Enhanced fault diagnosis of roller bearing elements using a combination of empirical mode decomposition and minimum entropy deconvolution

4 authors, including:

Zhenhe Zhang
University of Birmingham
3 PUBLICATIONS **49** CITATIONS

Mani Entezami
University of Birmingham
38 PUBLICATIONS **551** CITATIONS

Clive Roberts
University of Birmingham
190 PUBLICATIONS **3,521** CITATIONS

Some of the authors of this publication are also working on these related projects:

S-CODE project under grant agreement No 730849 View project

More Energy Efficient Trams -Stage 2- View project

Enhanced fault diagnosis of roller bearing elements using a combination of empirical mode decomposition and minimum entropy deconvolution

Z Zhang, M Entezami, E Stewart and C Roberts

Abstract

This paper introduces a new signal processing algorithm for vibration-based fault detection and diagnosis of roller bearings. The methodology proposed in this paper is based on the combination of two data-adaptive techniques which are further enhanced through the use of an automatic feature identification mechanism. The new technique, introduced as empirical mode envelope with minimum entropy, combines elements from the empirical mode decomposition (EMD) and minimum entropy deconvolution (MED) approaches with an energy moment technique to improve the feature selection stage of the EMD algorithm. This improvement allows the processing chain to identify early stage roller bearing faults in noisier signals. The energy moment technique is used to automatically identify the most appropriate intrinsic mode function from the EMD process prior to the MED algorithm being applied. This is in contrast to conventional approaches which tend to use the first mode or make selections based on traditional energy techniques. The combination of the adaptive techniques of EMD and MED allows the development of an improved technique for fault detection and diagnosis of signals. Combining these techniques with the energy moment approach allows further improved fault detection in complex non-stationary conditions. The processing chain has been tested using data obtained during laboratory testing. From the experimental results, it is shown that the new technique is capable of the detection of early stage (minor) roller and outer race defects found in tapered-roller-bearings rotating at a variety of speeds and noise scenarios.

Introduction

Rolling element bearings have widespread domestic and industrial applications. They are extensively applied in the transportation and manufacturing industries, where they perform an important role. In the vast majority of rotating machines, roller bearings are one of the most critical mechanical components.

Defects in bearings may arise during operation or in the manufacturing process. A quiet and smooth operation of the bearings is essential to the correct performance of the machinery.[1] In 2011, an in-service train operating on the Belfast to Dublin line suffered a bearing failure which was confirmed when an axlebox was found to be smoking at Connolly Station.[2] Fortunately, there were no casualties in this accident as the way-side hot axlebox detection equipment had identified the failure causing the train to be stopped. Having a suitable monitoring system to detect faults in early stages of development can lead to the prevention of catastrophic accidents and reduction in maintenance costs. Vibration monitoring has been established as the most common and reliable method for the inspection of bearings. Several studies[3–8] have been conducted to explain the mechanism of vibration and noise generation in bearings.

Normally, the vibration signal is collected from a surface-coupled vibration sensor mounted on the housing of the rotating machine. The working principle of the vibration monitoring method is generally based on the extraction of key frequencies

School of Electronic, Electrical and Systems Engineering, University of Birmingham, Edgbaston, Birmingham, UK

Corresponding author:
Z Zhang, School of Electronic, Electrical and Systems Engineering, University of Birmingham, No.52 Pritchatts Rd, Gisbert and Kapp Building, Birmingham, B15 2SA, UK.
Email: zxz007@bham.ac.uk

corresponding to characteristics of the rolling elements. The bearing itself is a source of vibration and noise due to its mechanical design and the interactions between the rolling elements and races.[9] The energy contained in these vibrations usually then increases in the presence of defects.[8]

As an element of a rotating system, a bearing itself is a source of vibration and noise due both to its fundamental design and the interactions between the rolling elements and races.[9] However, the sources of bearing vibration are not only limited to bearing operation or mechanical defects, but could be from the system in which the bearing is mounted. Lynagh et al.[10] present an analytic model of bearing vibration to clarify the various sources of bearing vibration, both from inside of bearing and from the supporting mechanical system. The main sources of bearing vibration correspond to the rolling element characteristic frequencies. These characteristic frequencies are: the fundamental train frequency, the ball to outer race frequency, the ball to inner race frequency and the ball frequency. These characteristic frequencies will be explained in section 'Rolling-element characteristic frequencies.' Researchers working in the field of bearing diagnostics traditionally use these characteristic frequencies in the diagnosis of bearing faults.[11,12] Another significant vibration source is the out-of-balance rotation caused by the variance between the centre of the shaft axis and the geometric centre of the bearings.[10] Additionally, there are other sources of bearing vibration regarded as being of secondary significance. These include the waviness of rolling elements and the existence of abnormal ball elements and races.[10] These anomalies contribute to the vibration spectrum of the bearing, but are of greater significant at higher frequencies. When defects occur in a bearing, the vibration energy at the corresponding characteristic frequencies significantly increases.[8]

The non-stationary of the bearing signal is resulted by the random fluctuations of the bearing signals during the experiment.[13] Each bearing element has specific characteristic frequencies, which is the geometric property of the bearing. And this property is widely used as a direct faulty information in the bearing diagnosis, called characteristic defect frequencies when it is used in the bearing fault diagnosis.[14] Due to the non-stationarity of the bearing signal, it is difficult to identify significant peaks at the characteristic frequencies in the spectrum unless the defect is predominant.[15] Hence, preparation (pre-processing) of the vibration data for standard envelope analysis techniques[5] is very important and can be complex.

Time synchronous average (TSA),[16] envelope analysis, wavelet transform,[11,43] the auto-regression moving average (ARMA),[12] cepstrum analysis[17] and, more recently, spectral kurtosis[13] are the most widely used signal processing techniques for bearing inspection and fault detection. These techniques are based on a number of specific parameters which must be correctly defined for the technique to operate most effectively. For example, filtering techniques such as wavelet transform, ARMA and other time-frequency domain analysis techniques need to specify the central frequency and bandwidth of their component filters. Du et al.[6] noticed that in some cases, particularly when signal is surrounded in severe noise, it is difficult to define parameters such as the central frequency and bandwidth of the band-pass filter. In these cases, spectral kurtosis can be used to identify the appropriate central frequency and bandwidth.[18] In contrast, the technique proposed in this paper is categorised as an adaptive technique. This means that parameters such as the central frequency and bandwidth do not need to be explicitly set. Elsha et al.[19] suggest improved effectiveness in identifying a naturally degraded bearing under conditions of 'relatively large background noise' when using an adaptive filter.

Adaptive techniques, also called data-driven techniques, are applied as a pre-processing step because of their ability to overcome the problem of periodic noise being injected into the signal from one rotating element by another. Empirical mode decomposition (EMD) is one such adaptive technique. It is based on the decomposition of the signal in order to generate less complex data sets.[6] Minimum entropy deconvolution (MED) is another adaptive signal processing technique.[20,21] It is effective at de-convolving impulsive sources from a mixture of signals.[7,20,22–24]

The EMD method is a time-domain processing technique which can be applied to reduce the level of noise within a signal.[25,44] Conversely, MED can be used to emphasise an impulsive element within a signal.[7] These two techniques can therefore be combined to both reduce the noise within a signal and to emphasise the components of the signal that are of interest. Such processing has been applied in the area of roller bearing diagnostics.[26,27] In these cases, the result of the first EMD iteration is used in the processing chain. In other EMD applications, it is common to consider multiple iterations of the process and select the result demonstrating the greatest level of energy.[28] The method used to select the most appropriate EMD iteration result is particularly important when combining multiple processing techniques. In addition to methods such as traditional energy and energy entropy, the energy moment technique is particularly suited to signals with non-stationary characteristics.[28]

This paper presents a novel processing chain for roller bearing vibration signal analysis. The system uses a combination of empirical mode envelope and minimum entropy techniques and so is referred to as EMEME. EMEME is the combination of EMD, an energy moment technique for iteration selection, and MED. Used together; the elements in this processing chain provide noise filtering and allow the extraction of key signal features. The combined techniques are shown to allow bearing fault detection in signals with

high noise levels. In the upcoming section, the fundamentals of EMD are explained. Unlike time-frequency processing techniques, EMD can decompose a signal without historical data or pre-knowledge of the signal. Next, the fundamentals of the energy moment technique are explained. The energy moment technique is capable of identifying the most appropriate EMD iteration in order to choose the most representative decomposed component of the original signal. Later, the fundamentals of MED are explained. In this case, MED is used to emphasise the impulsive element within a signal. Subsequently, the combined EMEME process is explained using a flow chart to demonstrate the working procedure. Roller bearing faults and geometry are discussed, followed by experimental introduction, testing results and analysis in a later section. The verification of the proposed technique and the novel enhancement are demonstrated and analysed in a sub-section. Conclusions are drawn in the last section.

Fundamentals of EMD

EMD is an adaptive technique, also known as a data-driven technique. It was originally proposed by Huang et al.[29]; encouraging results of the application of EMD to signals obtained from rotational mechanical systems are reported in various studies.[25,30,31] The EMD method decomposes the signal into different sub-signals, which are the intrinsic mode functions (IMFs). A detailed introduction to the EMD method can be found in Huang et al.[29]

EMD focuses on extracting the stationary points of signal.[26] Assuming a generic signal $x(t)$, the maxima and the minima of which are interpolated by means of a spline to obtain $Max(t)$ and $Min(t)$ respectively.

These two curves are able to represent the envelop of $x(t)$. The average signal from these two splines is called $mean(t)$. $x_1(t)$, shown in equation (1), can be calculated by subtracting the $mean(t)$ from the original signal $x(t)$

$$x_1(t) = x(t) - mean(t) \qquad (1)$$

If the resulting signal $x_1(t)$ is an intrinsic mode function (IMF), the algorithm ends. Otherwise, the previous steps are repeated as per equation (2). This iterative procedure is known as the 'sifting process.'

$$x_n(t) = x_{n-1}(t) - mean_{n-1}(t) \qquad (2)$$

To be an IMF, the signal must meet the following requirements:

1. The number of extrema in $x_n(t)$ equals the number of zero crossings.
2. The mean between the local maxima and local minima splines in $x_n(t)$ is equal to zero at any point.

