Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Sound and Vibration 294 (2006) 269–277
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

A roller bearing fault diagnosis method based on EMD energy entropy and ANN

Yang Yu*, YuDejie, Cheng Junsheng

College of Mechanical and Automotive Engineering, Hunan University, Changsha, 410082, PR China

Received 26 January 2005; received in revised form 8 September 2005; accepted 11 November 2005
Available online 27 December 2005

Abstract

According to the non-stationary characteristics of roller bearing fault vibration signals, a roller bearing fault diagnosis method based on empirical mode decomposition (EMD) energy entropy is put forward in this paper. Firstly, original acceleration vibration signals are decomposed into a finite number of stationary intrinsic mode functions (IMFs), then the concept of EMD energy entropy is proposed. The analysis results from EMD energy entropy of different vibration signals show that the energy of vibration signal will change in different frequency bands when bearing fault occurs. Therefore, to identify roller bearing fault patterns, energy feature extracted from a number of IMFs that contained the most dominant fault information could serve as input vectors of artificial neural network. The analysis results from roller bearing signals with inner-race and out-race faults show that the diagnosis approach based on neural network by using EMD to extract the energy of different frequency bands as features can identify roller bearing fault patterns accurately and effectively and is superior to that based on wavelet packet decomposition and reconstruction.

1. Introduction

While the roller bearing with faults is operating, its vibration signals will present non-stationary characteristics, and how to extract the fault characteristic information from the non-stationary vibration signals is the crux of the roller bearing fault diagnosis [1,2]. The traditional diagnosis techniques perform this from the waveforms of the fault vibration signals in the time or frequency domain, and thus construct the criterion functions to identify the working condition of roller bearings. However, because the nonlinear factors such as loads, clearance, friction, stiffness and so on have distinct influence on the vibration signals due to the complexity of the construction and working condition of roller bearings, it is very difficult to make an accurate evaluation of the working condition of roller bearings through the analysis in time or frequency domain only [2,3]. Wavelet analysis can provide the local features of the signal in both the time and frequency domain, so, it has been widely used in the roller bearing fault diagnosis [4,5]. However, the wavelet analysis is essentially an adjustable windowed Fourier transform. Due to the limitation of the length of the wavelet bases energy leakage will occur in wavelet transformation. Furthermore, once the wavelet bases and the decomposition

*Corresponding author.
E-mail address: signalp@tom.com (Y. Yu).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2005.11.002

scales are determined, the results of wavelet transform would be the signal under a certain scale, whose frequency components related only to the sample frequency but not to the signal itself. Therefore, wavelet analysis is not a self-adaptive signal processing method in nature [6,7]. Recently, a new signal analysis method, namely empirical mode decomposition (EMD, as defined in Section 1) developed by Huang et al., has been based on the local characteristic time scale of the signal and can decompose the complicated signal into a number of intrinsic mode functions (IMFs, as defined in Section 1) [8]. By analyzing each resulting IMF component that involves the local characteristic of the signal, the characteristic information of the original signal could be extracted more accurately and effectively. In addition, the frequency components involved in each IMF not only relates to sampling frequency but also changes with the signal itself; therefore, EMD is a self-adaptive signal processing method that can be applied to nonlinear and non-stationary process perfectly, and which has overcome the limitation of the Fourier transform and has high SNR as well.

In this paper, EMD is applied to the roller bearing fault diagnosis. First, the original acceleration vibration signal is decomposed by EMD and some IMF components are obtained, then the concept of EMD energy entropy is introduced, which can reflect the real work condition and the fault pattern of the roller bearing. The EMD energy entropies of different vibration signals illustrate that the energy of acceleration vibration signal in different frequency bands will change when bearing fault occurs. In order to identify the work condition of roller bearing further, in this paper, artificial neural network (ANN) [9] served as a classifier and the extracted energy features of the stationary IMFs are taken as network input vectors, and then the fault bearing and the normal bearing can be distinguished. Meanwhile, to verify the superiority of the EMD method, it is compared with the wavelet packet analysis method as well. Similarly, the original signal is decomposed by the wavelet packet, and then the energy features are extracted accordingly from the time series that are obtained after the wavelet coefficients are reconstructed. These resulting features are also used as input vectors to ANN to identify work condition of roller bearing. The experimental results show that the diagnosis approach of neural network based on EMD energy entropy has higher network identification ability.

