The fault detection and severity diagnosis of rolling element bearings using modulation signal bispectrum

Ibrahim Rehab, Xiange Tian, Fengshou Gu and Andrew Ball
University of Huddersfield
Queensgate, Huddersfield HD1 3DH, UK
Ibrahim.rehab@hud.ac.uk, u1178848@hud.ac.uk, F.Gu@hud.ac.uk

Abstract

The rolling element bearing is a key part in many mechanical equipment. The accurate and timely diagnosis of its faults is critical for predictive maintenance. Vibration signals from a defective bearing with a localized fault contain a series of impulsive responses, which result from the impacts of the defective part(s) with other elements and inevitable noise. Most researches carried out have focused on fault location identification. However, limited work has been reported for fault severity estimation, which is critical to make decision for maintenance actions. To improve current diagnostic capability,. This paper presents a new approach to detection and diagnosis of bearing fault severity based on vibration analysis using modulation signal bispectrum (MSB). It models the vibration sources from bearing defects as an impact process with constant size but three different lengths corresponding to outer race fault, inner race fault and roller fault, respectively. The results shows that MSB has a better and reliable performance in extract small changes from the faulty bearing for accurate fault detection and diagnosis for different bearing fault severity.

1. Introduction

Rolling element bearings are at the heart of almost every rotating machine. Therefore, they have received a lot of attention in the field of vibration analysis as they represent a common source of faults [(1)]. In order to keep machinery operating at its best and avoid catastrophic failure, financial cost and personal injuries, different methods bearing fault diagnosis have been developed and used effectively to detect the machine faults at an early stages, among which vibration signal processing is the most frequently applied one [(2)].

Vibration based condition monitoring have been widely used for detection and diagnosis of bearing defects for several decades [(3)]. A limited amount of work has been undertaken in using the vibration signals to investigate the potential of using the bearing faults with different fault severities. Ocak et al. [(4)] introduced a new bearing fault detection and diagnosis scheme based on hidden Markov modelling of vibration signal, the data was gathered for four different conditions (normal, inner race fault, outer race fault and ball fault) with three different severity. Ocak et al. [(5)] presents two separate algorithms for estimating the running speed and the bearing key frequencies of an induction motor using vibration data of ball bearing with inner and outer race defects with two different severity level. Xu et al. [(6)] presented different bearing fault conditions feature parameters using a modified fuzzy ARTMAP (FAM) network model based on

the feature-weight learning of inner race, outer race and ball faults using four different defect diameter. Zhang et al. [7] proposed a new method based on multi-scale entropy and adaptive neuro-fuzzy interference system conducted on electric motor bearings with three different fault categories of outer race, inner race and ball faults and four levels of fault severity. de Moura et al. [8] combined signal processing and pattern recognition techniques to diagnose three severities of outer fault as well as no-fault class was also considered. Muruganatha et al. [9] proposed a simple time series method for bearing fault feature extraction using singular spectrum analysis of the vibration signal of inner race, outer race and ball defects with four different fault size. Zhu et al. [10] proposed a new fault feature extraction method based on Intrinsic Mode Function, envelope sample entropy for rolling bearings fault diagnosis. Single point faults with four different fault diameters were introduced into test bearing with four bearing fault types which are normal, with inner race fault, with outer race fault and with ball fault. Jin et al. [11] presented a fuzzy ARTMAP (FAM) ensemble approach based on the improved Bayesian belief method and applied to the fault diagnosis of rolling element bearings with four different defect diameter and introduced into the inner race, the ball and outer race.

However, a new signal processing technique called modulation signal bispectrum analysis is used to develop pre-processing procedures for bearing condition monitoring, because of its ability to provide improved diagnostic capability compared with conventional power spectrum analysis due to its capability to suppression the noise.

2. Characteristic frequencies of bearing faults

Rolling element bearing consists of an inner race, an outer race, rolling elements and a cage, which holds the rolling elements in a given relative position, as presented in Fig. 1. In order to find the characteristics of the vibration responses due to faults, the bearing rings are assumed to be isolated continuous systems. It is further assumed that: (1) All rollers are equal in diameter; (2) There is in pure rolling contact between rollers, inner race and outer race; (3) There is no slipping between the shaft and the bearing; (4) Outer race is stationary and inner race rotates. The relative velocity between rollers, inner race and outer race are zero because they are in pure rolling contact.

