Wavelet Analysis And Envelope Detection For Rolling Element Bearing Fault Diagnosis – A Comparative Study

Sunil Tyagi

Center of Marine Engineering Technology
INS Shivaji, Lonavla – 410 402

ABSTRACT

Envelope Detection (ED) is traditionally always used with Fast Fourier Transform (FFT) to identify the rolling element bearing faults. The inability of FFT to detect non-stationary signals makes Wavelet Analysis (WA) an alternative for machinery fault diagnosis as WA can detect both stationary and non-stationery signals. A comparative study of ED with FFT and WA techniques for bearing fault diagnosis is presented in this report and it is shown with help of experimental results that the WA is a better fault diagnostic tool.

INTRODUCTION

Components that often fail in rolling element bearing are outer-race, inner-race and the ball. The bearing defects generate a series of impact vibrations every time a running ball passes over the surfaces of the defects. These impacts recur at Bearing Characteristic Frequencies (BCF), which are estimated based on the running speed of shaft, the geometry of the bearing, and the location of the defect **[1]**. Because the impact generated by a bearing fault distributes its energy over wide frequency range thus the BCF has relatively low energy, it is often overwhelmed by high-energy noise and vibrations generated from other macro structural components.

To allow for, the easier detection of such faults, the Envelope Detection (ED) technique has been used together with FFT **[2]**. The impact vibrations are difficult to be identified in low frequency range due to their low energy and interference, the usual practice is to view these micro-structural vibrations at the bearing resonance frequency range **[3]**. The modulated amplitudes of repetitive impacts are often excited at the bearing structural resonance frequency. Hence, the amplitude demodulation provided by ED allows the detection of localized detects **[4]**. Due to the inability of FFT to detect faults, which exhibit non-stationary impact signals, there is a need to seek for other alternatives.

Three types of analyses for non-stationary signals, the Short-time Fourier Transform (STFT) **[5]**, the Wigner-Ville Distribution (WVD) **[6]** and Wavelet Analysis (WA) **[7]** have recently been introduced. These analyses offer simultaneous representations in both time and frequency domains, which are essential for the analysis of nonstationary signals. For

transforms used for vibration-based machinery fault diagnosis, displaying of multi-resolution in time-frequency distribution diagram is an important requirement. In STFT, the resolutions in time and frequency are always constant and the WVD may lead to the emergence of cross terms, which causes misinterpretation of the signal. Hence, both STFT and WVD are deemed unsuitable. Since wavelet analysis can provide multi-resolution in both time and frequency, it is considered suitable to detect bearing faults **[8]**.

ROLLING ELEMENT BEARING FAULTS AND ENVELOPE DETECTION

Fault Related Bearing Characteristic Frequencies

The bearing characteristic (defect) frequency (BCF) depends on the geometry of the bearing, and on type of bearing defect, due to the different frequencies with which these components rotate relative to their neighboring components. The characteristic defect frequencies are defined by the equations of the form shown below:

Outer-race defects are characterized by Ball Pass Frequency Outer-race (BPFO) :

$$BPFO = \frac{n}{2} f \left(1 - \frac{d}{D} \cos \beta \right) \quad (1)$$

Inner-race defects are characterized by Ball Pass Frequency Inner-race (BPFI) :

$$BPFI = \frac{n}{2} f \left(1 + \frac{d}{D} \cos \beta \right) \quad (2)$$

Ball Pass Frequency Roller (BPFR) characterizes rolling element or ball defects:

$$BPFR = \frac{D}{d} f \left[1 - \left(\frac{d}{D} \cos \beta \right)^2 \right] \quad (3)$$

where 'f' is shaft frequency, 'n' is the number of balls, 'β' is the contact angle between inner and outer races, 'd' is the ball diameter and 'D' is the bearing pitch diameter.

Envelope Detection

Fundamental to the ED is the concept that each time a defect in a rolling element bearing makes contact under load with another surface in the bearing, an impulse is generated. This impulse is of extremely short duration compared with the interval between impulses, and its energy is distributed at a very low level over a wide range of frequencies. It is this wide distribution of energy, which makes bearing defects so difficult to detect by conventional spectrum analysis in the presence of vibration from other machine elements. Fortunately, the impact usually excites a resonance in the system at a much higher frequency than the

vibration generated by the other machine elements, with the result that some of the energy is concentrated into a narrow band near bearing resonance frequency. As a result of bearing excitation repeated burst of high frequency vibrations are produced, which is more readily detected. Take for example the bearing that is developing a crack in its outer race. Each time a ball passes over the crack, it creates a high-frequency burst of vibration, with each burst lasting for a very short time. In the simple spectra of this signal one would expect a peak at BPFO instead we get high frequency "haystack" because of excitation of bearing structural resonance. The signal produced is an amplitude-modulated signal with bearing structural resonance frequency as the carrier frequency and the modulation of amplitude is by the BCF (message signal). Envelope Detection, which the technique for amplitude demodulation is always, used to find out the repeated impulse type signals. The ED involves three main steps.

