of decreasing frequency, get the inversely reordered terminal nodes T_j $(j = 1,2,\cdots,N,\ \ N \le 2^l)$ with T_1 corresponding to the highest frequency band and T_N corresponding to the lowest frequency band. Reconstruct $T_j\ (j = 1,2,\cdots,N)$ to produce $u_j(t)$ $(j = 1,2,\cdots,N)$ in the time domain. Then we obtain $x(t) = \sum_{j=1}^{N} u_j(t)$.

(2) Check if $h(t) = \sum_{j=a}^{b} u_j(t)$ (for the first iteration, a and b are initialized as $a = 1$, $b = N$) satisfies the IMF conditions or not. If $h(t)$ does not satisfy the IMF conditions, set $b = b - 1$ and repeat this step (2) until $h(t)$ satisfies the IMF conditions. As discussed above, when b moves towards a (here, for the first iteration, $a = 1$), there should be at least one IMF. That is, for a certain value p_1, $h(t) = \sum_{j=1}^{p_1} u_j(t)$ should be satisfied with the conditions for an IMF.

If $h(t)$ satisfies the IMF conditions, it is designated as $c_1(t) = h(t) = \sum_{j=1}^{p_1} u_j(t)$, the first IMF of the signal $x(t)$.

(3) Separate $c_1(t)$ from $x(t)$ to get $r_1(t) = x(t) - c_1(t)$. If $r_1(t)$ is still oscillating signal, $r_1(t)$ is treated as the original data $h(t)$, and then, by resetting $a = p_1 + 1$, $b = N$ and repeating the above step (2), the second IMF $c_2(t) = \sum_{j=p_1+1}^{p_2} u_j(t)$ can be obtained. Here, resetting $a = p_1 + 1$, $b = N$ means that the second IMF will

be searched in the remaining frequency band except the higher frequency band spanned over by the first IMF.

(4) Repeat the steps (2)-(3) n times to get n IMFs.

$$\left.\begin{aligned} c_1(t) &= \sum_{j=1}^{p_1} u_j(t), \quad r_1(t) = x(t) - c_1(t) \\ c_2(t) &= \sum_{j=p_1+1}^{p_2} u_j(t), \quad r_2(t) = x(t) - \sum_{i=1}^{2} c_i(t) \\ &\cdots \quad \cdots \quad \cdots \\ c_n(t) &= \sum_{j=p_{n-1}+1}^{p_n} u_j(t), \quad r_n(t) = x(t) - \sum_{i=1}^{n} c_i(t) \end{aligned}\right\}$$

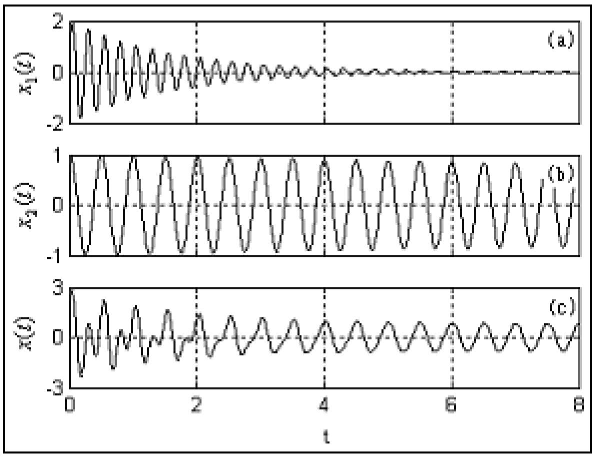
(5) If $r_n(t)$ is non-oscillating signal, terminate the loop. $r_n(t)$ is the residue term.

Performing the above proposed algorithm, we can get the IMF decomposition formula of the signal $x(t)$ as

$$\begin{aligned} x(t) &= \sum_{j=1}^{p_1} u_j(t) + \sum_{j=p_1+1}^{p_2} u_j(t) + \cdots \\ &\quad + \sum_{j=p_{n-1}+1}^{p_n} u_j(t) + \sum_{j=p_n+1}^{N} u_j(t) \\ &= \sum_{i=1}^{n} \sum_{j=p_{i-1}+1}^{p_i} u_j(t) + r_n(t) \\ &= \sum_{i=1}^{n} c_i(t) + r_n(t) \end{aligned} \tag{1}$$

Here $p_0 = 0$.

