Available online at www.sciencedirect.com
SCIENCE DIRECT®

Mechanical Systems and Signal Processing 19 (2005) 974–988

Mechanical Systems and Signal Processing

www.elsevier.com/locate/jnlabr/ymssp

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

Z.K. Peng[a,*], Peter W. Tse[a], F.L. Chu[b]

[a]*Smart Asset Management Laboratory, City University of Hong Kong, Tat Chee Ave, Hong Kong, PR China*
[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

*Corresponding author.
E-mail address: mepengzk@cityu.edu.hk (Z.K. Peng).

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2004.01.006

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