In the sifting process, $Max(t)$, $Min(t)$ and $mean(t)$ are recomputed so that $mean(t)$ can be subtracted from the previous signal $x_{n-1}(t)$. When the first IMF, $C_1(t)$, is extracted the sifting process stops.

Then, $C_1(t)$ is subtracted from the original signal $x(t)$

$$r_1(t) = x(t) - C_1(t) \qquad (3)$$

The signal $r_1(t)$ is treated as the original signal for the second IMF calculation by means of the sifting process

$$r_2(t) = r_1(t) - C_2(t) \qquad (4)$$

The EMD algorithm generates the final IMF, $C_N(t)$, when the residual signal $r_N(t)$ is a constant or monotonic function. The threshold to stop the procedure is described as

$$\delta(r_N(t)) < \delta_{stop} \qquad (5)$$

where $\delta(r_N(t))$ is the residual signal of $r_N(t)$ and δ_{stop} is the threshold to stop.

The IMFs extracted by the EMD algorithm can be considered to be sub-signals of the original signal. These sub-signals expose components that can be used for latter signal processing.

IMF selection by energy moment

EMD can be used to identify a number of IMFs corresponding to the key sub-signals within an original signal. Some of these IMFs will be more appropriate for use in further processing than others. There are a number of methods to identify the most appropriate IMF such as the rotating frequency phenomena of rolling elements,[26] energy of the IMFs[3] and energy entropy of the IMFs.[32] However, these methods disregard the distribution features of each IMF signal's energy.[28] Therefore, the extracted components of the signal may not accurately represent the nature of the fault.

In order to enhance the performance of feature extraction algorithms, IMF energy moment can be used as a method for the selection of a suitable IMF to represent the most important elements of the original signal. The energy moment technique calculates the energy distribution of each IMF in the time-domain. Bin et al.[28] illustrates the effectiveness of the energy moment technique by evaluating it using multiple sets of simulated signals.

Equation (6) is the energy moment calculation (EM) for continuous signals and equation (7) is used for discrete signals: a series of IMFs generated by EMD is $IMF_i(t)$, $i = 1, 2, 3, \ldots \ldots$

$$EM_i = \int^t \cdot |IMF_i(t)|^2 dt \qquad (6)$$

$$EM_i = \sum_{k=1}^{m}(k \cdot \Delta t)|IMF_i(k \cdot \Delta t)|^2 \qquad (7)$$

where Δt is the sampling period; m is the total number of data samples.

Then, the normalised eigenvector of the signal based on the energy moment is

$$E = [E_1, \ E_2, \ E_3, \ \ldots\ldots, E_m]/\ \sum_{i=1}^{m} E_i \qquad (8)$$

where each element of E corresponds to the percentage of the energy moment of the $IMF_i(t)$. In this work, the IMF with the highest energy moment is then selected as the representative IMF.

Fundamentals of MED

The minimum entropy deconvolution (MED) algorithm improves the peakedness of the signal by emphasising the transient components. It aims to reduce the randomness of a signal by minimising its entropy.

A generic signal $g(t)$ can be considered as two component parts as per the following equation

$$g(t) = \underline{h(t) * w(t)} + \underline{\eta(t)} \qquad (9)$$
$$\quad\quad\;\; \nearrow \quad\quad\;\; \nearrow$$
$$\quad\quad \text{Part 1} \quad \text{Part 2}$$

Part 1 is the convolution between the component related to the system behaviour $h(t)$ and its excitation $w(t)$. Part 2 represents randomly distributed noise.

When the signal $g(t)$ passes through a filter $f(t)$, it is represented as $u(t)$

$$u(t) = g(t) * f(t) = h(t) * w(t) * f(t) + \eta(t) * f(t) \qquad (10)$$

This can also be presented in the discrete domain

$$u(n) = g(n) * f(n) = \sum_{i=1}^{M-1} f(i)v(n-1) \qquad (11)$$

with $n = 0,1,2,\ldots, T + M - 2$; where T is the length of the convolution between the signals, and M is the length of the filter.

Key to this stage of the MED process, as discussed in Gonzalez et al.,[21] is choosing the value of the filter length (M). In the paper by Pennacchi et al.,[33] the filter length is selected based on the characteristic period of the system (i.e. for rotating components the filter length chosen to be the same as the period of rotation).

The objective of this filtering is to retain (or maximise) the contribution from the first part of the signal, while minimising that from the components associated with the noise. The result of this action is that the signal entropy is reduced. By iteratively repeating the process the algorithm identifies the output $u(n)$ that has an optimal match to $h(t)$, and thereby minimises the signal entropy. The signal entropy is evaluated using the Varimax norm[20]

$$V(u(t)) = \sum_{j=0}^{P} u^4(j) \Big/ \left(\sum_{j=0}^{P} u^2(j)\right)^2 \qquad (12)$$

with $P = T + M - 2$.

MED is particularly effective at separating frequent impulses resulting from shocks generated by localised defects.[7]

Empirical mode envelope with minimum entropy

The EMEME method proposed in this paper results from the combination of the EMD and MED algorithms with the use of the energy moment technique for IMF selection. The first stage is to apply the EMD algorithm to the experimental signals, to generate a series of IMFs, and then to select the IMF with the greatest energy moment value as described in section 'IMF selection by energy moment.' The second stage is then to apply the MED algorithm to minimise the signal's entropy, thereby emphasising key periodically occurring impulsive features.

In roller bearing diagnostics, the signals of interest are generally analytic signals, i.e. in the frequency domain representation any negative components can be disregarded. Such signals are affected by slippage occurring between the rolling elements and the races due to the presence of localised faults.[26] The analytic signal of the representative IMF, $r_{n_0}(t)$, is calculated and then used in the MED algorithm.

The Hilbert transform of the representative IMF $r_{n_0}(t)$ is

$$H[r_{n_0}(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} r_{n_0}(t')/(t - t')\mathrm{d}t' \qquad (13)$$

The analytic signal of the representative IMF $r_{n_0}(t)$[34] is

$$z(t) = r_{n_0}(t) + jH[r_{n_0}(t)] = a(t)\exp(j\varphi(t)) \qquad (14)$$

where

$$a(t) = \sqrt{r_{n_0}(t)^2 + H[r_{n_0}(t)]^2} \qquad (15)$$

$$\varphi(t) = arctan(H[r_{n_0}(t)]/r_{n_0}(t)) \qquad (16)$$

In which $a(t)$ is the envelope, and $\varphi(t)$ is the angular characteristic of the analytical signal.

The behaviour of the envelope $a(t)$ is important as it preserves the energy of the measured signal. The MED process is applied to this envelope.

Figure 1 demonstrates the stages involved in applying the EMEME technique to a vibration signal.
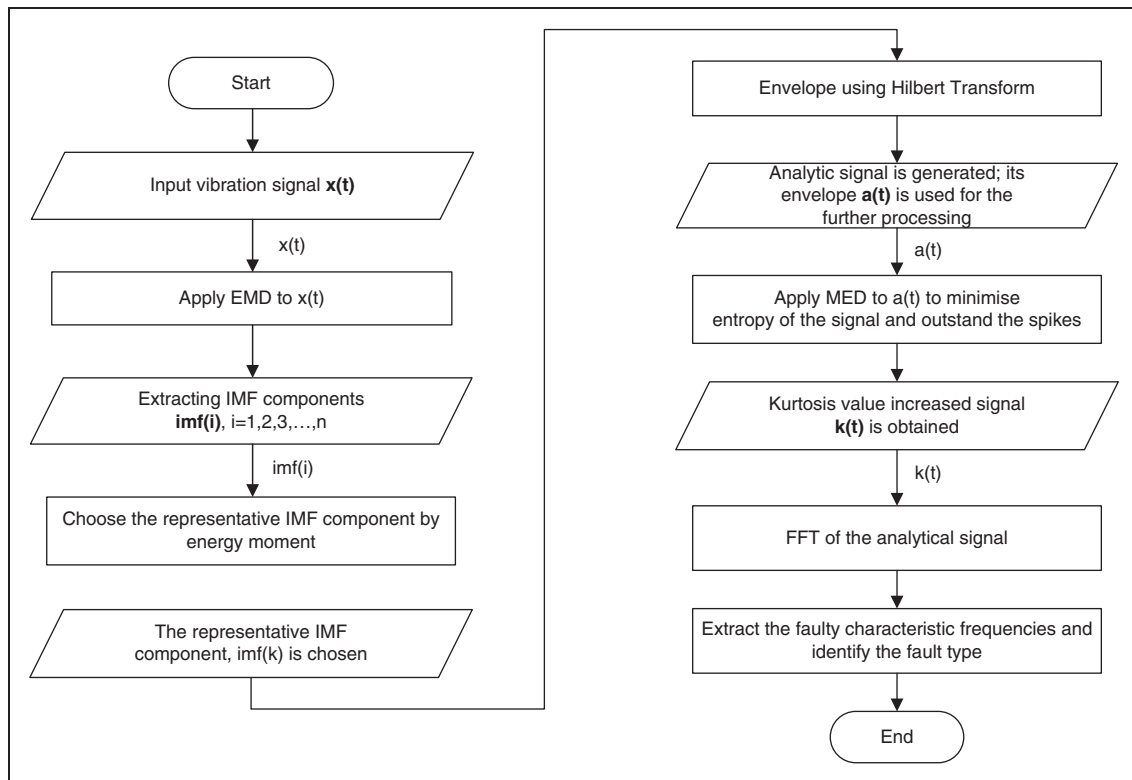
Figure 1. The flowchart of the proposed method process.

Typical bearing faults

Bearing defects can be caused by manufacturing errors, improper installation or abrasive wear, such as surface roughness, waviness, misalignments and off-size rolling elements.[4,9,35–37] Other types of bearing faults are localised, such as cracks, pits and spalls on the rolling surfaces. Defects are generally categorised into four types: roller defect, damaged cage, inner and outer race faults.[38]

A number of faults were artificially induced into the automotive bearings used in this study. Examples of these induced faults are presented here:

1. Roller fault; scratch and spalling of the roller, shown in Figure 2(a).
2. Outer race fault; scratching over approximately 30% of the outer race circumference, shown in Figure 2(b).
3. Damaged cage and roller fault; a small scratch on the roller along with a broken section of the cage, shown in Figure 2(c).

