NEW TECHNIQUES FOR VIBRATION CONDITION MONITORING: VOLTERRA KERNEL AND KOLMOGOROV-SMIRNOV

A Thesis submitted for the degree of Doctor of Philosophy

by:

Francisco Arruda Raposo Andrade

Department of Mechanical Engineering, Brunel University.

November 1999

Abstract

This research present a complete review of signal processing techniques used, today, in vibration based industrial condition monitoring and diagnostics. It also introduces two novel techniques to this field, namely: the Kolmogorov-Smirnov test and Volterra series, which have not yet been applied to vibration based condition monitoring.

The first technique, the Kolmogorov-Smirnov test, relies on a statistical comparison of the cumulative probability distribution functions (CDF) from two time series. It must be emphasised that this is not a moment technique, and it uses the whole CDF, in the comparison process.

The second tool suggested in this research is the Volterra series. This is a non-linear signal processing technique, which can be used to model a time series. The parameters of this model are used for condition monitoring applications.

Finally, this work also presents a comprehensive comparative study between these new methods and the existing techniques. This study is based on results from numerical and experimental applications of each technique here discussed.

The concluding remarks include suggestions on how the novel techniques proposed here can be improved.

Table of Contents

List of figures

List of tables

Abbreviations

AC	Alternating current
ANN	Artificial neural network
CDF	Cumulative density function
CF	Crest factor
CM	Condition monitoring
DC	Direct current
DFT	Discrete Fourier transform
D*n*	Daubechies wavelet with *n* coefficients
F1	Gear with smallest fatigue crack (fault 1)
F2	Gear with medium fatigue crack (fault 2)
F3	Gear with largest fatigue crack (fault 3)
FF	Form factor
FT	Fourier transform
FFT	Fast fourier transform
GMDH	Group method data handling
IDFT	Inverse discrete Fourier transform
KS	Kolmogorov Smirnov
MSM	Mean square map
PWD	Pseudo-Wigner distribution
PWVD	Pseudo Wigner-Ville distribution
RE1	Brand new gear
RE2	Normal operating gear
RE3	Normal operating gear
RMS	Root mean square
STFT	Short time fourier transform, spectrogram
SWD	Smoothed Wigner distribution
SWS	Stress wave sensor
TF	Time frequency
TS	Time scale
WD	Wigner distribution
WO	Worn out gear
WT	Wavelet transform
WVD	Wigner-Ville distribution
WWD	Weighted Wigner distribution

supervised by:

Prof. Ibrahim Esat

To my family, my girlfriend and my supervisor.

Acknowledgements

The author wishes to acknowledge the invaluable financial support from the Department of Mechanical Engineering (Brunel University), and CAPES, Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
Also the support of Dr. M. N. M. Badi (University of Hertfordshire), on the experimentation aspects of this research.

Declaration

No part of the work presented in this thesis has been included in support of an application for another degree or qualification of this or other university or institution of learning.

Chapter 1

Introduction

All mechanical equipment in industrial environment requires routine maintenance to prevent failures while they are in use. Various maintenance techniques have been devised, most of which rely on the statistical analysis of the life span of a component and/or on visual inspection. The former approach requires a vast amount of information about the components previous life cycles, while the latter will usually involve disassembly of components (intrusive maintenance). Unfortunately, these methods do not always form a basis for safe and effective maintenance programmes.

Also, the intrusive approach to maintenance is extremely time consuming and expensive. Even when a company performs routine predictive intrusive maintenance, the fault is never localised prior to the maintenance procedure. Not allowing the maintenance team to prepare according to the situation.

In the last twenty years, a third approach to maintenance has become very popular. This is based on non-intrusive continuous condition monitoring of mechanical devices. The condition monitoring (CM) approach aims at evaluating the condition of the equipment without having to stop it. In effect it aims at not only eliminating unexpected catastrophic failures, but also at maximising the system availability (the ratio of the actual running time over the desired running time – ideally 100%).

In aiming to improve system availability, two crucial points must be considered:

- Fault must be detected, localised and diagnosed before failure occurs.
- For repetitive faults eventual system reconfiguration is needed.

The achievement of these aims, will ensure that a device is able to perform its task with reasonable safety (avoiding catastrophic failures) and not affecting the quality of the output product. This forms the basis of industrial condition monitoring.

Table 1.1 shows the many techniques, and their possible applications, available to the maintenance engineer for achieving the above aims.