The paper is organized as follows. Section 1 is dedicated to the EMD method. In Section 2, the concept of EMD energy entropy is proposed and the EMD energy entropies of different vibration signals are calculated, which illustrates that the energy of acceleration vibration signal in different frequency bands changes when bearing fault occurs. In Section 3, the fault diagnosis method based on EMD and ANN is given, in which energy features extracted from a number of IMFs are used as input vectors of ANN. In Section 4, the fault diagnosis method is applied to roller bearing diagnosis and is compared with the wavelet packet analysis method. The conclusion of this paper is given in Section 5.

2. EMD method

The EMD method is developed from the simple assumption that any signal consists of different simple intrinsic modes of oscillations. Each linear or nonlinear mode will have the same number of extrema and zero-crossings. There is only one extremum between successive zero-crossings. Each mode should be independent of the others. In this way, each signal could be decomposed into a number of intrinsic mode functions (IMFs), each of which must satisfy the following definition [6]:

(1) In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one.
(2) At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero.

An IMF represents a simple oscillatory mode compared with the simple harmonic function. With the definition, any signal $x(t)$ can be decomposed as follows [4]:

(1) Identify all the local extrema, then connect all the local maxima by a cubic spline line as the upper envelope.
(2) Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them.

(3) The mean of upper and low envelope value is designated as m_1, and the difference between the signal $x(t)$ and m_1 is the first component, h_1, i.e.

$$x(t) - m_1 = h_1. \tag{1}$$

Ideally, if h_1 is an IMF, then h_1 is the first component of $x(t)$.

(4) If h_1 is not an IMF, h_1 is treated as the original signal and (1)–(3) are repeated; then

$$h_1 - m_{11} = h_{11}. \tag{2}$$

After repeated sifting, i.e. up to k times, h_{1k} becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}; \tag{3}$$

then, it is designated as

$$c_1 = h_{1k}, \tag{4}$$

the first IMF component from the original data. c_1 should contain the finest scale or the shortest period component of the signal.

(5) Separating c_1 from $x(t)$, we get

$$r_1 = x(t) - c_1, \tag{5}$$

r_1 is treated as the original data, and by repeating the above processes, the second IMF component c_2 of $x(t)$ could be obtained. Let us repeat the process as described above for n times, then n-IMFs of signal $x(t)$ could be obtained. Then,

$$\left.\begin{array}{c} r_1 - c_2 = r_2 \\ \vdots \\ r_{n-1} - c_n = r_n \end{array}\right\}. \tag{6}$$

The decomposition process can be stopped when r_n becomes a monotonic function from which no more IMF can be extracted. By summing up Eqs. (5) and (6), we finally obtain

$$x(t) = \sum_{j=1}^{n} c_j + r_n. \tag{7}$$

Thus, one can achieve a decomposition of the signal into n-empirical modes, and a residue r_n, which is the mean trend of $x(t)$. The IMFs $c_1, c_2, \ldots, c_n$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of signal $x(t)$, while r_n represents the central tendency of signal $x(t)$.

Fig. 1 shows the vibration acceleration signal of the roller bearing with out-race fault. The decomposed results are given in Fig. 2, which has 22 IMFs in common but only nine of them are shown in the figure because of limited space. It can be seen from the figures that the signal is decomposed into some IMFs with different time scales by which the characteristics of the signal can be presented in different resolution ratio.

3. EMD energy entropy

While the roller bearing with different faults is operating, the corresponding resonance frequency components are produced in the vibration signals, and here the energy of fault vibration signal changes with the frequency distribution. To illustrate this change case as mentioned above, the EMD energy entropy is proposed in this paper.