The particular bearing faults used in this study are characterised as shown in Table 1.

A basic requirement of any bearing condition monitoring system is to identify faults, but not necessarily to categorise them. More advanced bearing monitoring systems, such as those presented in this paper, allow different bearing faults to be individually identified and categorised.

## Rolling-element characteristic frequencies

Envelope detection or amplitude demodulation is a technique that extracts fault-related frequency components resulting in a modulated signal. The fast Fourier transform (FFT) spectrum of the modulated signal is known as envelope analysis.

An impact event appears within the signal when either a rolling element strikes a race fault, or a fault on a rolling element strikes a race. The geometric relationships between the roller and race components determine the timings of these impacts, and thus the characteristic fault frequencies of the bearing. Each different type of fault (as previously described) is illustrated by a different characteristic fault frequency. The presence of faults can also increase the vibration energy.

For a bearing with a stationary outer race, the following equations define the characteristic frequencies.[1]

FTF, fundamental train frequency, fault on the cage or mechanical looseness

$$f_c = \frac{\omega_r}{2}\left(1 - \frac{RD}{PD}\cos\beta\right) \tag{17}$$
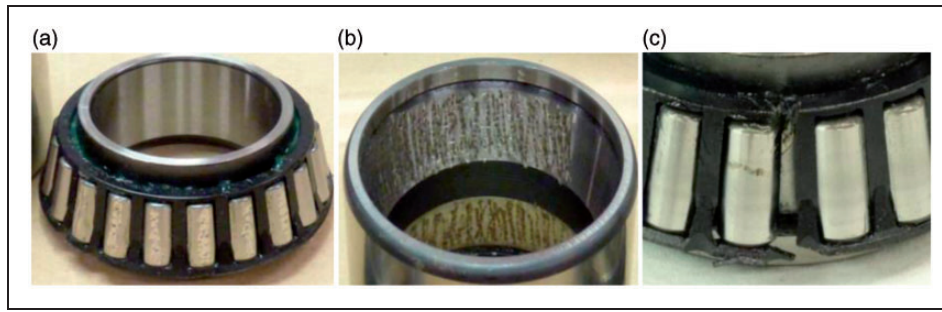
Figure 2. Typical faults of tapered-roller-bearing.

Table 1. Bearing test fault conditions.

Bearing test ID	Description
F0	Good condition, no introduced defects.
F1	Minor damage (single small shallow fault) to one small area on each outer race, consisting of surface roughening by means of an electrical discharge engraver. Fault length 1.7% of circumference.
F2	Minor damage to one small area of one roller in each cage, again consisting of surface roughening by electrical discharge engraver. Fault length 10% of circumference.

BPFO, ball passing frequency outer race, local fault on outer race

$$f_{bo} = \frac{R}{2}\omega_r \left(1 + \frac{RD}{PD}\cos\beta\right) \qquad (18)$$

BPFI, ball passing frequency inner race, local fault on inner race

$$f_{bi} = \frac{n}{2}\omega_r \left(1 - \frac{RD}{PD}\cos\beta\right) \qquad (19)$$

BFF, ball fault frequency = 2 * BSF, ball spin frequency, local fault on rolling element

$$f_{bf} = \frac{PD}{2RD}\omega_r \left(1 - \left(\frac{RD}{PD}\cos\beta\right)\right)^2 \qquad (20)$$

where $f_c, f_{bo}, f_{bi}, f_{bf}$ are the frequency of the item (Hz); ω_r is the shaft rotation rate (rev/s) or (Hz); RD is the roller diameter (m); PD is the mean roller race diameter (m); β is the contact angle (rad); and R is the number of rollers.

Experiments and testing results

Experiment set-up

All the data were collected in the BCRRE laboratory at the University of Birmingham. In order to carry out the laboratory tests on the roller bearings a test rig was built, this is shown in Figure 3.

The key features of the test rig are as follows:

1. A rotating shaft which is mounted on two support bearings and can be loaded with two different bearing sizes: automotive (up to ¼ tonne), and railway (up to 5 tonnes).
2. Single 2.2 kW electric motor to drive the shaft, via a belt for isolation, at varying speeds.
3. A hydraulic jack under each test bearing to apply a vertical load.
4. Pressure gauges (analogue and digital) to measure and monitor the applied load.
5. A tachometer to measure the rotational speed of the shaft.

The type 801023AB automotive tapered roller bearing are used as the unit under test. The specifications of the test bearing are shown in Table 2.
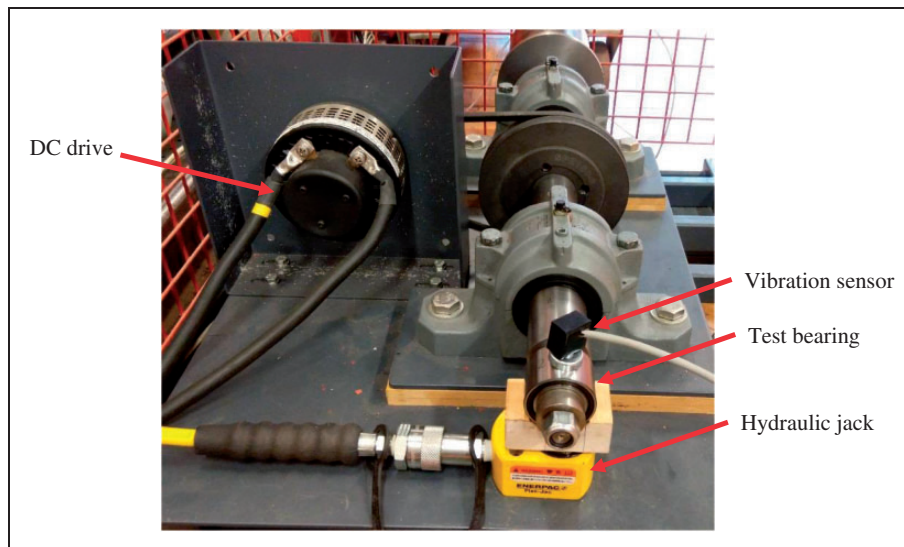
Experiments were carried out at various speeds to support the evaluation of the processing techniques. At the development stage, preliminary 500 r/min and 1000 r/min tests were undertaken; for the main tests a wider range from 250 r/min to 1000 r/min was used. The characteristic frequencies for the test bearing are shown in Table 3 for each of the different test speeds.

Testing results

This section describes the stages of application of the EMEME algorithm to the recorded data.

Feature extraction using EMEME
1. EMD applied to experimental signal

Figure 3. Bearing test rig.

Table 2. Bearings specification.

	Value
PD	42.5 mm
RD	5.3 mm
R	21
β	13.8°

Table 3. Example of characteristic frequencies of the bearings.

Scaled bearing (Hz)	Speed (r/min)					
	250	300	400	500	600	1000
FTF	1.8	2.2	2.9	3.6	4.4	7.3
BPFO	38.4	46.1	61.5	76.9	92.3	153.8
BPFI	49.05	58.8	78.5	98.1	117.7	196.2
BFF	32.9	39.5	52.7	65.85	79	131.7

FTF: fundamental train frequency; BPFO: ball passing frequency outer race; BPFI: ball passing frequency inner race; BFF: ball fault frequency.

In this section, EMD is applied to signals recorded from a healthy bearing at a speed of 500 r/min. By applying EMD, a series of 20 IMFs were extracted. IMF components extracted earlier in the process have a broader range of frequencies, and therefore cover a wider range of the frequency components from the original broadband signal. The first five IMFs along with the original signal are shown in Figure 4.

Figure 5 shows the original signal, and the IMFs and residuals arising from the application of EMD. The original signal and the residuals (obtained by removing the IMF from the original) are shown in the left hand column while the IMFs themselves are shown on the right.

The IMF component with the greatest apparent similarity is not always selected as the representative IMF because the relevant part of the signal can sometimes be masked by noise. The EMD method decomposes the signal into different oscillation modes in the time domain, defined by the time between the local maxima or minima.[39] Hence, the IMF components extracted earlier in the process are not always those with the greatest energy due to the alignment of the IMF with the periodicity of the fault.

The representative IMF should have the highest time-related energy, earlier introduced as the energy moment. The energy moment of each IMF signal is

used to select the representative IMF. In this case, IMF(2) is chosen to be the representative IMF.

2. MED applied to the representative IMF envelope

MED is applied to the representative IMF of F1 and F2 and the results are shown in Figures 6 and 7.