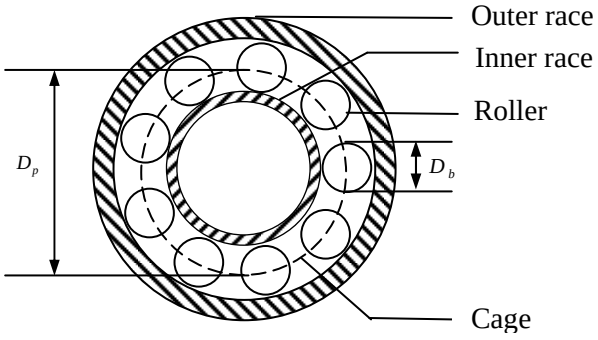

Figure 1. Rolling element bearing components

Race surface fatigue results in the appearance of spalls on the inner race, outer race or rolling elements. If one of the races has a spall, it will almost periodically impact with rolling elements. The fault signature is represented by successive impulses with a

repetition rate depending on the faulty component, geometric dimensions and the rotational speed. The period between impacts is different for all the listed elements and depends on the geometry of the bearing, the rotational speed and the load angle. For a fixed outer race bearing, the theoretical characteristic fault frequencies can be calculated using Eq. (1)-(4), and a derivation of these equations is presented in [12].
Fundamental cage frequency:

$$f_c = \frac{1}{2} F_s \left(1 - \frac{D_r}{D_c} \cos\phi\right) \quad \text{.....(1)}$$

Outer race defect frequency:

$$f_o = \frac{N_r}{2} F_s \left(1 - \frac{D_r}{D_c} \cos\phi\right) \quad \text{.....(2)}$$

Inner race defect frequency:

$$f_i = \frac{N_r}{2} F_s \left(1 + \frac{D_r}{D_c} \cos\phi\right) \quad \text{.....(3)}$$

Roller defect frequency:

$$f_r = \frac{D_c F_s}{2 D_r} \left(1 - \frac{D_r^2}{D_c^2} \cos^2\phi\right) \quad \text{.....(4)}$$

where, D_c is pitch circle diameter, D_r is roller diameter, ϕ is contact angle, N_r is number of roller and F_s is shaft rotational frequency.

In practice there is always slight sliding and slippage, especially when a bearing is under dynamic loads and with severe wear. Therefore, these frequencies may have a slight difference from calculated ones above.

3. Vibration responses to different sizes of fault

The bearing frequency equations provide a theoretical estimate of the frequencies to be expected when various defects occur on the bearing elements, based upon the assumption that an ideal impulse will be generated whenever a bearing element encounters the defect. For localised bearing faults such as spalling and pitting, sharp force impacts will be generated. These impacts will excite structural resonances and the resulting vibration will be measured by the transducer mounted externally on the machine casing [13].
However, due to the different geometry of the contact between the localised fault and the bearing component, the contact-stiffness can change because of the different geometrical properties in contact zones. On the other hand, a damaged bearing (particularly a small damage at an early stage of damage development) usually produces small amplitudes of vibration in high frequency bands due to impulsive impacts [14].

