ARMY RESEARCH LABORATORY

A Primer on Vibrational Ball Bearing Feature Generation for Prognostics and Diagnostics Algorithms

by Kwok F Tom

ARL-TR-7230 March 2015

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1138

ARL-TR-7230 **March 2015**

A Primer on Vibrational Ball Bearing Feature Generation for Prognostics and Diagnostics Algorithms

Kwok F Tom
Sensors and Electron Devices Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
March 2015	Final	10/2012–09/2013

4. TITLE AND SUBTITLE
A Primer on Vibrational Ball Bearing Feature Generation for Prognostics and Diagnostics Algorithms

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kwok F Tom

5d. PROJECT NUMBER
R.0005886.8

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-SER-E
2800 Powder Mill Road
Adelphi, MD 20783-1138

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7230

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report is the result of a prognostic and diagnostic program involving roller bearings. The objective of the effort was to develop techniques that could be used to detect the initial fault and predict the remaining useful life of a roller bearing. There are many techniques from digital signal processing, statistical, and machine learning fields that can be for fault detection and prediction. In this report, a description of roller bearing faults and life are presented. From this starting point, the report leads into various techniques that can be applied to vibrational data in order to generate features that can be used for fault detection. Feature generation is an important step in the prognostic and diagnostic development. This overview of possible features is intended to provide sufficient information to pursue feature selection and algorithm development for roller bearings prognostic and diagnostic techniques.

15. SUBJECT TERMS
feature generation, ball bearing fault frequency, signal processing techniques

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kwok F Tom
Unclassified	Unclassified	Unclassified	UU	54	19b. TELEPHONE NUMBER (*Include area code*) 301-394-2612

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

The purpose of this primer is to provide information and insight into the features that may be used to develop prognostic and diagnostic algorithms for determining the health of a ball bearing. Condition-based maintenance (CBM) is the new paradigm for the Army. CBM is a change of maintenance operation where the fault is detected and a failed component is replaced when necessary. To correctly infer the health status of a piece of machinery, it would be ideal to embed a "built-in test" capability into the hardware during its development cycle. As of this writing, such capability does not typically exist, and even if it did, the applied diagnostics techniques would likely be developed under the ideal usage case, which will not necessarily cover the full domain of usage cases. It would be wonderful if the output of the sensor could be read out directly and a corresponding decision determined based on the reading, but in practice this is usually not the case for mechanical systems. Sensors will be needed, as well as the correct interpretation of the sensor information they provide. Furthermore, vast amounts of statistical data will need to be collected in order to develop and train algorithms for diagnostics and prognostics. A very critical piece of information that needs to be collected in the development of prognostic and diagnostic algorithms is the "ground truth," which permits correlation to the actual health condition of the hardware that is being monitored. This data-driven methodology paradigm is necessary in order to develop the proper detection of a fault, its meaning, and its remaining useful life.

A data-driven approach was studied at the US Army Research Laboratory (ARL) for a mechanical system. In order to develop the appropriate algorithms, a well-controlled test needed to be executed that measured the sensor response as the hardware was exercised from a health operational state into its end of life with clear health states during the test. The applied techniques and algorithms are derived from the many fields of statistics and probability, digital signal processing, pattern recognition, data mining, and machine learning.

The development of prognostic and diagnostic algorithms involves exploring many techniques that can be used to provide anomaly detection, classification, and regression analysis. One can develop these algorithms based on evaluation of these techniques, but a domain expert may be needed to interpret the operational status of the hardware. These algorithms are highly dependent on the sensor information. An accelerometer device was the primary sensor employed to provide data for the oil cooler bearing evaluation under a previous project. The sensor device should be capable of capturing information related to the hardware platform as correlated to fault, degradation, and end of life indicators. In many applications, this information cannot be determined directly from the sensor data. The raw sensor data will need to be mapped to other feature sets that provide clear indicators of health status.

This report provides insight into the life of a roller bearing as it degrades. These components have been studied over many decades and its characteristics described. One important aspect has been the development of features that relate bearing degradation in its various phases of life. It is important to the algorithms that these features correlate very well with the bearing's telltale signs. These features should correlate with a measureable progression as the bearing operates over its life. Typically, the features need to have sufficient a signal-to-noise ratio and some growth related to the component's degradation.

2. Bearing Construction

A bearing serves the purpose of a load-carrying member that allows a component to rotate with respect to a mechanical assembly. The mechanical coupling is provided through a shaft that engages the inner race component of the bearing. An example of a bearing component and construction is illustrated in Fig. 1. In this case, a ball bearing is shown. There are other bearing constructions that use taper cylinder components in place of the roller ball, but the principles are the same. Given the manufacturer and model number, the mechanical specification can be determined from the manufacturer's catalog.

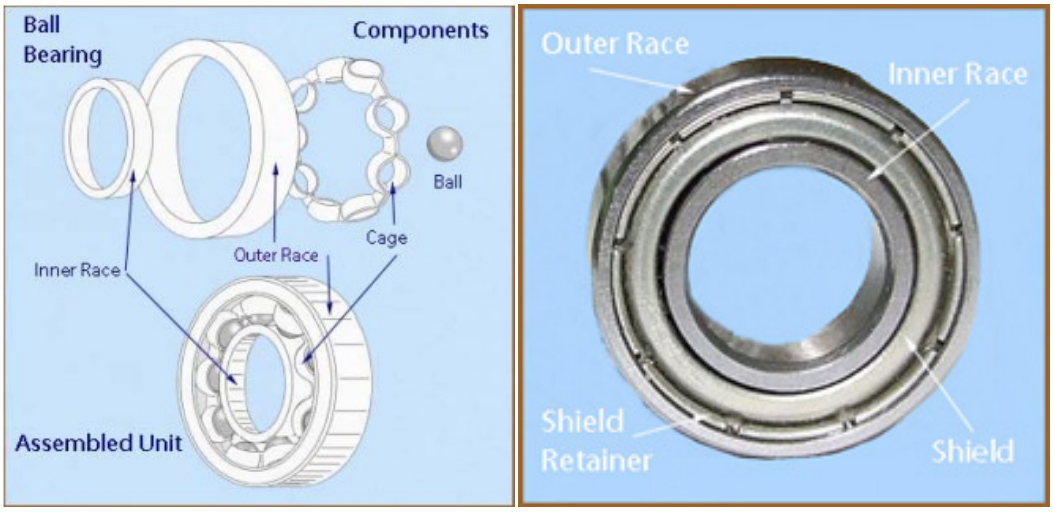

Fig. 1 Ball bearing illustration[1]

3. Failures

A summary of analysis from the literature in terms of deterioration of the bearings resulted in the identification of the following failure modes:[2]

- Fatigue – the degradation of the material due to normal usage over time. Minute cracks develop in the bearing surface and eventually progress to the surface where the material will separate. Also known as pitting, spalling, or flaking.
- Wear – normal degradation caused by dirt and foreign particles causing abrasion of the contact surfaces over time resulting in alterations in the raceway and ball bearings.
- Plastic deformation – alterations in the contact surfaces as a result of excessive loading while stationary or during small movements.
- Corrosion – the degradation as a result of water or other contaminants in the lubrication of the bearing. Oxidation rust products are formed on the surfaces and interfere with the lubrication and rolling operation of the bearing. The subsequent abrasion results in wear, flaking, and spalling.
- Brinelling – formation of regularly spaced indentations distributed over the raceway corresponding to the Hertzian contact area. Possible causes are static overloading or vibration and shock loads when in a stationary position. This can lead to spalling.
- Lubrication – the lack of sufficient lubricant that leads to skidding, slip, increased friction, heat generation, and sticking. This can also anneal the bearing elements reducing their hardness and fatigue life.
- Faulty installation – includes excessive preloading in either radial or axial directions, misalignment, tight fits, loose fits, or damage in the installation process.
- Excessive loads – self explanatory.
- Overheating – self explanatory.
- Seizing – self explanatory.

In most of these faults, a spall develops that indicates a fault in the bearing. Spalling of the bearing components provides mechanical responses that can be transduced by an accelerometer from a mechanical vibration into an electrical signal. The first occurrence of a spall indicates an incipient fault, but does not necessarily mean that it should be immediately replaced. The goal and hope is that one can detect this fault early enough in order to monitor the condition and replace component at a convenient time if possible.

4. Bearing Life

Engineers use the L_{10} or basic life model of a bearing, as part of the design process in selecting the appropriate bearing for the intended application. The International Organization for

Standardization (ISO) and American Bearing Manufacturers Association (ABMA) defines L_{10} as $L_{10} = \left(\frac{C}{P}\right)^p$ for 1 million revolutions where

$$C = basic\ dyamic\ load\ rating, lb$$
$$P = equivalent\ dynamic\ bearing\ load, lb$$
$$p = life - equation\ exponent$$
$$(p = 3\ for\ ball\ bearings; and\ p = \frac{10}{3}\ for\ roller\ bearings)$$

This basic life or L_{10} represents a life where 90% of a sufficiently large group of identical bearings can be expected to reach or exceed. This is a first step in predicting service life based on known operating conditions using a "model" based approach, but should be used cautionary. Other definitions of bearing life are summarized in Table 1.

Table 1 Definition of bearing life[3]

Life Type	Definition
Basic or L_{10}	When the bearing has reached 90% of its life as defined as 1 million revolutions
Median or average or Mean Time Between Failure (MTBF)	About 5 times Basic life
Service	Life of under actual operating conditions before it fails or needs to be replaced
Specification	Similar to Basic life. Manufacturer's requirement for bearing.

It would be nice if the bearing life process was "linear." The assumption could be made that all faults develop in the same way. In this instance, there would be a gradual degradation of the bearing condition and faults would occur similarly every time. This is not the real world.

As listed above, there are many ways that a bearing may fail: cracks; true and false brinelling; rust and corrosion; etc. Bearing degradation has been studied for decades and a general model has been made to illustrate the life of a bearing through its 4 stages.[4] One would expect to see a progression through each of its degradation stages, but that is not necessarily the case. It may actually skip some stages of its life.

A noted by the Mobius Institute,[4] as the bearing fails (depending upon the type of failure), there will be moments when cracks appear, pieces of metal flake away, and so on. At that moment, the vibration pattern and amplitude may change due to sharp edges to impact against the rolling elements and a piece of metal inside the bearing. The vibration measurement at that time may lead one to believe that the fault is quite severe. However, as the rolling elements continually strike the sharp surface, the edges will become rounded, and the metal pieces may be carried away by the lubricant. Therefore, the vibration will therefore change and lead one to think that the situations where the vibration appears to improve, but not really.

Per Barlov and Barkova,[5] from decades of bearing evaluations, a bearing model has been developed that provides indicators of end of life at its 80% point onward. It has been proposed that bearing lifetime prediction be broken down into a 2-step process. A long-term life monitoring up to 20% of a bearing's specific service life may be possible with low computational techniques and algorithms. Predicting the remaining service life at any point in time is very approximate and can be estimated only by introducing other computationally complex feature sets.

5. Bearing Fault Stages

Typically, rolling element bearings operate for approximately 80% of their useful life defect free. When failure occurs, there are generally 4 distinct stages of failure indicators. An early fault in the bearing does not necessarily mean that the bearing life is at hand. The bearing fault indicators or features are clearly detected in the frequency domain if the signal is sufficiently above the noise level. Figure 2 is an ideal depiction of frequency response at these various stages. In a bearing that is considered "healthy," the frequency response should be Gaussian or flat. There may be some frequency components that correspond to the shaft rotation. In the real world, the healthy bearing will actually have some non-flat shape to its frequency response. The 4 stages represent the last 20% of bearing life and is not a linearly proportioned in terms of its remaining life cycle. The frequency response is divided into 4 zones or areas of interest in the frequency domain, and one should be aware that the frequency axis is not linearly drawn. The spectral content is just a snapshot in time and is not meant to imply constant features throughout those stages.

