Effects of limiting the bandwidth of the vibration signal on bearing fault detection and diagnosis using state of the art techniques

Nader Sawalhi[1], Robert B. Randall[1]

[1] School of Mechanical and Manufacturing Engineering,
The University of New South Wales,
Sydney 2052, Australia

Abstract

This paper discusses the effect of limiting the bandwidth of the vibration signal on detecting a bearing fault using novel processing algorithms. When the bandwidth of a signal is limited, these algorithms fail to extract the symptoms of the fault. This is due to the fact that the developed algorithms are based on maximizing the impulsiveness of the signal. If the bandwidth of the signal is limited, only one or two harmonics of the impulse train remain and the result will be close to sinusoidal, thus non-impulsive and no longer detectable using envelope analysis. Genuine discrete frequency components can be separated by using separate time synchronous averaging (TSA) adjusted for every harmonic family in the signal instead of using the discrete/random separation technique (DRS). The analysis shows that it can be possible to locate bearing defect frequencies, and modulation sidebands, by inspecting the power spectrum density (PSD) of the signal, but the results are not so definitive.

Keywords: Rolling element bearings, Envelope analysis, Bandwidth, Spectral kurtosis, Minimum entropy deconvolution, Discrete random separation, Time synchronous averaging, Power spectrum density.

Introduction

Vibration signals of a defective bearing with a localised fault contain a series of impulse responses, which results from the impacts of the defective part(s) with other elements. These impulses are generated almost periodically and their characteristics depend on the location of the defect; that is, whether it is on the inner race, outer race or rolling elements. In practice, the spacings between the impulses vary randomly to a certain extent due to varying load angle and slip which leads to the smearing of the defect harmonics at higher frequencies (defect frequencies will appear as discrete components in the low frequency region but will be smeared in the high frequency region). In the vicinity of a resonance (high frequency region), this information can be extracted when no random fluctuation exist, but is impossible with a small amount of random fluctuation; as the harmonics smear into one other [1]. This problem has been solved by frequency analysing the envelope of the response signal (envelope analysis or high frequency resonance technique (HFRT)) obtained by amplitude demodulation [1]. This enveloping is usually applied to a frequency region where the signal-to-noise ratio (SNR) is the highest, for example around a structural resonance frequency excited by the bearing fault (previously best determined by spectrum comparison). In order to benefit from the HFRT and other available techniques, which build on extracting the hidden impacts in the system, the signal must be acquired at a reasonable sampling frequency, high enough to include excited resonances away from the interaction and dominance of other components. If the sampling frequency (determining the bandwidth of the signal) is limited, the essence of using the HFRT technique will be violated and it will no longer be of use to extract the defective bearing signature.

In this work, which has been inspired by an imposed limitation on the bandwidth of vibration signals taken from gas turbine engines, we discuss the effect of restricting the bandwidth of an acquired signal on the effectiveness of detecting a bearing fault using a range of novel signal processing techniques. The discussion is based on the analysis of a signal with a known inner race fault taken from a high-speed test rig and sampled at 50 kHz. The novel signal processing algorithms employed in the analysis include discrete/random separation (DRS), minimum entropy deconvolution (MED), spectral kurtosis (SK) and envelope analysis. These algorithms are tested at the high sampling frequency of the signal (50kHz) and again on a decimated version of the signal (with new sampling frequency 6.25 kHz). The use of power spectrum density (PSD) as an alternative diagnosis tool to locate the fault is discussed and illustrated.

A brief description of the novel techniques

Following is a summary of a set of new analysis techniques [2] and their relevance to detecting faults in rolling element bearings.

Removal of Discrete Frequency Components

The slightly random nature of bearing signals, caused by slip, allows them to be separated from gear signals, and other discrete frequency components, which may mask the weak bearing signals. Three effective ways of doing this are SANC (self adaptive noise cancellation) [1], DRS (Discrete/Random Separation) [3] and linear prediction based on AR modelling [4].

SANC relies on the fact that the correlation length of the deterministic components is much greater than that of the bearing signals, so when an adaptive filter receives a delayed version of the signal it recognises only the deterministic part and separates it from the random part (the bearing signal).

DRS is a more efficient way of achieving the same thing by generating the linear transfer function (using FFT methods) between the signal and a delayed version. In principle, this produces a value of 1 where there are discrete frequencies and zero otherwise. The amplitude of this transfer function is then used to filter the signal (again using efficient FFT methods) separating out the discrete frequency components and giving the random part by subtraction. In order to generate the linear transfer function and implement the DRS, the analyst has to select a suitable time delay and decide on the filter length. The minimal filter length should be inversely proportional to the minimal frequency spacing between two discrete components in the power spectrum, which are to be enhanced independently. The minimal time delay must be longer than the autocorrelation length (memory) of the random signal to be filtered out.

A final method which is sometimes useful, as it simultaneously whitens the spectrum of the bearing signal, making it more impulsive, is linear prediction. This is based on autoregressive (AR) modelling, and predicts the deterministic continuation of the signal from previous samples, and leaves the random (and impulsive) part by subtraction.

MED (Minimum Entropy Deconvolution)

This method was developed in the 80's for sharpening reflected pulses in seismic and underwater signals, by removing the effect of the transfer path from the source to the receiver transducer. It generates an inverse filter which maximises the kurtosis of the filter output (corresponding to the input of the original transmission path). It was applied to gear signals in ref [5] and then to bearing signals [6].