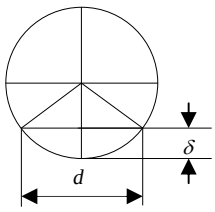

Figure 2. Schematic diagram of geometry deformation

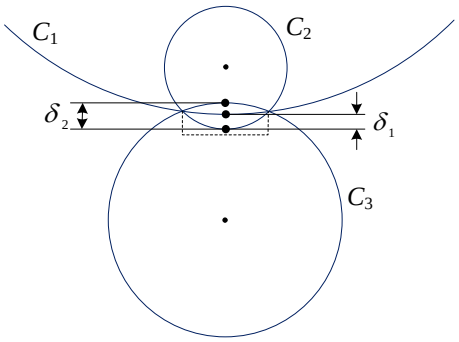

Figure 3. Geometry deformation for two kinds of contact

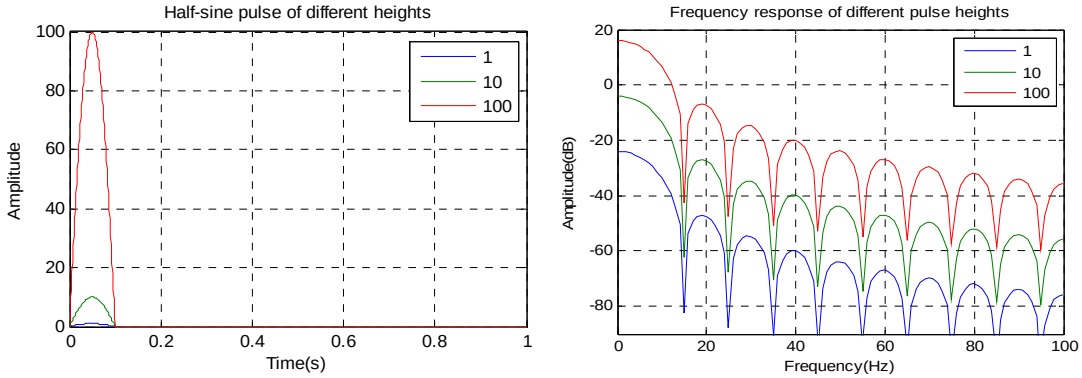

Figure 4. Frequency responses of difference pulse heights

The contact deformation is composed of geometric deformation and elastic deformation. Elastic deformation occurs along the contact surfaces of a bearing's rolling elements and raceway surfaces under loading. Geometric deformation caused by defect is related to defect location and size. While elastic deformation is related to load and defect size.

The total deformation Δ includes geometry deformation δ_g and elastic deformation δ_e.

$$\Delta = \delta_g + \delta_e \quad \text{.....(5)}$$

As shown in Fig. 2, if the width of fault is d and radius of circle is r, the chord height can be expressed as

$$crd = r - \sqrt{r^2 - \left(\frac{d}{2}\right)^2} \quad \text{.....(6)}$$

There are two kinds of contact model between bearing components as shown in Fig. 3, which are contact between a convex and a concave surface (C1 and C2) and contact between two convex surfaces (C2 and C3). If the chord height for C1, C2 and C3 are defined as $crd1$, $crd2$ and $crd3$. The geometry deformation of concave-convex contact δ_1 and concave-concave contact δ_2 are given out by Eq. (7) and Eq. (8), respectively.

$$\delta_1 = crd2 - crd1 \quad \text{.....(7)}$$

$$\delta_2 = crd2 + crd3 \quad \text{.....(8)}$$

Furthermore, considering that the fault on inner race creates concave-concave contact whereas the fault on the roller has both concave-concave and concave-convex contact. The vibration impact from the inner race defect may create the highest responses when

the sizes of the faults are the same over different races. From the relationship it is easy to understand that geometry deformations on different components have a relationship as shown in Eq. (9).

$$\delta_{go} < \delta_{gr} < \delta_{gi} \quad \text{.....(9)}$$

Based on this relationship, Fig. 4 illustrates half-sine pulses of three different heights: 1, 10 and 100 and their frequency responses. It is obvious that when the amplitude is higher, the frequency response increases. This shows that when the fault size is the same, the fault on inner race may produce the highest responses whereas the fault on the outer race will cause the lowest responses.

Taking into account the elastic deformation, it is easy to understand the impulsive differences between different fault severities. As fault degree increases, the load area will decrease while elastic deformation δ_e will increase, which will result in the growth of impulse height.

4. Signal processing based on MSB

4.1 Power spectrum (PS)

The power spectrum method is generally used to describe the power distribution of signal in the frequency domain. In the same environment, the amplitude of PS increases with the growth of fault severity. So it is widely applied for fault severity diagnosis. Usually it is calculated using Fourier transform (FT) by:

$$p(f) = E\left\langle X(f)X^*(f)\right\rangle \quad \text{.....(10)}$$

where, $X(f)$ and its conjugate $X^*(f)$ are the Fourier Transform of the sequence $x(n)$, and $E\langle.\rangle$ is the expectation operation [15].