Table 1.1 – Condition Monitoring Method Selector [1]

	Vibration analysis	Noise analysis	Acoustic emission	Debris monitoring	Oil condition monitoring	Winding monitoring	Optical detection systems	Pressure monitoring	Temperature monitoring	Stress/Strain analysis	Corrosion monitoring	Performance monitoring
Bearings	■	■		■	■		■		■		■	
Belts										■		■
Blowers, Fans	■	■				■	■	■	■			■
Boilers, Heat Exchangers			■					■	■	■	■	■
Brazing, Welding Equipment						■		■				■
Casting, Forging Machines									■			■
Compressors, Pneumatic Machines	■	■		■	■	■	■	■	■			■
Couplings	■	■					■			■		■
Guillotines, Cutting Machines	■	■		■	■	■		■		■		■
Earthmoving, Excavating Plant	■	■		■	■			■	■		■	■
Electric Motors, Generators	■	■				■			■			■
Elevators, Conveyors	■	■				■			■	■	■	■
Escalators	■					■				■		
Filters, Separators, Valves				■	■			■			■	■
Gearboxes	■	■		■	■		■	■	■	■	■	■
Incinerators, Furnaces, Autoclaves								■	■		■	■
Internal Combustion Engines				■	■			■	■		■	■
Loader / Stackers				■	■	■		■			■	■
Machine Tools Mechanical	■	■		■	■							■
Machine Tools Hydraulic	■	■		■	■			■				■
Pressure Vessels, Accumulators			■					■	■	■	■	■
Pumps	■	■						■	■		■	■
Structures, Rigging	■									■	■	
Transformers				■	■	■		■	■			■
Turbines, Aero Engines	■	■		■	■		■	■	■		■	■
Vacuum Equipment	■	■		■	■			■	■		■	■
Wire, Cable Making										■		■
Winding, Lifting Machinery			■							■		■

Today vibration and acoustic analysis are very popular methods for monitoring the condition of rotating devices. It is particularly used for monitoring dynamic (i.e. rotating) devices such as gearboxes, bearing, rotating shafts and cutting lathes amongst others [1]. This popularity is directly related to the flexibility of this method (see the number of applications where it can be used in Table 1.1) and also to the ease with which these signals can be collected. Acoustic analysis only needs a microphone, while vibration signatures can be easily obtained with vibration transducers coupled to data loggers. The actual type of transducer is dependent on the frequency being monitored [2, chapter 7], *low frequencies is best sensed by displacement, medium frequency is best sensed by velocity, and high frequencies is best sensed by acceleration.*

For the above reasons this research will focus on signal processing techniques which can be used for vibration analysis of machines.

Finally, the main thrust behind this research comes from the urgent need to improve on the existing vibration based condition monitoring methods, and also to develop new methods for CM. This will aim at meeting two basic industry requirements:

- Increase of equipment uptime. Effective CM does increase the operational availability of various industrial systems. This will not only increase the production (hence revenue), but also reduce costs due to unexpected breakdowns.
- Increase of industrial safety levels. Machinery safety is today a key issue in industry. Effective CM helps to reduce unexpected failures, which can lead to unsafe situations.

This report is formatted as follows:

Chapter 2, Literature survey. Is a review of the state of the art literature on the different aspects and techniques for vibration condition monitoring and fault diagnostic.

Chapter 3, Numerically simulated and experimental data. This chapter describes how the simulated and experimental data were obtained. This data is used as the input to the different techniques being studied in this research.

Chapter 4, Statistical methods for condition monitoring. This chapter gives an introduction to the theory behind the several statistical tools used in industrial vibration analysis. It also introduces the KS test as a novel technique for vibration condition monitoring. These tools are used to process the experimental and numerical data.

Chapter 5, Frequency and Time-frequency approach to condition monitoring. This chapter describes two frequency methods (pure spectral analysis and cepstrum analysis) and two time-frequency techniques (Spectrogram, Wigner distribution) used for analysing time-series. These techniques are used to process the numerical and theoretical data. Also, it reviews other time-frequency techniques, which were commonly used in the past.

Chapter 6, Wavelet approach to condition monitoring. This chapter describes the theory behind wavelets and their usage in condition monitoring. Also, it compares the performance of different wavelet families (Harmonic wavelets and Daubechies series). These wavelets are also used to process the numerical and experimental data.

Chapter 7, Non-linear methods for condition monitoring. In this chapter non-linear signal processing techniques are discussed. The Volterra series is then studied in detail, and used (for the first time) as a tool for vibration signal analysis.

Chapter 8, Discussion. In this chapter a summary and a review with a direct comparison between all the techniques discussed in the previous chapters will be included. This aims to serve as a guide for choosing appropriate techniques for specific applications.

Chapter 9, Conclusions and further work. This chapter concludes this research and also observes topics, which could, and should, be studied further.

1.1. Research Objectives

The main objective of this research is to establish a basis of comparison between the different signal processing techniques used today for the condition monitoring of industrial rotating devices.

Furthermore, it aims at introducing the usage of two novel techniques to vibration condition monitoring: the KS test and the Volterra kernel estimation in vibration analysis. To achieve these objectives the following programme was devised:

- Perform a complete literature review of today's available techniques for the condition monitoring of rotating devices. This review should include not only the theoretical but also practical techniques. This step is essential to give an insight of the field being researched.

- Collect experimental data related to a specific type of fault (fatigue cracks) in rotating gears. These data sets will form the basis for a true comparison to be made between the different techniques under study. Three fatigue cracks of different sizes, to simulate fault advancement, were selected. The geometry of these cracks is fully described in chapter 3.

- Introduce the KS test to vibration condition monitoring and compare its effectiveness with existing techniques.

- Adapt the usage of Volterra series (so far primarily used in biomedical applications), so that it can be applied to industrial condition monitoring. This will allow an assessment of its prospects as a signal processing technique for condition monitoring and diagnostic.