The Process On the "typical" time waveform shown in Fig 1 (a), the large wave indicates a large low-frequency component (perhaps due to misalignment, or unbalance). On top of the low-frequency component are superimposed the bursts of high frequency vibrations from the impacts due to bearing defect.

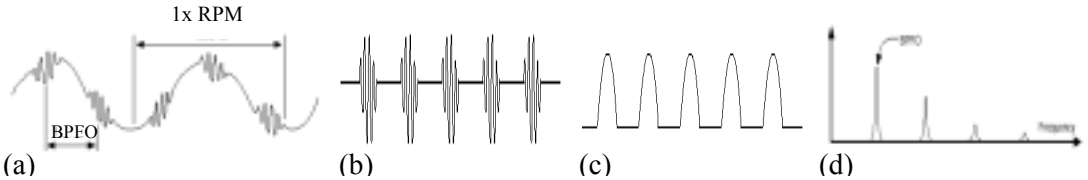

Fig. 1 : Envelope detection process (a) Unfiltered Time Signal (b) Bandpassed Time Signal (c) Envelope of Bandpased Signal (d) Envelope Spectrum

First step is to apply a band-pass filter, which removes the large low-frequency components as well as the high frequency noise only the burst of high frequency vibrations remains as shown in Fig. 1 (b). Next, we trace an "*envelope*" around the bursts in the waveform (Fig. 1 (c)) to identify the impact events as repetitions of the same fault. In third step, FFT of this *enveloped* signal is taken, to obtain a frequency spectrum. It now clearly presents the BPFO peaks (and harmonics) as shown is Fig. 1 (d).

The bearing structural resonance frequency is selected as the central frequency of the bandpass filter. Traditionally, impact tests are carried out on bearing to identify the resonant frequency. However, impact tests are not a necessity; the resonant frequency can be identified from inspection of the unfiltered signal's spectrum **[4]**. There are different ways to extract the envelope, traditionally bandpass filtering, rectifying and lowpass filtering is used to carry out the demodulation. However, Hilbert transform has also been used very effectively for ED **[9]**. In present work, Hilbert Transform is used to extract the envelope.

Hilbert transform The Hilbert Transform is a 90-degree phase shifter **[10]**. All negative frequencies of a signal get a $+90^{\circ}$ phase shift and all positive frequencies get a

–90˚ phase shift. The frequency response function of a Hilbert Transformer has the following property.

$$H(\omega) = \begin{cases} -j & \text{for } \omega > 0 \\ j & \text{for } \omega < 0 \end{cases} \tag{4}$$

Hilbert transform in Time Domain for a given a signal x(t), is defined as

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{5}$$

It is the convolution of the signal x(t) with function $1/\pi t$, which is the impulse response function of the Hilbert Transformer. The phase shifted and original signals are summed up to obtain an analytic signal $x_+(t)$ defined as follows.

$$x_+(t) = x(t) + j\,\hat{x}(t) \tag{6}$$

The Hilbert transform is typically implemented as an FIR filter so the original signal must be delayed to match the group delay of the Hilbert transform. This process can be followed by taking the absolute value of analytic function to generate the envelope V(t).

$$V(t) = \sqrt{x_+(t)^2 + x(t)^2} \tag{7}$$

WAVELET ANALYSIS

The continuous wavelet transform The continuous wavelet transform (CWT) was developed as an alternative approach to the STFT to overcome the resolution problem. The idea behind the CWT is to scale and translate the basic wavelet shape by very small steps in relation to a continuous signal and to compute the wavelet coefficient at each step **[11]**. This process of generation of CWT coefficients (transforming) is represented by the following equations:

$$W_\psi(\tau, s) = \int_{-\infty}^{+\infty} x(t) \cdot \psi^*_{\tau,s}(t)dt \tag{8}$$

Where $\psi^*_{\tau,s}(t)$ is the complex conjugate of $\psi_{\tau,s}(t)$ which is the scaled and shifted version of the transforming function, called a "mother wavelet", which defined as

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{9}$$

The transformed signal is a function of two variables, **τ** and **s**, the translation and scale parameters, respectively. The mother wavelet is a prototype for generating the other wavelet (window) functions. The scale parameter performs scaling operation on the mother

wavelet i.e. stretching and compressing the mother wavelet function, which in turn can be used to represent the signal in different frequency range. Each scale would represent a frequency band. Scale may roughly be considered as inverse of frequency as low scales represents high frequency band and high scales represents low frequencies. The term translation corresponds to time information in the transform domain; it shifts the wavelet along the time axis to capture the time information contained in the signal.