A non-dimensional quantity that measures the relative 'peakedness' of a signal relative to the Gaussian distribution (known as kurtosis) is calculated and shown in the figures to demonstrate the effectiveness of the MED. For normally distributed data, such as the distribution of the natural vibrations of a healthy bearing,[40] the kurtosis value is 3. For signals with repetitive impulsive forces, the kurtosis can be much greater. Although kurtosis cannot identify a particular type of defect, it can be used to indicate if there is defect present in a bearing. In this case, the vibrations resulting from the motion arising from the irregularities in the bearing's internal components can lead to a significantly increased kurtosis value. In Figures 6 and 7, the filter length of the MED is set to correspond to the bearing's rotational speed (in Hz). The figures show that the kurtosis values of the signals
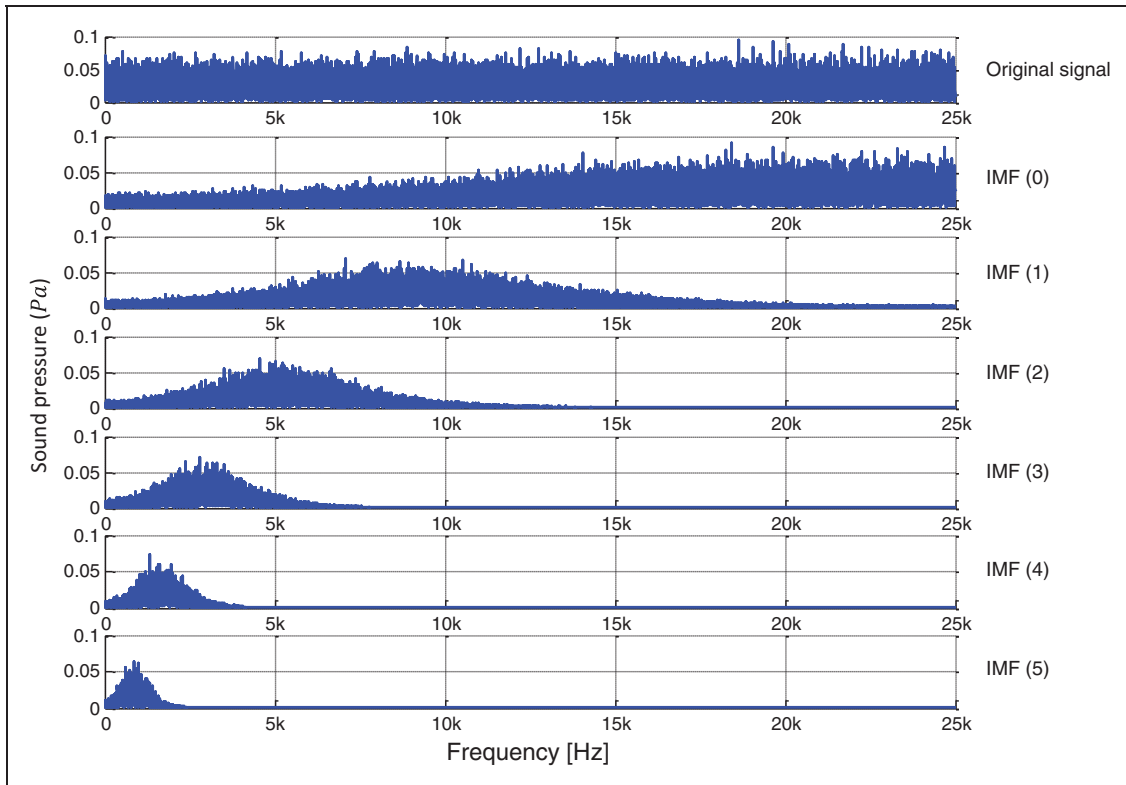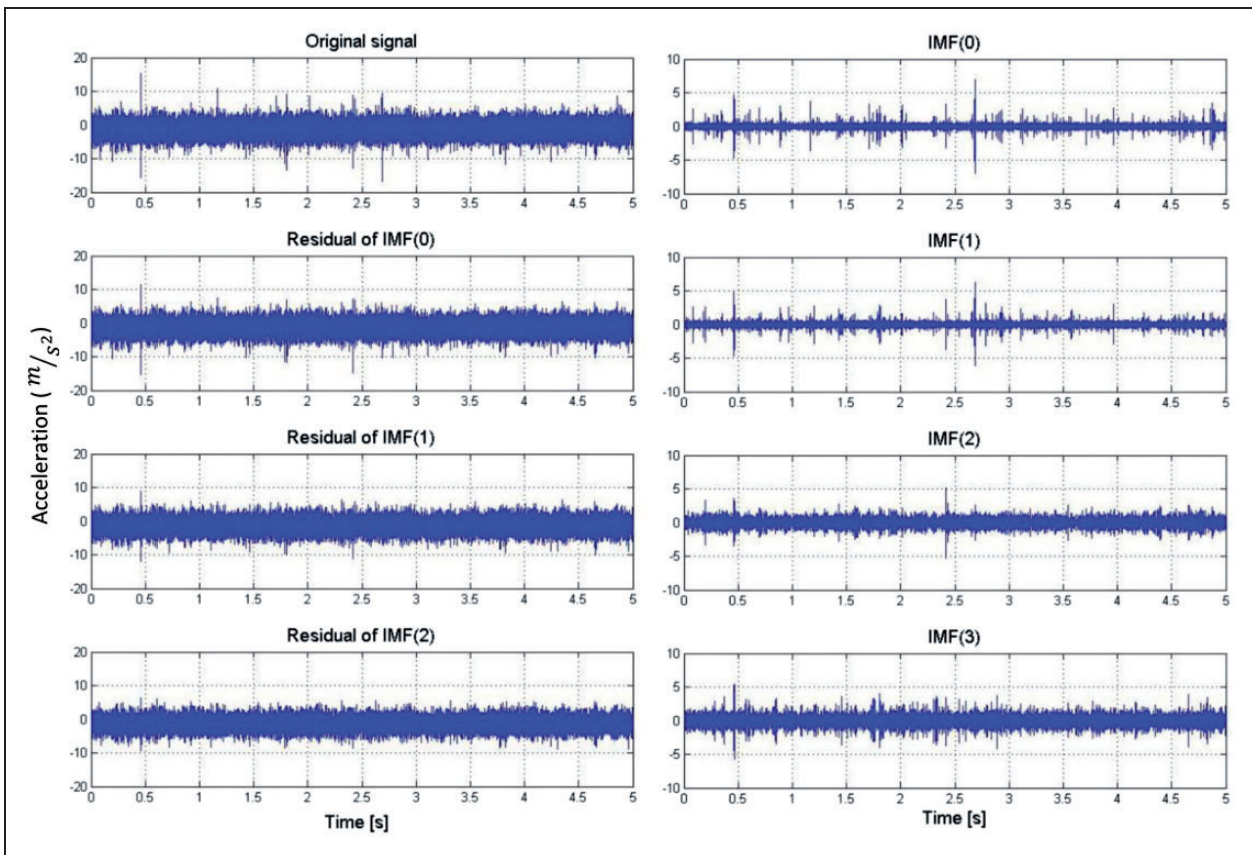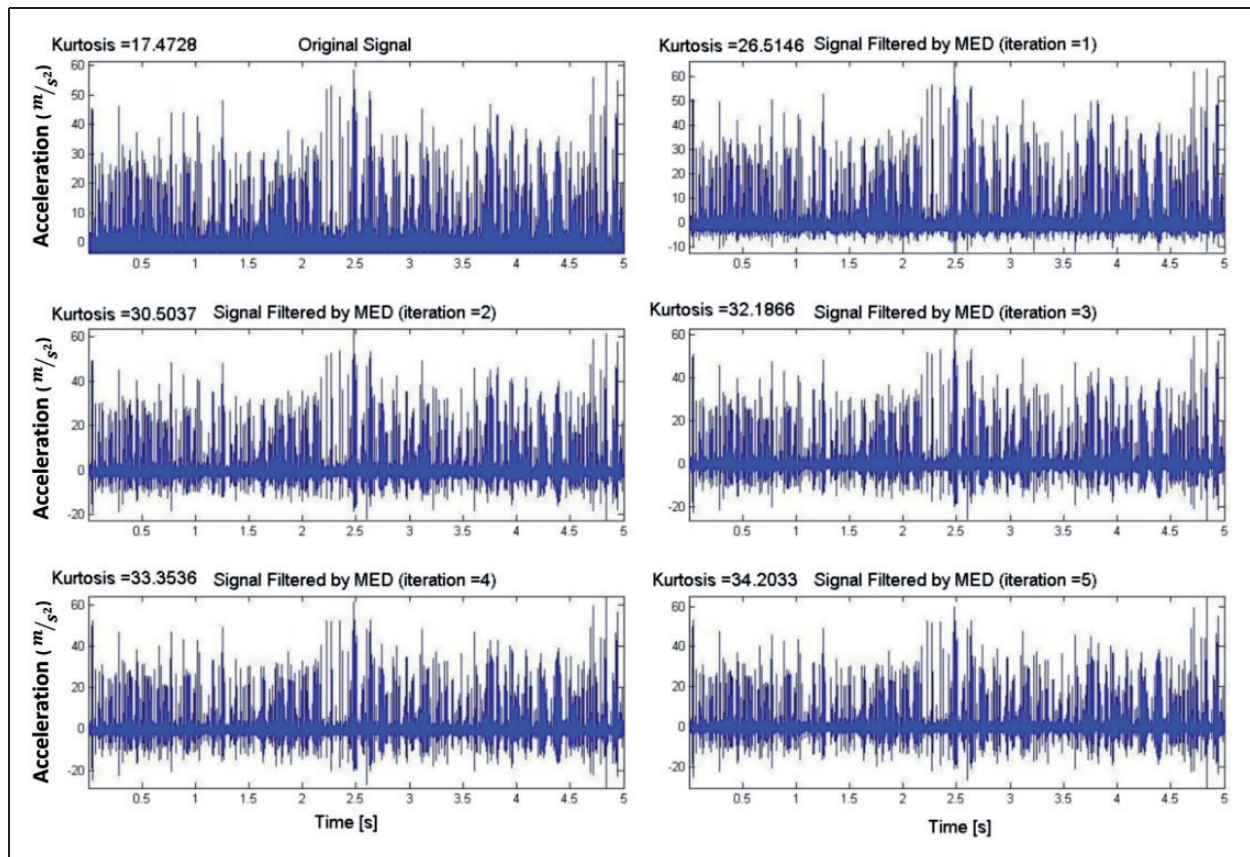
Figure 4. IMFs frequency ranges.

Figure 5. EMD procedure of the healthy bearing.

Figure 6. Effect of MED iterations when applied to F1 faulty vibration signal.

filtered by MED increase, this indicates that the spikes related to the bearing faults are enhanced.

Verification of EMEME results. Frequency domain analysis is used to verify the EMEME result. The results of the healthy and faulty bearings are shown in this section where the feasibility of the method is also evaluated under induced noisy conditions. Vibration data recorded from faulty bearings with minor damage on the outer race and on the roller (F1 and F2), spinning at speeds from 250 r/min to 1000 r/min, are used to evaluate the method.

The results of the lowest rotational speed of 250 RPM (i.e. 4.17 Hz) and the highest speed of 1000 r/min (i.e. 16.7 Hz) are presented. The former case illustrates the bearing diagnosis results when the bearing vibration is weak and the latter shows the results when the vibration noise is high. An FFT has been applied to the signals obtained with EMEME. The spectrums are shown in Figures 8 and 9. The equivalent results from F0 healthy bearing are also presented for comparison purposes, in order to clearly indicate the abnormal behaviour of the faults.