- Use the different techniques under study to process the numerical and experimental data; observing the advantages and disadvantages of each method.

- Collect the main traits and perform a direct comparison between each technique under study, and how these fit with the published literature on the subject.

Chapter 2

Literature survey

Condition Monitoring and Diagnostics of industrial equipment are techniques, which complement any industrial maintenance programme. In fact, its main purpose is to determine the optimum amount of preventive maintenance required by different components and machines. It is this optimisation which will lead to the lowest machine running costs.

The objective of this chapter is to present a review of the previous work on this subject. This review is grouped into subsections, namely:

- Maintenance and condition monitoring;
- Vibration condition monitoring;
- Gearbox Vibration;
- Statistical descriptors;
- Frequency and time-frequency methods for condition monitoring;
- Wavelet decomposition for Condition monitoring;
- Non-linear methods for condition monitoring;
- Summary of literature survey;

At the end of the chapter a summary is included.

2.1. *Maintenance and Condition Monitoring*

The notion of maintenance is, today, familiar to most people. Its main objective is to avoid, or reduce to the lowest possible level the number of unexpected failures. The ever-increasing complexity of new machine is also increasing the complexity of maintenance systems. Therefore, the need to develop faster, cheaper and more accurate diagnostic systems for industrial maintenance is always present.

Maintenance lies within the field of Terotechnology [3, chapter 1] and it can be subdivided according to Figure 2.1, below.

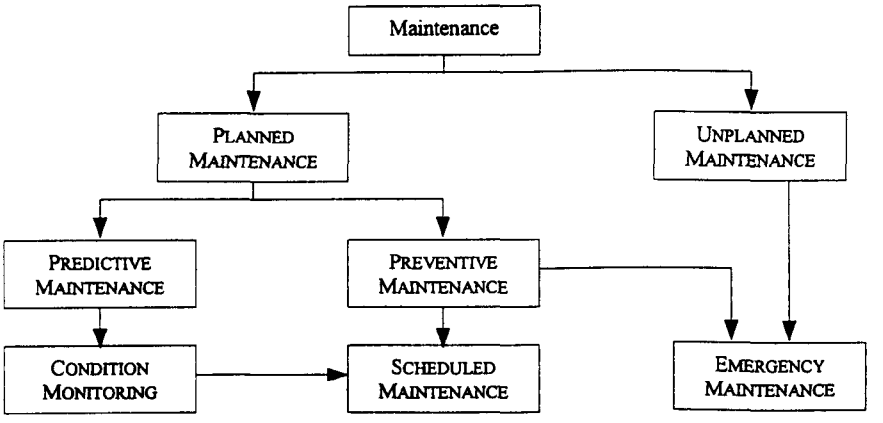


Figure 2.1 – Approaches to maintenance

As it can be seen there are several approaches to maintenance. All approaches are valid, however each will be associated with a different cost. Unplanned maintenance (i.e. do nothing until failure occurs) is the simplest method, however it also tends to be the most expensive. On the other hand, planned maintenance requires a certain degree of organisation, which is paid off by the system reliability.

Condition monitoring lies within the scope of planned maintenance and aims at establishing the optimum level of preventive maintenance that should be directed at each component within different machines. This technique has several inherent advantages. The machinery is inspected periodically to assess the actual state of its components, this ensures that the equipment is only maintained when needed, maximising uptime. Also this inspection is performed while the machine is on-line, again, maximising its uptime. Finally, observation of fault development allows for intrusive maintenance to be planned in advance.

Today a reasonable number of publications in the field of maintenance and condition monitoring can be found [1,2,3,4,5,6]. In [1] a detailed description of the different condition monitoring techniques and their respective applications can be found. Including vibration, acoustic, debris, corrosion, thermal and pressure analysis amongst others.

From all the techniques observed above, vibration analysis was chosen due to its generality (Table 1.1 shows some of the applications where vibration analysis can be used), and its ability to analyse machinery with highly dynamic characteristics. Furthermore, it has already been shown that different faults lead to different vibration characteristics. Table 2.1 summarises some common mechanical faults together with its characteristic vibration frequency and amplitude. Further details in this area can be seen in [7].

2.2. *Vibration Analysis*

Vibration analysis, and its counterpart: acoustic analysis, are amongst the most popular forms of condition monitoring. The vibration signature of a device contains detailed information not only about the dynamic behaviour of its moving parts, but also about its structural components (as these contribute to the overall vibration transfer function of the device). From the vibration signature a full diagnosis of the condition of the device is obtainable.

For most practical applications an effective condition monitoring program will consist of four basic steps [8]. These are:

- ***Identification of check points.*** In this step one needs to ascertain which are the relevant data collection points. These collection points must be chosen so that they enhance the signal distortions, which might arise from a fault. Ideally, sensors should be placed as close as possible to the specific component being monitored.
- ***Vibration Measurement.*** In this step one needs to determine how much data will be collected, select appropriate vibration transducer, sampling rates and sampling periods. This step is of utmost importance to an effective condition monitoring

system as the sampling rate is very much related to the range of recorded frequencies. The collected data is effectively a time-series, as successive observations are not independent and that the time order of the observations must be observed [9].