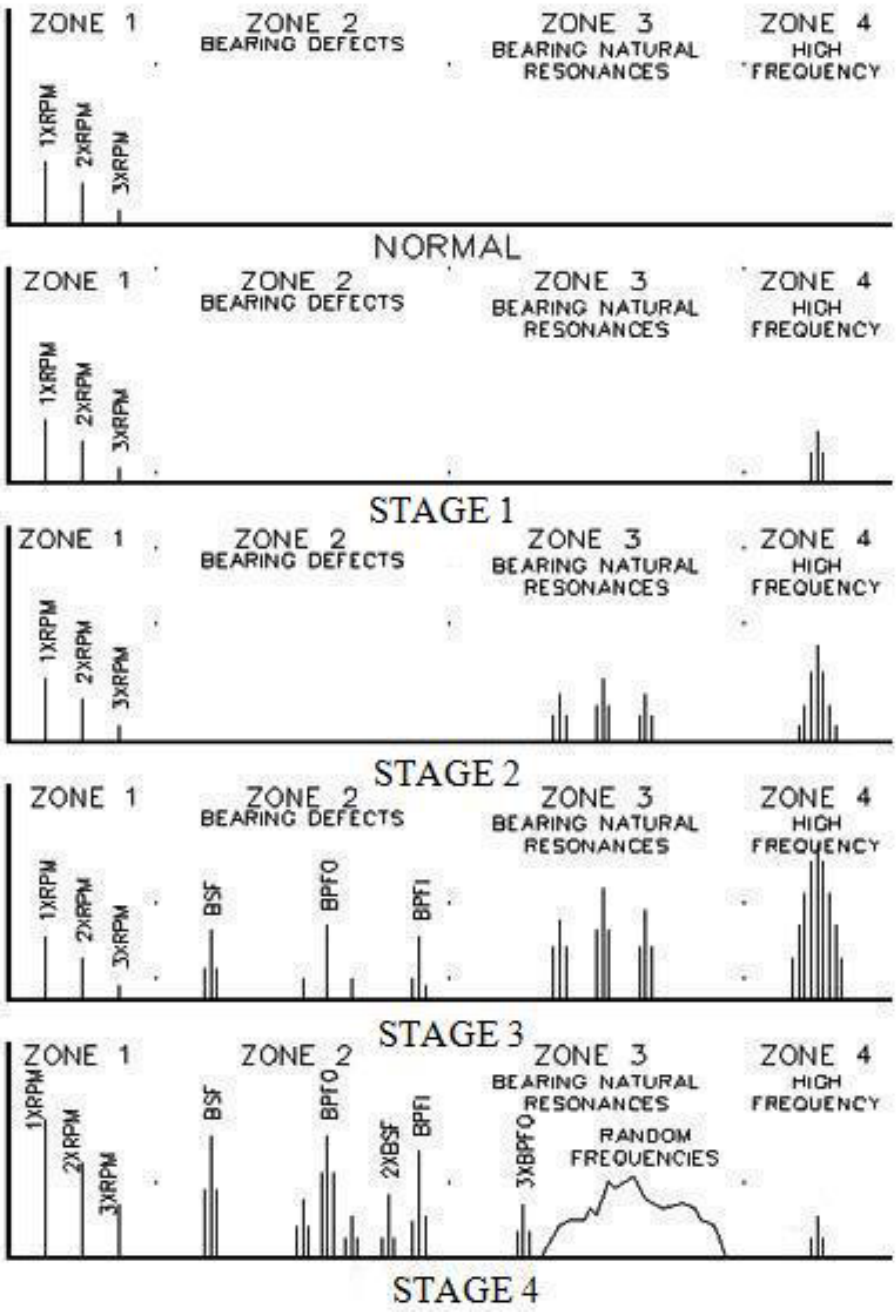

Fig. 2 Bearing life model (reproduced with permission from Mike Howard, STI Vibration Monitoring, Inc)[6]

5.1 Stage 1

Stage 1 represents 10% to 20% of the bearing's remaining life. In Fig. 3 the bearing is still considered a good bearing. Failures in this stage normally occur below the surface so a visual examination would not be revealing. They normally begin 4 to 5 thousandths of an inch (0.1 to 0.125 mm) below the surface of the raceway. There are many techniques developed by various vendors to detect the energy in this part of the frequency spectrum. Sub-surface cracking

generates very low amplitude stress waves in the 300 to 500 kHz range. A stress wave sensor would be used to detect the energy in this part of the frequency spectrum. Earliest indications of bearing problems appear in ultrasonic frequencies ranging from approximately 20–60 KHz. An approximate frequency span for the high frequency region is 2 to 120 KHz. Techniques that have been developed by commercial vendors are spectra emission energy (SEE), spike energy spectrum (gSE), high frequency detection (HFD), and shock pulse method (SPM):

- SEE:
    - Developed by SKF Condition Monitoring Group.
    - Uses a high frequency acoustic emission sensor with an enveloping technique:
        - The signal is bandpassed (250–350 KHz).
        - The signal is filtered and enveloped.
        - The signal is lowpassed to remove high frequency content.
        - The signal is transformed into the frequency domain.
- gSE:
    - Developed by IRD.
    - The signal is highpassed to remove low frequencies.
    - The signal is rectified to capture an impact response.
    - The signal is digitalized and lowpassed to obtain an envelope.
    - The signal is transformed into the frequency domain.
- HFD:
    - Developed by SKF and CSI.
    - The signal processed in the 5–60 KHz range.
    - The signal uses a sensor resonant for amplifying the bearing defect.
    - The signal is converted to a value that represents the level of the bearing defect
- SPM:
    - Developed by SPM Instruments.
    - The signal is highpassed above 32 KHz to obtain the transient waveform.
    - The signal is converted into a series of analog pulses corresponding to the transient condition.

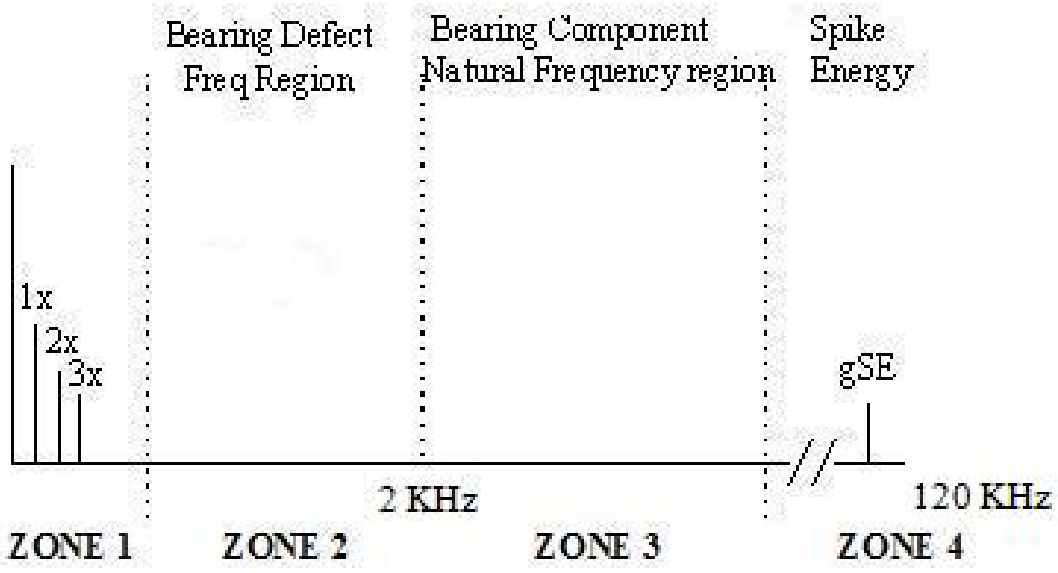

Fig. 3 Stage 1 fault (reproduced with permission from David Stevens, IEng, AV Technology) [9]

5.2 Stage 2

Stage 2 represents 5% to 10% of the bearing remaining life, as shown in Fig. 4. As the fault progresses microscopic pits (<40 microns) occur on the surface of the failed component. The defects are invisible to the naked eye and require magnification to observe the fault. As the fault continues to develop, the defects will evolve into spalls, cracks, flakes, etc. In the very early stages, the impacting created by the microscopic pits causes the bearing components to vibrate at their natural frequencies. These natural frequencies are in the 2 to 8 KHz range. The vibration pattern gradually changes as a result. The force of the impacts is greater, and periodicity is seen in the vibration measurements. Enveloping (demodulation) is also effective, with peaks visible at the bearing defect frequencies along with harmonics. Harmonics of the bearing defect frequencies may also be visible in the acceleration spectrum, and time waveform analysis may show signs of the fault. Depending on the data acquisition system period, it may have to be decreased to capture the fault progression growth. Bearing degradation is usually linear for a period to time and can be trended, but as the lifetime ends, it becomes nonlinear. At the end of stage 2, bearing defect frequencies appear, and sideband frequencies may also be present above and below the defect frequencies.[4,7–10]

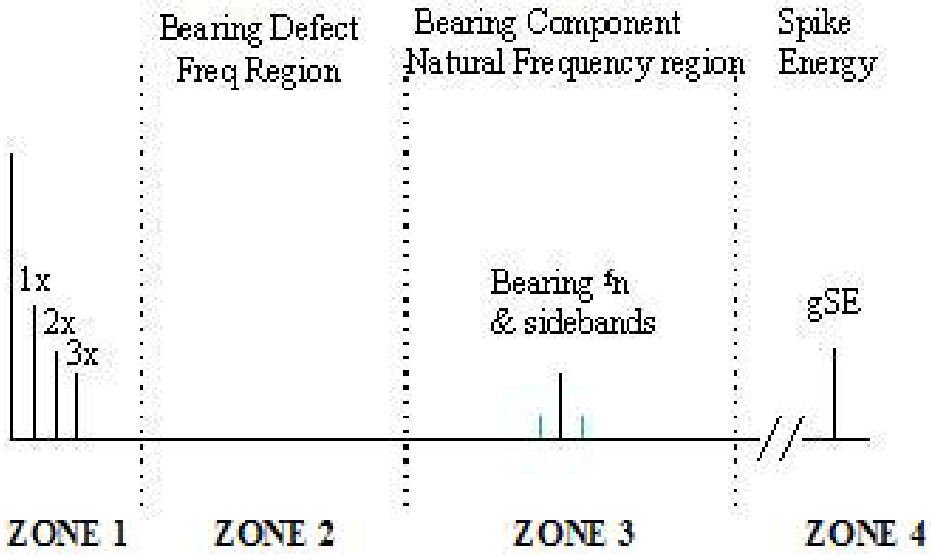

Fig. 4 Stage 2 fault (reproduced with permission from David Stevens, IEng, AV Technology)[9]

5.3 Stage 3

Stage 3 represents 1% to 5% of the bearing remaining life, as shown in Fig. 5. Further failure progression causes the initial flaking, cracking and/or spalling that is commonly associated with rolling element bearing failures. The fault is visually apparent in the raceways and/or rolling elements, yet the damage is still confined to the bearing itself. These faults cause the signal to be strong enough to generate signals at the bearing defect frequencies. This is the point in failure where acceleration measurements can first positively identify a bearing defect.[4,7–10]

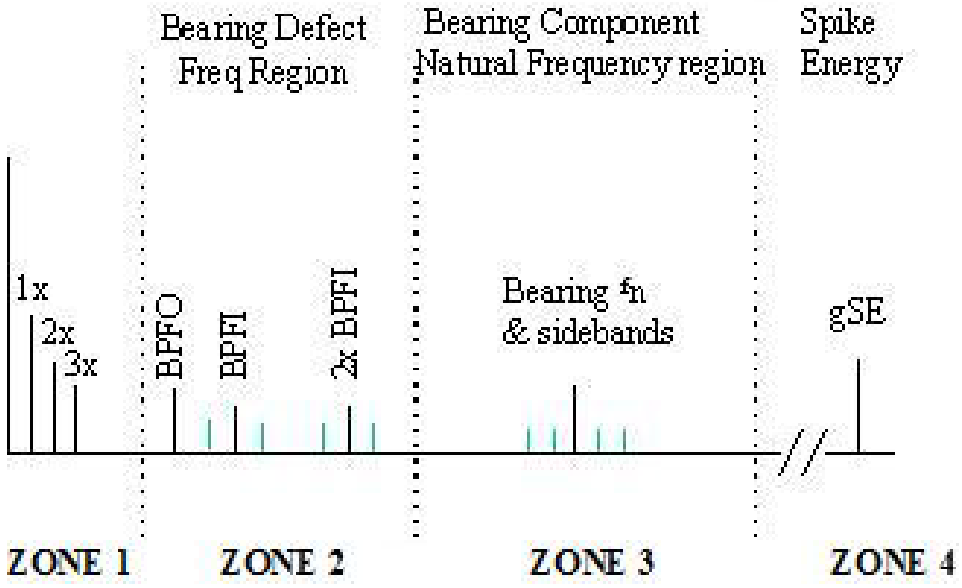

Fig. 5 Stage 3 fault (reproduced with permission from David Stevens, IEng, AV Technology)[9]

In Stage 3, the amplitude of the bearing defect frequency region increases and harmonics appear. At some point, other bearing defect frequencies are noticeable in both the bearing defect and natural frequency portions of the frequency domain. In addition to the increasing harmonics of the bearing fault frequencies, there are modulations or sidebands associated with the shaft rotating speed. The calculations for the bearing defect frequencies are for an ideal bearing. This may not necessarily be the situation for the bearing that one monitors; as the bearing degrades, the tolerances will change from their original state. These variations may have impact on the observed bearing defect frequencies resulting in slightly different frequencies from the ideal calculated bearing defect frequencies.[4,7–10]

At this point, there is significant damage as indicated by the bearing defect frequency and the damage can easily be seen through a visual inspection of the bearing. The high frequency techniques such as SPM and gSE will trend upward. Depending on where the fault originates, the bearing defect frequencies will exhibit different characteristics. A summary is as follows:[4,7–10]

- For an outer race bearing fault (horizontally oriented machine), there are harmonics at the ball pass frequency outer (BPFO). These harmonics are typically lower than the main or fundamental harmonic of the BPFO. As the bearing degrades, the amplitude of the harmonics will increase in amplitude to be greater the fundamental harmonic of the BPFO.
- For an inner race bearing fault, the spectral frequency content also displays harmonics of the ball pass frequency inner (BPFI) along with sidebands of the shaft rotational speed. These sidebands are observed at the fundamental and harmonic frequencies. As the shaft rotates, the inner race fault will rise and fall as it moves through the loading zone of the bearing. The sidebands are created due to amplitude modification of the inner race bearing fault signal. This phenomenon is known as amplitude modulation (AM).
- For a rolling element bearing fault, the fault may be noticed at the ball spin frequency (BSF) fundamental frequency. Since the fault of the rolling element will impact the inner and outer race per revolution of the shaft, the peak magnitude may be at twice the BSF frequency. In addition, there will be sidebands around the BSF harmonics, but the sidebands will be generated at the fundamental train frequency (FTF), also known as the cage frequency. The rolling element bearing fault travels through the bearing load zone with integration of the cage rotation.
- Examination of the actual sensor time data exhibits the impacts and associated modulation in the time waveform.