The CWT is a reversible transform. Even though the basis functions are, in general may not be orthonormal. The reconstruction is possible by implementing the following formula:

$$x(t) = \frac{1}{c_g} \int_{-\infty}^{+\infty} \int_{0}^{\infty} W_{\psi}(\tau, s) \psi_{\tau,s}(t) \frac{1}{s^2} d\tau \, ds \tag{10}$$

where C_g is a constant that depends on the wavelet used. The success of the reconstruction depends on this constant called, the admissibility constant. The admissibility constant for each type of wavelet should satisfy the admissibility condition **[12]**.

There are many types of wavelet functions available for different purposes, such as the Harr, Dabechies, Gaussian, Meyer Mexican Hat, and Morlet functions. Generally, continuous WA is preferable for vibration-based machine fault diagnosis, as the resolution is higher compared to the dyadic type of WA. In this study, the continuous type of Meyer wavelet is used.

RESULTS OF USING WA AND ED FOR BEARING FAULT DETECTION

Experimental Setup

For test purposes, Spectra Quest Machinery Fault Simulator was used to generate vibration patterns caused by a variety of bearing faults. The machine was composed of a variable speed drive, an AC motor driving a shaft rotor assembly; shafts were rested on two ball bearings, which were induced with faults. The instruments used for the experiments include a Brüel & Kjær piezoelectric accelerometer 4347, a charge amplifier (BK2626), a Stroboscope for speed calibration and a digital data recorder system of nSoft make, for storing the vibrations signals to a PC for data transfer and analysis. The ED with FFT was performed by Matlab codes using Hilbert transform and WA was performed using Continuous Wavelet Transform GUI tool of Matlab.

Experiment A variety of artificially fault induced ball-bearing type MB 204 was used. The types of faults included a defective outer-race, a defective inner-race, and a defective ball. Although the motor was set to rotate at 30 Hz, the actual rotating speed monitored by the stroboscope was found to be 28.85 Hz. The BCF was calculated using Eqs. (1) to (3), the geometric parameters of the bearings are d =0.3125”, D=1.319”, n=8

and β~2 °. The calculated BCF for each type of fault are presented in Table 1. The vibration signals were measured from the accelerometers at a sampling rate of 62.5 kHz. Table 1 also shows the time period between the impacts (inverse of BCF) and number of sample points between these impacts, which is equal to time period multiplied by sampling frequency.

Table 1: Calculated BCF for different faults

Fault Type	Outer-Race (BPFO)	Inner-Race (BPFI)	Roller (BPFR)
BCF for Shaft frequency f=28.85 Hz	86.49 Hz	144.3 Hz	107.9 Hz
Time Interval of Impacts	11.6 ms	6.93 ms	9.27 ms
Samples between impacts	723	433	579

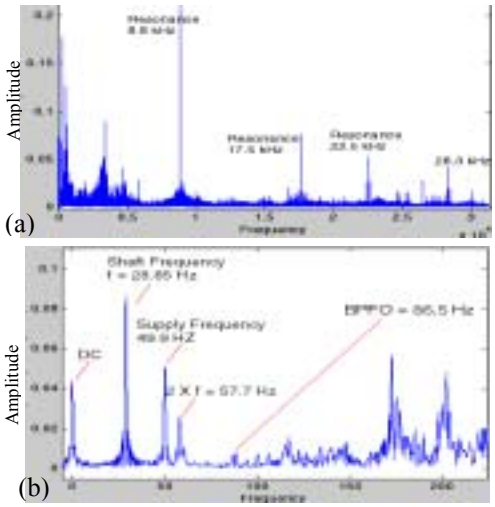

Fig. 2: (a) Complete spectrum of Bearing with outer race defect (b) Zoomed view

Results of Fault Diagnosis Using FFT with ED

Fig. 2 shows the simple spectrum of bearing with defective outer-race. Fig. 2 (a) shows the complete spectrum where as Fig 2 (b) shows the spectrum zoomed near low frequency region where the BCFs are likely to be found, the peak at BCF for outer race defect BPFO = 86.5 Hz is hardly discernible and is most likely to be missed out in presence of more noise. Further, the peaks are so small that trending of amplitude is also quite difficult, thus it is evident that simple spectrum analysis is not a suitable technique for bearing faults analysis.