In the spectrums presented here, the arrows indicate the calculated principle bearing frequencies (see section 'Rolling-element characteristic frequencies') for both cases. A certain amount of out-of-balance rotation can be clearly seen in the two graphs,

which is the primary unbalance at the shaft speed (ω_r).[10] The bearing inherent frequencies have also been calculated and are shown in the two graphs. These are the fundamental train frequency (f_c), the ball frequency of the outer race (f_{bo}), the ball frequency of the inner race (f_{bi}) and the ball fault frequency (f_{bs}).

Comparing the healthy results for the two cases; the difference between the average background noise level and the characteristic frequencies shown in the 250 r/min situation is of greater magnitude than in the 1000 r/min scenario. This implies a greater level of background noise in the measurement taken at higher speed.

Comparing the healthy (F0) and faulty bearings (F1 and F2); the out-of-balance level, indicated by the shaft speed, is always the highest component in the spectrum of F0, while the appropriate fault frequencies (ball frequency of the outer race for F1 and ball fault frequency for F2) are highlighted in their respective spectrums. This demonstrates that the EMEME approach is capable of extracting and emphasising the faulty information from defective bearing signals without pre-knowledge of the fault type.

The peaks numbered 1, 2, and 4 in Figures 8 and 9 correspond to the frequencies of the principle 1st harmonic of the shaft speeds. The peaks numbered 3 and 5 are at the frequencies of the principle 1st harmonic
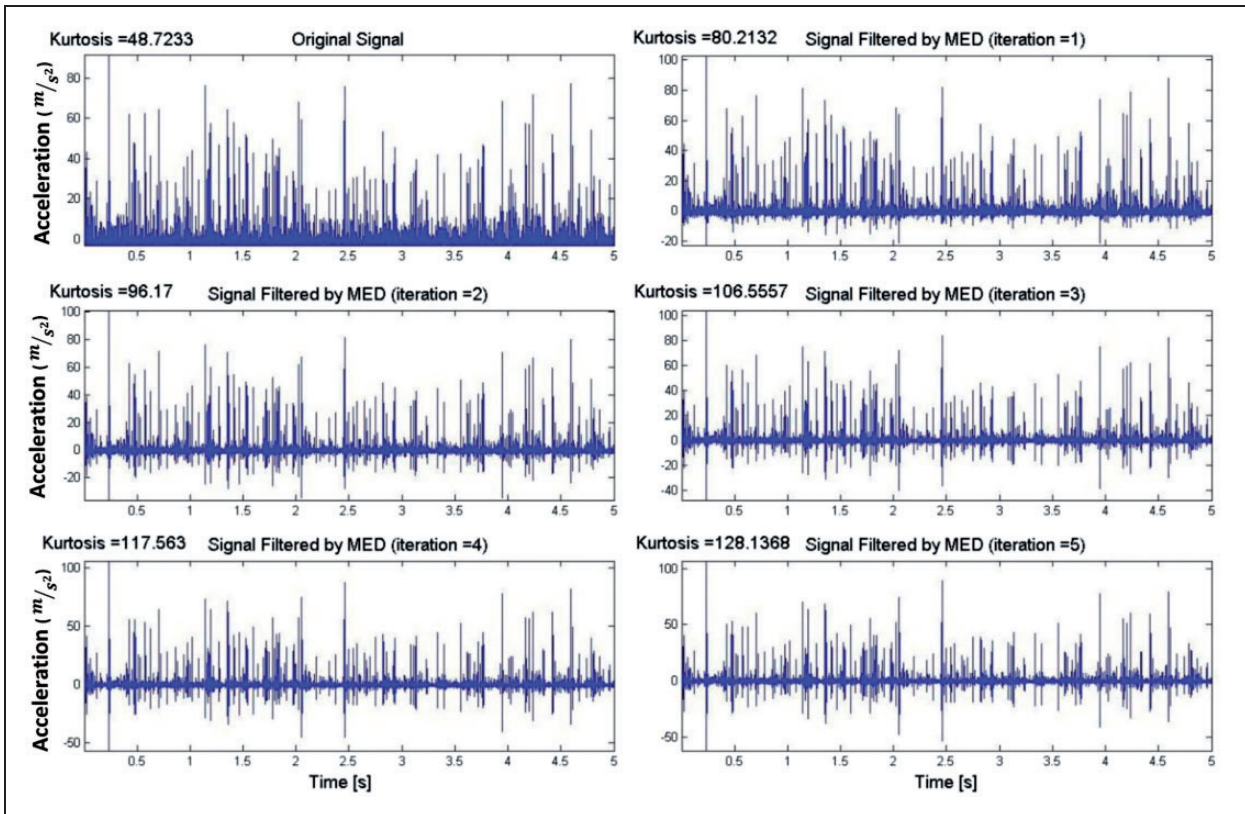
Figure 7. Effect of MED iterations when applied to F2 faulty vibration signal.
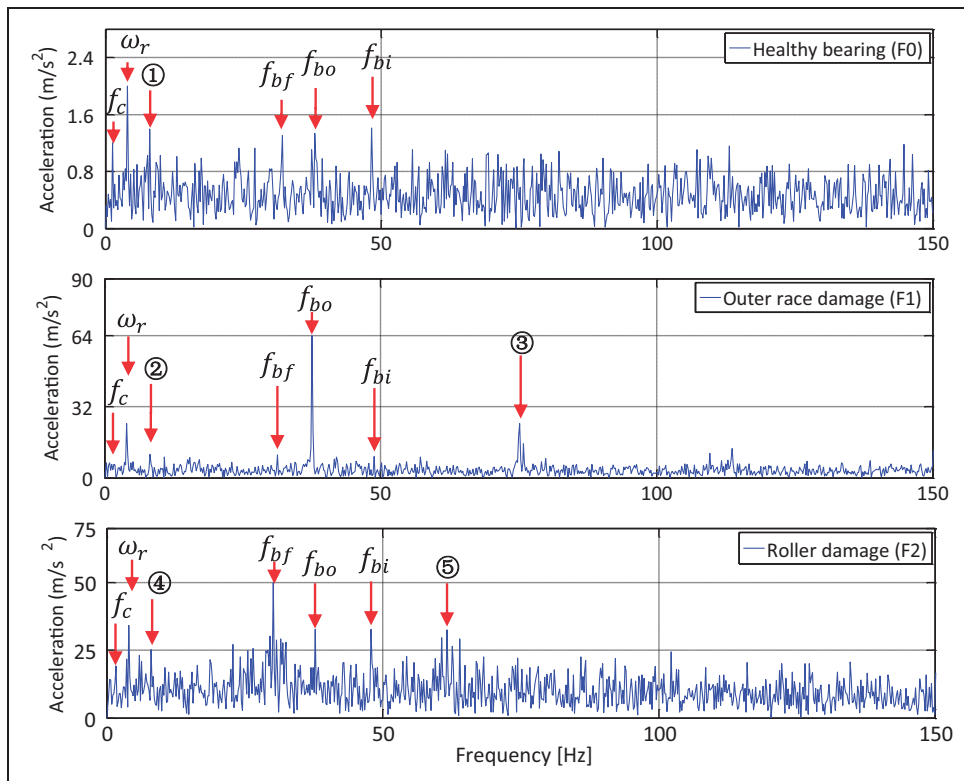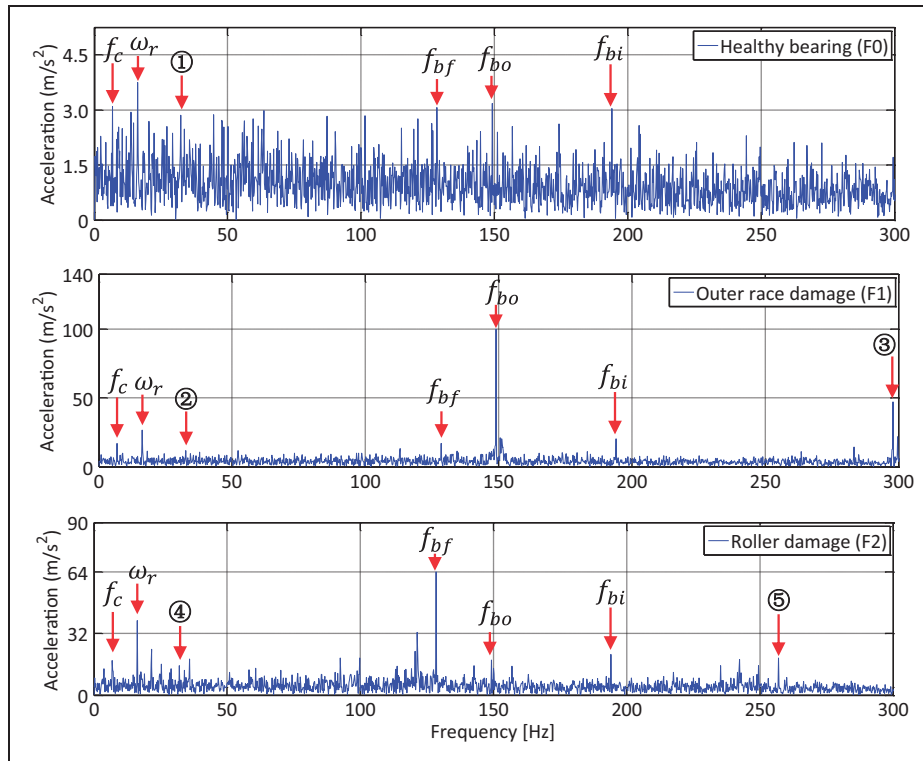

Figure 8. FFT envelope of healthy and faulty bearing spinning at 250 r/min (Top) healthy bearing (Middle) outer race damaged bearing (Bottom) roller damaged bearing's EMEME algorithm processed data.