It is obvious that the obtained IMFs are orthogonal to each other due to the orthogonality of the wavelet packet base functions. For more details, consider a scalar product of arbitrarily selected two IMFs.



$$\begin{aligned} \int_T c_i(t) c_j(t) dt &= \int_T \sum_{m=p_{i-1}+1}^{p_i} u_m(t) \sum_{n=p_{j-1}+1}^{p_j} u_n(t) dt \\ &= \int_T \sum_{m=p_{i-1}+1}^{p_i} \sum_k q_{mkl} w_{ml}(t-k) \\ &\qquad \sum_{n=p_{j-1}+1}^{p_j} \sum_k q_{nkl} w_{nl}(t-k) dt \\ &= \int_T \sum_{m,n} \left(\sum_k q_{mkl} w_{ml}(t-k) \sum_k q_{nkl} w_{nl}(t-k) \right) dt \\ &= \int_T \sum_{m,n} \sum_k q_{mkl} q_{nkl} w_{ml}(t-k) w_{nl}(t-k)\, dt \end{aligned} \tag{2}$$

Here, $w_{jl}(t)$ are wavelet packet functions, q_{jkl} are wavelet packet coefficients

Eq. (2) equals to zero, for $w_{ml}(t)$ is orthogonal to $w_{nl}(t)$ for $m \neq n$. That is, $c_i(t)$ is orthogonal to $c_j(t)$ for $i \neq j$.

The uniqueness of the decomposition into Eq. (1) is guaranteed by virtue of the method itself and it obvious that the reconstructed data from the sum of all the IMFs equals to the original data.

Some numerical examples show the effectiveness of the proposed method in comparison with the previous methods. In the following examples, Daubechies wavelet, db 43 in MATLAB wavelet packet toolbox, was used as wavelet base function.

First, the proposed method is compared with the

EMD method for a signal of which some local extrema is not enough apparent to be distinguished. Fig. 1 shows the time histories of two vibration modes $x_1(t)$, $x_2(t)$ and the sum of $x_1(t)$ and $x_2(t)$ with the expressions of

$$\left.\begin{aligned} x_1(t) &= 2e^{-2\pi 0.025\cdot 4\cdot t}\sin(2\pi\sqrt{1-0.025^2}\cdot 4\cdot t+0.3) \\ x_2(t) &= e^{-2\pi 0.002\cdot 2\cdot t}\sin(2\pi\sqrt{1-0.002^2}\cdot 2\cdot t+1.5) \\ x(t) &= x_1(t)+x_2(t) \end{aligned}\right\} \tag{3}$$

The damping ratio of $x_1(t)$, 0.025, is much bigger than that of $x_2(t)$, 0.002, and thus, as shown in the Fig. 1-c, the first vibration mode is so rapidly damped off that the local extrema can not be distinguishable in the trail of the summed signal $x(t)$. The IMFs obtained by the EMD method and the proposed method are shown in Figs 2 and 3 respectively.

Fig. 1. Two vibration modes and their summed signal:
(a) first mode $x_1(t)$; (b) second mode $x_2(t)$; and (c) summed signal $x(t)$.

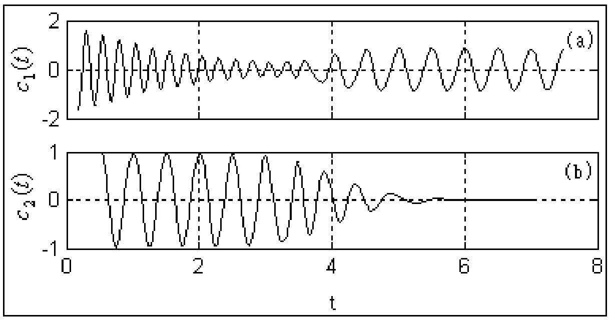


Fig. 2. The IMFs obtained by the EMD method: (a) first IMF; (b) second IMF.

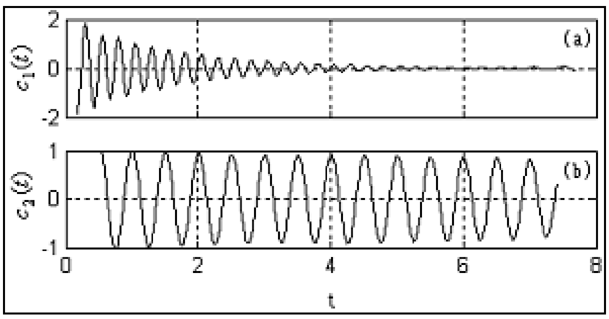


Fig. 3. The IMFs obtained by the proposed method: (a) first IMF; (b) second IMF.