If n IMFs and a residue r_n are obtained by using the EMD method to decompose the roller bearing vibration signal $x(t)$ where the energy of the n IMFs is $E_1, E_2, \ldots, E_n$, respectively; then, due to the orthogonality of the EMD decomposition, the sum of the energy of the n IMFs should be equal to the total energy of the original signal when the residue r_n is ignored. As the IMFs $c_1(t), c_2(t), \ldots, c_n(t)$ include different frequency components, $E = \{E_1, E_2, \ldots, E_n\}$, forms an energy distribution in the frequency domain of roller

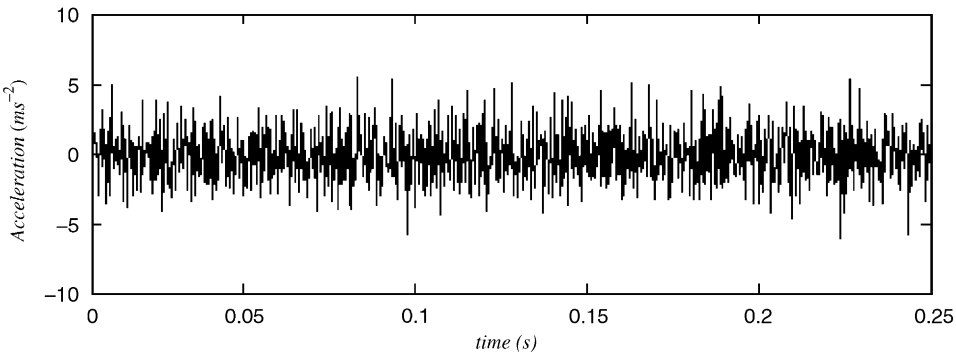

Fig. 1. The vibration acceleration signal of the roller bearing with out-race fault.

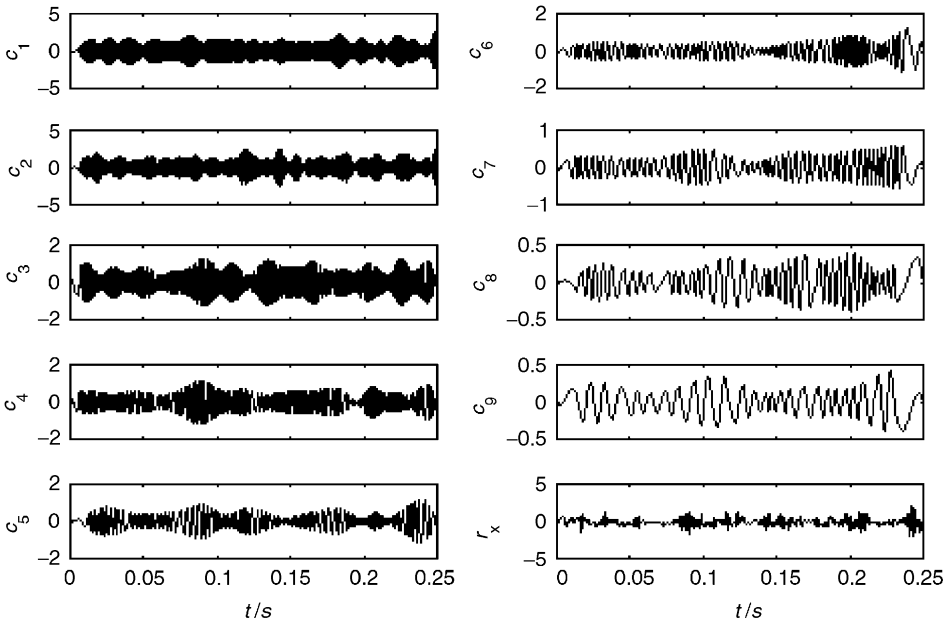

Fig. 2. The EMD decomposed results of vibration signal of the roller bearing with out-race fault.

bearing vibration signal, and then the corresponding EMD energy entropy is designated as

$$H_{EN} = -\sum_{i=1}^{n} p_i \log p_i, \tag{8}$$

where $p_i = E_i/E$ is the percent of the energy of $c_i(t)$ in the whole signal energy ($E = \sum_{i=1}^{n} E_i$).