5.4 Stage 4

Stage 4 represents 1% to 1 revolution of the bearing remaining life, as shown in Fig. 6. When multiple cracks, excessive flaking, or spalling occur, this is the 4th and final stage of bearing failure. Often, the rolling elements begin to deform and the cage may disintegrate or break. In this final stage, the usual suspects (bearing defect frequency, its harmonics, and sidebands) may

actually disappear. The shaft has more freedom to move around inside the bearing. Additionally the noise floor of the entire spectrum may increase since the generated frequencies will no longer occur at exactly the same time interval.[4,7–10]

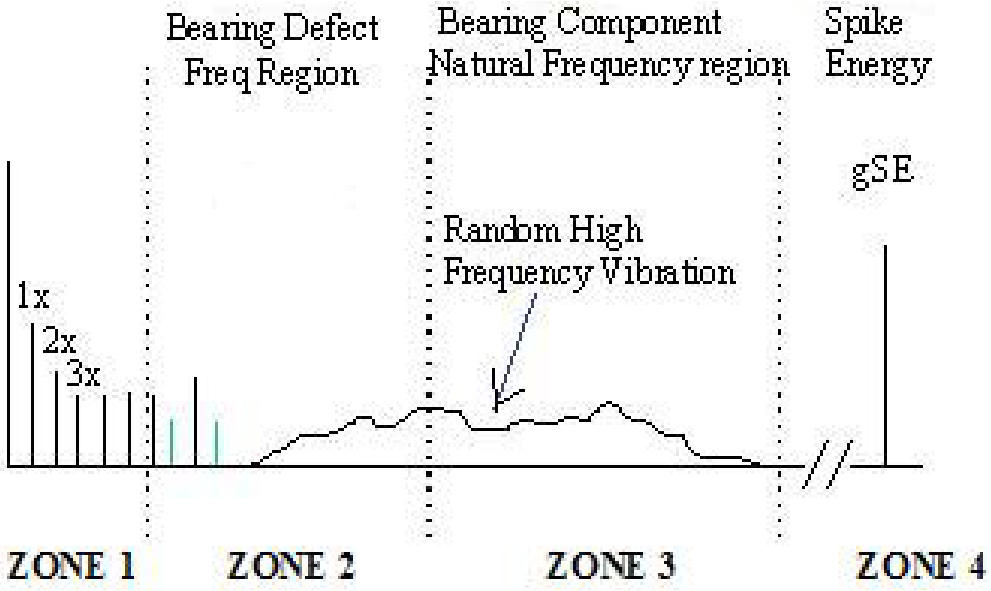

Fig. 6 Stage 4 fault (reproduced with permission from David Stevens, IEng, AV Technology) [9]

Numerous modulated frequencies and harmonics provide an indication that the defects are distributed around the bearing races and no longer localized. With the increase in degradation, the shaft vibrates more due to the degradation of the bearing internal clearances. This results in the increase in amplitude of the shaft rotation frequency. In addition, harmonics of the shaft rotational speed are produced. The bearing defect frequencies in both zones 2 and 3 start to decrease in amplitude and are replaced with a random broadband high frequency component. As it progresses toward the end of life, the amplitude of the high frequency noise and spike energy may decrease, but just prior to failure, the spike energy usually grows to excessive amplitudes. The following is summary of the progression in stage 4:[4,7–10]

- The impact point is worn out, so that there are no sharp edges; this reduces the bearing defect frequencies in the component high frequency resonance region.
- The bearing impact response does not show its periodicity. The bearing defect frequencies start to decrease in amplitude.
- The clearance of the bearing increases. This results in an increase in the looseness of the shaft with the bearing housing. As a result, the shaft rotational speed and its harmonics increase and are noticeable.
- The root mean square (RMS) value of the sensor time data increases. It is in this final stage that the time-domain feature displays sensitivity to the bearing fault.

6. Bearing Fault Frequency

In this report, the terms "fault" and "defect" are used synonymously. The spectral characteristics of bearing faults have been derived and defined as shown in Table 2. There are 5 basic motions that can be used to describe the dynamics of a bearing movement and each motion generates a unique frequency response. The following definitions of these fault frequencies are provided:[2, 12]

- Shaft rotational frequency (f_{ROT}) – the rotational frequency of the rotor or shaft is fundamental to the movements of bearings. In a steady-state operating condition, the bearing outer raceway can be assumed to be stationary, while the inner raceway is rotating at the speed of the shaft. If the inner ring is fixed (stationary) while the outer rotates with the shaft, the minus sign must be changed to plus sign within the parentheses of Cage Frequency equation.
- Fundamental Cage frequency (f_{FTF}) – the rotational frequency of the cage of the bearing.
- Ball Pass Outer Raceway frequency (f_{BPFO}) – this is defined as the rate at which the balls pass a point on the track of the outer raceway.
- Ball Pass Inner Raceway frequency (f_{BPFI}) – this is defined as the rate at which the balls pass a point on the track of the inner raceway.
- Ball Spin frequency (f_{BSF}) – this is the rate of the rotation of a ball about its own axis.

The equations to calculate these frequencies are based on characteristics of the bearing, as outlined in Table 2. Figure 7 defines each of the parameters.

Table 2 Bearing defect frequency equations[2, 12]

Fault Type	Fault Frequency
FTF, also known as Cage	$f_{FTF} = \frac{f_{ROT}}{2}\left[1 - \frac{d}{D}\cos\alpha\right]$
BPFO	$f_{BPFO} = \frac{f_{ROT}}{2}N_b\left[1 - \frac{d}{D}\cos\alpha\right]$
BPFI	$f_{BPFI} = \frac{f_{ROT}}{2}N_b\left[1 + \frac{d}{D}\cos\alpha\right]$
BSF, also known as Roller	$f_{BSF} = \frac{f_{ROT}}{2}\left(\frac{d}{D}\right)\left[1 - \left(\frac{d}{D}\right)^2 cos^2\alpha\right]$
$where\ f_{ROT} = shaft\ rotational\ frequency$ $\alpha = contact\ angle$ $D = Pitch\ diameter$ $d = ball\ bearing\ diameter$ $N_b = number\ of\ ball\ bearings$	

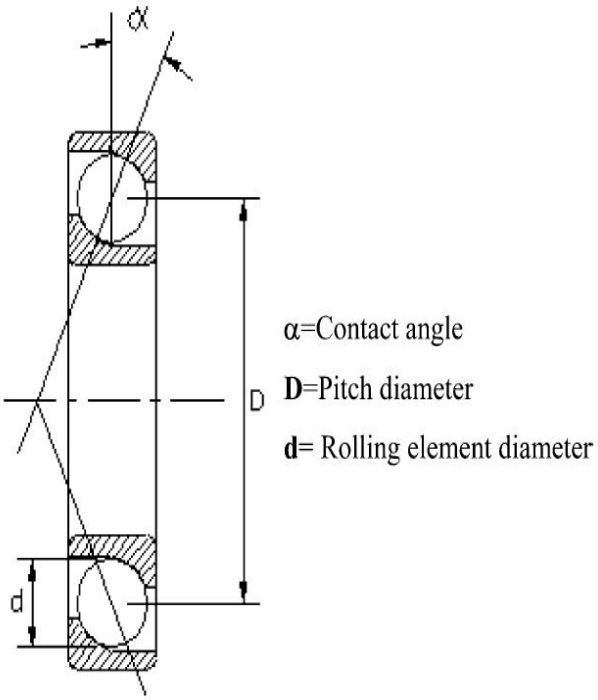

Fig. 7 Bearing parameters[2, 12]

The amplitude of the signal varies depending on not only the size of the defect, but the location of the defect in relation to the placement of the sensor, the tightness of the bearing flat, and the balance and/or alignment condition of the machine. If the fault in on the inner race, it will roll in and out of the load zone as the shaft rotates. When the impact occurs in the load zone, the strength of the signal is greater than when the fault occurs opposite the load zone.

6.1 Understanding Bearing Fault Frequencies

The following should be kept in mind when discussing bearing fault frequencies:[7]

1. The bearing fault frequencies are based on the kinematics of the bearing, i.e., the physical bearing geometry and rotational shaft speed.
2. There are multiple sources of vibration in a complex machine of which the bearing is a subcomponent. For an accelerometer mounted very close to the bearing, the potential source is the shaft. Typically, the bearing fault frequencies are not integer multiplies of the rotating speed.
3. Ideal conditions are assumed in the calculations. The shaft and bearing are rotating in uniformly. Operation of the internal bearing components (outer race, inner race, and ball elements) is pure rolling contact.
4. Calculation of the bearing fault frequencies may not necessarily be exact. Differences between the calculated frequency and measured frequencies may be due to skidding versus rolling contact. Variation between the calculated and measured bearing fault frequencies could be in the order of 5% to 10% different.
5. The magnitude of the bearing fault frequency does not necessarily indicate the severity of the fault, but the presence of a fault.

There are "rule of thumbs" for the calculation of bearing fault frequencies when the bearing geometry is not known. These are summarized in Table 3.

Table 3 Bearing defect frequency estimates[12–14]

Fault Type	Fault Frequency Estimate
FTF, also known as Cage	$0.4\frac{RPM}{60}$
BPFO	$0.4N_b\frac{RPM}{60}$
BPFI	$0.6N_b\frac{RPM}{60}$
BSF, also known as Roller	$0.2N_b\frac{RPM}{60}$
$where\ RPM = shaft\ rotational\ speed$ $N_b = number\ of\ ball\ bearings = \frac{(BPFO + BPFI)}{\frac{RPM}{60}}$	

7. Data Acquisition Parameters

The vibrational signature is transduced by an accelerometer into an electrical signature. This electrical signature is sampled by a data acquisition system, which digitizes the analog signal into a digital signal. The digital time signature extracted from vibration data is represented with N sampling points having amplitudes $\{x_1, x_2, x_3, \cdots, x_N\}$.

7.1 Signal-to-Noise Ratio

The vibrational signature is converted from the analog signature into a digital signature by means of an analog-to-digital (A/D) converter. There are many specifications for an A/D converter with 2 of the important ones being the signal-to-noise ratio (SNR) and sampling rate. One of the A/D converter specifications is the number of bits. Quantization or amplitude resolution size is established by the number of bits. Typically, the larger number of bits provides a finer amplitude resolution of the vibrational signature. SNR is an important parameter in detecting and identifying a signal embedded in noise. An ideal A/D converter has a digitization error of $\pm\frac{1}{2}$ least significant bit (LSB).

The SNR calculation of an A/D is summarized below. A maximum error of an ideal A/D converter is $\pm\frac{1}{2}$ LSB or represented by "q" below. "B" is the number of bits of the A/D converter. The following is the summary of the relationship leading to SNR of an A/D converter:[15]

- Full scale (FS) sinewave = v(t) = $\left[\frac{q2^B}{2}\right]\sin(2\pi ft)$
- RMS value of FS sinewave = $\frac{q2^B}{2\sqrt{2}}$
- RMS value of quantization noise = $\frac{q}{\sqrt{12}}$
- SNR = $20\log_{10}\left[\frac{RMS\ Value\ of\ FS\ Sinewave}{RMS\ Value\ of\ Quantization\ Noise}\right] = 20\log_{10}2^B + 20\log_{10}\sqrt{\frac{3}{2}}$
- SNR = $6.02B + 1.76\ (dB)$

7.2 Sampling Rate

There are various factors that influence the selection of the sampling frequency. As explained in the bearing fault stages, the bearing fault can reside in different parts of the frequency domain and have different frequency characteristics as it degrades. One can unambiguously resolve a signal over the Nyquist band, which is defined up to $\frac{1}{2}$ of the sampling rate. As the bearing fault progresses, the frequency domain analysis becomes complex as the number of frequency components increases. Sampling to match the necessary criteria of the highest bearing fault

frequency is not adequate. There are other mechanical noises, which can make the bearing fault frequency difficult to detect until the fault degrades to the final stage of life.