To implement ED, the center frequency of bandpass filter should be selected to coincide with the center frequency of the resonance to be studied, which generally falls in the range 10 to 50 kHz **[4]**. In Fig 2(a) resonance are visible at 8.8 kHz, 17.5 kHz, 22.5 kHz and 26.3 kHz so the center frequency of the bandpass filter was set to each of these frequencies and results were analysed. It was found that best results were obtained when bandpass filter's central frequency was set to17.5 kHz.

The envelope spectra for various types of bearing defects obtained by FFT with ED are shown in Fig. 3. Fig. 3(a) shows the spectrum of the enveloped signal for bearing with outer race defect. From the spectrum the impact repetition, frequency at 86.6 Hz and its second harmonic at 173 Hz can be clearly recognized. The frequency 86.6 Hz is very close to the calculated BPFO at 86.5 Hz as listed in table 1. Hence, the defect is identified as outer-race defect. In Fig 3(b) the impact repetition frequency at 144 Hz can be recognized however; its second harmonic at 288 Hz is not very evident. As the frequency 144 Hz is very close to the calculated BPFI at 144.3 Hz, hence the defect can be identified as inner-race defect. By inspection of Fig. 3(c) ball defect can also be identified as the impact repetition frequency at 108 Hz can be recognized as well as the second harmonic at 216 Hz.

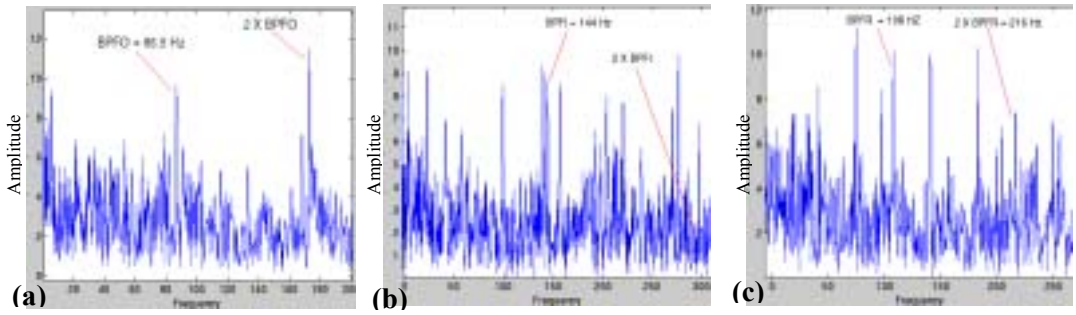

Fig. 3: (a) Envelope spectrum of bearing with outer-race defect (b) Envelope spectrum of bearing with inner-race defect (c) Envelope spectrum of bearing with ball defect

From above results, it is evident that the FFT with ED technique is definitely able to reveal the defects but the results are not consistent for all types of defects. The technique is able to give excellent results for bearings having outer race defect. However, detection of roller and inner race defect may not be easy.

Results of Fault Diagnosis Using Wavelet Analysis

In this report, are shown some typical results obtained from the vibration signals obtained at the radial direction of the bearings running at 28.85 Hz. Figs. 4 (a) to (d) shows the results of bearings running under the conditions of normal, inner-race fault, outer-race fault and roller fault respectively in the time-scale distribution diagrams generated by WA.

The top figure in each plot is the acquired signal giving amplitude Vs time (samples). The sampling time is $1.6*10^{-5}$ sec, the plots represents 8192 samples equal to 131 m sec. The lower plots are the plots of CWT coefficients plotted on a time (samples) & scale (frequency) grid, the brightness of colour indicates the amplitude at respective point on the sample - scale grid. The parameter scale as explained earlier may be considered as inverse

of frequency as low scales represents high frequency band and high scales represents low frequencies.