SK (Spectral Kurtosis)

In recent years it has been shown [7,8] that SK provides an excellent solution to the problem of choosing a suitable band for bandpass filtration and demodulation for envelope analysis, without requiring historical data. In fact, it has been shown to produce almost identical results to the dB spectrum difference. The SK is obtained by first performing a time/frequency analysis, and then calculating the kurtosis in the time direction for each frequency line. Antoni showed [7,9], that the optimum combination of centre frequency and bandwidth could be found using the "kurtogram". In [4] it was shown that complex Morlet wavelets could alternatively be used for the time/frequency analysis, the filters then corresponding better to second order responses with constant damping factor, and defined a "wavelet kurtogram" based on this.

Experimental work

This case is taken from a bearing test facility (FAG-Germany), which runs at around 11760 rpm (196 Hz) and is belt driven by a motor. The running speed of the bearings is usually constant. The bearings under test are single row angular contact bearings (FAG 7205) with an angle of contact of 25^o, a ball diameter of 8.0 mm and a pitch diameter of 38.5 mm. The number of rolling elements is 12 and the ball pass frequency of the inner race (BPFI) for the bearing under test at the running speed is estimated at 1397.5 Hz. When performing the measurements, an accelerometer was attached to the outer test bearing using a magnetic base (because there was no access to the signal internal to the machine). Vibration signals were collected using a National Instruments card (NI PX1-4472), which was programmed to capture a data length of four seconds at a sampling rate of 50 kHz. A spall, which had developed on the inner race in one of the test bearings (after 52 hours), is shown in Fig.1.

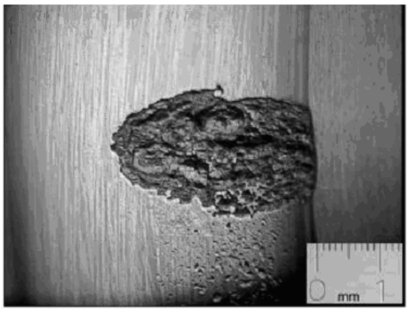

Fig 1: spall defect size

Signal analysis at sampling frequency of 50 kHz

This is a case where the spacing of the individual pulses from the faulty bearing is shorter than the length of the impulse response (exacerbated by the relatively low resonance frequency of the magnet mounting). Thus the MED technique was applied to separate the individual impulses. The wavelet kurtograms for the signals before and after the application of the MED technique (Fig. 2) show that the SK has been increased from 2.5 to 20 (kurtosis values calculated for the envelope signals), making the impulsiveness apparent.

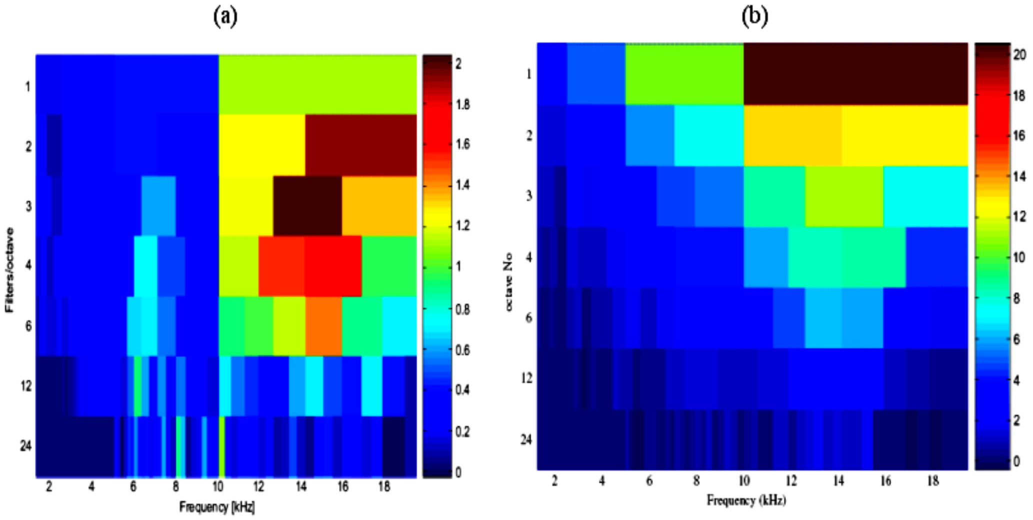

Fig 2: Wavelet kurtograms [6] (a) Before application of MED (b) After MED

Fig. 3 shows the time signals after various stages of the signal processing. Linear prediction in (3.b) increases the impulsivity slightly, but the kurtosis is still only 1.25. SK filtering in (3.c) (using the filter characteristics of Fig. 2.a) lifted the kurtosis value to 3.08. However, the separation of impulses is not complete and the kurtosis value is still very low compared with the result of such a fault if it were in a low speed machine. After application of MED in (3.d) the kurtosis increases dramatically to 38.5, showing that the extended impulse responses have been considerably reduced. Further filtering by the filter determined in Fig 2.b does not greatly change the kurtosis, but makes the pattern of response pulses clearer, and more typical of an inner race fault. This is confirmed very clearly by the envelope spectrum of Fig. 4, which shows a typical pattern for an inner race fault, with harmonics of BPFI, and also sidebands spaced at shaft speed.