4.2 Modulation signal bispectrum

Although PS can describe the power level of frequency components, it also contain noise in the spectrum which will affect the fault diagnosis accuracy. MSB has the capability to detect nonlinear components and suppress random noise by detecting phase coupling in modulation signal. The definition of MSB can be described by Eq. (11). [15, 16]

$$B_{MS}(f_1, f_2) = E\left\langle X(f_2 + f_1)X(f_2 - f_1)X^*(f_2)X^*(f_2)\right\rangle \quad \text{.....(11)}$$

where, $X^*(f)$ is the complex conjugate of $X(f)$ and $E\langle.\rangle$ is the statistical expectation operation. f_1, f_2 and $f_1 + f_2$ indicates three individual frequency components achieved from Fourier series integral. The magnitude and phase of MSB can be expressed as Eq. (12) and (13).

$$A_{MS}(f_1, f_2) = E\left\langle \left|X(f_2 + f_1)\right|\left|X(f_2 - f_1)\right|\left|X^*(f_2)\right|\left|X^*(f_2)\right|\right\rangle \quad \text{.....(12)}$$

$$\phi_{MS}(f_1, f_2) = \phi(f_2 + f_1) + \phi(f_2 - f_1) - |\phi(f_2)| - |\phi(f_2)| \quad \text{.....(13)}$$

It takes into account both $(f_2 - f_1)$ and $(f_2 + f_1)$ simultaneously in Eq. (11) for measuring the nonlinear effects of modulation signals. If $(f_2 - f_1)$ and $(f_2 + f_1)$ are both due to the nonlinear effect between $(f_2 - f_1)$ and $(f_2 + f_1)$, there will be a bispectral peak at bifrequency $B_{MS}(f_1, f_2)$. On the other hand, if these components are not coupled but have random distribution the magnitude of MSB will be close to nil. In this way it allows the wideband noise in bearing vibration signals to be suppressed effectively so that the discrete components can be obtained more accurately.

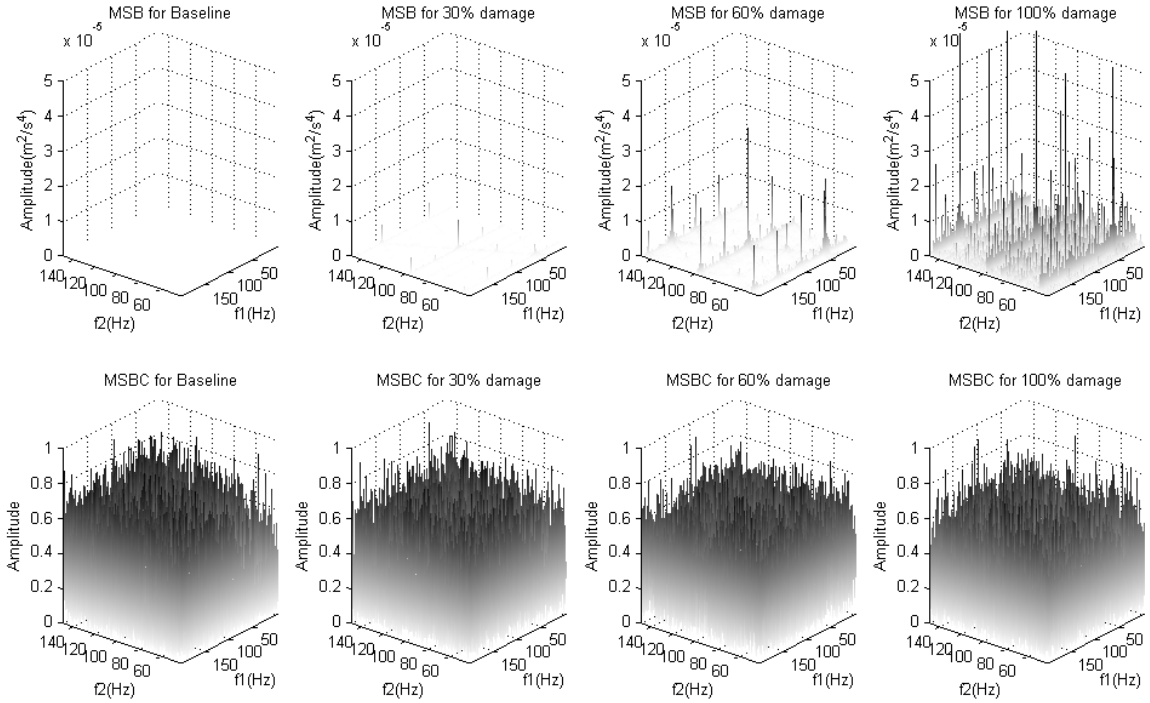

Figure 5. MSB and MSBc

To measure the degree of coupling between three components, a Modulation Signal Bicoherence (MSBc) can be used and calculated by Eq. (14).