If detection of the bearing defect frequency is the goal, then it has been suggested that the optimum bandwidth of observing harmonics of the defect frequencies calculated should be limited to the 4th through the 10th orders. It was suggested that vibration harmonics with multiples of less than 4 may be present due to manufacturing tolerances – deviations between actual and ideal bearing rolling surfaces. The magnitude of harmonic components below the 4th order may be different by only 3–6 dB during the beginning and final stages of a bearing's life. More accurate and responsive indicators are derived from the 4 to 10 times the bearing defective frequencies. Magnitudes of the bearing defect frequencies typically increase the initial stages of wear on these surfaces.[5]

Sampling is the process by which the mechanical vibrational signature is converted into discrete time samples. The rate or sampling frequency (f_s) determines the frequency content that the data acquisition system can provide unambiguous information about the maximum signal frequency. According to the Nyquist theorem, if a signal is perfectly band limited and sampled at f_s, one can identify a signal up to $\frac{1}{2}f_s$ correctly. This does not mean that one can necessarily reproduce the time shape response accurately, which is very important if one were to detect the bearing defect fault in the bearing's natural frequency region or in its resonant region.[16]

7.3 Resonance

The measured frequency response of a ball bearing is not necessarily flat over the frequency band. It is a complex, composite mixture from various sources such as the accelerometer itself, the mechanical structure that retains the ball bearing, and the actual bearing itself. Resonance of a complex system offers a benefit in detecting bearing fault frequencies. When a structure has resonance frequency regions, an input signal in that frequency is amplified. It is not obvious why a bearing fault frequency shows up in resonance frequency regions that many times higher in frequency than indicated by the bearing fault frequencies. At the beginning of the bearing fault, the time-domain signature appears as a series of impulse functions. From a Fourier transform of a series of ideal impulse functions, the resulting frequency transformation yields a series of frequencies spaced at the reciprocal of the impulse period, as shown in Fig. 8.

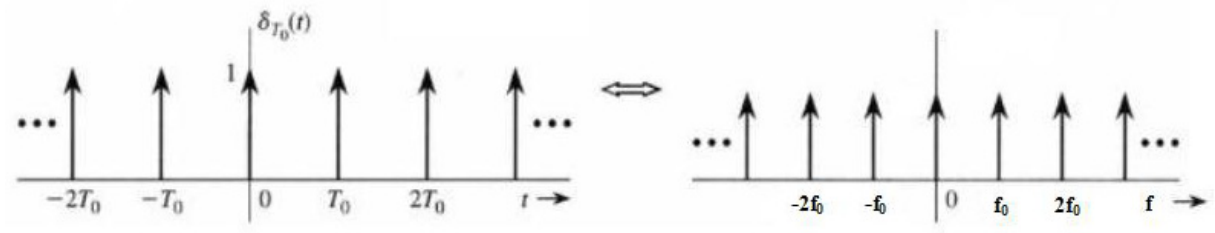

Fig. 8 Impulse train and its frequency transformation

The lower frequency region of the bearing fault frequency is typically overwhelmed by other mechanical noise. It is not until late in the bearing fault stage that the bearing fault frequency can be easily detected in this frequency region. On the other hand, the noise above the bearing fault frequency tends to be lower than in the bearing fault frequency. If the impulsive signature happens to occupy a frequency resonance region, the signal will be amplified. Taking advantage of this combined SNR improvement may provide an early detection of a bearing fault. This should be done with caution as the bandwidth of the resonance may not be sufficiently wide enough to allow for a sufficient reproduction of the impulse nature of the bearing fault.

8. Signal Processing Techniques

Signal processing techniques are applied to the accelerometer signature to gain further insight into the interpretation of the measured data set. Analysis of the sensor data is very difficult in its raw form. Many signal processing techniques have been developed to process the data into a form more readily interpreted or provide a significant improvement of the SNR of the signal feature set. Signal processing techniques have been developed for a broad range of fields such as statistics, communications, digital signal processing, radar, and sonar applications. Many of these techniques are applicable to the analysis of the vibrational data from a bearing. These techniques are from the areas of statistical, time-domain, frequency-domain, and time-frequency domain analysis. The digital vibrational time signature extracted from vibration data is represented with N sampling points having amplitudes represented as $\{x_1, x_2, x_3, \cdots, x_N\}$. As a precursor to the data signature analysis, a data quality check should be performed to determine that the data acquisition system is functioning properly.

8.1 Data Quality Check

The following checks should be performed to ensure the quality of the data:[17]

- The number of data points is consistent among the data files.
- The monotonic increases or decreases in the overall signature file indicate a drifting condition in the data acquisition system.
- An excessive noise amplitude increase from one file to the next file could indicate a data acquisition problem.
- Amplitude saturation (clipping) of the vibrational signature creates nonlinearity in the signature that would make it harder to analyze.
- Constant amplitude in the data file would indicate a malfunction in the data converter hardware.

- A shift in mean and standard deviation within the data file could possibly indicate a data acquisition problem.

8.2 Statistical Analysis

Statistical analysis is the mathematical science dealing with the analysis or interpretation of data. The data analyst uses a few straightforward statistical techniques as means of summarizing the collected data from the sensors. These statistical techniques are under the area of descriptive statistics, which is a methodology to condense the data in quantitative terms.

Statistical techniques that are mainly used for alarm purposes in industrial plants are the statistical moments of order 2, 3, and 4. The probability density function (PDF) of the vibrational time series of a good bearing has a Gaussian distribution (also known as a Normal distribution), whereas a damaged bearing results in non-Gaussian distribution with dominant tails because of a relative increase in the number of high levels of acceleration. An important fact to remember about a Gaussian distribution is that the first 2 moments define the distribution.

8.2.1 Histogram – Discrete Probability Density Function

A histogram or a discrete PDF provides a tool to characterize the amplitude of the vibrational time series data into a form that provides a visualization profile that can relate to the statistical features. The calculation is as follows:

- Let d be the number of divisions that are needed to divide the range into; let h_i with $0 \leq i \leq d$ be the columns of the histogram, then

$$h_i = \sum_{j=0}^{n} \frac{1}{n} r_i(x_i), \forall i, 0 \leq i \leq d$$

$$r_i(x) = \begin{cases} 1, if\ \frac{i(max(x_i)-min(x_i))}{d} \leq x < \frac{(I+1)(max(x_i)-min(x_i))}{d} \\ 0, otherwise \end{cases} \quad (1)$$

- The histogram upper bound (h_U) and lower bound histogram (h_L) are defined as

$$h_U = max(x_i) + \frac{\Delta}{2}$$

$$h_L = max(x_i) - \frac{\Delta}{2}$$

$$where\ \Delta = max(x_i) - min\left(\frac{x_i}{[n-1]}\right) \quad (2)$$

8.2.2 Moments

If these moments are calculated about the mean, they are called central statistical moments. The first and second moments are well known, being the mean and the variance, respectively. These

are analogous to the first and second area moments of inertia with the area shape defined by the PDF. The third moment is the termed skewness and the fourth moment is termed kurtosis. The general equation for the order of moment is as follows:[2]

$$M_p = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^p \tag{3}$$

$$where\ p\ is\ the\ order\ of\ the\ moment$$
$$N\ is\ the\ number\ of\ data\ value$$
$$i\ is\ the\ index\ of\ the\ data\ value$$
$$\bar{x}\ is\ the\ mean\ value\ of\ the\ data\ set$$

8.2.3 Mean

Mean is the most common measure of a statistical distribution. In this case, mean is the arithmetic average for a set of measurements:[2]

$$\bar{x} = \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

8.2.4 Variance

Variance is a measure of the dispersion of a waveform about its mean, and is called the second moment of the signal:[2]

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{5}$$

8.2.5 Skewness

Skewness is the statistical moment of the third order normalized by the standard deviation to the third power. This moment indicates the asymmetry of the PDF or degree of deviation from the symmetry of a distribution:[2]

$$\gamma = \frac{M_3}{\sigma^3}$$
$$\gamma = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3}{\sigma^3}$$
$$\gamma = \frac{1}{N\sigma^3}\sum_{i=1}^{N}(x_i - \bar{x})^3 \tag{6}$$

8.2.6 Kurtosis

Kurtosis is the fourth statistical moment, normalized by the standard deviation to the fourth power. It is a measure of whether the data are peaked or flat relative to a normal distribution. Most background signals measured by a data acquisition are considered to have a normal distribution in amplitude. The normal distribution has a value of 3:[2]

$$\kappa = \frac{M_4}{\sigma^4}$$

$$\kappa = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4}{\sigma^4}$$
$$\kappa = \frac{1}{N\sigma^4}\sum_{i=1}^{N}(x_i - \bar{x})^4 \quad (7)$$

8.2.7 New Statistical Moments

Typically, the variance and kurtosis have played an important role in evaluating bearing health monitoring. These features are even moments that relate to the spread of the distribution. Odd moments provide relationship between the mean and peak of the distribution. For the vibrational time series as it relates to a good bearing, the amplitude distribution will be very close to the ideal Gaussian distribution. In this case, all odd moments will be 0 and the even moments take on finite values. For the case of a good health bearing, the skewness moment is 0. This limits the importance of this feature for monitoring, but with transformations, other statistical moments can be formulated that would be useful features to monitor in a both normal and degraded state.

If the mean value is removed or subtracted from the data, the general equation for the moment $M_p = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^p$ can be rewritten as

$$M_n = \frac{1}{N}\sum_{i=1}^{N}(x_i)^{\beta} \ , \beta \ \geq 1 \quad (8)$$

In addition, 2 other transformations can be applied to convert the vibrational time series into positive values, such as absolute or square value. Now let

$$Y = \ |X| \ or \ Y = \ X^2 \quad (11)$$

then

$$(M_n)_r = \frac{1}{N}\sum_{i=1}^{N}(y_i)^{\beta} \ , \beta \ \geq 1 \quad (12)$$

where the subscript r represents the moment of the unidirectional data about the origin.

Let the equality $r = a$ represent the vibrational time series data by getting its absolute value, and $r = s$ represent the square value, respectively. Consequently, the new normalized central statistical moments can be defined as

$$NM_r^{\beta} = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i)^{\beta}}{\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right)^{\beta}} \ , \beta \ \geq 1 \quad (13)$$

New normalized statistical moments can be formulated as features for bearing condition monitoring for theses transformed vibrational time series. Four normalized central statistical moments are given:

Normalized second moment absolute NM_a^2 (namely A2)

$$NM_a^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i)^2}{\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right)^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2}{\left(\frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2}{\left(\frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^2} \quad (14)$$

Normalized third moment absolute NM_a^3 (namely A3)

$$NM_a^3 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}y_i\right)^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}|x_i|^3}{\left(\frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}|x_i|^3}{\left(\frac{1}{N}\sum_{i=1}^{N}|x_i|\right)^3} \quad (15)$$

Normalized 3/2 moment squared ($\mathbf{NM_s^{3/2}}$)

$$NM_s^{3/2} = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i)^{3/2}}{\left(\frac{1}{N}\sum_{i=1}^{N}y_i\right)^{3/2}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left((x_i)^2\right)^{3/2}}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i)^2\right)^{3/2}} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i)^3}{\sigma^3} \quad (16)$$

Normalized second moment squared ($\mathbf{NM_s^2}$)

$$NM_s^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i)^2}{\left(\frac{1}{N}\sum_{i=1}^{N}y_i\right)^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left((x_i)^2\right)^2}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i)^2\right)^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i)^4}{\sigma^4} \quad (17)$$

8.3 Time-Domain Analysis

Time-domain analysis is simpler than frequency-domain analysis and less computationally intensive. Various features can be formulated from the time series waveform to characterize its overall shape. It is hoped that one can differentiate a good bearing from a defective one based on evaluating the differences in feature values. These features are typical waveform characterization or features that been developed throughout various studies on the degradation of the ball bearings.