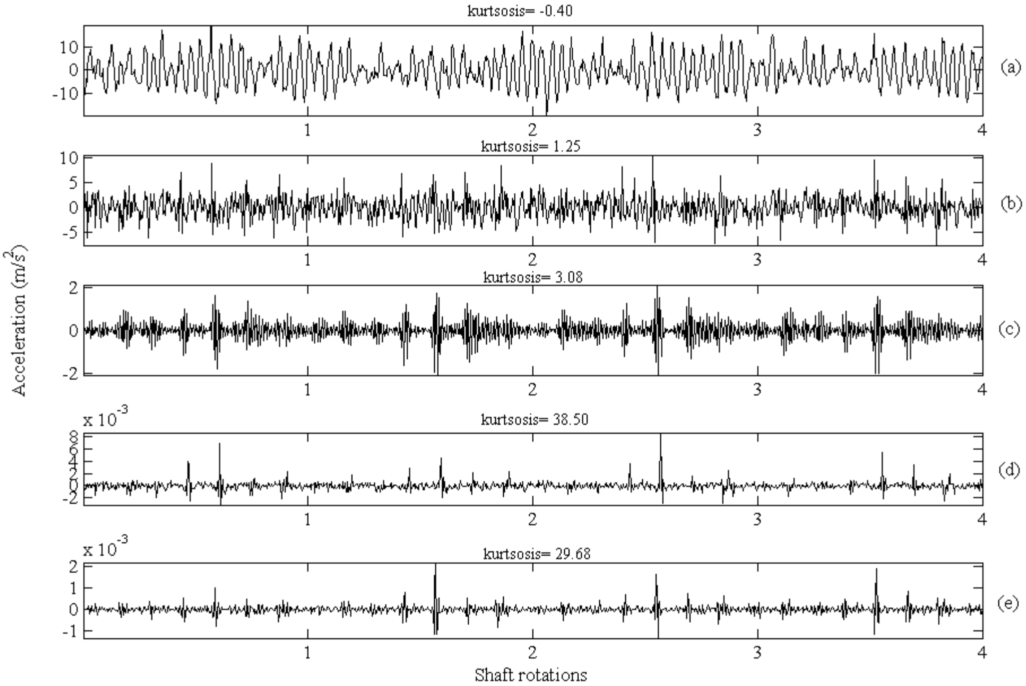

Fig 3: (a) (a) Raw signal (b) Residual of the AR (29) linear prediction filter (c) signal b filtered using SK (d) signal b filtered using MED (e) signal d filtered using SK [6]

To further demonstrate the improved capability of surveillance using SK combined with MED, the kurtosis values were trended against the development of fault size and the results are presented in Fig. 5.

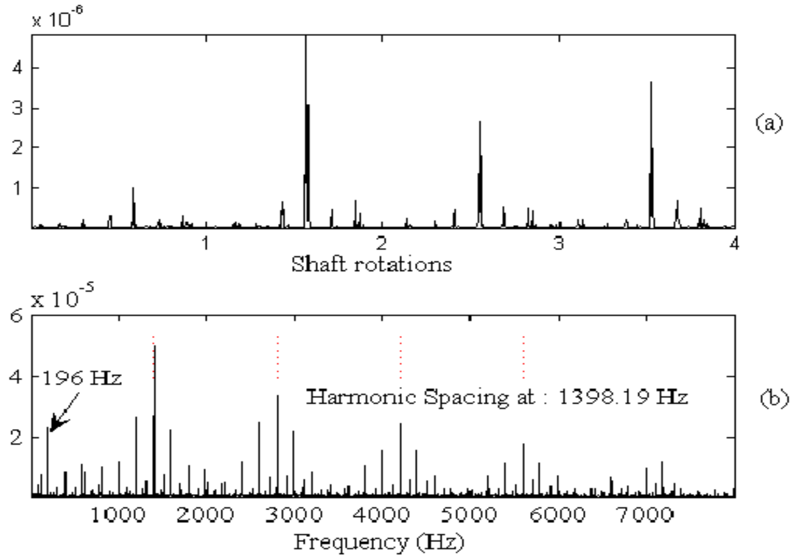

Fig 4: (a) Squared Envelope and (b) squared envelope spectrum of the signal obtained using SK optimum filter after being pre-whitened and filtered using MED [6]

The introduction of the MED filtering as a processing step, not only increased the kurtosis values and clarified the impulses, but also made it well correlated with the fault size (constant low kurtosis values when MED is not used) as illustrated in Fig. 5. The results presented indicate the importance of combining the SK and the MED techniques in order to detect faults at an early stage, and show the possibility of trending the fault development using kurtosis values of the filtered signals.