- ***Digital Signal Processing (DSP)***. Here, a signal processing technique, which will enhance the patterns caused by existing faults, must be chosen. The specific DSP technique to be used is very much dependent on the problem at stake. Today there is a wide range of techniques available. This choice is crucial to the performance of the monitoring system [10].
- ***Post-process analysis and interpretation***. In this final step, the results obtained after processing the collected data are analysed. This analysis can be either visual, which requires human interaction, or it can be automated. In the latter case, an automated pattern recognition system must be implemented.

Previous work on the usage of vibration analysis to condition monitoring of mechanical devices has shown that each type of fault introduces different components to the vibration signature of a device [7]. The table below summarises how the vibration signal components can be related to different faults.

Table 2.1 – Faults and vibrations

Fault	Frequency	Vibration amplitude
Imbalance	1 x rpm	Proportional to imbalance. (radial)
Misalignment or Bent shaft	1 x rpm (usual) 2 x rpm (often) 3, 4 x rpm (seldomly)	Mainly axial, usually more than 50% of radial vibration
Damaged rolling element bearings	High frequencies not multiple of rpm	Unsteady
Journal bearings loose in housing,	Sub-harmonics of shaft rpm (½ , 1 x rpm)	
Lubrication	High frequencies not multiple of rpm	
Oil-film whirl/whip in journal bearings	0.42 to 0.48 x rpm	
Damaged or worn gears	Tooth meshing frequency + harmonics (no. of teeth x shaft rpm)	Low
Mech. Looseness	2 x rpm	Second harmonic > 2x fundamental f
Defective belt-drives	1,2,3, and 4 x belt rpm	Erratic or pulsating
Electrical defects (magnetic fields).	1 x rpm or 1, 2 x synchronous frequency	Disappears when power is turned off

The usage of vibration signatures as a basis for condition monitoring can be found in a wide range of applications including: cutting tools, drilling and rotating devices (gearboxes, bearings, shafts) [6,11,12,13,14,15]. The actual number of successful applications is very large, and it is still growing very fast. Only a few examples are given here.

This research aims at comparing different digital signal processing techniques, which can be used to monitor the condition of a spur gear. More specifically to identify the presence of tooth fatigue cracks in its very early stages.

2.3. Gearbox Vibration

Gearboxes play a vital role in rotating machinery. It is present in most modern machinery ranging from automobiles to chemical plants. Its main purpose is to modify the parameters of power being transmitted. Today there is a wide range of types and configurations of gearboxes, however some basic wear symptoms can be found in most of these configurations. This opens the way for the usage of digital signal processing techniques, which aim to maximise the vibration features arising from wear and faults on gearboxes.

Gearbox failure will usually occur on the load carrying elements (i.e. shafts, bearings and gears). There are 3 root causes for gearbox failures.

- Gearbox operation beyond design limits.
- Lack of preventive maintenance (including lubrication and seals).
- Improper original assembly.

This research focuses on gear failures alone, not considering bearing and shaft failure. The most common types of gear failure are related to tooth geometry irregularities (tooth scoring and pitting, shaved tooth, cracked tooth and broken or missing tooth amongst others), or gear misalignments. All these failures will invariably introduce changes in the vibration signature of the gearbox, causing an overall increase (or reduction) in the vibration level of the device.

There are several signal processing techniques available for the analysis of time series (vibration signature). The methods covered by the scope of this research were divided into four categories, namely: statistical descriptors, frequency and time-frequency analysis, wavelet decomposition and non-linear methods. These are discussed in the next sections.

2.4. Statistical descriptors for CM

The usage of statistical descriptors in condition monitoring has been very much limited to the time domain analysis of a vibration signature. This analysis is based on the comparison of some useful vibration parameters [1], namely: mean, variance, standard deviation, absolute deviation, root mean square, skewness, kurtosis, peak to peak value, crest factor, form factor, spike energy, etc. (see section 4.1)

It must be noted that two moments in particular, namely: skew and kurtosis, are widely and successfully used in condition monitoring applications. Their main advantage is related to their mathematical simplicity. Unfortunately, these measures alone cannot be used to fully diagnose a mechanical device.

The crest factor (CF) technique has also shown potential for industrial condition monitoring applications [16,17]. In [17] Badi et al have shown how it can be used in assessing the condition of a drive line (including a simple gearbox). Finally, the spike energy technique is now used commercially by IRD Mechanalysis for the condition monitoring of bearings [11].

This research suggests the usage of a well known statistical comparison test, as another technique, which could be added to the hall of statistical descriptors for condition monitoring. This technique is known as the Kolgomorov-Smirnov (KS) test. Although it has been successfully used to compare distributions in other fields such as astronomy [18], biology [19] and identification of periodicity in signals [20], to the best of the author's knowledge, the KS test has not yet been applied to the analysis of vibration signatures from rotating devices. (see section 4.2)

This test works by comparing the whole cumulative probability density function of two time series; these, usually are a reference signal (good signal), and a second signal which indicates the current state of the rotating device. The outcome of this comparison is a similarity percentage (indicating how similar the two signals are).