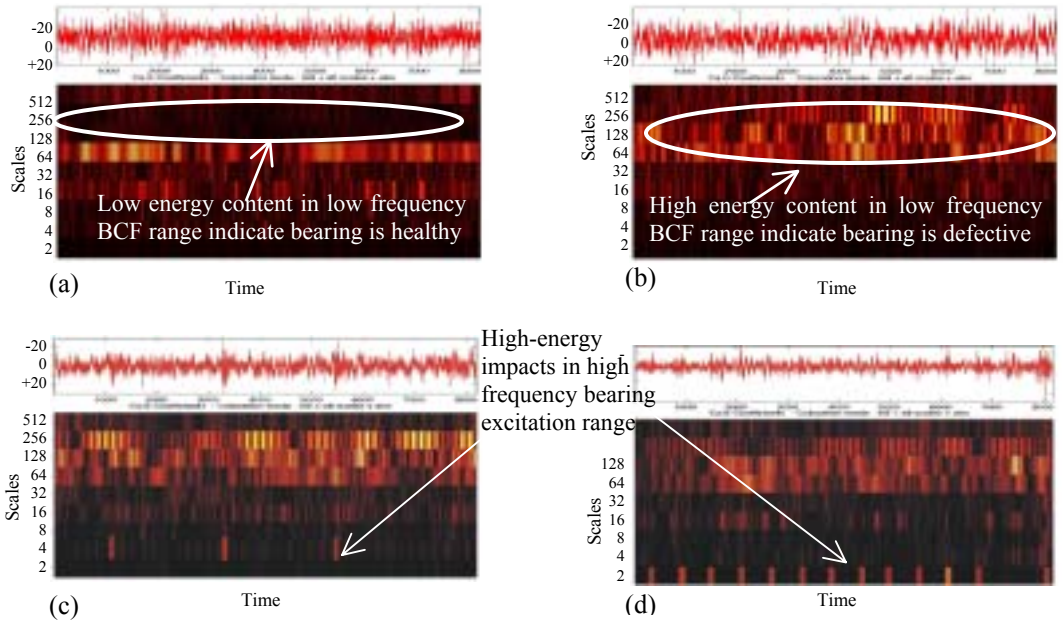

Fig. 4 : Results of wavelet analysis (a) CWT of healthy bearing (b) Bearing with defective outer-race (c) Bearing with defective inner-race (d) Bearing with ball defect

When the bearing is operating in normal condition, the energy level in the low frequency range, where most of the macro- structural vibrations are located, should be low. As shown in Fig. 4 (a), this is indicated by dark color in the low frequency range (scales 128 to 512). Meanwhile, in Fig. 4 (b), where an outer-race fault exists in the bearing, the energy levels of vibrations in the low frequency range are high as highlighted by bright colored contours in low frequency range. Further evidence to prove the occurrence of a fault could be obtained by inspecting the energy level at the high frequency range, where the bearing excitation range is located i.e. scales 6 to 2. Figs. 4 (c) and (d) show the high-energy impacts generated by the defects in the inner-race and the roller respectively. If high concentrations of energy exist in the high frequency bearing excitation range, which are highlighted by strips of bright color along with increase in energy at low frequency range, then it confirms that faults are occurring in the inspected bearing.

It is always desirable to know the cause of a fault. The length of time interval of the impacts is inversely proportional to the BCF, the cause of the fault can be determined by measuring the time interval of each impact at the high frequency bearing excitation range in the time-frequency distribution diagrams and then comparing the time intervals listed in Table I. To let the operator performs the visual inspection of the time intervals of impacts,

the results of WA are displayed as plot of CWT coefficients in scale 2 representing central frequency about 20 kHz is presented in Figs.5 (a) to (c). The scale 2 was selected since the bearing excitation range is embedded in the high frequency range.

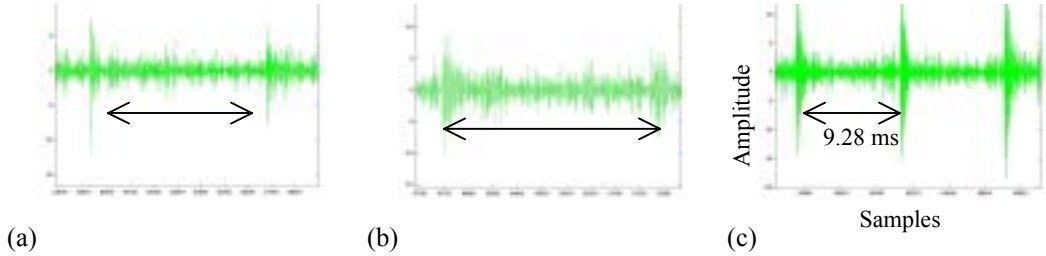

Fig. 5 : CWT coefficients at scale 2 **(a)** Bearing with outer-race defect **(b)** Bearing with inner-race defect **(c)** Bearing with ball defect

In the plots of CWT coefficients at scale 2 in Fig. 5, the impacts are clearly identifiable as well as the time period between the impacts can also be easily measured. The time intervals of impacts in Figs. 5 (a), (b) and (c) are estimated as 7.04 ms, 11.68 ms and 9.28 ms. These intervals are approximately equal to the inverse of the calculated BPFI (144.3 Hz), BPFO (86.49) and BPFR (107.9 Hz), corresponding to the defects of inner-race, outer-race and roller respectively as given in Table 1.