Figs. 3, 4 and 5 show, respectively, the vibration acceleration signals of the roller bearing that is normal, with out-race fault, and with inner-race fault. If these acceleration signals are decomposed by the EMD method, the EMD energy entropies shown as Table 1 would be obtained. It can be concluded from the table that the energy entropy of the vibration signals of normal roller bearing is bigger than that of the other because the energy distribution of this kind of signals in each frequency band is comparatively even and uncertain. When the out-race fault occurs in the roller bearing, the corresponding resonance frequency components are produced, therefore, the energy entropy would reduce because the energy distributes mainly in the resonance frequency band and the distribution uncertainty is relatively less. Moreover, if the inner-race fault occurs in the roller bearing, the higher resonance frequency components are produced and the impact is

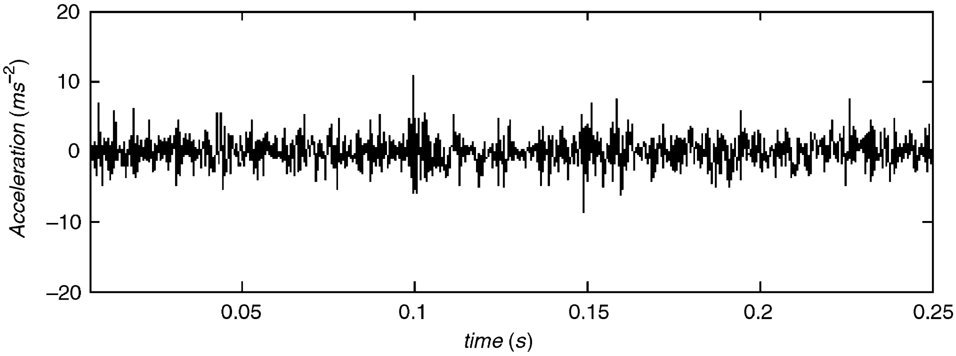

Fig. 3. The vibration acceleration signal of the normal roller bearing.

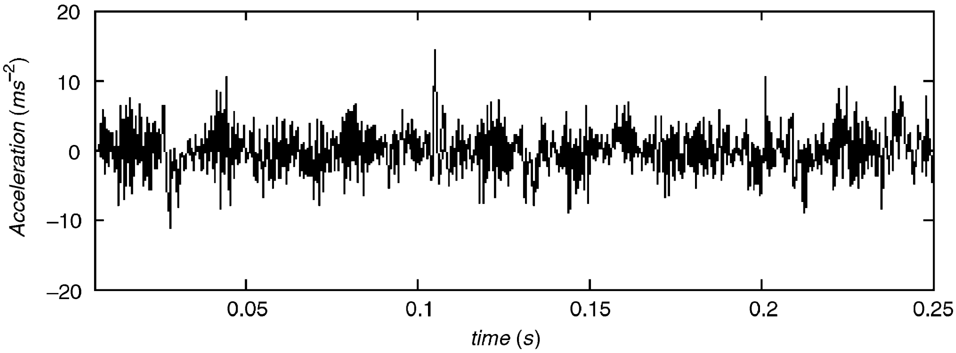

Fig. 4. The vibration acceleration signal of the roller bearing with out-race fault.

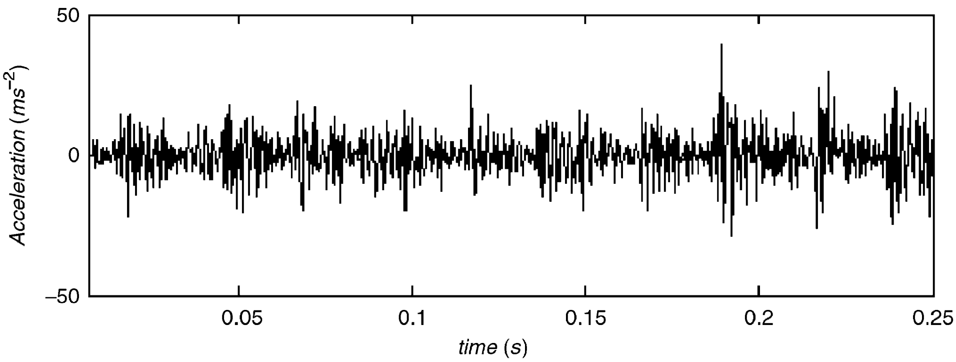

Fig. 5. The vibration acceleration signal of the roller bearing with inner-race fault.