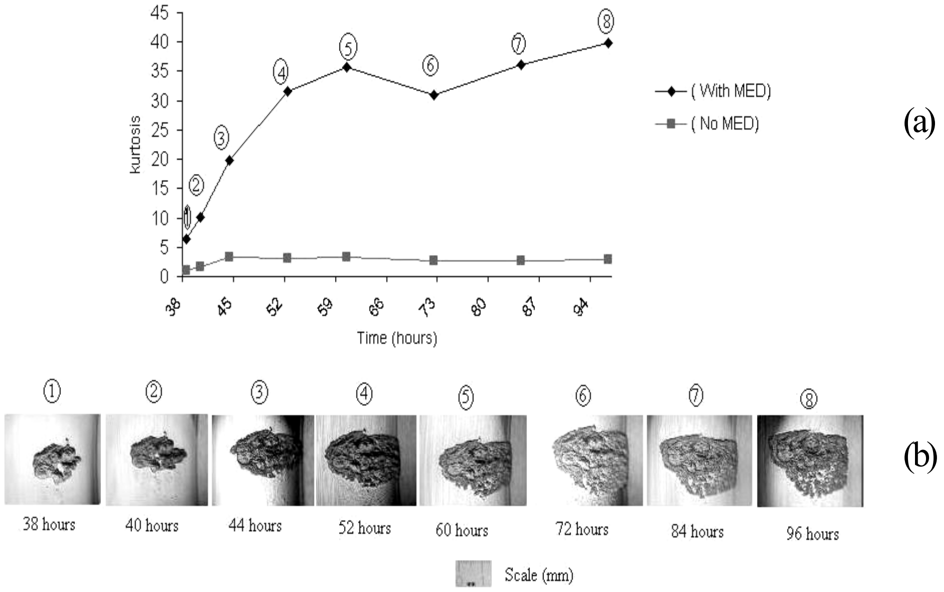

Fig 5: (a) Trending using the kurtosis of filtered signals with and without MED (b) Fault sizes [6]

Signal analysis at sampling frequency of 6.25 kHz

The same processing steps have been applied to a decimated version of the signal (sampled down to 6.25 kHz). The wavelet kurtogram for the decimated signal is shown in Fig. 6. This has very low kurtosis values (no impulsiveness detected) and is in agreement with the values of kurtosis in the 0-3 kHz of Fig 2.b.

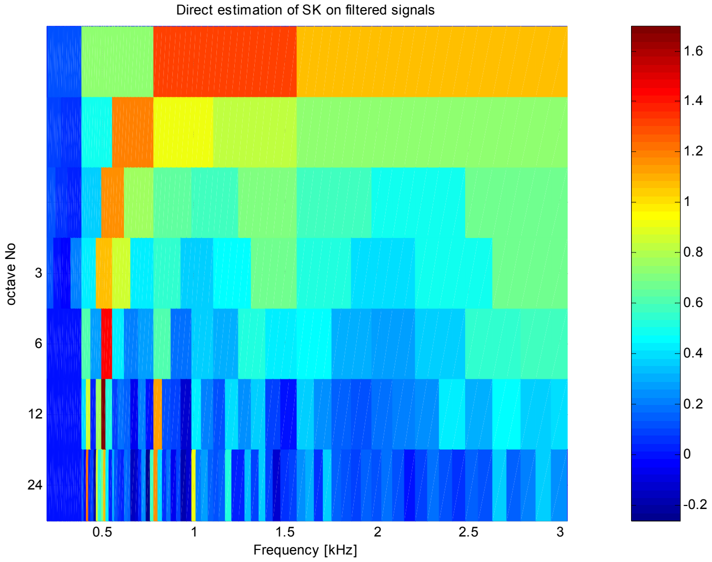

Figure 6: Wavelet kurtogram of the decimated signal

The filtered signal using the filter that maximizes the kurtosis fails to extract the BPFI and gives a trivial result as shown in Fig. 7. This is because of the limited frequency range of the signal, which means that only one or two harmonics of the impulse train may remain. To generate a periodic train of impulses requires 8-10 harmonics, all aligned in phase, so if all but one or two are filtered away, the result will be close to a sinusoidal signal, and thus non-impulsive and unlikely to be detectable using envelope analysis

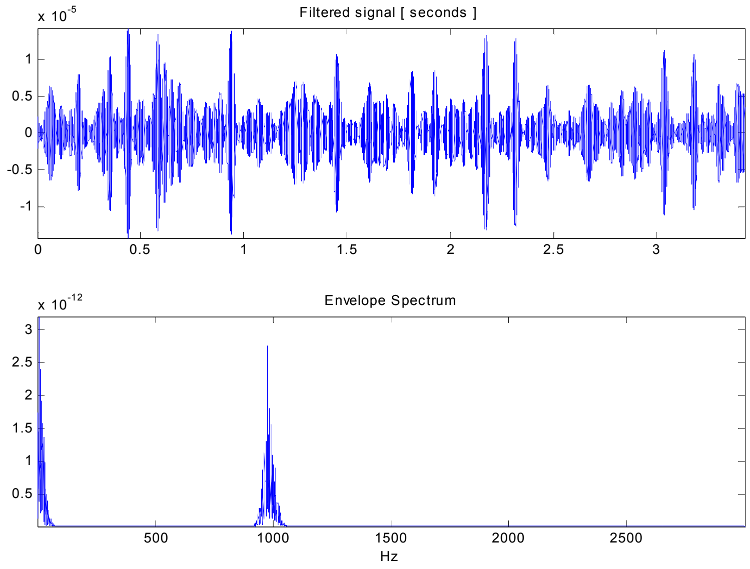

Figure 7: (a) Squared Envelope and (b) squared envelope spectrum of the decimated signal obtained using SK optimum filter after being pre-whitened and filtered using MED

Power spectrum density (PSD) as a means of locating the defect frequency

Both previous sections show clearly that by limiting the bandwidth of a signal, the applicability of a range of extremely useful techniques for detecting faults in rolling element bearings is greatly affected. We show in this section that in this low frequency region, power spectrum density (PSD) does give the possibility of distinguishing a good and faulty bearing. However, although this is shown to be possible by the use of PSD, this needs a great amount of time and effort to inspect the content of the PSD signal (as opposed to only inspecting the squared envelope spectrum of a filtered signal) and the possibility of trending the fault development is greatly reduced. Note that the term PSD is used even though it does not apply strictly to discrete frequency components, except by invoking delta functions. It is shown in the following discussions, that for this signal the BPFI is not necessarily the dominant harmonic (2nd harmonic is clearer) and that the sidebands spaced at the shaft speed are stronger than the BPFI itself. This adds to the complexity of using the PSD on its own to locate the BPFI, especially if the frequency range is limited so that it only encompasses the first harmonic of the BPFI. As the fault severity increases, there is a better chance of locating the fault as its harmonics and sidebands increase noticeably. However, the trending of the fault is much more difficult than using the SK as in Fig. 5 and presumably varies from case to case, because the individual frequency components are greatly influenced by a variable transfer function, very sensitive to speed changes. Finally, it is shown that unlike its use on a signal with a high sampling frequency, the discrete/random analysis technique (DRS) removes or minimizes the harmonics of the BPFI and their sidebands, because the low harmonics are not sufficiently smeared to be treated as random.