$$b^2_{MS}(f_1, f_2) = \frac{|B_{MS}(f_1, f_2)|^2}{E\left\langle \left|X(f_2)X(f_2)X^*(f_2)X^*(f_2)\right|^2 \right\rangle E\left\langle \left|X(f_2 + f_1)X(f_2 - f_1)\right|^2 \right\rangle} \quad \text{.....(14)}$$

In addition it can be also seen that for the case of $f_1 = 0$, the MSB is degraded to power spectrum as shown in Eq. (15).

$$PS(f_2) = \sqrt{B_{MS}(0, f_2)} = \sqrt{E\left\langle X(f_2)X^*(f_2)X(f_2)X^*(f_2)\right\rangle} \quad \text{.....(15)}$$

The phase of power spectrum for any components is nil and thus it is not possible to suppress random noise by averaging operation. Therefore, MSB has better performance in suppressing random noise compared to PS.

Fig. 5 shows representative results of MSB and MSBc for the vibration signals with roller fault case under four conditions, such as baseline, 30% damage, 60% damage and 100% damage. The four graphs in the first row of Fig. 5 give out MSB magnitude results while the four graphs in the second row illustrate corresponding MSBc results. Moreover, it can be seen that there are many distinctive peaks in the roller fault frequency and its harmonics. Both the number and amplitude of peaks increase with the growth of fault severity. Especially, their nonlinear coupling can be further identified by the corresponding MSBc amplitude.

Fault diagnosis feature is obtained by extracting and averaging MSB peaks in the low frequency range according to Eq. (16).

$$A_B = \frac{1}{4k^2} \sum_{j=m-k}^{m+k} \sum_{i=n-k}^{n+k} A_{ij} \quad \text{.....(16)}$$

where (m, n) is the index of the MSB peak position, k=2 is the number of the spectral lines around peak (m, n). The peak amplitude of MSBc is also calculated in this peak area to get the degree of phase coupling.

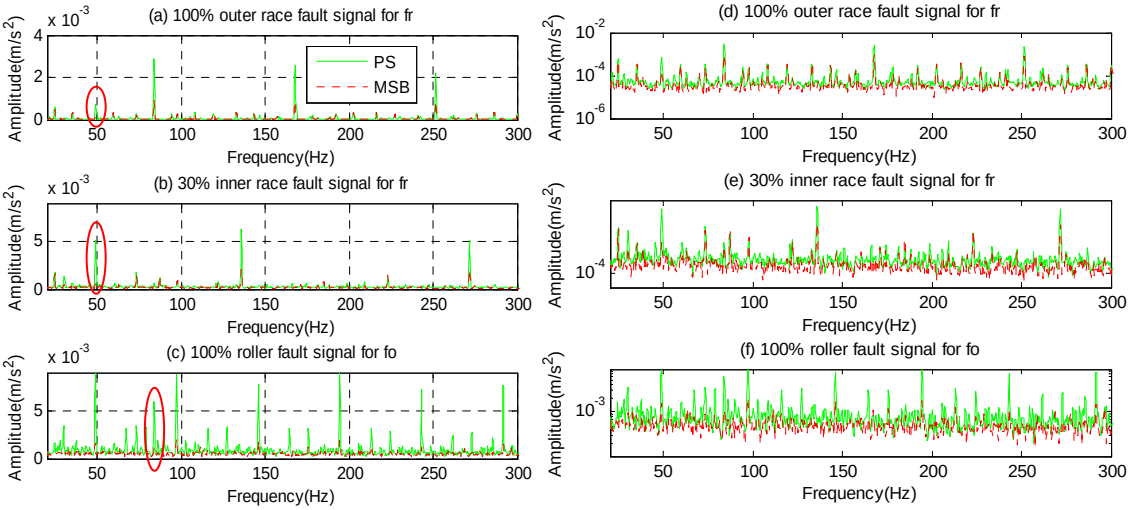

Figure 6. Denoise results of MSB

To show the noise suppress performance of MSB, three signals for 100% outer race fault signal, 30% inner race fault signal and 100% roller fault signal are examined. The spectrum amplitude should be very small at the non-fault frequencies. Fig. 6(a) shows the PS and MSB spectrum of outer race fault signal at roller fault frequency. It is obvious that the amplitude of roller fault frequency is much higher in PS than in MSB. It is contributed by the second harmonic of shaft frequency and is reduced in MSB because the frequency is not generated by roller fault and does not have nonlinear characteristics. In addition, the fault-free frequency will influence the accuracy of fault