The results obtained in this research were very encouraging, and suggest that this method can be efficiently used for fault detection/advancement analysis[21, 22].

2.5. *Frequency and time-frequency methods for condition monitoring*

Frequency and time-frequency methods rely very much on the Fourier transform [23,24] to convert the time domain vibration signature into frequency and/or time-frequency domain. This conversion can be performed by means of well established Fast Fourier Transform (FFT) algorithms [25,26].

Pure spectral analysis is the first frequency method under consideration, as it has been widely used for monitoring the condition of rotating devices. This technique describes a vibration signature in terms of all its frequency components (and their respective phases), within the Nyquist theorem. Faults are detected by visual comparison between 'good' (gearbox in good operating condition) and 'faulty' (gearbox with a fault) spectrums. Unfortunately, this technique is not sufficiently robust to reliably detect and diagnose gearbox faults. This can be related to the fact that the vibration components from some types of faults are localised, hence when transforming the whole time series the fault contribution is overshadowed by the rest of the signal. (see section 5.1)

Note, however, that although this technique in itself is not sufficiently robust to reliably detect faults, it served as a basis for describing how different faults affect the vibration spectrum [7,27]. Also, this technique forms the basis of other condition monitoring methods, such as cepstrum and time-frequency (spectrogram) decompositions.

The cepstrum (anagram of spectrum) is a direct extension of the pure spectral analysis; however, it focuses on the detection of harmonic components in a time series. The cepstrum is a non-linear signal analysis technique [28], which was first used for echo detection [29] in 1963.

The cepstrum works by performing the inverse Fourier transform of the logarithm of the Fourier transform of the input sequence. There are two variations of this technique, namely: the complex cepstrum and the real cepstrum. (see section 5.2)

In the complex cepstrum, the complex logarithm is used after the input series is Fourier transformed. This is useful when phase information of the input signal needs to be retained.

In the real cepstrum, the logarithm of the magnitude of the input signal Fourier transform is evaluated. This is much simpler, but it ignores and discards the phase information in the time series. A discussion between these two variants of the cepstrum method can be found in [30].

After the successful applications of cepstrum for speech analysis [31], and geophysical data processing [32], Randall used it as a tool for condition monitoring and diagnostics of rotating devices. Randall's successful results [27,33,34,35], led many other researchers to use it in a wide range of applications [4,36,37,38] including gear condition monitoring [34,39,40,41].

In this research the real cepstrum is used to analyse the vibration signatures. Also cepstral subtraction is performed between good and faulty signals. This has proved to be an efficient identification technique. Work has also been done on using the cepstral reconstruction (inverse of cepstrum transform) to diagnose the type of fault.

The remaining techniques under the scope of this section can be grouped as time-frequency methods. These methods aim to model a signal by describing which frequencies were present at which instants in time. The output from this method is a 3 dimensional map, containing time information, frequency information and energy

content. In this research the two time-frequency methods were studied, namely: spectrogram and Wigner-Ville Distribution.

The spectrogram (also known as short-time Fourier transform or windowed Fourier transform) was one of the first time-frequency distributions. It was first used in the 1940's to analyse human speech [42]. Since then, subsequent developments [43,44,45] turned this technique into a powerful tool for time-frequency analysis (see section 5.3). A review comparing the different approaches to the spectrogram calculation can be found in [46].

The spectrogram works by viewing segments of the time domain data through windows, and then transforming these segments to frequency domain, *representing the energy distribution of the signal over the frequency domain at every instant of time* [47].

The most commonly used window function is the Gaussian function, as it possesses similar shapes in the time and frequency domain. Therefore it does not introduce unwanted sidelobes into the spectrogram. Other window functions, such as Hamming, Hanning, rectangular and triangular have also been used. A review of the properties of different window functions can be seen in [48].

Some applications demonstrating the effectiveness of the spectrogram as a time-frequency analysis tool can be seen in the works of:

- Wang and McFadden on gear condition monitoring [47,49];
- Fazio and Molinaro [50] in echo detection, and Kim et al [51] in geophysics;
- and the work of Kingsbury et al in speech analysis [52].

These are only a few examples of the successful applications of the spectrogram. The work by Wang and McFadden on gear condition monitoring is directly related to this research and must be particularly noted. In [47] it was found that:

- *the spectrogram has advantages over the Wigner-Ville distribution for the analysis of vibration signals for the early detection of damage in gears.*

This agrees with the results obtained in this research when analysing the performance of the spectrogram.

Finally, it must also be noted that the spectrogram has one main handicap related to the window function choice. If the window is chosen so that it focuses on time resolution (narrow window in time domain), then once the Fourier transform is performed the window function will have low frequency resolution (wide window in frequency domain). As stated by Cohen [53], *there is an inherent trade-off between time and frequency resolutions in the spectrogram.* Therefore, it cannot be used to analyse signals with large and small patterns simultaneously.