Table 1
The EMD energy entropies of the vibration signals of the roller bearing with different fault

Normal	Out-race fault	Inner-race fault
1.9499	1.5217	1.1492

more severe, so the energy would focus on the resonance frequency band all the more and the energy entropy would be the least.

It can be seen from the above analysis that the energy entropy based on EMD can basically reflect the work condition and the fault pattern of the roller bearing. But it is not enough if we distinguish the work condition and the fault pattern only according to the EMD energy entropy; further analysis is desirable.

4. Roller bearing fault diagnosis method based on EMD and ANN

It can be seen from the above analysis that the EMD energy entropies of the vibration signal of the roller bearings with different work conditions and fault patterns are obviously different, which shows that the energy of each IMF changed when roller bearing went wrong. In this paper, it is adopted that taking the energy feature of each IMF component as the ANN input vector, the work condition and fault patterns of the roller bearing can be identified effectively. The flow chart of the roller bearing fault diagnosis method based on EMD and ANN is shown in Fig. 6.

The fault diagnosis method is given as the following:

(1) Some signals are collected as samples under the three circumstances that the roller bearing is normal, that the roller bearing has out-race faults, and that the roller bearing has inner-race faults.
(2) The original vibration signals are decomposed into some IMFs and the first m IMFs which include the most dominant fault information are chosen to extract the feature.
(3) Calculate the total energy E_i of the first m IMFs;

$$E_i = \int_{-\infty}^{+\infty} |c_i(t)|^2 \mathrm{d}t \quad (i = 1, 2, \ldots, m). \tag{9}$$

(4) Construct a feature vector T with the energy as element,

$$T = [E_1, E_2, \ldots, E_m]. \tag{10}$$

Considering that the energy is sometimes biggest, T is adjusted by normalizing the feature for the convenience of the following analysis and processing.
Let

$$E = \left(\sum_{i=1}^{m} |E_i|^2 \right)^{1/2}. \tag{11}$$

Then

$$T' = [E_1/E, E_2/E, \ldots, E_m/E]. \tag{12}$$

And the vector T' is a normalized vector.
(5) The procedure of training an ANN is carried out by utilizing the most commonly used algorithm known as back-propagation (BP). The number of nodes in the input layer is determined by the number of feature vector (T'). By trial and error method, the number of nodes in the hidden layer can be established. The output of nodes is decided by the number of the fault patterns as given below: pattern 1—normal bearing [1 0 0]; pattern 2—bearing with out-race fault [0 1 0]; pattern 3—bearing with inner-race fault [0 0 1]. After

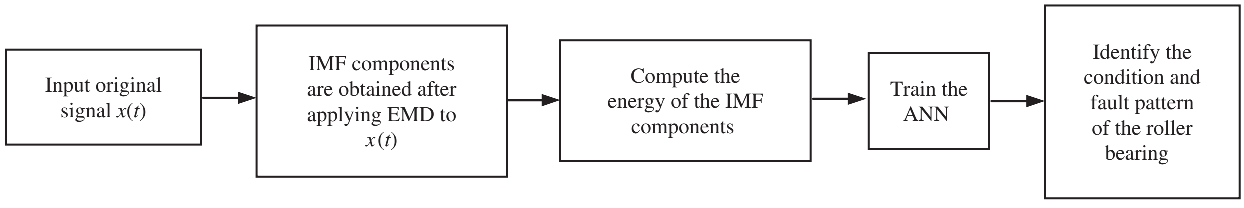

Fig. 6. The flow chart of the roller bearing fault diagnosis method based on EMD and ANN.

the ANN is successfully trained, it would be ready to test samples to identify the different work conditions and fault patterns.