Locating the BPFI harmonics and modulation in the PSD

Fig. 8 shows the PSD of the decimated signal and the harmonics of the shaft speed (195 Hz). In Fig. 9, it is shown that the BPFI (1398.53) and its second harmonic can be located in the PSD, with the second harmonic being the strongest. Note however that the 2nd harmonic is above the lowpass filter cutoff frequency (2.5 kHz) for this sampling frequency. Figs. 10 and 11 show the modulation of the BPFI and its second harmonic at the shaft speed. The sidebands are stronger and more prominent than the carriers, which could easily distract the analyst.

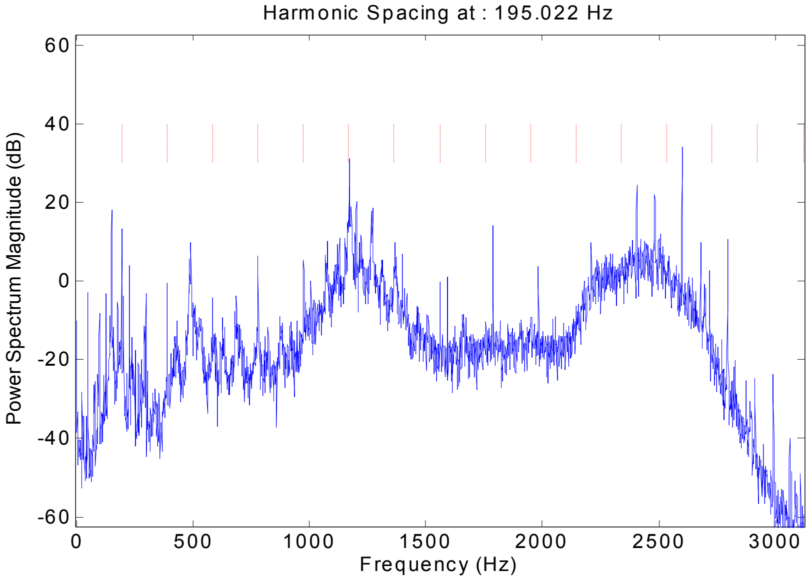

Figure 8: PSD of the decimated signal and the harmonics of the shaft speed

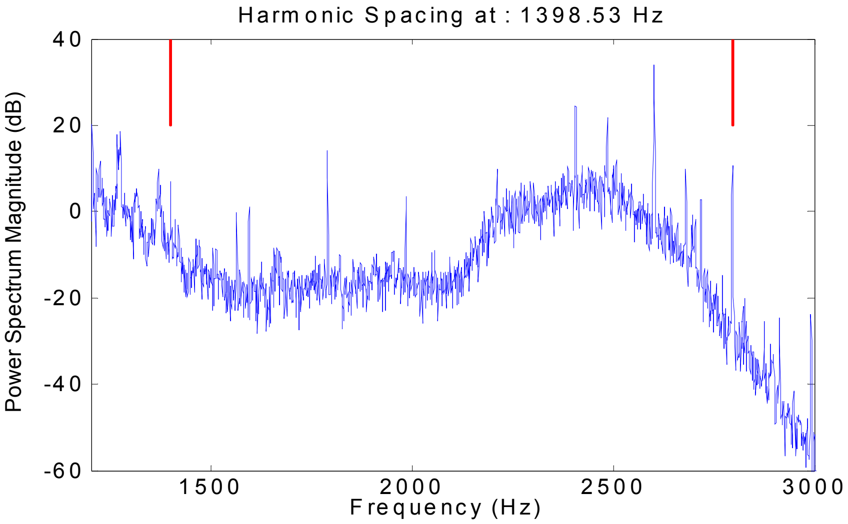

Figure 9: A zoom-in showing the 1st and 2nd harmonic of the BPFI (Note the strength of the 2nd harmonic)

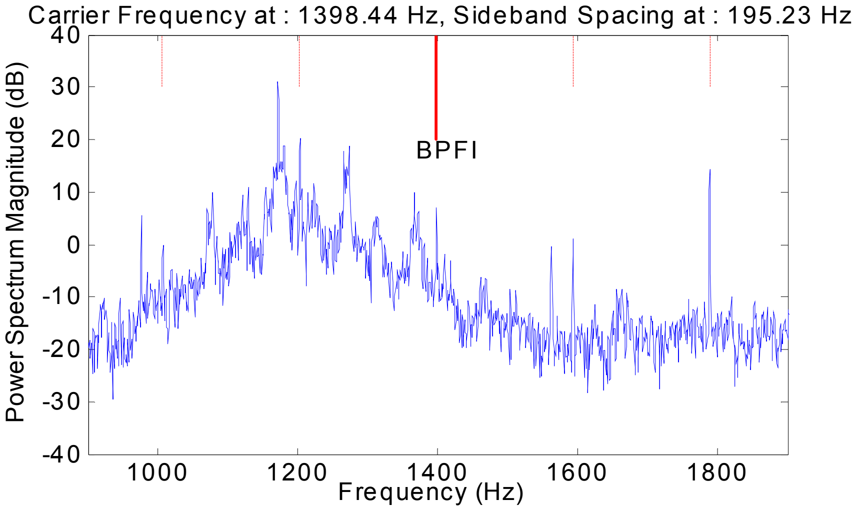

Figure 10: The strong modulation of the BPFI at the speed of the shaft (Note that the sidebands in general are more prominent)