The second time-frequency method studied in this research programme is the **Wigner distribution** (WD). This technique was first used for signal analysis in 1948 [54]. Also, like the spectrogram, the WD is a member of the Cohen's class of time-frequency distributions [53,55]. A comprehensive review of this technique was published by Claasen and Mecklenbrauker [56,57,58] in 1980.

The WD works by *adding up pieces made up of the product of the signal at a past time multiplied by the signal at a future time, the time into the past being equal to the time into the future* [59]. Note that as this 'overlap' is done, the WD will weight far away times equally to near times. This basic property can be modified, by using a window function to emphasise the signal around a given time 't'. This modified WD is also known as pseudo WD (PWD), smoothed WD (SWD) [59] or weighted WD (WWD) [60]. (see section 5.4)

In this research the PWD will be used, as it is more localised than the pure WD. This is of uttermost importance in reducing the cross-term interference, which is an intrinsic disadvantage of PWD.

The PWD has also been successfully used in many applications [14,15,60,61,62]. These cover a wide range of areas from echo detection to rotating machinery condition monitoring.

The work of Staszewski [60] et al is directly related to this research, and two points must be noted:

- *The WD is capable of detecting local tooth faults in spur gears*
- *The PWD reduces cross-term interference in the time-domain. This simplifies the results.*

This justifies the inclusion of a theoretical description of the WD in this research and also justifies the choice to only analyse data (experimental and simulated) with the PWD (see section 5.4.2 to 5.4.5).

Finally, it must be noted that cross-term interference is not the only handicap of the PWD, when compared to other TF representations. The introduction of the window function, also affects the frequency resolution, which can cause *a loss of sensitivity regarding fault detection*, as stated by Staszewski [60].

2.6. *Wavelet decomposition for Condition Monitoring*

In the recent years wavelets have been successfully applied to a wide variety of condition monitoring applications. Part of the power behind wavelets lie in the fact that it comes from a wide variety of research backgrounds. As a consequence wavelets seems to have many different definitions. In [63], Sweldens attempts to outline a common denominator for various developments, which have been called wavelet.

The wavelet transform (WT) is an example of a time-scale decomposition of a given signal. It originated in the 1980's with the works of Morlet and Grossman [64,65]. This method is still evolving [63], and so far it has been applied with some success to areas as diverse as: image and sound processing [66,67,68], biosignal analysis [69], and vibration analysis [70,71,72,73,74] (including rotating machinery condition

monitoring). The wavelet has also been used, with some success, as a pre-processing method in automated condition monitoring systems [75].

The wavelet transform works by expanding the time series in terms of a family of functions (wavelets) generated from the dilation and translation of a single function, the 'mother wavelet' [76,77]. The translation of the 'mother wavelet' allows for the analysis of the signal at different instants in time. While, the dilation of the 'mother wavelet' has the effect of narrowing, or widening, the time window depending on whether the frequency is high or low. This is one of the most important characteristics of the WT, as it enables the analysis of very short lived transients coupled to longer lived transients in time domain. (see section 6.2)

In this research the Daubechies series and harmonic wavelets will be used (and compared). The choice for using these wavelets is based not only on their orthogonal properties, but also on the results which leading researchers (Newland, Wang and McFadden, and Staszewski and Tomlinson, amongst others) have already obtained.

In [78], one of the first applications of wavelets to vibration condition monitoring, Wang and McFadden concluded: *it is possible to detect a fault in one of the gears by visual inspection of the time-scale distribution.* In the following year McFadden [79] added: *the wavelet transform enables better separation in the time dimension of signals of short duration at high frequencies than the Gabor spectrogram.* Newland [70] has shown the effectiveness of wavelets, analysing vibration due to road and rail traffic on four-storey building. Also, he compares harmonic wavelets [80] to the dilation wavelets (daubechies series), concluding, amongst other things: *Harmonic wavelets can be described by a simple analytical formula, while dilation wavelets cannot be expressed in functional form.*

2.7. *Non-linear methods for CM*

The techniques previously discussed are all widely used for linear systems. Methods for the analysis of non-linear systems have not received that as much attention in the past years as the previous methods. Although it is possible to represent a non-linear

system by dividing it into smaller linear sections and use the already well established theory for linear systems; it has been recognised that non-linear processes can only be analysed as a whole by non-linear methods [81, 82].

For this reason, the scope of this research included a non-linear signal processing technique, namely, **the Volterra Series**. This technique was first introduced by Volterra in 1887, and subsequent developments to it, especially those by Frechet [83] and Wiener [84,85] pushed this method into a field of mathematics whose main aim is the study of functionals and the representation of non-linear systems. Also, it must be noted that although this technique has been available for more than one century, it has not yet (to the best of the author's knowledge) been used for monitoring the condition of rotating devices.

There are also other non-linear techniques, such as: artificial neural networks [86], quadratic detectors [87], parameter estimation [88] and block-oriented systems [89], amongst others. A review of these methods can be seen in [81]. These above techniques, however, rely on many heuristic decisions, and are also very much dependent on the system to be modelled.