5. Application

The test rig is shown in Fig. 7, which consists of a motor, a coupling, a rotor and a shaft with two 6311-type roller bearings. The test rig was used for modeling different fault types such as imbalance, misalignment and different types of bearing faults. The shaft's rotating frequency is 25 Hz and the rotor's moment of inertia is 0.03 kg m^2; after the impulse excitation experiment to the roller bearing, the first three resonance frequencies of the roller bearing are determined as 420, 732, and 1016 Hz, respectively, so the sample frequency can be taken as 4096 Hz; the vibration signal is collected by the acceleration sensor fixed on the bearing housing when the shaft's rotating frequency is steady. As the roller bearing usually works at a constant speed, the start-up and stop process are not under consideration. By laser cutting in the inner-race or out-race of the bearing the fault is introduced, and the slot width and depth is 0.15 and 0.13 mm, respectively. Due to the restriction of the experimental condition, it is not possible to introduce fault in the rollers. Roller bearings in three conditions (normal bearing, bearing with inner-race fault and bearing with out-race fault) are tested and 15 vibration signals of roller bearings in each condition are obtained in which 10 groups are drawn out at random as the sample data and the rest are taken as test data.

First, after the original vibration signals are decomposed into some IMFs by EMD, the first eight IMFs that include the most dominant fault information are chosen and arranged from high to low according to the frequency components as $c_1(t), c_2(t), \ldots, c_8(t))$; then, the fault feature vector T' are obtained according to (6), (7) and (8); finally, ANN is adopted to identify the various patterns. The ANN has three layers in which the fault feature vector T' in the three patterns respectively are taken as the ANN inputs; the hidden layer includes 18 nodes and the output are the three patterns accordingly, i.e., normal, out-race fault and inner-race fault. So the final network structure consisted of two layers: input layer, 8 nodes; hidden layer, 18 nodes; output layer, three nodes. Each pattern is trained by 10 samples and the cut-off error is 0.0001; the learning speed of the ANN training algorithm is 0.12 and the network is kept being trained till convergence. By applying ANN that has been trained to the test samples, all the test samples are identified successfully. Due to limited space, only the identification results of three test samples (corresponding to the three patterns) based on EMD preprocessing are shown in Table 2.

By applying three layers wavelet packet decomposition to the original signal with Daubechies 10(D10) wavelet base, the wavelet packet decomposition coefficients of eight frequency bands of the third layer are obtained, which are reconstructed to form a new time series. Also, they are arranged from high to low according to the frequency components as $c_1(t), c_2(t), \ldots, c_8(t)$, and then the energy of the eight reconstructed series is extracted according to Eqs. (6)–(8), which would be taken as the feature vector to train the ANN and the training process is as above. Apply the trained ANN to identify the test samples and the overall average classification rate (93%) is achieved. Due to limited space, only the identification results of three test samples (corresponding to the three patterns) based on wavelet packet preprocessing are shown in Table 2.

Although both the methods based on EMD or wavelet packet analysis as preprocessor to extract the energy in each frequency band as network input vector are accessible to identify the fault bearings, it can be seen from Table 2 that the network method based on EMD is superior to that based on wavelet packet analysis in

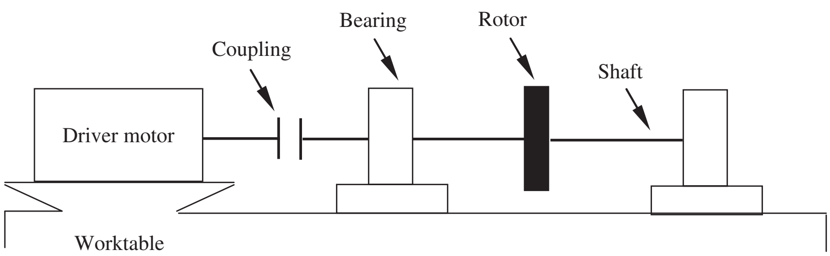

Fig. 7. Test rig.

Table 2
Roller bearing fault diagnosis results based on EMD or wavelet packet analysis

Signal	Preprocessor	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	ANN output
Normal	EMD	0.9133	0.3282	0.1730	0.1207	0.0917	0.0557	0.0360	0.0355	(0.9624,0.0412,0.0393)
	Wavelet	0.4538	0.6747	0.2059	0.2090	0.1594	0.3350	0.2302	0.2491	(0.8854,0.0291,0.2713)
Out-race fault	EMD	0.9383	0.3399	0.0568	0.0269	0.0089	0.0048	0.0029	0.0023	(0.0894,0.9405,0.0010)
	Wavelet	0.3783	0.2730	0.7733	0.4253	0.0159	0.0560	0.0066	0.0075	(0.0663,0.7591,0.4267)
Inner-race fault	EMD	0.9536	0.2485	0.1499	0.0714	0.0265	0.0203	0.0090	0.0012	(0.0038,0.0043,0.9376)
	Wavelet	0.3565	0.5836	0.4500	0.2199	0.2748	0.4416	0.0644	0.0690	(0.0784,0.3987,0.7589)

network identification ability, which is due to that the wavelet packet decomposition is not self-adaptive, that is, the frequency components after decomposition would not change with the vibration signal. Although the EMD decomposition is a self-adaptive one according to the signal itself, its decomposition process relies on the change information of signal and thus is more sensitive to the faults.