8.3.1 RMS

The RMS is related to the energy of the signal. The presence of defects are directly detected by the increase in vibration level:[18,19]

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} \quad (18)$$

8.3.2 Maximum Amplitude Value

Maximum amplitude value indicates the severity of a bearing defect:[19]

$$max(x_i)$$

8.3.3 Minimum Amplitude Value

The minimum amplitude value is designated as

$$\min(x_i)$$

8.3.4 Peak Value

The peak value is

$$Peak = \frac{1}{2}\{max(x_i) - min(x_i)\} \quad (19)$$

8.3.5 Peak to Peak

The peak to peak value is

$$Peak\ to\ Peak\ = \{max(x_i) - min(x_i)\} \tag{20}$$

8.3.6 Crest Factor

Crest factor is a measure of how much impacting is occurring in the time waveform. Impacting in the time waveform may indicate rolling element wear or cavitation.[19] This feature is sensitive to the incipient spall, but not very effective as the failure develops. As the fault develops, the RMS value increases while the peak value remains constant. So in this case of bearing degradation, the crest factor will decrease over time from the incipient fault:

$$Crest\ Factor = \frac{Peak\ Level}{RMS} \tag{21}$$

8.3.7 K Factor

As stated previously, the peak value is not a linear function of the fault degradation. The peak value will increase slowly or decrease from the time of the incipient fault; this corresponds to a wearing away of the sharpness of the spall and increase in the spall size. This will result in the increase of the RMS value of the signal. In this case, either an increase in RMS or peak value will result in an increase of the K factor feature:[5]

$$K\ Factor = Peak\ x\ RMS \tag{21}$$

8.3.8 Square Mean Root Absolute

The following is the square mean root absolute:

$$\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2$$

8.3.9 Mean Absolute

The following is the mean absolute:

$$\frac{1}{N}\sum_{i=1}^{N}|x_i|$$

8.3.10 Weighted SSR Absolute

The following is the weighted SSR absolute:

$$\frac{1}{N}\left(\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2$$

8.3.11 Clearance Factor

The clearance factor is similar to kurtosis, but is robust to operating conditions. It is also sensitive to early fatigue spalling.

$$\frac{Peak\ value}{\frac{1}{N}\left(\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2} = \frac{Peak\ value}{Square\ mean\ root\ absolute} \tag{22}$$

8.3.12 Impulse Factor

The impulse factor is sensitive to early fatigue spalling and is very similar to crest factor feature:

$$\frac{Peak\ value}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} = \frac{Peak\ value}{Mean\ absolute} \tag{23}$$

8.3.13 Shape Factor

The following is the shape factor equation:

$$\frac{RMS\ value}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} = \frac{RMS\ value}{Mean\ absolute} \tag{24}$$

8.3.14 Shannon Entropy

Shannon entropy is one of the important metrics in information theory. Claude F Shannon introduced his concept in the paper *A Mathematical Theory of Communication* (1948). This entropy provides a measure of the amount of randomness of the measured vibrational data. Entropy estimation is a two stage process; first, a histogram is estimated, and then the entropy is calculated. The Shannon entropy is calculated using following formula:

$$S(X) = \sum_{i=1}^{n} h(x_i)I(x_i) = \sum_{i=1}^{n} h(x_i)\log_b\frac{1}{h(x_i)} = -\sum_{i=1}^{n} h(x_i)\log_b h(x_i) \tag{25}$$

8.3.15 Normal Negative Log Likelihood

The following is the normal negative log likelihood:

$$-\sum_{i=1}^{N} log\left[\frac{1}{\sigma\sqrt{2\pi}}exp\left\{\frac{-[(x_i-\mu)^2]}{2\sigma^2}\right\}\right]$$

8.4 Frequency-Domain Analysis

In this process, the time series data are transformed into sums of simpler trigonometric functions such as the sine and cosine functions. Through this conversion or transformation, the detection of faults is generally easier since it improves the SNR of the data and the fault signature tends to be more visually observable. Signals, in general, can be divided into two classes: stationary and non-stationary. Signals whose average properties do not change with time and that are independent of the particular sample record used to determine them are said to be stationary. Non-stationary signals are those whose average properties change with time.[20]

Stationary signals are characterized by time-invariant statistical properties, such as the mean value or autocorrelation function. Analysis of stationary signals has largely been based on well-known spectral techniques such as the Fourier transform, which identifies the constituent frequency components within the signal.[21] Traditional data analysis methods such as Fourier analysis are all based on linear and stationary assumptions, i.e., the signal to be processed must be linear and temporally stationary.[22] In frequency domain analysis, the mathematical process is to transform the time measurements into the frequency domain, where it may be easier to analyze and interpret the sensor measurements. This transformation of the vibrational time series into the frequency domain along with the knowledge of the bearing fault frequency and its progression through the fault stages provide a power interpretation of frequency spectral data.

8.4.1 Fast Fourier Transform

The fast Fourier transform (FFT) is an efficient algorithm for calculating the discrete Fourier transform. A discrete Fourier transform converts a series of time discrete measurements into its components in the frequency domain. There are many forms for implementing FFT algorithms. The FFT provides the same result as the discrete Fourier transform, but with a main difference in the speed of the overall calculation. For a large size of data measurements, this can reduce the computational time by several orders of magnitude.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \qquad (26)$$
$$k = 0, \dots, N-1$$

8.4.2 Frequency Resolution (FFT)

The vibrational signature in the time domain is typically transformed into the frequency domain through the use of the processing algorithm known as the FFT. Since there is the possibility of different bearing fault frequencies, which are dependent on where the fault actually occurs, one must have a frequency resolution that can discriminate these fault frequencies from other mechanical-related signals. As bearing degradation continues after the incipient fault, the frequency response becomes more complex and requires a finer frequency resolution if one needs to identify the associated relationship of the frequency to the actual fault mechanism. As shown in Fig. 9, the FFT provides a frequency resolution $\frac{f_s}{N}$. So in order to get a frequency resolution in the order of a few Hz, the sampling rate and number of samples need to very high:[23]

- Sampling Frequency f_s
- Number of Points in FFT, N
- Frequency Resolution = $\frac{f_s}{N}$
- FFT Processing Gain = $10 \log_{10}\left(\frac{N}{2}\right)$

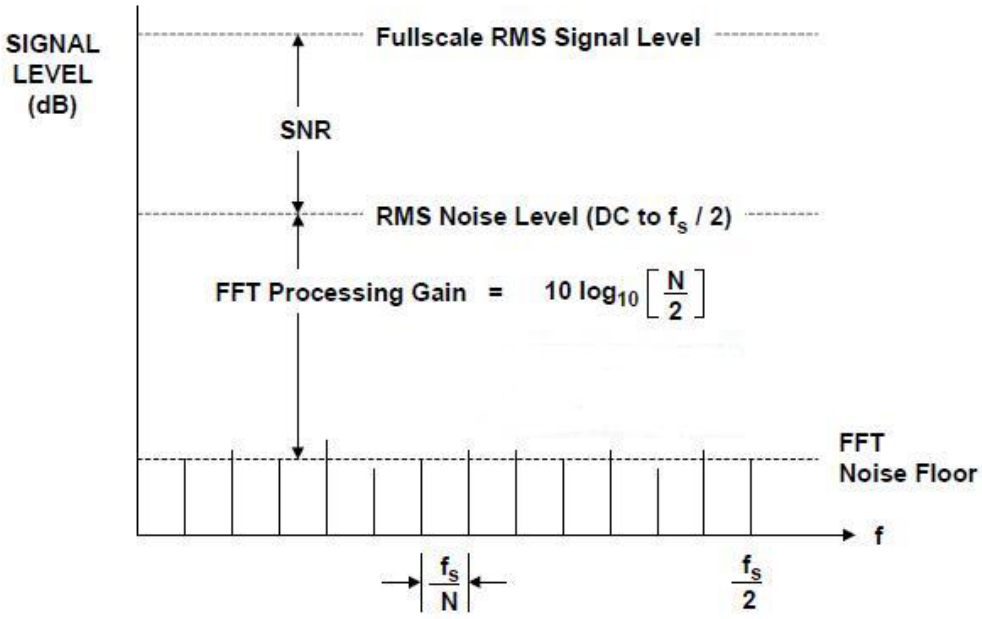

Fig. 9 FFT resolution and processing gain

8.4.3 FFT Processing Gain

The detectability of the bearing's defect frequency is the SNR of the bearing's defect frequency to the overall noise level. One of the advantages of the FFT is the FFT processing gain, which provides an improvement in the SNR as a function of the data file size N. For a very large number of data sample in a data file, there is a very sizeable improvement obtained, as shown in Fig. 9. So the effective SNR is improved over what was stated in Section 7.1. The SNR with FFT processing gain is as follows:[23]

$$\text{SNR} = 6.02B + 1.76 + 10\log_{10}\left(\frac{N}{2}\right)(dB) \tag{27}$$

8.4.4 Hilbert Transform

The Hilbert transform is defined as follows for a continuous signal:

$$H[x(t)] = \tilde{x}(t) = \pi^{-1}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{28}$$

The application of this transform to vibrational data provides some additional information about amplitude, instantaneous phase, and frequency of vibrations.[24]

The mathematical term "analytic" is applied to $\tilde{x}(t)$ because it is an analytic function of a continuous complex variable. A formal analytic signal is a complex-valued continuous-time function with a Fourier transform that vanishes for negative frequencies. In the discrete domain, the analytic-like discrete-time signal can be generated with some similar properties as the continuous-time analytic signal in the signal processing sense, but it is not an analytic function in

the mathematical sense. A technique for creating a Hilbert transform of a discrete time sequence is as follows:[25]

- Compute the N-point transform using an FFT of N real data samples

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k\frac{n}{N}} \tag{28}$$

- Form the N point 1-sided, discrete-time "analytic" signal transform

$$Z[m] = \begin{cases} X[0], for\ m = 0 \\ 2X[m], for\ 1 \leq m\ \leq \frac{N}{2} - 1 \\ X\left[\frac{N}{2}\right], for\ m = \frac{N}{s} \\ 0, for\ \frac{N}{2} + 1\ \leq m\ \leq N - 1 \end{cases} \tag{29}$$

- Compute, using an N-point inverse FFT

$$X[n] = \frac{1}{NT}\sum_{m=0}^{N-1} Z[m] e^{\frac{j2\pi mn}{N}} \tag{30}$$

8.5 Envelope Analysis

The envelope of the waveform results in an integral curve that represents the overall shape of the signal. It may be constant (a waveform is a continuous harmonic) or may vary with time.

The form or shape of the variation of the instantaneous amplitude is called a wave envelope. In using the Hilbert transform, the rapid oscillations can be removed from the amplitude-modulated signal to produce an outline shape of the slow envelope alone. This is illustrated in Fig. 10. For a healthy bearing, the vibrational time amplitude waveform is considered a Gaussian function that is "white." Its envelope should be fairly flat and the frequency spectrum somewhat constant. When a fault occurs, there is an impact response each time the ball bearing hits the fault. In this illustration, the fault is located on the outer race. An envelope of the overall vibrational signature is shown in red with the frequency spectrum. In this case, the frequency spectrum clearly shows the bearing defect frequency and its harmonic.

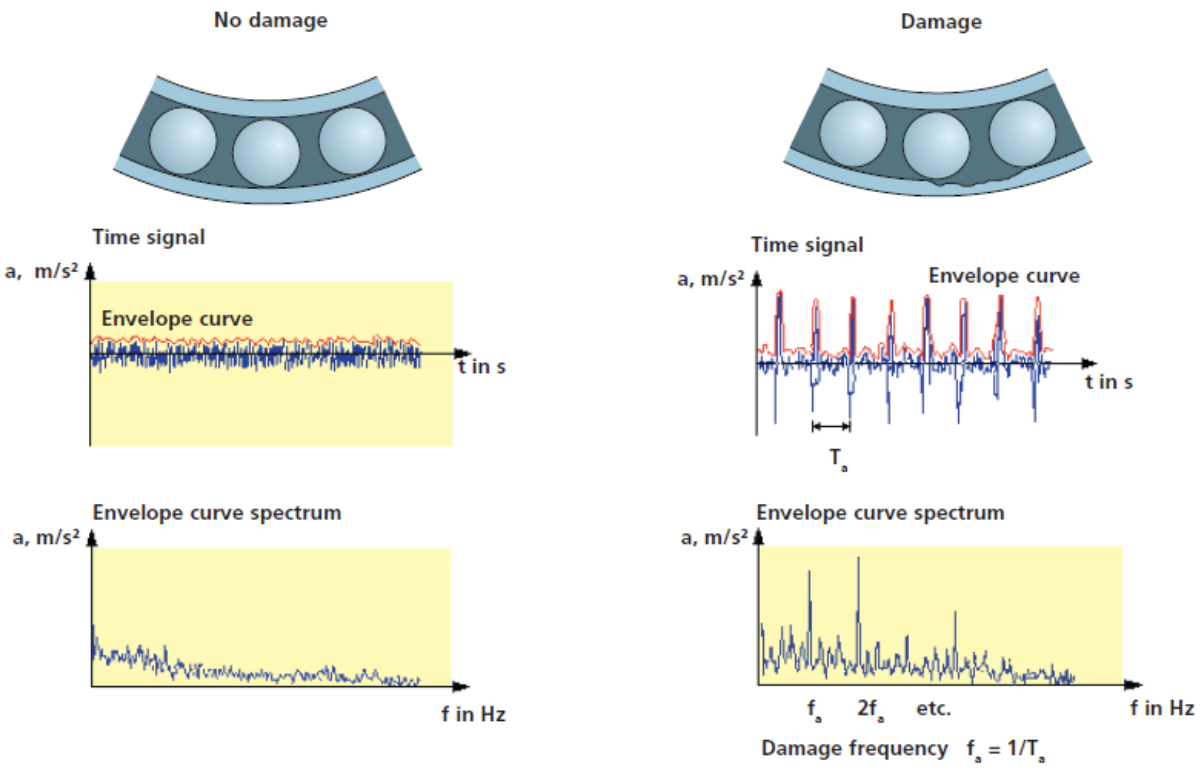

Fig. 10 Envelope spectrum of good and faulted bearing (reproduced with permission from Pruftechnick AG)[26]

The envelopes of different types of bearing faults do not result in the same overall shape pattern. Figure 11 is an idealized illustration of the vibrational time series and the envelope of the different bearing faults. From this illustration, it should be clear that the pattern of the outer race fault vibrational time waveform should be the easiest to detect since it is quite repetitive. In the case of the inner race fault, the waveform has impacts that occur at the inverse of BPFI, but it is also modulated by the shaft speed. A frequency analysis of these waveforms will show corresponding frequencies that correlate to the correspond fault, respectively. Only the outer race fault will show frequency spectral lines that relate to 1 fault frequency. The other types of faults will show the bearing defect frequency with multiple spectral lines corresponding to the particular fault and their mixing products of the envelope curve. In practical rolling bearing applications, the frequencies of the cyclic impacts vary in a random manner, due to slip, varying speeds, and varying load angle. Consequently, the inconsistency in the cyclic impacts may complicate the detection of the nominal bearing defect frequencies from the raw vibration spectrum. However, the envelope spectrum or its square can be successfully used for indicating the defect frequencies even with random slip.