Now, Volterra series do not suffer from the above handicaps, and lead to a more general model, which can be used for a wide range of inputs and combination of waveforms. The higher the dimensionality of the series the higher its generality. Unfortunately these higher order series are computationally expensive, and [102] *not easy to assimilate and deal with.*

Comprehensive reviews about Volterra series have been compiled by Hung and Stark [90], and by Barret [91], but it must be noted that this method is still evolving as it is starting to be used in practical applications, such as:

- The work of Koh and Powers [92] shows that a Volterra filter can be used to model and predict the non-linear dynamic behaviour of offshore structures (excited by irregular sea waves), more accurately than other linear methods.
- The work of Korenberg and Morin [93] on myoelectric signal discrimination shows that Volterra series can be used to distinguish biological signals arising from different muscle contractions.
- The work of Korenberg and Paarman [94], which contains an adapted orthogonal algorithm to process time series. In this study it was found that this technique was able to accurately estimate the relevant frequencies in a signal, even if this signal is *heavily corrupted by noise* and has *randomly data missing*.

The Volterra series methods work by measuring the Volterra kernels [95] (a constant, a one-dimensional and multidimensional weighting functions). These kernels are at the core of the non-linear system representation. (see section 7.3)

There are different algorithms to accomplish the kernel estimation [90,96,97,98,99,100, 101]. In this research the Lee-Schetzen cross-correlation algorithm [102] will be used due to its simplicity when compared to other methods. A review with the properties, advantages and disadvantages of the many approaches to Volterra kernel estimation can be found in [103].

In this research it was found that the second order kernel does indicate the presence of a fault in the vibration signature of the rotating device.

Therefore, although the Volterra series is a very mathematically biased and computationally expensive technique it can also be seen as an available tool for the modelling and condition monitoring of industrial equipment.

2.8. *Summary of Literature survey*

Sections 2.1 to 2.7 reviewed the existing literature available for the different techniques studied under this research. As it can be seen when applying these techniques in an industrial environment, the maintenance engineer primarily focuses on magnifying the vibration patterns induced by the fault, so these can be visually observed. This observation is usually done against a reference (or good) signal. Hence condition monitoring becomes a comparison exercise between two different images (or vectors).

Some work has been done aiming at automated condition monitoring systems. Unfortunately these systems are extremely dependent on the feature extraction method, which in turn is a heuristic process. Therefore, although it is possible to achieve good results in specific applications, the generality of the system is very restricted.

Statistical descriptors, pure spectral analysis, cepstrum, spectrogram, Wigner-Ville distributions and wavelets all have their advantages and disadvantages. These were already described in brief, and will be discussed in detail in the later chapters. All these techniques have already been used successfully in a variety of applications.

Now this research suggests two other techniques for condition monitoring and diagnostics of rotating devices, namely:

- The Kolmogorov-Smirnov test.
- The Volterra series.

Although these techniques are relatively well established in other fields, to the best of the author's knowledge, they have not yet been yet tried in condition monitoring applications.

The first technique is a tool for time domain condition monitoring. It works by comparing the cumulative probability density function of two distributions. These

distributions usually are a reference signal (good signal), and a second signal which indicates the current state of the rotating device. The outcome of this comparison is a similarity percentage (indicating how similar the two signals are).

The second technique also works in time domain. It uses the time series to model the vibration signature of the device. During this modelling, kernels are estimated. These serve as a basis for comparing different vibration signatures.

Both these techniques have proved very effective in monitoring the condition of gearbox faults by vibration analysis.

Chapter 3

Experimentation

This chapter gives a full description of the experimental setup and of the data (simulation and experimental) used throughout this research. As suggested in the previous chapter, this research uses a gearbox as the main tool to test the hypotheses formulated in the main objectives of this research. Hence, all data files are related to gear vibration condition monitoring

This chapter is divided into three main sections. The first describes the gear faults which will be studied in this research. The second describes the apparatus used in the experimentation and the third describes how the collected vibration signatures were grouped and labelled. The latter is of utmost importance as the results from the different analysing techniques refer to the labels introduced in this section.

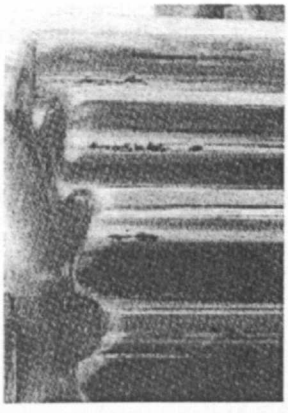
3.1 Gear faults

All major faults in gearboxes can be classified into shaft (misalignment, unbalance), tooth (wear, scuffing and cracking) and bearing (rolling element defect, bearing-housing fit) related problems [1]. The early gear defects on a single tooth can be called local tooth defect. The faults studied in this research are wear effects and cracks caused by bending fatigue.

3.1.1. Wear effects

The meshing contact between gear teeth will invariably cause wear on the gear tooth face. This wear appears in the form of pitting and scuffing. A common cause for

these defects is gear misalignment. This overloads one end of the tooth, causing stress concentrations at specific areas of the tooth. This leads to a fast deterioration of the active profile of the gear and a rapid loss of the tooth involute profile shape. This gives rise to dynamic loads, which causes excessive localised wear.