The simulation result shows that the proposed method is more capable of decomposing signal into vibration modes than the EMD method. The key to success lies in the fact that the proposed method is thoroughly based on the filtering by wavelet packet analysis and, therefore, can separate more accurate and reliable vibration modes independent of whether the local extrema of the signal exist or not. As a matter of fact, the EMD method, in itself, can not be said to be an algorithm for decomposing signal into vibration modes. It is but the method for IMFs. That is, the EMD method allows the mode mixing in the decomposed IMF because it only considers the IMF conditions during the processing of signal decomposition. As can be seen in Fig. 2-a, the first IMF obtained by the EMD method is not a vibration mode, but, though allowing mode mixing, it surely is an IMF. On the other hand, the natural vibration modes of linear structure are, in general, harmonic functions with concentrated frequencies. If damping is heavy, the spectrum distribution of a vibration mode is not very ideal line spectrum but is narrow-banded one with a certain central frequency while satisfying the IMF conditions. So, the natural vibration mode can be regarded as a special IMF with a concentrated spectrum. The method proposed in this paper is more superior to the EMD method in separating the vibration modes of linear structures. That's because the method can generate the IMFs on the principle of the complete separation of the frequency bands. Nevertheless, it does not mean that the proposed method is only suitable for the vibration mode problems. In later section, we will show that the method has a good capacity of processing the non-stationary signals, too, by illustrating an example of the time-frequency spectrum analysis for a real voice sound signal. For the sake of abbreviation, the proposed method is referred to as "GIFWPD" method standing for "Generating IMFs by Filtering based on Wavelet Packet Decomposition".

Next, the proposed GIFWPD method is compared with the FFDSI method [10] for a non-stationary signal with a strong trend. Fig. 4 shows two signals $x_1(t)$, $x_2(t)$ and the sum of $x_1(t)$ and $x_2(t)$ given by the following equation:

$$\left.\begin{aligned} x_1(t) &= 3e^{-2\pi 0.008\cdot 2\cdot t}\sin(2\pi\sqrt{1-0.008^2}\cdot 2\cdot t) \\ x_2(t) &= 10\cdot t \\ x(t) &= x_1(t)+x_2(t) \end{aligned}\right\} \quad (4)$$

As seen in Fig. 5, the FFDSI method failed to separate the correct IMF and residue component. This is due to the assumption of regularity, the characteristic peculiar to the Fourier transform on which the FFDSI method is based. The result of decomposition by the GIFWPD method for the same signal is shown in Fig. 6. As can be seen in the Fig. 6, the GIFWPD method successfully separated the correct IMF and strong trend.

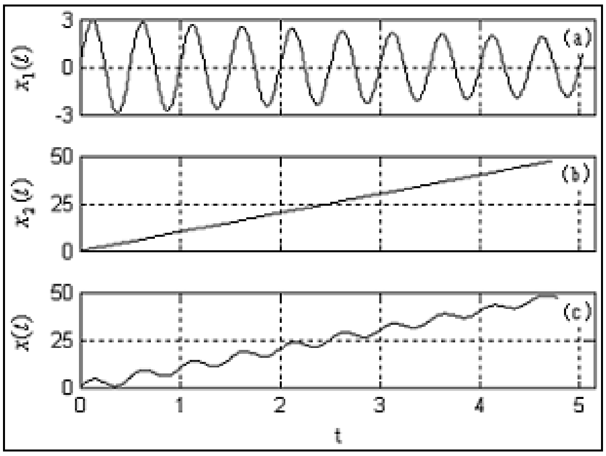


Fig. 4. Two signal components and their summed signal:

(a) damping sinusoidal signal $x_1(t)$; (b) trend signal $x_2(t)$; and (c) the summed signal $x(t)$.

Fig. 5. An IMF and residue component obtained by the FFDSI:

(a) first IMF; (b) residue component.

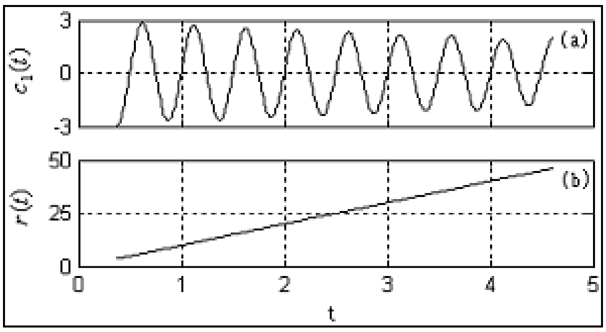
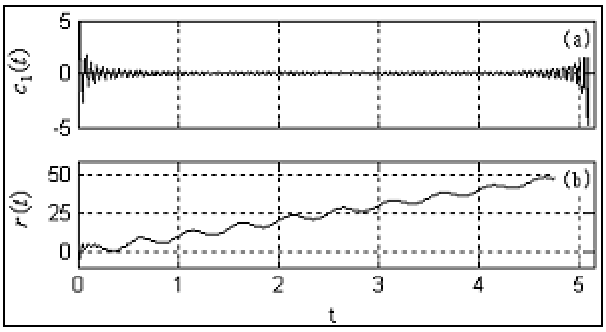


Fig. 6. An IMF and residue component obtained by



the GIFWPD method:

(a) first IMF; (b) residue component.