6. Conclusion

According to the non-stationary characteristics of roller bearing faulty signals, a fault diagnosis method based on EMD and ANN is put forward in this paper. First, EMD was utilized to preprocess different types of vibration signals. Then ANN was used on the preprocessed data in order to determine the work condition of roller bearing. When the work condition of roller bearing changes, the EMD energy entropy varies as well, which indicates that the energy of each frequency component changes when the roller bearing with a different fault is operating. Therefore, the energy of each IMF component is adopted as the ANN input features to identify the work condition of the roller bearing. From the theory analysis and experiment results, it can be concluded that

(1) EMD is a self-adaptive signal processing method that can be applied to nonlinear and non-stationary processes perfectly.
(2) The combination of EMD with ANN successfully identified the work condition and fault patterns of roller bearing and provided a useful tool for intelligent diagnosis of faults in roller bearing.
(3) The ANN method that took the energy of each frequency component based on EMD as the input features has higher identification ability than that based on wavelet packet analysis.

The algorithm proposed in this paper is based on the EMD method and ANN. ANN is a mature pattern reorganization method and poses mature algorithm. However, although EMD method has been used in the analysis of the non-stationary signals such as wave data, earthquake signals and structure, bridge state-monitoring signal, as a new signal process method, its application has many problems such as that the end effect need to be solved. Presently, there has been some efficient ways to restraint the end effect [10]. It is sure that after the problems in the application of EMD method are solved, the method proposed in this paper will be applied to the fault diagnosis for roller bearing more efficiently.

Acknowledgment

The support for this research under Chinese National Science Foundation Grant (No.50275050) is gratefully acknowledged.

References

[1] D. Ho, R.B. Rand, Optimization of bearing diagnostic techniques using simulated and actual bearing fault signal, *Mechanical Systems and Signal Processing* 14 (5) (2000) 763–788.

[2] Z. Peng, F. Chu, Y. He, Vibration signal analysis and feature extraction based on re-assigned wavelet scalogram, *Journal of Sound and Vibration* 253 (5) (2002) 1087–1100.

[3] C.J. Li, S.M. Wu, On-line detection of localized defects in bearings by pattern recognition analysis, *ASME Journal of Engineering for Industries* 111 (1989) 331–336.

[4] P.W. Tse, Y.H. Peng, R. Yam, Wavelet analysis and envelope detection for rolling element bearing for rolling element bearing fault diagnosis-their affectivities and flexibilities, *Journal of Vibration and Acoustic* 123 (2001) 303–310.

[5] J. Ling, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, *Journal of Sound and Vibration* 234 (1) (2000) 135–148.

[6] N.E. Huang, Z. Shen, S.R. Long, A new view of nonlinear water waves: the Hilbert spectrum, *Annual Reviews of Fluid Mechanics* 3 (1999) 417–457.

[7] H.T. Vincent, S.-L.J. Hu, Z. Hou, Damage detection using empirical mode decomposition method and a comparison with wavelet analysis.in: *Proceedings of the Second International Workshop on Structural Health Monitoring*, Stanford, 1999, pp. 891–900.

[8] N.E. Huang, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London Series A* 454 (1998) 903–995.

[9] B.A. Paya, I.I. Esat, M.N.M. Badi, Artificial neural network based fault diagnosis of rotating machinery using wavelet transforms as processor, *Mechanical Systems and Signal Processing* 11 (5) (1997) 751–765.

[10] Y. Deng, W. Wang, Boundary-processing-technique in EMD method and Hilbert transform, *Chinese Science Bulletin* 46 (11) (2001) 257–263.