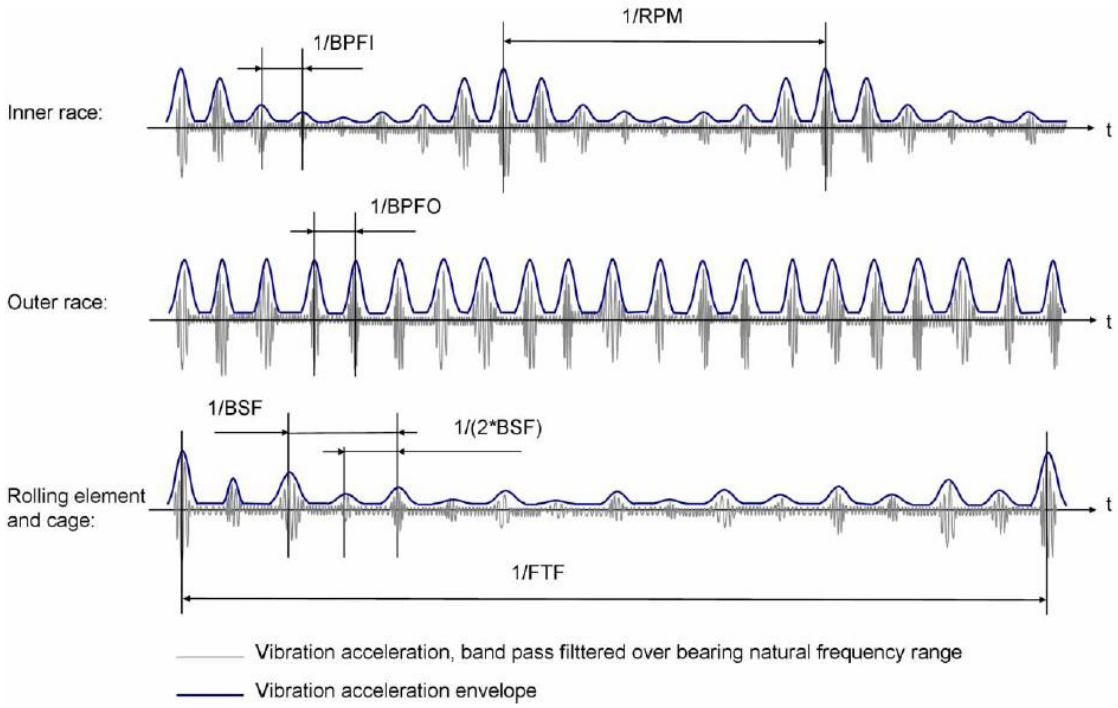

Fig. 11 Bearing faults and envelope waveforms (reproduced with permission from SAGE Publications, Ltd)[27]

8.5.1 Modulation of Fault Frequencies

The variation of the waveform is the result of load modulation. Strength of the impact from rolling over a fault is a function of the bearing radial load. If a fault is located on the stationary component of the bearing, the fault will encounter the same impact force as each ball bearing rolls over the fault. Therefore, all the impact responses will be equal in amplitude. If the fault is not located on a stationary component, then the fault will encounter a varying impact force that is repetitive with the rotation speed. This will result in the vibrational impact responses being amplitude modulated at the fault frequency with the rotational speed. Which bearing fault frequency gets modulated by the rotation speed will depend on which race is stationary. Only the non-stationary race fault will get modulated by the rotation speed. Similar modulation will occur with a ball fault, being amplitude modulated by the cage and rotation speed.[28]

In terms of bearing analysis, the signal processing of enveloping and amplitude demodulation describe very similar process. As discussed previously, the process is to remove high amplitude, low mechanical vibration, which masks the low bearing fault vibration, and translate the high frequency resonant impact response into a low frequency signal so that time and frequency analysis can be performed. The envelope spectrum will contain noise if the bearing is healthy and spectral lines at the corresponding bearing defect frequency if there is a fault. These lines will increase as it progresses through the bearing fault stages, but at the end, the bearing fault frequencies will decrease with increasing noise in the envelope spectrum.[4]

8.5.2 Quadratic Phase Coupling (QPC)

For a linear system, the frequency of an output signal will be the same as the input signal. When 2 input sinusoids of different frequencies are passed through a linear system, the output frequencies are preserved. If these same 2 sinusoids are passed through a nonlinear system, the output signal will have components at the sum and difference frequencies of the 2 sinusoids as well. The term quadratic phase coupling (QPC) is used to describe the resultant components for this nonlinear mixing process, which has quadratically coupled frequency pairs.

Initially, the bearing fault frequency will show up as a single harmonic. As the fault progresses, harmonics will relate to the fault. For example, as an outer race fault grows, the fault will show at $f_{BPFO}, 2f_{BPFO}, 3f_{BPFO}, \cdots$. As the fault progression continues, there will be a mixing of this fault with other fault frequencies related to the bearing. For example, the outer and inner race fault frequencies might show up, i.e., $f_{BPFO} \pm f_{BPFI}$. A general expression for these frequencies is as follows:

$$mf_1 \pm nf_2$$
$$where\ n = 1,2,3 \ldots\ and\ m = 1,2,3, \ldots$$
$$f_1 = f_{BPFO}, f_{BPFI}, f_{BSP}, f_{FTF}, f_{ROT}$$
$$f_2 = f_{BPFO}, f_{BPFI}, f_{BSP}, f_{FTF}, f_{ROT}$$

Figure 12 illustrates the complex nature of the spectral mixing of the bearing fault frequencies. This situation could possibly exhibit at some point late in the degradation progression. In this figure, the actual frequencies are not important. The important concept is the various spectral lines and their relationship. There are many combinations that could be generated and some terms may overlap or be located very close to other terms.

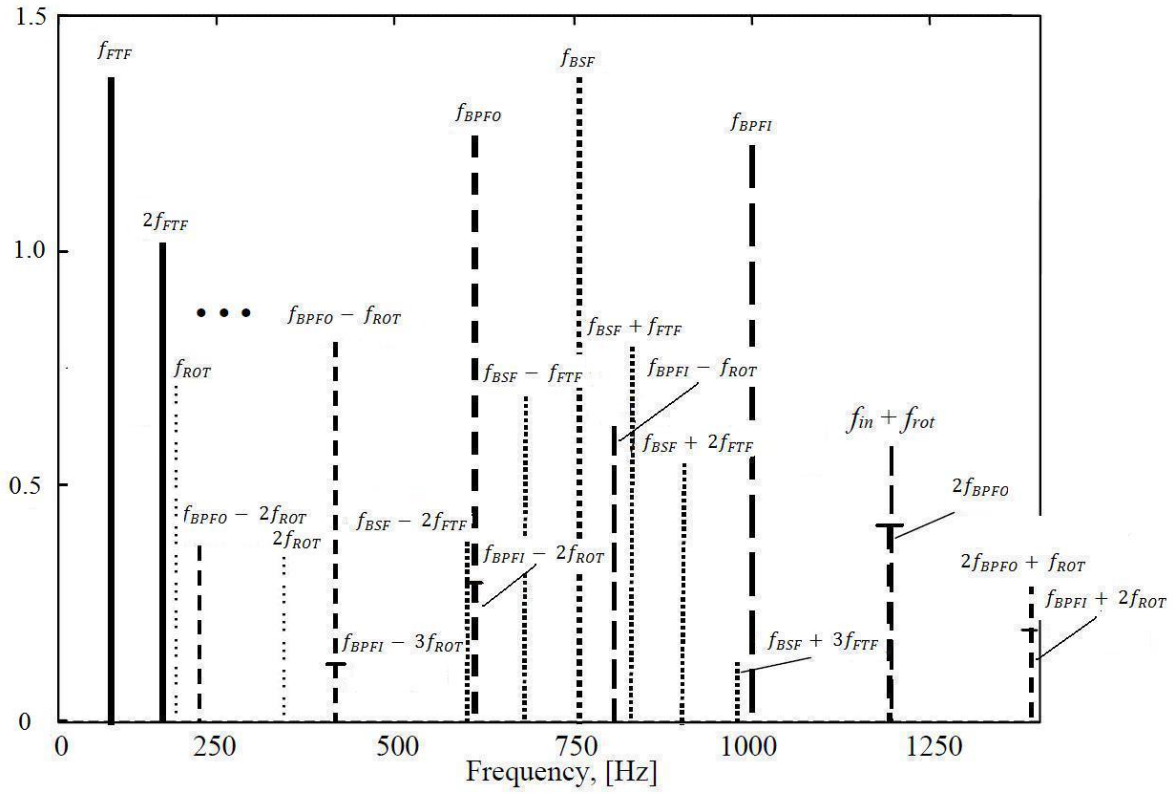

Fig. 12 Complex mixing of bearing fault frequencies

8.6 Higher-Order Spectra Analysis

The vibrational signal from the accelerometer has a Gaussian distributed noise for a healthy bearing. As stated, a Gaussian signal can be characterized by its mean and variance. Moreover, the first-order cumulant of a Gaussian process is equal to the mean, the second-order cumulant to the variance, and all higher-order cumulants are identically 0.

The power spectrum is a member of the higher-order spectral class known as the second-order spectrum. In the power spectrum, the transformation of the vibrational time series treats the conversion as a superposition of statistically uncorrelated harmonic components. In the case of a Gaussian process, this power spectrum is sufficient to describe its characteristics. As the bearing degrades, the vibrational signal undergoes a nonlinear process. As this continues, the detection of nonlinearities becomes important in determining the bearing's health status. The second-order statistical methods are not sufficient for the analysis. The phase relations between frequency components are not taken into account by the statistical measures up to order 2.

The higher-order spectra (HOS) technique has been developed fairly recently to account for the nonlinearity and phase couplings that occur in non-Gaussian and nonlinear systems. In the literature, HOS is also known as "higher order statistics." For a linear system, the frequency of an output signal is the same as the input signal. When 2 input sinusoids of different frequencies are passed through a linear system, the output frequencies are preserved. If these same 2

sinusoids are passed through a nonlinear system, the output signal will have components at the sum and difference frequencies of the 2 sinusoids as well. Thus, QPC is used to describe the resultant components for this nonlinear mixing process that has quadratically coupled frequency pairs.

8.6.1 Bispectrum

QPC provides a powerful indicator in nonlinear system identification. In our case, detection of the QPC is a very good indicator that the bearing is in a fault state. Detection of the QPC is achieved through the use of HOS (order greater than 2). HOS use the higher-order cumulants of the data. Bispectrum is known as the third-order spectrum and trispectrum used to label the fourth order.

A Fourier transform of a real discrete zero-mean process x_n is given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k\frac{n}{N}} \tag{31}$$

The power spectrum, (k), can be defined of from the discrete Fourier transform as follows:

$$P(k) = E[X(k)X^*(k)] \tag{32}$$

As such, the power spectrum decomposes the vibrational power over the frequency domain, i.e., the signal second moment. Since this operates on real data, the power spectrum has no phase information.

The definition of the bispectrum is as follows:

$$B(k,l) = E[X(k)X(l)X^*(k+l)] \tag{33}$$

Similar to the power spectrum, the bispectrum is related to the skewness of the vibrational signal. It decomposes the vibrational signal third moment over the frequency domain. This provides a tool to evaluate the non-symmetric nonlinearities of the signal. If the vibrational signal is not skewed, the bispectrum will be 0.

There are 2 methods of estimating the bispectrum: indirect and direct. The following are the steps for the direct method:

1. Divide the finite data set of length N into K (possibly overlapping) segments of M samples each, i.e., N ≥ KM. Detrend (i.e., subtract the mean value of each segment and appropriately window the data to provide some control over the effects of spectral leakage). If necessary, delete the last data points or add zeros to obtain a convenient length M = 2^n for FFT.

2. Generate the FFT coefficients for each data segment:

$$X_i(k) = \frac{1}{M}\sum_{l=0}^{M-1} x_i(l)e^{-j\frac{2\pi k}{M}l} \tag{34}$$

3. Form the raw spectral estimated based on the FFT coefficients:

$$\widehat{S_{3x,l}}(k,l) = X_i(k)X_i(l)X_i^*(k+l) \tag{35}$$

4. Compute the segment-averaged estimate of the bispectrum of the given data from the averages over K pieces.

$$B(k,l) = \frac{1}{K}\sum_{i=1}^{K}\widehat{S_{3x,l}}(k,l) \tag{36}$$

8.6.2 Trispectrum

The definition of the trispectrum is as follows:

$$T(k,l,m) = E[X(k)X(l)X(m)X^*(k+l+m)] \tag{37}$$

In a similar way to the power spectrum and bispectrum, the trispectrum decomposes the vibrational signal as the vibrational signal's kurtosis over the frequency domain. In this case, the trispectrum is a tool for detecting symmetric nonlinearities.