The tricks lie in the fact that the filtered signals through the wavelet decomposition don't have to be always stationary, and in wavelet analysis, as the scale factor increases, the temporal resolution decreases, producing a better estimate of the unknown trend. And, furthermore, in case that the trend has a polynomial form like the above, the polynomial part is suppressed in the details of the original signal and it comes into play only in the approximation potion of the signal, provided the number of vanishing moments of the wavelet (for this example, the number is 43 because db43 was used) exceeds the degree of the polynomial (for this example, the degree is 1). This means that the trend and the vibration modes are separated more effectively in wavelet analysis.

3 Numerical simulations for time-frequency analysis of non-stationary signals

The main purpose of the non-stationary signal process is to find the time-frequency laws of signals. In order to understand the spectrum varying with time, it is necessary to perform two-dimensional analysis of time and frequency of the signal.

In the same way as the case of the EMD method, the GIFWPD method can also be directly used in the computation of the Hilbert spectrum. Although the only distinction is that the IMFs obtained by the GIFWPD method, instead of the EMD method, are used in the computation of Hilbert spectrum, yet it is the first essential to the Hilbert spectrum analysis.

we will demonstrate that the GIFWPD method has a good capacity of processing the non-stationary signals by illustrating an example of the time-frequency spectrum analysis for a real voice sound signal. Comparative studies of time-frequency analysis by the short time Fourier transform (STFT), the Wigner-Ville distribution (WVD) and the GIFWPD method are presented.

A time history of voice sound of “a-i” and its Fourier spectrum are shown in Fig. 7.

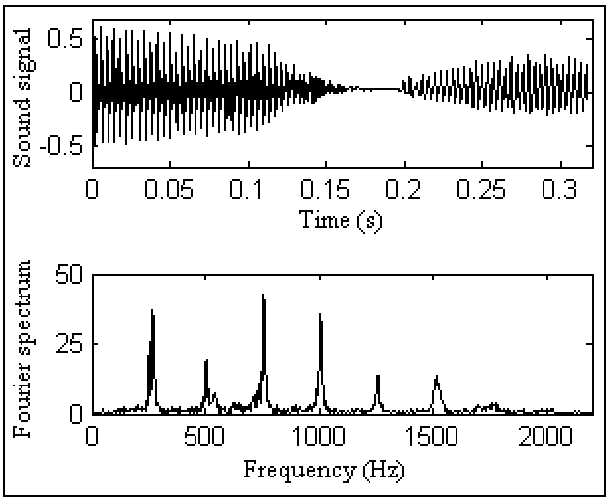


Fig. 7. Time history of real voice sound “a-i” and its Fourier spectrum.

Fig. 8 shows the results of time-frequency analysis obtained by the STFT, the WVD and the GIFWPD method. The time-frequency resolution of the spectrogram in Fig. 8-a is not so good and the Winger-Ville distribution in Fig. 8-b has the cross-terms, so, it is difficult to accurately define the real time-frequency curves.

The requirement for most time-frequency methods is to estimate the time-frequency laws as accurately as possible by making the curves more compact and clear in the time-frequency plane. In view of this requirement, the result of the Fig. 8-c gives a relatively satisfactory result. As seen in the Fig. 8-c, the Hilbert spectrum computed by the GIFWPD method provides a clear indication of how the main frequencies of the voice sound signal are change with the lapse of time while the sound of “a” is converted to the sound of “i". This result of time-frequency analysis illustrate that the computation of the Hilbert spectrum by the GIFWPD method can be a good approach for non-stationary spectrum analysis.

4 Conclusion

The proposed GIFWPD method has a good reliability of IMF decomposition even under some unfavorable conditions compared with the previous methods. The method has a good ability to decompose a vibration signal data of linear structure into the natural vibration modes. And the method can be perfectly used in the Hilbert spectrum analysis for non-stationary signals, too.

Some applications of the GIFWPD method illustrate that the method can be an effective and useful approach in stationary and non-stationary signal processing.