diagnosis which will be shown in section 6 when the diagnostic results are presented. Similar conclusions can be obtained from Fig. 6(b) and Fig. 6(c). In addition, many random noises which influence diagnosis accuracy can be reduced significantly as shown in Fig 6(c). Fig. 6 (d)-(f) are the representation of (a)-(c) by plotting logarithmic y axis to make the random noise more clear. It is obvious that the random noise is reduced. Especially, in Fig. 6(f) the effect is clear and some random peaks are also suppressed significantly.

5. Test rig facility and fault simulation

The experimental data analysed in this paper was collected from the bearing test rig illustrated in Fig. 7. It is composed of motor, coupling, shaft, bearings and brake.

Bearing type is NSK N406 cylindrical roller bearing and its geometric dimensions are listed in Table 1. One Sinocera piezoelectric accelerometer is mounted on the housing of the N406 bearing vertically to measure the vibration. The frequency range of accelerometer is from 0.5 Hz to 10 kHz and the sensitivity is 8.08 mv/ms^2.

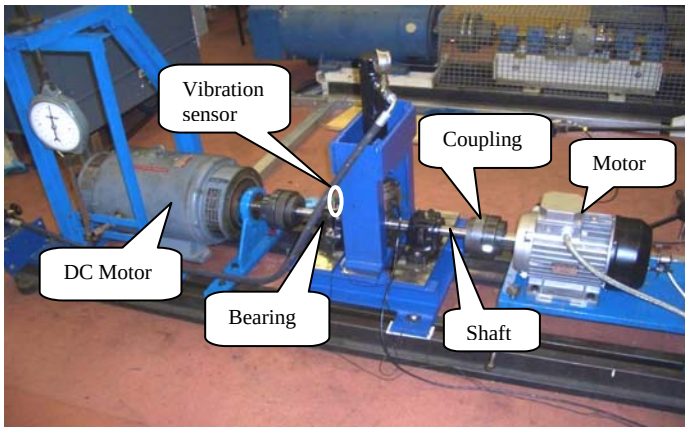

Figure 7. Photo of bearing test rig

Defect frequencies in experiment calculated according to equation (4)-(9) are listed in Table 2.

Table 1. Specification of NSK type N406 cylindrical roller bearing

Parameter	Measurement
Pitch Diameter	59 mm
Bore Diameter	30 mm
Roller Diameter	14 mm
Roller Number	9
Contact Angle	0

Table 2. Fault characteristic frequencies

Fault type	Defect frequency (Hz)
Outer race	85.8
Inner race	139.2
Roller	49.7

The experiment was carried out based on ten different rolling bearing. One bearing is healthy and taken as the baseline form comparison. Three bearings are induced outer race fault with three different fault severities, which has constant size but three different lengths, 30%, 60% and 100% of the bearing outer race width. In the same way, inner race fault and roller fault are induced to the other two groups of bearings, and each group includes three bearing for three different fault severity levels. Fig. 8 gives out the photo of defect rolling bearing with 30% roller fault, 60% inner race fault and 100% outer race fault, respectively.

(a) 30% roller fault (b) 60% inner race fault (c) 100% outer race fault

Figure 8. Photo of fault bearing

6. Signal processing results and discussion

The vibration signal processing procedure is illustrated in Fig. 9. Firstly, it calculates the envelope of vibration signal. Then, MSB analysis is carried out to enhance the nonlinear components in the envelope signal. The next step is extracting features at various bearing characterstic frequencies from MSB results as well PS results which is the MSB slice when $f_1 = 0$ as described in Eq. (15). Finally, the fault type and severity is examined using the features.