Figure 9. FFT envelope of healthy and faulty bearing spinning at1000 r/min (Top) healthy bearing (Middle) outer race damaged bearing (Bottom) roller damaged bearing's EMEME algorithm processed data.
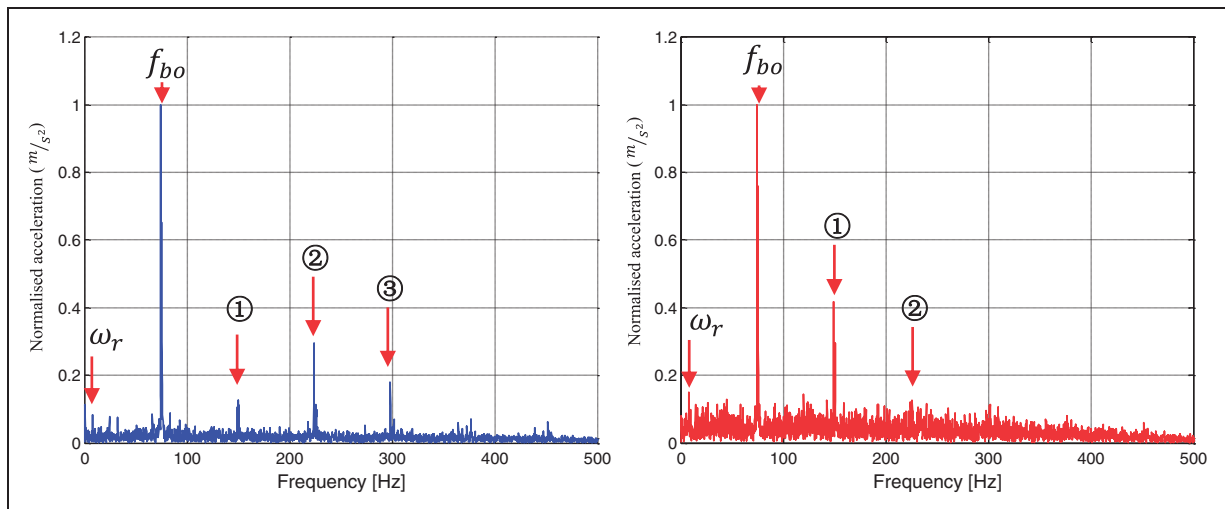
Figure 10. Conventional envelope method (left) vs. EMEME (right) at 500 r/min no added noise.

of the respective fault frequencies (f_{bo} for the bearing with the damaged outer race, and f_{bs} for the bearing with the damaged roller). This demonstrates that the EMEME approach can be used to identify not only the fundamental frequency, but also the first harmonic components of fault information in the frequency domain.

Comparison to conventional HFRT. The envelope analysis technique discussed in the introduction of this paper is also known as the high frequency resonance technique

(HFRT). HFRT operates by applying specific filters to emphasise particular frequency bands of interest, usually above 4 kHz, within the signal.[1,5] Low amplitude, high frequency, components are then shifted to a lower frequency range where the modulating signal can be extracted. An FFT is then applied to the resulting signals to give a final frequency characterisation.

In this section, traditional HFRT is compared to the new EMEME method presented in this paper. Figure 10 shows the results of applying the HFRT (left) and EMEME (right) processing methods to
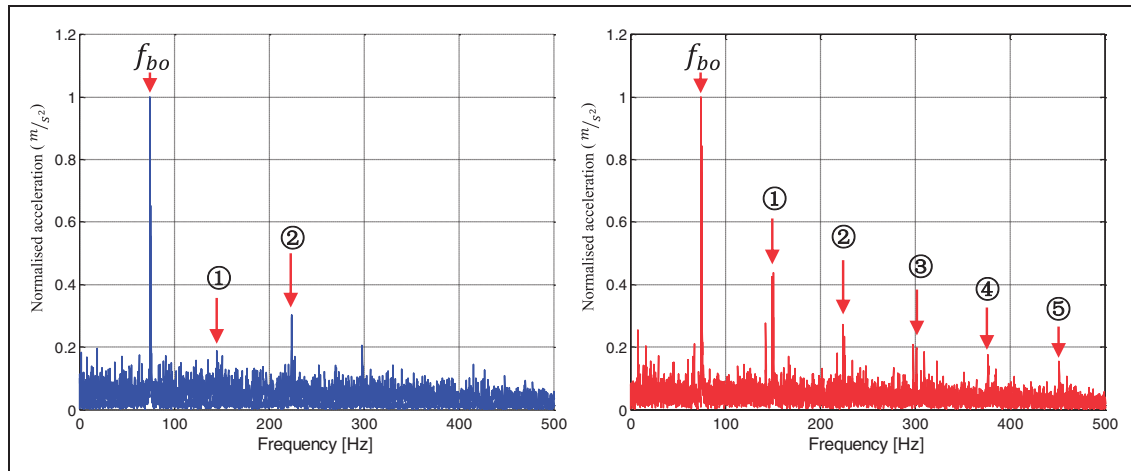
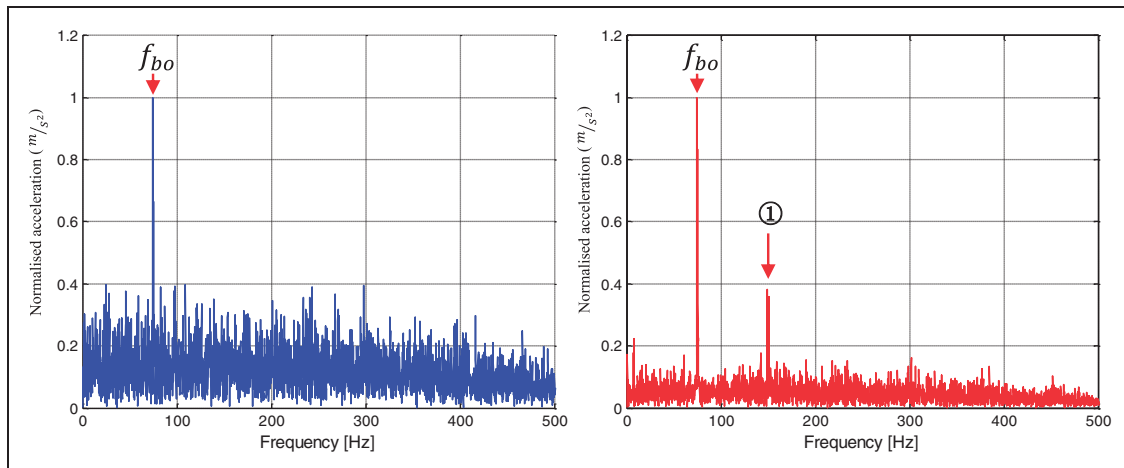**Figure 11.** Conventional envelope method (left) vs. EMEME (right) at 500 r/min with noise SNR of 5.

Figure 12. Conventional envelope method (left) vs. EMEME (right) at 500 r/min with noise SNR of 1.5.

data recorded from the test rig using a bearing with an induced outer race fault rotating at 500 r/min. Normalised acceleration is used to describe the amplitude of each frequency component, in order to simplify comparison of the waveforms. Both methods clearly identify the primary fault frequency and the ball frequency of the outer race (f_{bo}). The HFRT method also identifies the next three harmonics, while the EMEME technique highlights the 1st harmonic more clearly, but then does not indicate the higher harmonics as well.

The comparison presented in Figure 10 shows the use of both HFRT and EMEME in low noise (laboratory) conditions. In order to make the test more representative, and also to further challenge the algorithms, artificial Gaussian noise was introduced into the signals using MATLAB prior to processing.[41] The level of noise inserted into the signal generated a signal-to-noise ratio (SNR) of 5. The results of the algorithms applied to the noisy signal are shown in Figure 11 with the HFRT on the left and the EMEME on the right.

With the addition of the noise the standard HFRT processing performs less well, particularly in relation to the 1st harmonic which is barely visible in the spectrum. The EMEME technique, however, still clearly shows the primary fault frequency with the 1st harmonic, as well as a further four harmonics to a lesser extent.

Increasing the level of noise introduced into the signals from an SNR of 5 to an SNR of 1.5 further increases the challenge to the algorithms. The results, shown in Figure 12, indicate that the noise floor in the spectrum resulting from the HFRT processing is increased to a point where only the fundamental fault frequency is clearly identifiable. Conversely, even with this increased level of noise, the EMEME processing still provides clear indications of both the fundamental fault frequency and the 1st harmonic.

The enhancement of energy moment feature extraction method. To illustrate the advantages of using the energy moment technique to select the representative IMF, a comparison has been made between it,
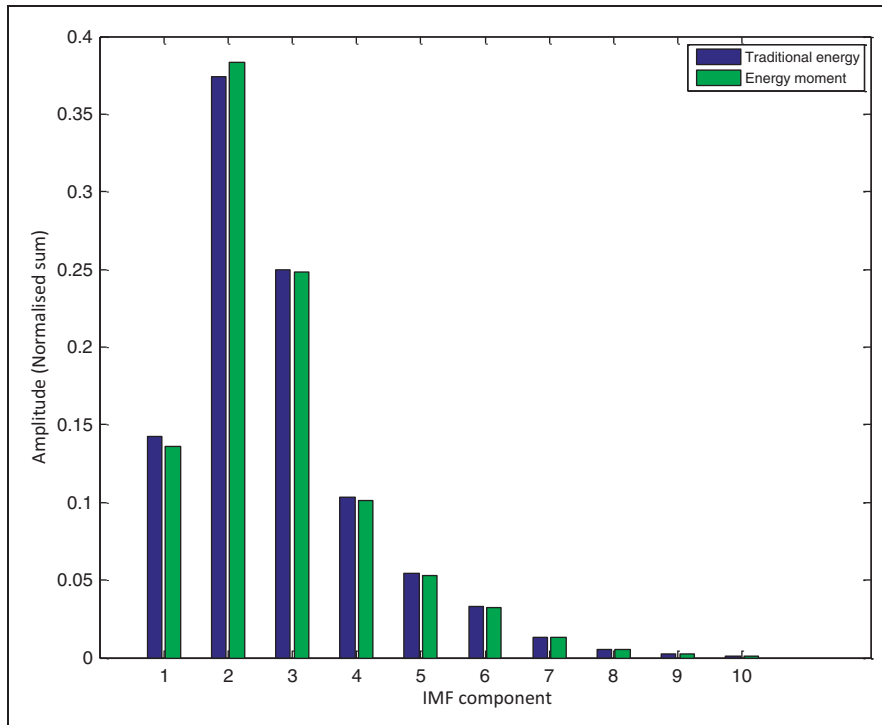
Figure 13. Distribution of amplitude values of IMF components generated using the energy moment and traditional energy techniques.
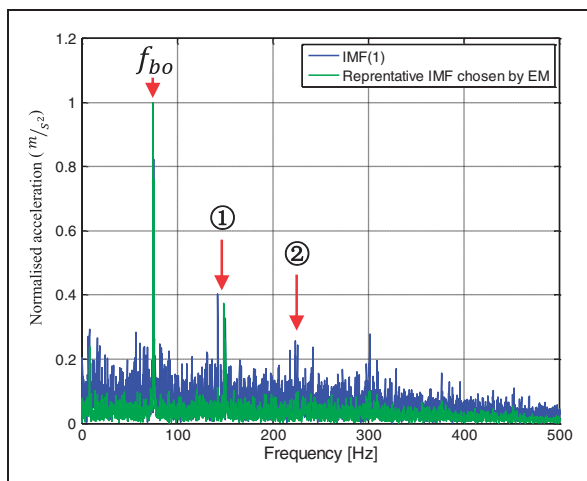
Figure 14. FFT envelopes generated using the energy moment feature extraction method and IMF(1).

traditional energy evaluation methods,[34] and selecting just the first IMF as used in Ricci et al.[26]

Taking the case of the F1 faulty bearing as an example; a chart showing the results of the traditional energy evaluation method and the energy moment technique for different IMFs is presented in Figure 13. The chart shows the distribution of the amplitude values for the top 10 IMFs, calculated as per equations (7) and (8) in the case of the energy moment and as per Yu et al.[34] for the traditional energy technique.