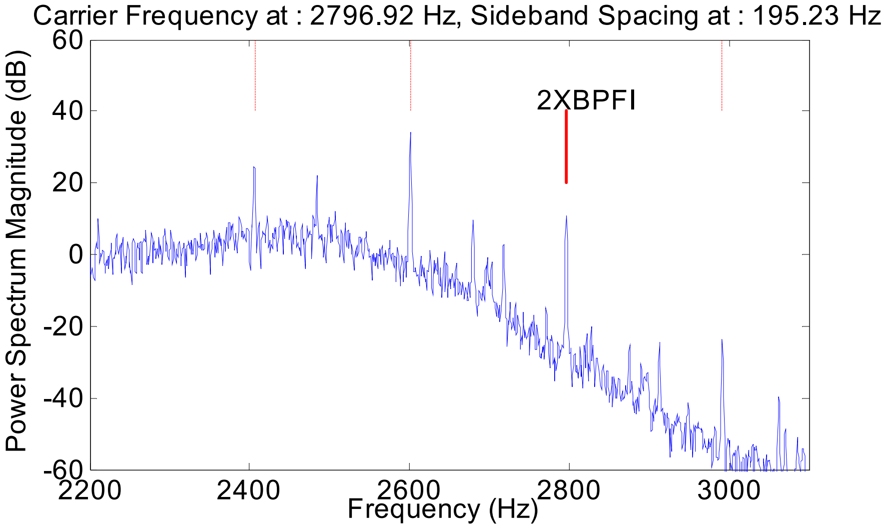

Figure 11: The strong modulation of the second BPFI harmonic at the speed of the shaft (Note the sidebands in general are more prominent)

Fault development effect on the BPFI harmonics and sidebands

Fig. 12 shows the effect of increasing the fault severity on the harmonics of the BPFI and their sidebands. The PSD shown in Fig. 12.b is for a signal taken 50 hours later than the one shown in Fig. 12.a. Note that the increase in fault severity has a major effect on the first harmonic of the BPFI (around 15 dB increase), while the second harmonic of the BPFI increased by 3 dB. Sidebands show a similar trend, with all increasing noticeably.

The conclusions drawn from Fig. 12 suggest that locating the fault at its early stage could be difficult if the 1st harmonic of the BPFI is the only harmonic in the frequency range. This might be very small or non-existent in the PSD at the early stages of the fault. It depends entirely on what resonances may be excited at that frequency.

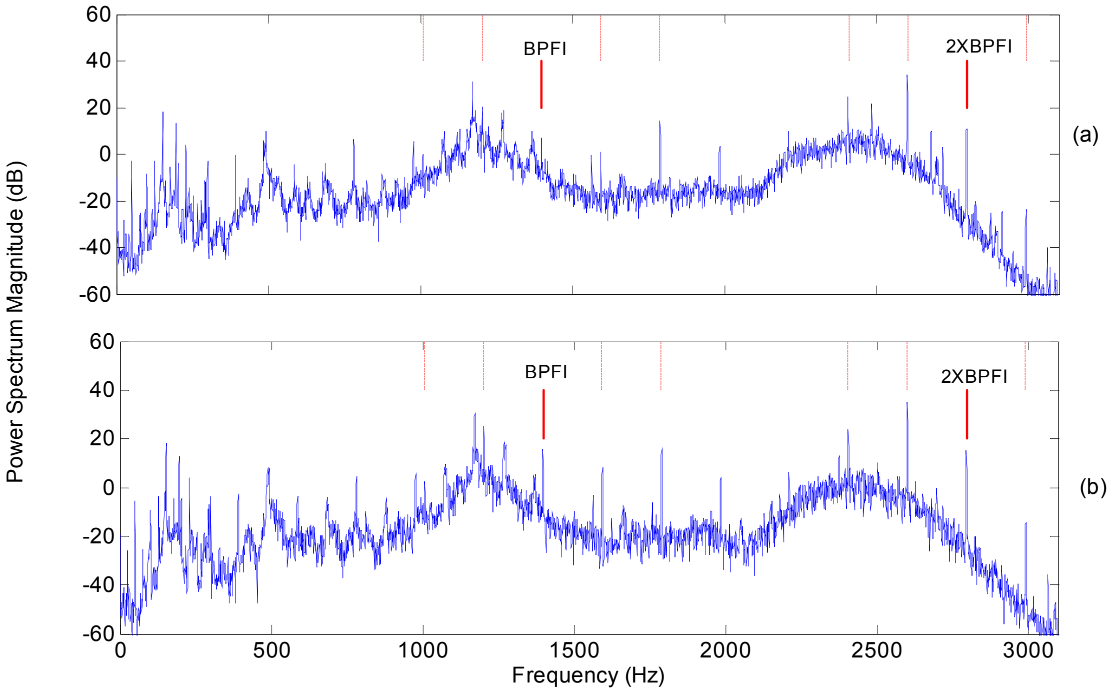

Figure 12: (a) PSD of the decimated signal with the fault size pictured in Fig. 1 (b) PSD of the decimated signal 50 hours after having the fault size pictured in Fig. 1