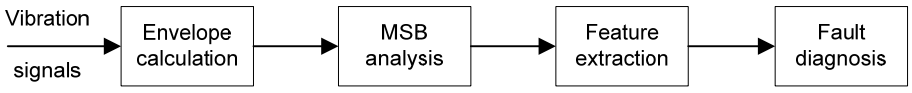

Figure 9. Signal processing procedure

It is assumed that the fault type is unknown in a signal measured. Therefore, feature values of each signal are calculated at all three potential fault frequencies predicted by Eq. (1) – (3) as outer race f_o, inner race f_i and roller f_r. Then the fault type can be diagnosed by comparing between these three values and from a predefined threshold. Fig. 10 shows the results of both PS and MSB for all the tested signals. It can be seen in Fig. 10 (a), (b) and (c) of that the feature amplitudes for nearly all the cases are distinctively higher for the fault cases at the corresponding fault frequencies, compared with that fault-free cases, allowing the fault types to be determined straightforward. However, there are several cases which may have uncertainties in the determining of the fault types based on PS results. As shown in Fig. 10(a), for 100% outer race fault the amplitude at f_r is also clearly higher than the baseline, which may lead to a diagnostic result that there is a defect on the roller. Similarly, the high amplitude at f_r for 30% inner race fault case in Fig. 10(b) and the high amplitude at f_o for 100% roller fault case in Fig. 10(c) can give rise to incorrect diagnosis of fault types. However, from Fig.

11, it can be seen that the frequencies mentioned above have lower bicoherence amplitudes in comparison with its own baseline, which means that the nonlinear effects are lower and hence confirm there should not be faulty on the components. In the meantime, it shows that random noise and non-relevant interferences are reduced effectively by MSB analysis. As a result, MSB analysis produces a result that has no confusion in differentiating different types of faults as shown in Fig. 10 (d) (e) and (f). In addition, MSB also show a smaller difference between the two cases of higher severities for the inner race fault, showing that MSB can produce better severity estimation.