8.6.3 Bicoherence and Tricoherence

A problem with the bispectrum is that the resulting transformation has a variance that is proportional to the cubic product of the power spectra. In this case, the second-order properties may have a very significant influence in the bispectrum and provide improper interpretations. It is for this reason that the bispectrum is normalized. This normalization process results in the bicoherence. Since the bicoherence is independent of the energy or amplitude of the signal, it can be used as convenient test statistics for the detection of non-Gaussian, nonlinear, and coupled processes:[29–32]

$$b^2(k,l) = \frac{|B(k,l)|^2}{E[|X(k)X(l)|^2]E[X(k+l)^2]} \tag{38}$$

This normalization process is also applied to the trispectrum yielding the tricoherence concept.

$$t^2(k,l,m) = \frac{|T(k,l,m)|^2}{E[|X(k)X(l)X(m)|^2]E[|X(k+l+m)|^2]} \tag{39}$$

The bicoherence and tricoherence can be estimated as follows:

$$\hat{b}^2(k,l) = \frac{\left|\sum_{j=1}^{N} X_j(k)X_j(l){X_j}^*(k+l)\right|^2}{\sum_{j=1}^{N}\left|X_j(k)X_j(l)\right|^2\sum_{j=1}^{N}\left|X_j(k+l)\right|^2} \tag{40}$$

$$\hat{t}^2(k,l,m) = \frac{\left|\sum_{j=1}^{N} X_j(k)X_j(l)X_j(m){X_j}^*(k+l+m)\right|^2}{\sum_{j=1}^{N}\left|X_j(k)X_j(l)X_j(m)\right|^2\sum_{j=1}^{N}\left|X_j(k+l+m)\right|^2} \tag{41}$$

8.7 Time-Frequency Analysis

Time-frequency analysis is an attempt to overcome some of the shortcomings of Fourier analysis. Time-frequency analysis is a series of signal processing techniques for analyzing

signals that are transient in nature or non-stationary. The statistical properties of a non-stationary signal change over time. The time-averaging properties of a non-stationary signal change over time, thus making the time-averaging approach adopted in the Fourier transform ineffective. Time-frequency analysis techniques are methods for analyzing nonlinear and non-stationary data.[33]

Time frequency techniques decompose 1-dimensional time-series signals into a 2-dimensional (2-D) plane by exposing the time-dependent variations of characteristic frequencies within the signal, thus presenting a valid and effective tool for non-stationary signal analysis.[21]

There are many techniques that have been developed to perform time-frequency analysis. Unfortunately, these techniques generate artifacts in cross-term products and the user must be aware of how the application of the particular technique will contaminate the results. All the methods are designed to modify the global representation of the Fourier analysis, but they all are limited in one way or another. Necessary conditions for the basis to represent a nonlinear and non-stationary time series:[33]

1. complete – guarantees the degree of precision of the expansion
2. orthogonal – guarantees positivity of energy and avoids leakage
3. local – crucial for non-stationarity, for in such data there is no time scale
4. adaptive – adapting to the local variations of the data can allow the decomposition to fully account for the underlying physics.

The following is a list of time-frequency analysis that is described:

1. Short-time Fourier transform (STFT)
2. Wavelet
3. Cohen
4. Wigner-Ville
5. Choi-Williams
6. Zhao-Atlas-Marks
7. Hilbert-Huang

8.7.1 STFT

The STFT is the windowing or dividing of the raw data into frames and applying the Fourier transform. This division can be overlapping or non-overlapping data. It is an attempt to analyze the non-stationary characteristics of the signal. The resulting 2-D signal is typically visually

displayed as a spectrogram, which represents the magnitude squared of the STFT or power variation in the signal over time.

$$STFT(\tau, f) = \int x(t)g(t-\tau)e^{-j2\pi ft}dt \tag{42}$$

The time-frequency decomposition of the signal is generated with a constant frequency and time resolution. A 2-D and 3-dimensional (3-D) image of the STFT are illustrated in Fig. 13.

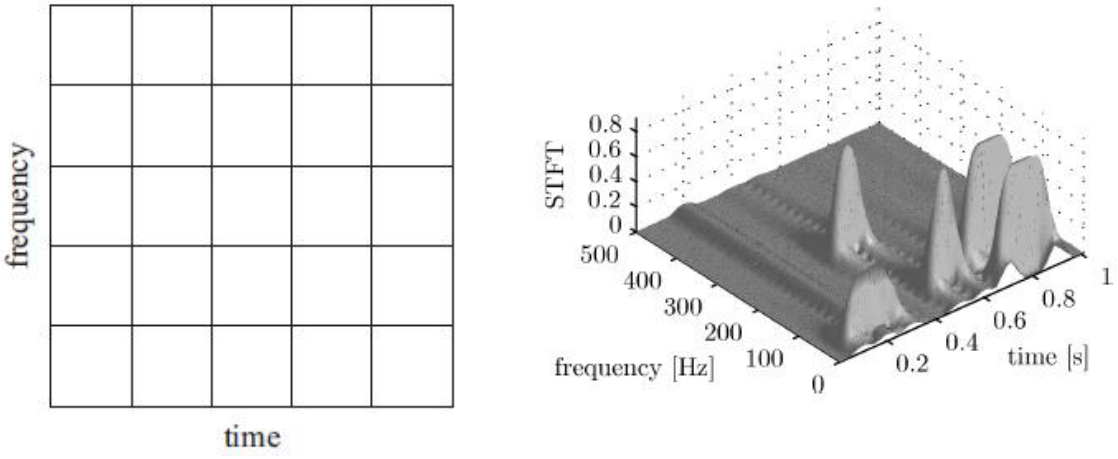

Fig. 13 STFT (reproduced with permission from Prof. Dr. Ir. Maarten Steinbuch, Eindhoven University of Technology)[34]

8.7.2 Wavelet Transform

The wavelet transform is a technique similar to the STFT. This technique decomposes the data with various functions that are scaled in amplitude and time. Wavelet transform is a windowing technique with variable sized regions The terms "dilations and translations" are the processes applied to the basis function or mother wavelet as it is scaled in width and location as applied to the data set. As with the STFT, it provides the advantage of temporal resolution in addition to the frequency information. Wavelet analysis allows use of long time intervals where more precise low frequency information is needed and shorter regions where high frequency information is needed:[35]

$$WT(s, \tau) = \int x(t)\psi^*_{s,\tau}(t)dt \tag{43}$$

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{44}$$

$$where\ s\ is\ the\ scale\ variable$$

$$\tau\ is\ the\ translation\ variable$$

$$\psi^*_{s,\tau}(t)\ is\ the\ complex\ conjugation\ of\ the\ mother\ wavelet$$

For the discrete case, the discrete wavelet (DWT) is written as

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}}\psi\left(\frac{t-k\tau_0 s_0^j}{s_0^j}\right) \tag{45}$$

There are various selections of mother wavelets to choose from and performance is dependent on the correlation of the mother wavelet to the signal characteristics. In the case of the DWT, the decomposition is performed through the use of filter banks and a down-sampling process. A 2-D and 3-D image of the wavelet transform is illustrated in Fig. 14. Note that the frequency and time resolution are not constant.

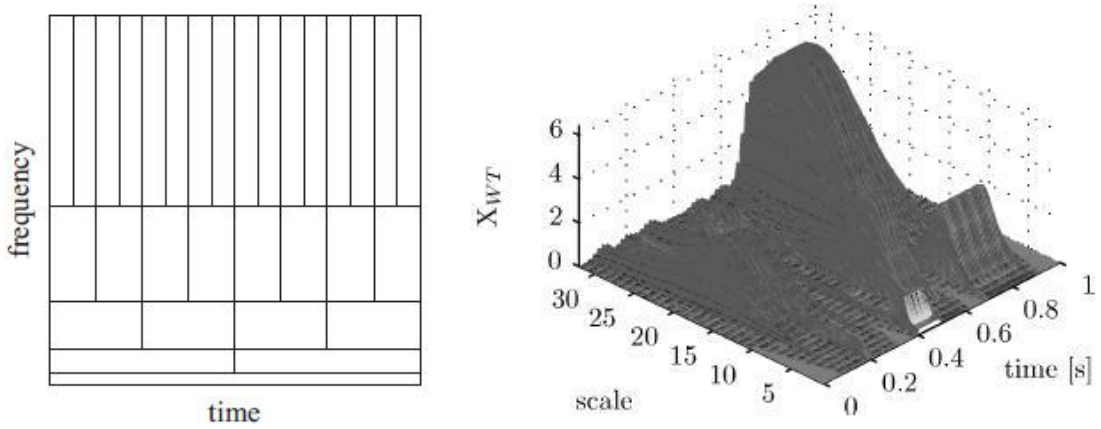

Fig. 14 Wavelet transform (reproduced with permission from Prof. Dr. Ir. Maarten Steinbuch, Eindhoven University of Technology)[34]

8.7.3 Cohen

This is a general class of processing that performs analysis in the time and frequency domains. It is an attempt to overcome some of the problems associated with the STFT. It uses bilinear transformation to transform measurement data into the frequency domain while accounting for the non-stationary aspect in the measured data set. The Wigner-Ville, Choi-Williams, and Zhao-Atlas-Marks are special implementations of the Cohen technique:

$$C_x(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_z(u,\tau)g(v,\tau)e^{(j2\pi[ut-f\tau])}du\,d\tau \tag{46}$$

$$where\ A_z(u,\tau)\ is\ the\ ambiguity\ function$$

$$A_z(u,\tau) = \int_{-\infty}^{\infty} z\left(t+\frac{\tau}{2}\right)z^*\left(t-\frac{\tau}{2}\right)e^{-j2\pi tu}\,dt \tag{47}$$

$$g(v,\tau)\ is\ the\ kernel\ function$$

8.7.4 Wigner-Ville

Wigner-Ville distribution is a very important quadratic-form, time-frequency distribution with optimized resolution in both the time and frequency domains. Wigner-Ville is computed by correlating the function with itself, the correlation being a product of the function at a past time with the function at a future time.[36] The Wigner-Ville is the Cohen technique with the kernel function defined as unity:

$$g(\nu,\tau) = 1$$

$$W_x(t,f) = \int_{-\infty}^{\infty} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau}\, d\tau \tag{48}$$

$$where\ z^*\ is\ the\ conjugate\ of\ z$$

8.7.5 Choi-Williams

The Choi-Williams distribution uses the exponential as the kernel function in the Cohen distribution to suppress the cross-term products in the transformation:

$$g(\nu,\tau) = e^{\left(-\frac{\nu^2\tau^2}{\sigma}\right)}$$

$$CW_x(t,f) = \iint \frac{\sqrt{\pi\sigma}}{|\tau|} e^{\frac{-\pi^2\sigma(t-u)^2}{\tau^2}} z\left(u+\frac{\tau}{2}\right) z^*\left(u-\frac{\tau}{2}\right) e^{-j2\pi f\tau} du\, d\tau \tag{49}$$

8.7.6 Zhao-Atlas-Marks (Cone-Shaped Kernel)

The Zhao-Atlas-Marks distribution uses a time-lag kernel as the kernel function in the Cohen distribution for suppression of the cross-term products:

$$g(\nu,\tau) = \omega(\tau)\frac{a}{2|\tau|} sinc\frac{2\upsilon\tau}{a}$$

$$ZAM(t,f) = \int_{-\infty}^{\infty}\int_{t-\left|\frac{\tau}{a}\right|}^{t+\left|\frac{\tau}{a}\right|} \omega(\tau) z\left(u+\frac{\tau}{2}\right) z^*\left(u-\frac{\tau}{2}\right) e^{-j2\pi f\tau} du\, d\tau \tag{50}$$

8.7.7 Hilbert-Huang Transform

The Hilbert-Huang transform is another time-frequency analysis technique that combines 2 processing techniques: empirical mode decomposition (EMD) and the Hilbert transform. The EMD is an algorithm where the signal is decomposed into a set of functions called intrinsic mode functions (IMF), which is almost monocomponent. The IMF represent simple oscillatory mode versus the harmonic output of the Fourier transform. EMD is empirical, intuitive, direct, and adaptive, with a posteriori defined basis derived from the data.[22]

EMD has shortcomings in resolving low frequencies. It can only resolve the signal when the spectral components differ by more than an octave. False artificial components are produced by EMD, but new techniques such as ensemble EMD have overcome this false mode decomposition.[24]

The Hilbert-Huang transform is defined as follows:[37]

Empirical Mode Decomposition

An IMF is defined as a function that satisfies the following requirements:

1. In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by 1.

2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is 0.