In both cases, the second (rather than the first) IMF is found to be that with the highest value.

To further evaluate the selection of the second IMF (as selected by the energy moment feature extraction method) compared to the first IMF, FFTs of the processed signals are shown in Figure 14. The noise level present in the spectrum corresponding to the chosen IMF, IMF(2), is reduced and hence the shaft speed and the 2nd harmonic of the fault frequency are more clearly visible.

For this example, and with minimal noise in the signal, Figure 13 shows only a very minor difference in amplitude between the 2nd IMF identified using the energy moment technique and the traditional energy approach. This difference is greater in scenarios with lower signal to noise ratios.

In order to evaluate the use of the energy moment IMF selection technique for noisy environments; data was collected from the same faulty bearing installed in a complex system including a Radicon Series J gearbox with a 4.94:1 ratio.[42] Figure 15 shows the signals recorded from both the faulty bearing (left), and the faulty bearing coupled to the gearbox (right).

Figure 16 shows the distribution of the amplitude values of the top 10 IMFs for the traditional energy and energy moment evaluation techniques in this case. In this case, with increased signal noise, the IMF selected by the two techniques is different. The energy moment technique selects the second IMF, while the traditional energy approach selects the fourth.

As with the original example, the FFTs of the processed signals are shown in Figure 17. All of the three techniques considered identify a key vibration
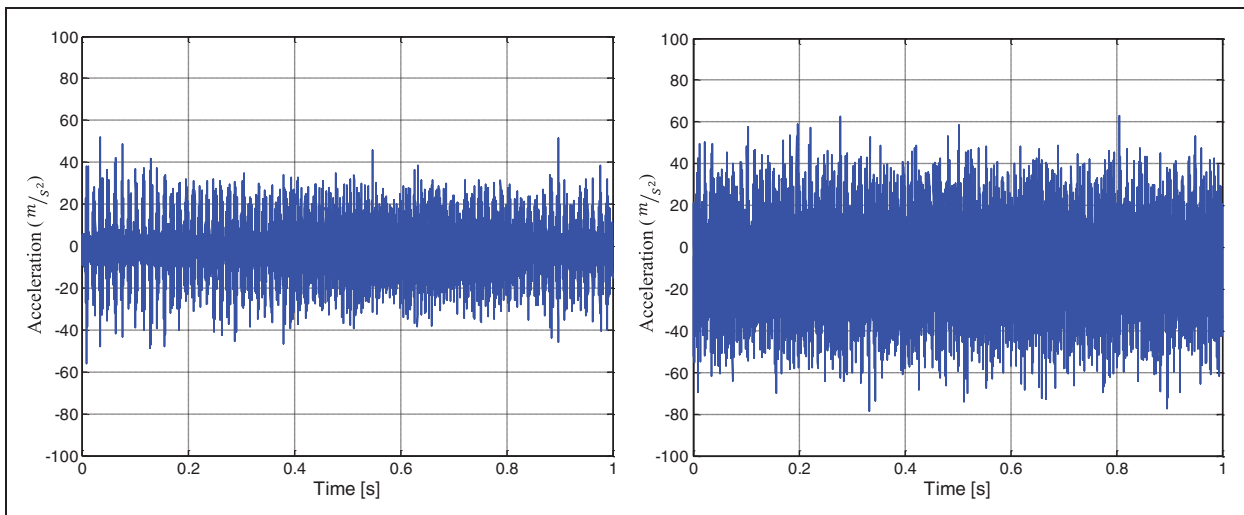
**Figure 15.** Signals from faulty bearing, (left) simple system (right) complex system including gearbox.
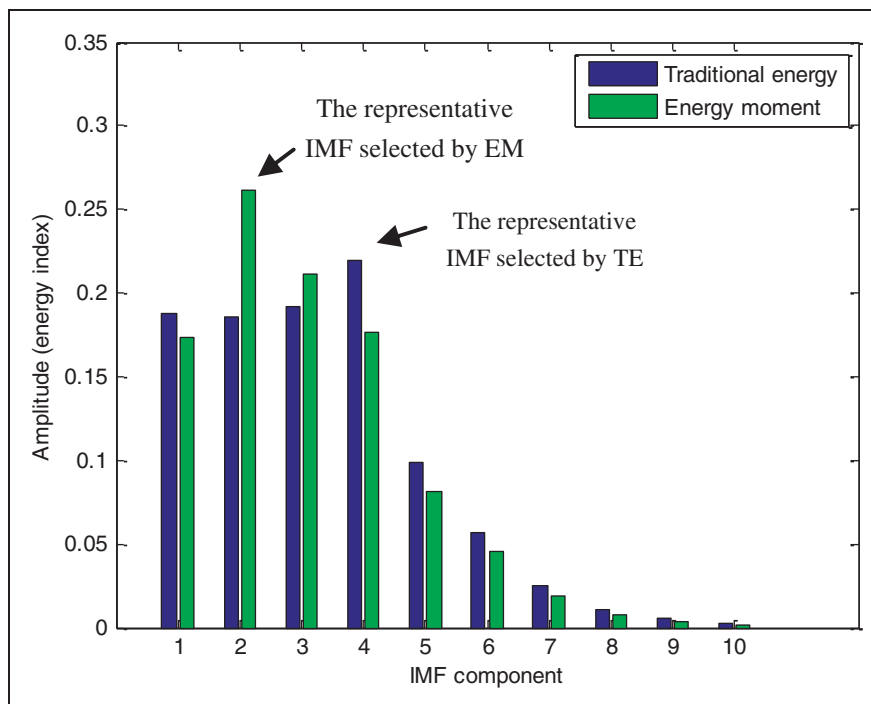
Figure 16. Distribution of amplitude values of IMF components generated using the energy moment and traditional energy techniques for the complex system including gearbox noise.

resulting from and outer race fault, the ball frequency of the outer race (f_{bo}). The graph on the left presents a comparison between the energy moment feature extraction method and selecting only IMF(1) (the technique used in Ricci et al.[26]). The graph on the right shows a comparison between the energy moment and traditional energy methods for IMF selection.

In the figure, the greatest fault frequency amplitude is seen to be in the envelope generated from the IMF selected using the energy moment processing technique. The traditional energy approach generates greater peak values than those generated from IMF(1).

Conclusions

This paper has described the development and application of a new algorithm for vibration-based bearing diagnostics to be applied to non-stationary experimental signals. The technique, referred to as empirical mode envelope with minimum entropy, combines elements of both the empirical mode decomposition and minimum entropy techniques with energy moment based automatic IMF selection routines in order to improve fault feature identification. The improved functionality of the technique has been demonstrated through its application to data collected from healthy
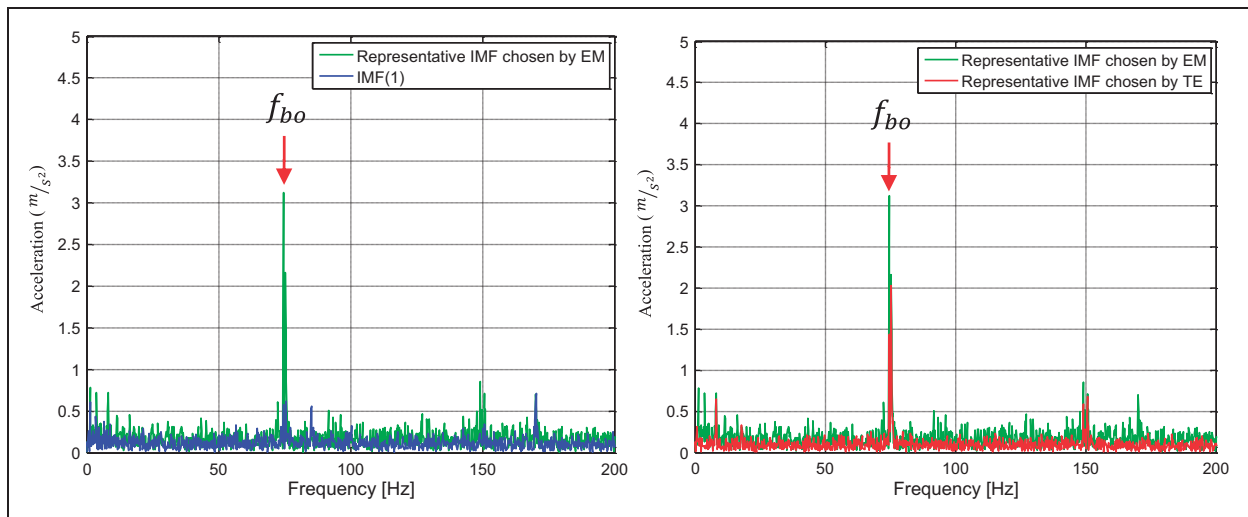
Figure 17. FFT envelopes for a signal including gearbox noise: (left) IMF selected using the energy moment technique (EM) and for IMF(1); (right) IMFs selected using the energy moment technique (EM) and the traditional energy (TE) method.

bearings, bearings with roller faults, and bearings with race faults; both in isolation and coupled to more complex systems, and at a range of rotational speeds.