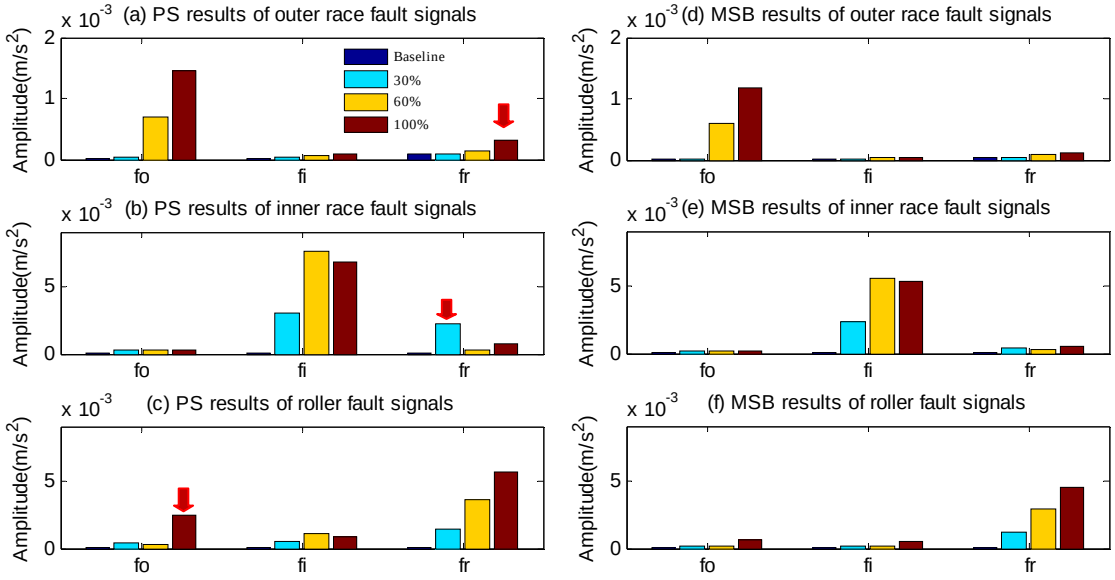

Figure 10. Comparing fault diagnosis results of PS and MSB

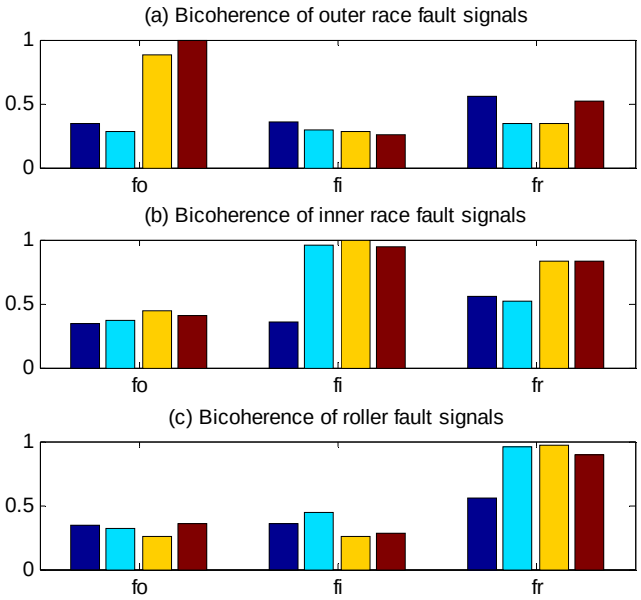

Figure 11. MSBc of vibration signals

From the results and discussion above, it can be concluded that MSB analysis have better performance than PS in both fault type diagnosis and severity estimation.

7. Conclusion

In this paper, a new method has been developed by combining MSB technique with envelope analysis for rolling bearing fault detection and diagnosis. MSB is effective in reducing random noise and interference components of shaft frequency, which is similar with the roller fault frequency, by detecting the phase coupling of sidebands in vibration signal. Therefore, MSB improve the signal to noise ratio (SNR) for obtaining a more reliable detection and diagnosis results. The effectiveness of the proposed method has been evaluated based on experimental data sets from three type of faults and three levels of damage severities. The experimental data analysis results show that MSB has better noise reduction performance compared with PS results. Moreover, not only the types of fault can be identified correctly but also the severity can be estimated more accurately.

References

1. McFadden, P. and J. Smith, Model for the vibration produced by a single point defect in a rolling element bearing. Journal of Sound and Vibration, 1984. 96(1): p. 69-82.
2. Lei, Y., Z. He, and Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery. Expert Systems with Applications, 2008. 35(4): p. 1593-1600.
3. Tandon, N. and A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology International, 1999. 32(8): p. 469-480.
4. Ocak, H. and K.A. Loparo. A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals. in Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on. 2001.
5. Ocak, H. and K.A. Loparo, Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data. Mechanical Systems and Signal Processing, 2004. 18(3): p. 515-533.
6. Xu, Z., et al., Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing. Expert Systems with Applications, 2009. 36(6): p. 9961-9968.
7. Zhang, L., et al., Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. Expert Systems with Applications, 2010. 37(8): p. 6077-6085.
8. de Moura, E.P., et al., Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. Mechanical Systems and Signal Processing, 2011. 25(5): p. 1765-1772.
9. Muruganatham, B., et al., Roller element bearing fault diagnosis using singular spectrum analysis. Mechanical Systems and Signal Processing, 2013. 35(1–2): p. 150-166.
10. Zhu, K., X. Song, and D. Xue, Fault Diagnosis of Rolling Bearings Based on IMF Envelope Sample Entropy and Support Vector Machine. Journal of Information & Computational Science 2013: p. 5189-5198.

11. Jin, M., et al., Reliable fault diagnosis method using ensemble fuzzy ARTMAP based on improved Bayesian belief method. Neurocomputing, 2014. 133(0): p. 309-316.
12. Tomasz Barszcz and Nader Sawalhi, “Fault Detection Enhancement in Rolling Element Bearings Using the Minimum Entropy Deconvolution”, Archives of Acoustics, vol. 37, 2012, pp.131-141.
13. T. Igarashi and H. Hamada, “Studies on the vibration and sound of defective rolling bearings (First Report: Vibration of Ball Bearings with One Defect)”, Bulletin of the Japan Society of Mechanical Engineers, vol. 25, 1982, pp.621-822.
14. N. Tandon and A. Choudtury, “An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect”, Journal of Sound and Vibration, vol. 205(3), 1997, pp. 275-292.
15. Alwodai, A., et al., A Study of Motor Bearing Fault Diagnosis using Modulation Signal Bispectrum Analysis of Motor Current Signals. Journal of Signal and Information Processing, 2013. 4: p. 72.
16. Gu, F., et al., Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment. Mechanical Systems and Signal Processing, 2011. 25(1): p. 360-372.