CONCLUSIONS

Both the methods of Wavelet analysis and FFT with ED are effective to identify outer race defects. However, the faults caused by defective inner-race and rollers are more difficult to identify by FFT with ED technique. On the other hand, when using the time-frequency distribution diagrams provided by WA, the high-energy impacts caused by inner-race and ball defects can be easily identified in the high frequency bearing excitation ranges.

In summary, to diagnose the faults WA is found to be more flexible as it gives time information. WA does provide good resolution in frequency at the low frequency range, and fine resolution in time at the high frequency range. Such a multi-resolution capability is an advantage for vibration-based machine fault diagnosis. WA is a simple visual inspection method and it does not require the analyst to have a lot of experience in Fault diagnosis.

The procedures of using WA in the fault diagnosis of rolling element bearings can be divided into two stages. In the first stage, if high-energy are observed in low frequency BCF range of the time-scale distribution diagram, then faults may be occurring in the bearing. Essential evidence to prove the existence of faults may be obtained by inspecting whether a high-energy impact is evident in low scales of time-scale distribution diagram. If

the cause of fault must be identified, then the second stage of inspecting the time interval of impacts in the high frequency bearing excitation range should be performed.

NOMENCLATURE

d	- Ball diameter.	D	- Bearing pitch diameter.
n	- Number of balls.	β	- Bearing contact angle.
$x(t)$	- Time domain signal.	$\hat{x}(t)$	- Hilbert transform of $x(t)$.
$x_+(t)$	- Analytic signal.	$V(t)$	- Envelope.
$W_\psi(\tau,s)$	- CWT coefficient.	τ	- Translation parameter.
s	- Scale parameter.	ψ	- Mother wavelet.

$\psi_{\tau,s}(t)$ - Scaled and shifted version of mother wavelet.
C_g - Admissibility constant.
$H(\omega)$ - Frequency response function of Hilbert Transformer.

REFERENCES

1. Norton M. P., Fundamentals of Noise & Vibration for Engineers, Cambridge University Press, 1989.
2. Hansen H. K., Envelope Analysis for Diagnosis of Local Faults in Rolling Element Bearing, Brüel & Kjær Application Note: BO 0501-11, Brüel & Kjær Ltd., Denmark.
3. Courrech J., Envelop Analysis for Effective Rolling – Element Fault Detection-Facts or Fiction? , Up Time Magazine, 2000.
4. McFadden P.D. and Smith J.D., The Vibration Monitoring of Rolling Element Bearing by High-Frequency Resonance Technique – A review, Technical Report of Engineering Department, Mechanics Division, Cambridge University, CUED/C-Mech/TR30, 1983.
5. Gade S. and Herlufsen H., Digital Filter Techniques vs. FFT Technique for Damping Measurements, Brüel & Kjær Technical Review No.1. 1994.
6. Ville, J., Theory et Application de la Notion de Signal Analytique, Cables et Transmissions, 20 A, 1984.
7. Newland D.E., Wavelet Analysis of Vibration Part 1: Theory, ASME Journal of Vibration and Acoustics, Oct 1994.
8. Wang W.J., McFadden P. D,, 1996, Application of Wavelets to Gearbox Vibration Signal for Fault Detection, Journal of Sound and Vibration, 192(5), 1996.
9. Proakis J.G. and Salehi M., Contemporary Communication Systems, BookWare Companion Series, 2000.
10. Proakis J.G. and Manolakis D. G., Digital Signal Processing, 3ed ed., Prentice-Hall of India Pvt. Ltd., 1997.
11. Samar, V. J., Bopardikar, A., Rao, R. and Swartz, K., Wavelet Analysis of Neuroelectric Waveforms, Brain and Language, 66, 1999.
12. Kaiser, G., A Friendly Guide to Wavelets, Birkhäuser Boston, 1997.