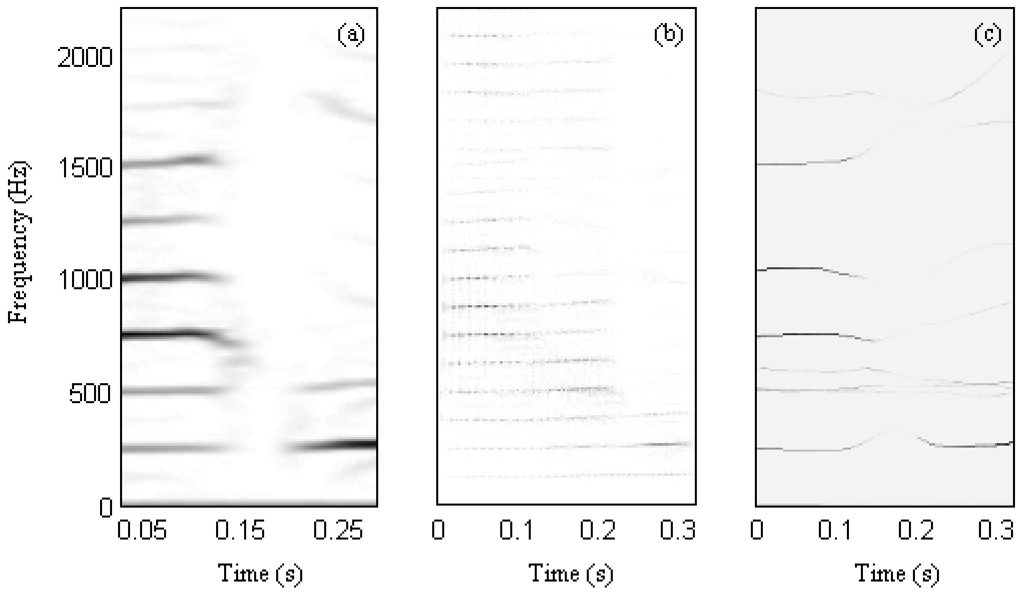


Fig. 8. Time-frequency analysis of real voice sound of “ai”:

(a) spectrogram; (b) the WVD; and (c) the Hilbert spectrum by the GIFWPD method.

References:

[1] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, Henry H. Liu, The empirical mode decomposition method and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proceedings of the Royal Society London A, 454 (1998) 903–995. doi: 10.1098/rspa.1998.0193.

[2] Q. Gao, C. Duan, H. Fan, Q. Meng, Rotating machine fault diagnosis using empirical mode decomposition. Mechanical Systems and Signal Processing 22 (2008) 1072–1081. doi: 10.1016/j.ymssp.2007.10.003.

[3] Ruqiang Yan, Robert X. Gao, Rotary machine health diagnosis based on empirical mode decomposition. ASME Journal of Vibration and Acoustics 130 (2008) 1-12. doi: 10.1115/1.2827360.

[4] Dejie Yu, Junsheng Cheng, Yu Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. Mechanical Systems and Signal Processing 19 (2005) 259-270 doi: 10.1016/S0888-3270(03)00099-2.

[5] S. Nagarajaiah, B. Basu, Output only modal identification and structural damage detection using time frequency & wavelet techniques. Earthquake Engineering and Engineering Vibration 8 (2009) 583-605. doi: 10.1007/s11803-009-9120-6.

[6] Z.K. Peng, Peter W. Tse, F.L. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis. Journal of Sound and Vibration 286 (2005) 187–205. doi: 10.1016/j.jsv.2004.10.005.

[7] Min Zheng, Fan Shen, Yuping Dou, Xiaoyan Yan, Modal identification based on Hilbert-Huang Transform of structural response with SVD preprocessing. Acta Mechanica Sinica 25 (2009) 883–888. doi: 10.1007/s10409-009-0289-8.

[8] Genda Chen, Zuocai Wang, A signal decomposition theorem with Hilbert transform and its application to narrowband time series with closely spaced frequency components. Mechanical Systems and Signal Processing 28 (2012) 258–279. doi: 10.1016/j.ymssp.2011.02.002.

[9] Zhaohua Wu, Norden E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis 1 (2009) 1-41.

[10] L.Z. Zhang, A method of signal decomposition into orthogonal IMFs. Journal of vibration and shock (in Chinese) 26(5) (2007) 27-32.

[11] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, Wavelets Toolbox User's Guide, The Mathworks, Inc., 1997.

[12] Ingrid Daubechies, Ten Lectures on Wavelets, Pennsylvania, The Society for Industrial and Applied Mathematics, 1992.