The EMEME technique has been applied to the data from the test rig to detect faults and to generate frequency spectra which demonstrate the key frequencies associated with each different fault type. The clarity of these elements of the spectra is indicative of the improved performance of the technique. Tests have been carried out with particularly minor levels of the selected faults in order to demonstrate the technique's ability to detect early stage (minor) symptoms. Additional comparisons with conventional envelope processing and EMD/MED techniques using a statically selected IMF have also been carried out. This has shown that the proposed technique is also suitable for use in complex systems and noisy environments, and provides a clearer identification of faults within the results spectra.

In addition to the suitability of the EMEME technique for identifying the fundamental fault frequency in a complex system, results presented in this paper indicate improved capability for the identification of harmonics associated with this fundamental fault frequency. These harmonics can be used to improve the reliability of the classification of a fault associated with the bearing.

Highlights

- A new vibration-based signal processing algorithm for fault detection and diagnosis of roller bearing is introduced.
- A deep investigation of the combination of two data-adaptive techniques which are empirical mode decomposition (EMD) and minimum entropy deconvolution (MED) is demonstrated in this paper. The energy moment technique is applied in the most appropriate intrinsic mode

function (IMF) selection, prior to the MED algorithm being applied.
- The new technique is verified using the experimental data of different fault types at a range of speeds using a bespoke test facility housed at the University of Birmingham.
- The new technique is compared with the conventional method, high frequency resonance technique (HFRT), to demonstrate the enhancement.
- The stage of the most appropriate IMF selection in the new technique is compared with the other selection methods, to demonstrate the enhancement.

References

1. Tandon N and Choudhury A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribol Int* 1999; 32: 469–480.
2. R. A. I. U. Ireland. Investigation report: Bearing failure on a train at Connolly Station. In: *RAIU*, Ireland, 2011.
3. Cheng J, Yu D, Tang J, et al. Application of frequency family separation method based upon EMD and local

Hilbert energy spectrum method to gear fault diagnosis. *Mech Mach Theory* 2008; 6: 712–723.

4. Choudhury A and Tandon N. A theoretical model to predict vibration response of rolling bearings to distributed defects under radial load. *Tran ASME, J Vib Acoust* 1998; 1: 214–220.

5. Darlow MS, Badgley RH and Hogg GW. Application of high frequency resonance techniques for bearings diagnostics in helicopter gearboxes. US Army Air Mobility Research and Development Laboratory, 1974.

6. Du Q and Yang S. Application of the EMD method in the vibration analysis of ball bearings. *Mech Syst Signal Process* 2007; 21: 2634–2644.

7. Endo H and Randall R. Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. *Mech Syst Signal Process* 2007; 21: 906–919.

8. Farshidianfar A, Abbasi AR and Abbasion S. Dynamic modelling of chaotic response of bearing systems due to surface defects. In: *16th annual (international) conference on mechanical engineering-ISME2008*, Shahid Bahonar University of Kerman, Iran, 14–16 May 2008.

9. Wardle F and Poon S. Rolling bearing noise - cause and cure. *Chart Mech Eng* 1983; 30: 36–40.

10. Lynagh N, Rahnejat H, Ebrahimi M, et al. Bearing induced vibration in precision high speed routing spindles. *Int J Mach Tools Manuf* 2000; 40: 561–577.

11. Vafaei S and Rahnejat H. Indicated repeatable runout with wavelet decomposition (IRR-WD) for effective deteriation of bearing-induced vibration. *J Sound Vib* 2003; 260: 67–82.

12. Vafaei S, Rahnejat H and Aini R. Vibration monitoring of high speed spindles using spectral analysis techniques. *Int J Mach Tools Manuf* 2002; 42: 1223–1234.

13. Antoni J and Randall RB. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mech Syst Signal Process* 2009; 23.

14. Rai V and Morhanty A. Bearing fault disgnosis using FFT of intrinsic mode functions in Hilbert-Huang transform. *Mech Syst Signal Process* 2007; 21: 2607–2615.

15. Johnston A and Stronach A. Bearing fault detection in a hostile environment. In: *Proceedings of international conference on condition monitoring*, Brighton, UK, May 1986.

16. McFadden P. Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration. *Mech Syst Signal Process* 1987; 1: 173–183.

17. Badaoui M, Antoni J, Guillet F, et al. Use of the moving ceptrum integral to detect and localise tooth spalls in gears. *Mech Syst Signal Process* 2001; 15: 873–885.

18. Randall RB and Antoni J. Rolling element bearing diagnostics-a tutorial. *Mech Syst Signal Process* 2011; 25: 485–520.

19. Elasha F, Mba D and Ruiz-Carcel C. Effectiveness of Adaptive filter algorithms and spectral kurtosis in bearing faults detection in a gearbox. In: Sinha JK (ed.) *Vibration engineering and technology of machinery.*

New York: Springer International Publishing, 2014, pp.219–229.

20. Wiggins R. Minimum entropy deconvolution. *Geoexploration* 1978; 16: 21–35.

21. Gonzalez G, Badra R, Medina R, et al. Period estimation using minimum entropy deconvolution (MED). *Signal Proceeding 41*, 1995.

22. Boumadhi M and Lacoume J. Blind identification using the kurtosis: Results of field data processing. *IEEE Trans Signal Process* 1995; 0-7803-2431-5: 1960–1983.

23. Lee J and Nandi A. Extraction of impacting signals using blind deconvolution. *J Sound Vib* 1999; 232: 945–962.

24. Nandi A, Mampel D and Roscher B. Blind deconvolution of ultrasonic signals in non-destructive testing application. *IEEE Trans Signal Process* 1997; 5.

25. Loutridis S. Damage detection in gear systems using empirical model decomposition. *Eng Struct* 2004; 26: 1833–1841.

26. Ricci R, Borghesani P, Chatterton S, et al. The combination of empirical mode decomposition and minimum entropy deconvolution for roller bearing diagnostics in non-stationary operation. In: *Proceedings of the ASME 2012 international design engineering technical conferences & computers and information in engineering conference*, Chicago, Illinois, USA, 2012.

27. Chatterton S, Ricci R and Pennacchi P. Signal processing diagnostic tool for rolling element bearings using EMD and MED. In: Dalpiaz G, Rubini R, D'Elia G, et al. (eds) *Advances in condition monitoring of machinery in non-stationary operations*. New York: Springer, 2014, pp.379–388.

28. Bin GF, Gao JJ, Li XJ, et al. Early fault diagnosis of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network. *Mech Syst Signal Process* 2012; 27: 696–711.

29. Huang N, Shen Z, Long S, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc Roy Soc Lond Ser A* 1998; 454.

30. Yang W and Tavner P. Empirical mode decomposition, and adptive approach for interpreting shaft vibration signals of large rotating machinery. *J Sound Vib* 2009; 321: 1144–1170.

31. Ricci R and Pennacchi P. Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions. *Mech Syst Signal Process* 2011.

32. Yu Y and Cheng J. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J Sound Vib* 2006; 1: 269–277.

33. Pennacchi P, Ricci R, Chatterton S, et al. Effectiveness of MED for fault diagnosis in roller bearings. In: *Vibration problems ICOVP*, 2011, pp.637–642.

34. Yu D, Cheng J and Yang Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of rooler bearings. *Mech Syst Signal Process* 2005; 19: 259–270.

35. Sunnersjo C. Rolling bearing vibrations - geometrical imperfections and wear. *J Sound Vib* 1985; 4: 455–474.

36. Meyer L, Ahlgren F and Weichbrodt B. An analytic model for ball bearing vibrations to predict vibration response to distruted defects. *Trans ASME, J Mech Des* 1980; 3: 195–207.

37. Washo M. A quick method of determing root causes and corrective actions of failed ball bearings. *Lubric Eng* 1996; 3: 206–213.

38. Ganeriwala S. Review of techniques for bearings & gearbox diagnostics. In: *IMAC conference*, Richmond, USA, 2010.

39. Parey A and Pachori RB. Modified empirical mode decomposition process for improved fault diagnosis. In: *The 8th IFToMM international conference on rotor dynamics*, Seoul, Korea, 2010.

40. Dron J, Bolaers F and Rasolofondraibe I. Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearing. *J Sound Vib* 2004; 270: 61–73.

41. The MathWorks, Inc. Generate white Gaussian noise, http://uk.mathworks.com/help/comm/ref/wgn.html (accessed 2 December 2014).

42. Radicon, "Radicon," http://www.radicon.com/_docs/salagears%20eng.pdf (accessed 12 February 2015).

43. Ohue Y, Yoshida A and Seki M. Application of the wavelet transform to health monitoring and evaluation of dynamic characteristics in gear sets. *Proc IMechE, Part J: J Engineering Tribology* 2004; 218.

44. Du Q and Yang S. Application of the EMD method in the vibration analysis of ball bearings. *Mech Syst Signal Process* 2007; 21: 2634–2644.

Appendix

Notation

$a(t)$	envelope
BPFO (f_{bo})	ball passing frequency outer race
BPFI (f_{bi})	ball passing frequency inner race
BFF (f_{bf})	ball fault frequency
BSF (f_{bs})	ball spin frequency
$C(t)$	intrinsic mode function
E	percentage of the energy moment in term of the series
f	frequency
$f(t)$	filter function
FTF (f_c)	fundamental train frequency
$g(t)$	a generic signal
$h(t)$	system function
H	Hilbert transform
k	index of data sample
m	total number of data samples
M	length of the filter
$Max(t)$	maxima in term of time
$Min(t)$	minima in term of time
$mean(t)$	average in term of time
n	number of iteration
n_0	index of the representative IMF
N	final number of iteration
PD	mean roller race diameter
$r(t)$	residual signal
R	number of rollers
RD	roller diameter
t	time
T	length of the convolution between signals
$u(t)$	filter output
V	vatimax norm
$w(t)$	excitation
$x(t)$	a generic signal
$z(t)$	analytic signal
β	contact angle
δ	threshold
Δt	sampling period
$\eta(t)$	randomly distributed noise
ω_r	shaft rotation rate
$\varphi(t)$	angular characteristic