DRS effect on the BPFI harmonics and their sidebands

DRS has been used to separate discrete and random signals and has proved of great use when separating gear and bearing frequencies as the former give discrete frequencies in the spectrum, while the latter will smear at higher frequencies (due to slippage). By applying DRS and then using envelope analysis to demodulate a certain band, the ability to detect a bearing fault and separate it from shaft related harmonics increases dramatically. When it comes to using the DRS technique on a signal with limited frequency range, where bearing related frequencies appear more like discrete frequencies, it is expected that this technique will have a negative effect by removing or minimizing the low order bearing related harmonics and side bands. In fact this is what can be seen from Figs. 13-15.

Fig. 13.a shows the PSD of the raw vibration signal, the harmonics of the BPFI and the sidebands at the shaft speed. The PSD of the DRS residual (discrete frequencies removed) in Fig.13.b shows the removal of the BPFI harmonics and sidebands. The zoomed plots of Figs. 14 and 15 show that very clearly.

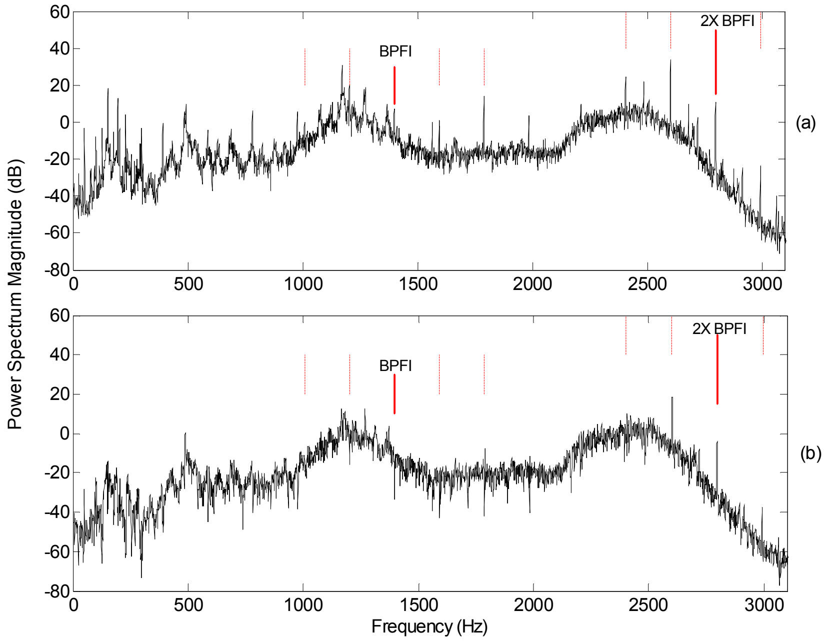

Figure 13: (a) PSD of the raw vibration signal (b) PSD of the residual of DRS

Genuine discrete frequency components can still be separated by using separate time synchronous averaging (TSA) adjusted for every harmonic family in the signal (typically two or three for a gas turbine engine). A separation algorithm (gear/bearing signal separation), which is based on re-sampling the signal to have an integer number of samples per revolution for a specific shaft [10,11] would be ultimately recommended to completely remove all shaft harmonics without much disruption of the vibration signal.

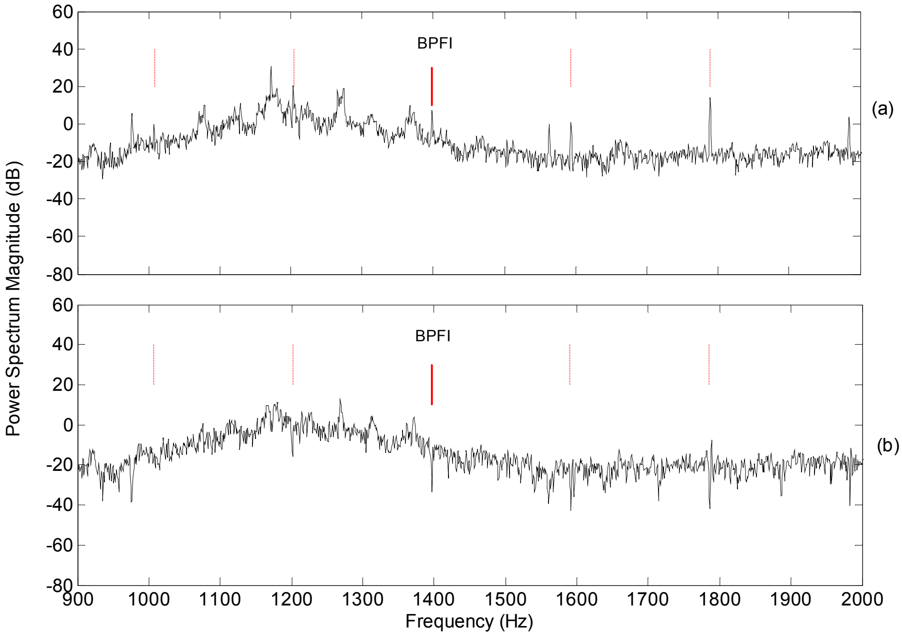

Figure 14: The effect of removing discrete frequencies using DRS on the BPFI and its sidebands. (a) Raw (b) DRS residual