Therefore, an IMF represents a simple oscillatory mode as a counterpart to the simple harmonic function, but it is much more general: instead of constant amplitude and frequency in a simple harmonic component, an IMF can have variable amplitude and frequency along the time axis. The procedure of extracting an IMF is called sifting. The sifting process is as follows:

1. Identify all the local extrema in the test data.
2. Connect all the local maxima by a cubic spline line as the upper envelope.
3. Repeat the procedure for the local minima to produce the lower envelope.

The upper and lower envelopes should cover all the data between them. Their mean is m_1. The difference between the data and m_1 is the first component h_1:

$$X(t) - m_1 = h_1$$

Ideally the construction of h_1 described above should have made it symmetric and have all maxima positive and all minima negative. Also, h_1 should satisfy the definition of an IMF. After the first round of sifting, the crest may become a local maximum. New extrema generated in this way actually reveal the proper modes lost in the initial examination. In the subsequent sifting process, h_1 can only be treated as a proto-IMF. In the next step, it is treated as the data, then

$$h_1 - m_{11} = h_{11}$$

After repeated sifting up to k times, h_1 becomes an IMF, that is,

$$h_{1(k-1)} - m_{1k} = h_{1k}$$

Then, it is designated as the first IMF component from the data:

$$c_1 = h_{1k}$$

The Stoppage Criteria of the Sifting Process

The stoppage criterion determines the number of sifting steps to produce an IMF. Two different stoppage criteria have been used traditionally:

1. The first criterion is proposed by Huang. It is similar to the Cauchy convergence test, and we define a sum of the difference, SD, as

$$SD_k = \frac{\sum_{t=0}^{T} |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^{T} h_{k-1}^2(t)} \tag{51}$$

   Then the sifting process is stopped when SD is smaller than a pre-defined value.

2. A second criterion is based on the number called the S-number, which is defined as the number of consecutive siftings when the numbers of zero-crossings and extrema are equal

or at most differing by 1. An S-number is pre-selected, and the sifting process will stop only if for S consecutive times the number of zero-crossings and extrema stay the same, and are equal or at most differ by 1.

Once a stoppage criterion is selected, the first IMF, c_1, can be obtained. Overall, c_1 should contain the finest scale or the shortest period component of the signal. We can, then, separate c_1 from the rest of the data by $X(t) - c_1 = r_1$. Since the residue, r_1, still contains longer period variations in the data, it is treated as the new data and subjected to the same sifting process as described above.

This procedure can be repeated to all the subsequent r_j's, and the result is

$$r_{n-1} - c_n = r_n$$

The sifting process stops finally when the residue, r_n, becomes a monotonic function from which no more IMF can be extracted. From the above equations, we can induce that

$$X(t) = \sum_{j=1}^{n} c_j + r_n \tag{52}$$

Thus, a decomposition of the data into n-empirical modes is achieved. The components of EMD are usually physically meaningful because the characteristic scales are defined by the physical data.

Hilbert Transform

Having obtained the IMF components, the instantaneous frequency can be computed by applying the Hilbert transform to each of the IMF. After performing the Hilbert transform on each IMF component, the original data can be expressed as the real part in the following form:

$$X(t) = Real \sum_{j=1}^{n} a_j(t) e^{if\omega_j(t)dt} \tag{53}$$

8.8 Cepstrum Analysis

Cepstrum analysis is a tool for the detection of periodicity in a frequency spectrum. The cepstrum is defined as the power spectrum of the logarithm of the power spectrum of the signal or spectrum of a spectrum. For a time series with multiple harmonic components, the cepstrum will yield a single component. For the case of the bearing fault, an impact response tends to generate a series of harmonics related to the fault type. Even though the impact is low energy, the combination of the harmonics enhances the detection process, but this is more difficult to interpret when more than 1 harmonics series exists.

Principal use of the cepstrum for bearing fault detection is in detecting periodicities associated with bearing frequency harmonics and associated sideband patterns. The cepstrum is the signal processing technique that takes the inverse FFT of the logarithm of the squared magnitude of the Fourier transform of the measured signal:[38]

$$C_p(\tau) = FFT^{-1}\left(\log\left[S_{xx}(f)\right]\right) \tag{54}$$

9. Conclusion

This primer was written to provide insight into the various techniques that would be necessary to generate features for the use in the development of roller bearing diagnostics and prognostics. These techniques are necessary to transform the time-domain vibrational data into domains that provide clear indication of the bearing fault and degradation. Feature generation is somewhat straightforward when considering the overall process of developing diagnostics and prognostics algorithms. The next step is feature extraction. These selected features need to strongly correlate to the incipient fault and trend with the degradation.

It is through these transformations that features will be formulated that have sufficient signal-to-noise values that permit easier and confident detection and classification of the fault. It would be an advantage if one could use the output of the sensor, in this case, an accelerometer, and determine the condition from the amplitude directly, but this is not the situation. Transforming the time data into the various domains aids in the fault detection, but the features are not necessary linear with degradation. In some cases, the features actual seem to indicated a recovery from the fault. These features can be sensitive over a period in the degradation progression or nonlinear with respect to the fault degradation cycle.

This report summarized the various techniques that can easily be found in the open literature and provides some intuition in terms of the application of these techniques as they relate to bearing failures progression. Some are very easy to compute and comprehend, such as statistical and time-domain analysis. On the other hand, the Hilbert-Huang transformation is very computational intensive and may not necessarily converge to a solution. The various frequency-domain techniques have their advantages and disadvantages. There are artifacts in the computations that one must be aware of in order to properly interpret transformations as they relate to the bearing fault degradation signatures. It is the proper interpretation of the various transform domains in relationship to the degradation cycle that is important in comprehending bearing failure.

The use of these techniques depends on having the proper sensor/data acquisition system that provide sufficient sensitivity, dynamic range, and frequency response in order to employ these techniques effectively. There are filtering techniques that will be necessary to improve the quality of the sensor data prior to using some of the techniques described in this report. Application of filtering is highly dependent on the actual hardware being monitored. The "resonance" is a function of the mechanical structure that the bearing is used in.

Effective use of these techniques is highly dependent on the final application. If an operator is used in the decision loop, these techniques would prove to be invaluable in the decision-making process. In this case, the trained/expert operator would need to have the necessary insight into the bearing failure stages and their correlation to the feature output. There are many techniques that can be used to interpret these features in an automated decision. Techniques from statistical analysis, pattern recognition, data mining, and machine learning have been developed over the last few decades that can be used. What is needed for the development of diagnostic and prognostic algorithms is a framework and methodology organization. Significant amounts of data will be necessary for the evaluation and developing these algorithms. In addition to the data, very good "ground truth" is necessary for the proper identification of the health state of a bearing.

10. References

1. http://www.bassfishingandcatching.com/fishing-reel-bearings.html.
2. Tom KF. Survey of diagnostic techniques for dynamic components. Adelphi (MD): US Army Research Laboratory (US); January 2010. Report No.: ARL-TR-5082.
3. Mraz S. The meaning of bearing life, Machine Design, http://machinedesign.com/bearings/meaning-bearing-life.
4. Detecting rolling element bearing faults with vibration analysis, http://www.mobiusinstitute.com/articles.aspx?id=2088.
5. Barlov A, Barkova N. Condition Assessment and Life Prediction of Rolling Element Bearings – Part 1, VibroAcoustical Systems and Technologies, http://www.vibrotek.com/articles/sv95/part1/index.htm.
6. Rolling Element Bearings, STI Field Application Note, http://www.stiweb.com.
7. Tracking rolling element bearing failures using mechanical and electrical vibration testing methods, ALL-TEST Pro, LLC.
8. Jones R. Enveloping for bearing analysis, sound and vibration, SKF Condition Monitoring, February 1996.
9. Stevens D. Rolling Element Bearings, http://www.vibanalysis.co.uk/vibanalysis/rolling/rolling.html.
10. Vibration Analysis Reporting – Bearing Failure Stages & Responses, http://reliabilityweb.com/index.php/articles/vibration_analysis_reporting_-_bearing_failure_stages_responses/.
11. Rolling Element Bearing, Reliability Direct, Field Application Note , http://www.reliabilitydirect.com/appnotes/reb.html.
12. Rolling Element Testing, Machinery Vibration Analysis Fundamentals, www.lifetime-reliability.com.
13. Graney BP, Starry K. Rolling element bearing analysis. Materials Evaluation. 70(1):78–85.
14. Ganeriwala S. Review of Techniques for Bearings & Gearbox Diagnostics, IMAC Conference, Feb 3, 2010.
15. Kester W, Sheingold D, Bryant J. Chapter 2 Fundamentals of Sampled Data Systems, Section 2.1 Coding and Quantizing.

16. Wescott T. Sampling: What Nyquist Didn't Say, and What to Do About it.

17. Hively L. Data Quality Analysis, ARL Presentation, Mar 2009.

18. Lee H-H, Nguyen N-T, Kwon J-M. Bearing Diagnosis Using Time-Domain Features and Decision Tree. School of Electrical Engineering, University of Ulsam, Korea

19. de Almedia, RGT, da Silva VSA, Padovese, LR. New technique for evaluation of global vibration levels in rolling bearings. Shock and Vibration. 2002;9:225–234.

20. Sawalhi N. Diagnostics, prognostics and fault simulation for rolling element bearings. The University of New South Wales, School of Mechanical and Manufacturing Engineering, Australia, Doctor of Philosophy thesis, 2007.

21. Gao RX, Yan R. Non-stationary signal processing for bearing health monitoring. Int. J. Manufacturing Research. 2006;1(1):18–40.

22. Raja JE, Loo CK, Rao MVC. The necessity of a non-linear & non-stationary data processing in engineering and the New Hilbert Huang Transform (HHT). Pay Attention, Gain Understanding. December 2007;1(3).

23. Kester W. Section 5 Fast Fourier Transforms.

24. Feldman M. Hilbert transform in vibration analysis. Mechanical Systems and Signal Processing 25. 2011;735–802.

25. Marple S Jr. Lawrence.Computing the Discrete-Time "Analytic" Signal via FFT, IEEE Transactions on Signal Processing. September 1999;47(9).

26. Machine diagnosis: Quick and easy through FFT analysis, Pruftechnik AG.

27. Halme J, Andersson P. Rolling contact fatigue and wear fundamentals for rolling diagnostics – state of the art. Proc. IMechE Vol. 224 Part J: J. Engineering Tribology.

28. Konstanin-Hansen H. Envelope Analysis for Diagnostics of Local Faults in Rolling Element Bearings, Application Note, Bruel&Kjae.

29. Rivola A, White PR. Detecting system non-linearities by means of higher order statistics.

30. Petropulu AP. Higher-Order Spectral Analysis, The Biomedial Engineering Handbook: Section Edition.

31. Supriya MH. Realisation of a Target Classifier for Noise Sources in the Ocean, Cochin University of Science and Technology, Department of Electronics, PhD thesis, 2007.

32. Jelali M. Control Performance Management in Industrial Automation, Appendix B, Springer.

33. Huang NE, Shen Z, Long SR, Wu MC, Shis HH, Zheng Q, Yen N-C, Tung CC, Liu HH. The Empirical Model Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. Proc. Royal Society London. 1998;903–945.

34. Merry RJE. Wavelet Theory and Applications, A literature study, Einhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, June 7, 2005.

35. Loutridis SJ. Instantaneous Energy Density as a Feature for Gear Fault Detection. Mechanical Systems and Signal Processing. 2006;20.

36. Cao Y. Adaptability and Comparison of the Wavelet-Based with Traditional Equivalent Linearization Method and Potential Application for Damage Detection. North Carolina State University, Mechanical Engineering, Master thesis, 2002.

37. Huang NE, Shen SSP. Hilbert-Huang Transform and Its Applications. Interdisciplinary Mathematical Sciences. 2008;5:2–14.

38. Randall RB. Cepstrum Analysis and Gearbox Fault Diagnosis, Application Notes, B&K Instruments, Inc.

List of Symbols, Abbreviations, and Acronyms

2-D	2-dimensional
3-D	3-dimensional
A/D	analog-to-digital
ABMA	American Bearing Manufacturers Association
AM	amplitude modulation
ARL	US Army Research Laboratory
BPFI	ball pass frequency inner
BPFO	ball pass frequency outer
BSF	ball spin frequency
CBM	condition-based maintenance
EMD	empirical mode decomposition
FFT	fast Fourier transform
FTF	fundamental train frequency
gSE	spike energy spectrum
HFD	high frequency detection
HOS	higher-order spectra
IMF	intrinsic mode functions
ISO	International Organization for Standardization
LSB	least significant bit
MTBF	Mean Time Between Failure
PDF	probability density function
QPC	quadratic phase coupling
RMS	root mean square
SEE	spectra emission energy

SNR	signal-to-noise ratio
SPM	shock pulse method
STFT	short-time Fourier transform

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR DTIC OCA

2 (PDF) DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT

1 (PDF) GOVT PRINTG OFC A MALHOTRA

8 (PDF) DIRECTOR US ARMY RESEARCH LAB RDRL SER E K TOM RDRL SER U C LY RDRL SER E A BAYBA RDRL SER E M CONN RDRL SER E R DEL ROSARIO RDRL SER E D WASHINGTON RDLR SER E K TOM

APG RDRL VTP A GHOSHAL