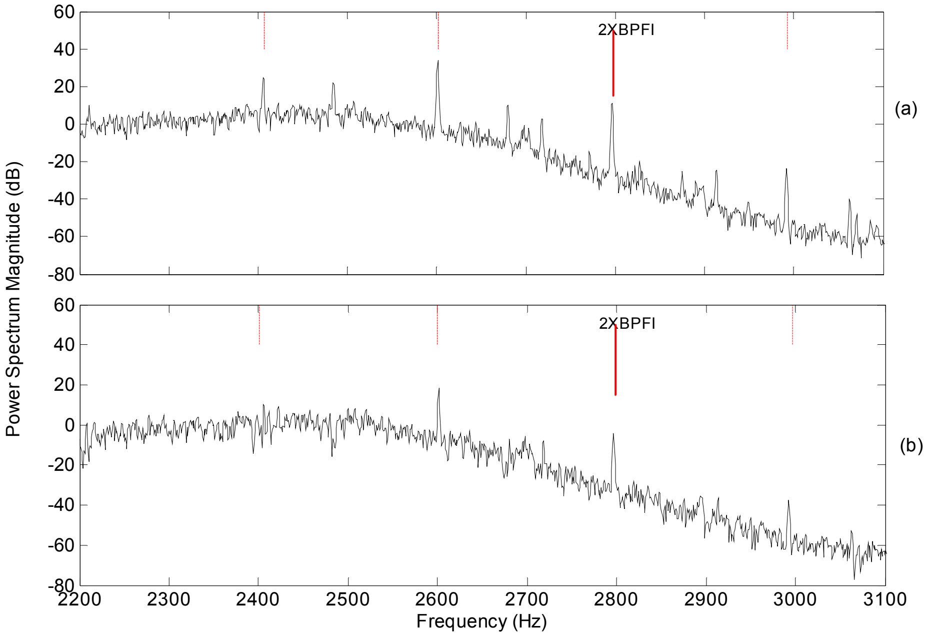

Figure 15: The effect of removing discrete frequencies using DRS on the second harmonic of the BPFI and its sidebands (a) Raw (b) DRS residual

Conclusions

If the frequency range of the signal is limited, only one or two harmonics of the impulse train may remain and the result will be close to sinusoidal, thus non-impulsive and no longer detectable using envelope analysis. DRS gives an undesirable effect by removing the low harmonics of bearing related frequencies from the spectrum as they are not sufficiently smeared, and appear as discrete components in the low frequency region. Genuine discrete frequency components can however still be separated by using separate time synchronous averaging (TSA) adjusted for every harmonic family in the signal (typically two or three for a gas turbine engine). It can be still possible to locate the ball pass frequency of the inner race (BPFI), and modulation sidebands, by inspecting the power spectrum density (PSD) of the signal. This alternative diagnosis approach is much more difficult when the bearing signals are not distinguished by their impulsiveness and random characteristics, making it difficult to recognize them and separate them from other discrete components, even after removal of shaft harmonics. This is partly because of the unknown amount of slip, causing variations in calculated bearing frequencies, and the lack of a family of harmonics to give added weight to the diagnosis of a series of periodic impulses.

Acknowledgments

This work is supported by the Australian Defence Science and Technology Organization (DSTO) as a part of their Centre of Expertise scheme. It was also supported by FAG and FAG Industrial Services (FIS) by facilitating the collection of the data.

References

1. Ho, D. and Randall R.B, “Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signals”, *Mechanical Systems and Signal Processing*, Vol.14, No.5, September 2000, pp763-788.

2. Sawalhi N., and Randall R.B, “Semi-automated Bearing Diagnostics – Three Case Studies”. Paper presented at Comadem conference, Portugal, June 2007.

3. Antoni, J. and Randall R.B, “Unsupervised Noise Cancellation for Vibration Systems: Part II – A Novel Frequency Domain Algorithm”, *Mechanical Systems and Signal Processing*, Vol 18, No.1, 2004, pp. 103-118.

4. Sawalhi, N. and Randall, R.B, “Spectral Kurtosis Optimization for Rolling Element Bearings”. ISSPA Conference, Sydney, August 2005.

5. Endo, H. and Randall, R.B, “Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter” *Mechanical Systems and Signal Processing*, Vol. 21,No. 2, 2007, Pages 906-919.

6. Sawalhi N., and Randall R.B. and Endo, H, “The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis”, *Mechanical Systems and Signal Processing*, 21, No. 6, 2007, pp. 2616-2633.

7. Antoni, J., Randall, R. B, **“**The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines**”**, *Mechanical Systems and Signal Processing*, Volume 20, Issue 2, 2006, Pages 308-331

8. Randall, R.B, “Applications of Spectral Kurtosis in Machine Diagnostics and Prognostics”, *DAMAS 2005*, Gdansk, July 2005.

9. Antoni, J, “Fast computation of the Kurtogram for the detection of transient faults”, *Mechanical Systems and Signal processing*, vol. 2,No. 1, 2007, pp. 108-124.

10. Groover, C.L, Tretheweya, M. W, Maynard, K P., and Lebold, M. S., “Removal of order domain content in rotating equipment signals by double resampling” Mechanical *Systems and Signal Processing*, vol. 19,2005, 483-500.

11. Sawalhi N., Randall R.B., “Localised fault diagnosis in rolling element bearings in gearboxes”, paper presented at the Fifth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Edinburgh, 15-18 July 2008.