In this research the effects of these faults on the overall vibration signature of a gear is analysed. For this a worn-out gear was used. This gear was obtained by running a new gear until it showed worn-out signs (i.e. pitting and scuffing). Figure 3.1 shows these defects on the worn-out gear. Note that the pitting on this gear is at its very early stages.

Figure 3.1 – Worn-out gear

3.1.2. Fatigue cracks

Fatigue crack is the fracture of metal under repeated (cyclic) stressing below the yield stress of the material. Fatigue cracks reduce the effective stiffness of a tooth. This affects the vibration signature of the gear.

Fatigue cracks can lead to tooth breakage under three main forms: bending fatigue, overload and random fracture [105]. Tooth breakage can cause catastrophic effects on machinery on a wide range of industries, and must be prevented. Fracture usually occurs at the root of the tooth, starting near the end of the tooth under tension (when the load is unidirectional). The crack propagates along the base of the tooth. Also, in some cases the fracture can be initiated by surface damage due to contact stress. The figures below show typical fatigue cracks, and how these can lead to tooth fracture.

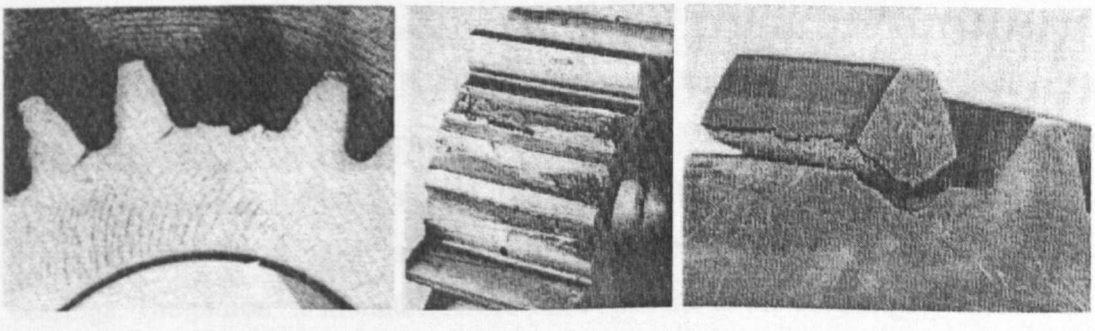
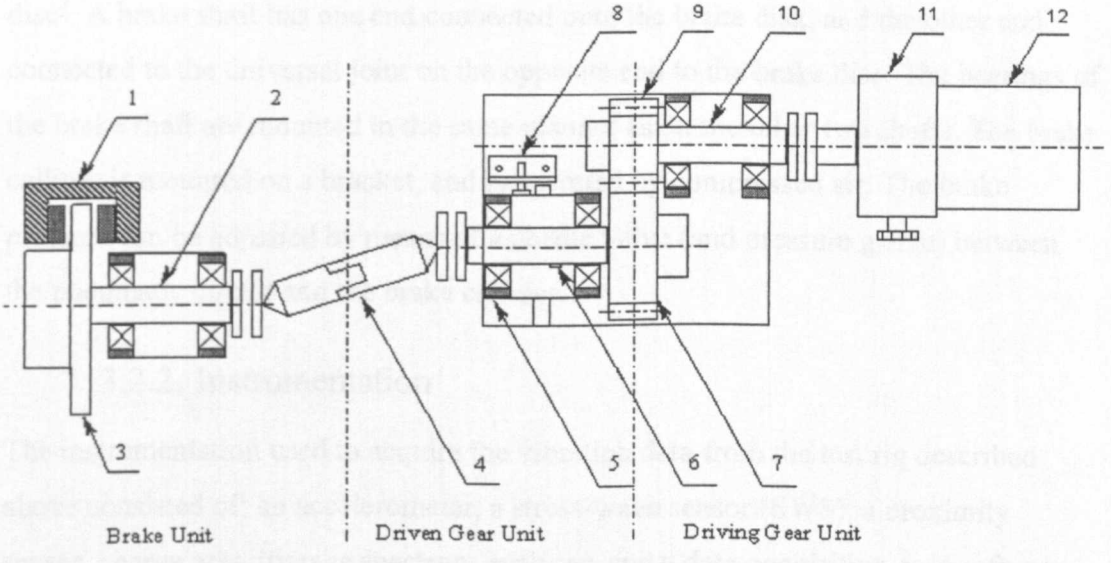


Figure 3.2 – Gears showing fatigue cracks

3.2. *Experimental Setup*

3.2.1. Test Rig

Aiming to investigate the previously described types of faults a test rig modelling a drive-line for the collection of vibration time data from rotating gears was used. This rig has been designed for a previous PhD research project. The test rig contains common components present in a real drive line using meshing gears (i.e. gears, shafts, bearings and couplings). The layout for the rig is shown in the figure below. Gear details are included in section 3.3 and [106] include further rig details.



1. Brake calliper
2. Brake shaft
3. Brake disc
4. Universal joint
5. Movable base plate
6. Output shaft
7. Wheel
8. Spacer block
9. Pinion
10. Input shaft
11. Kopp Variator®
12. AC motor

Figure 3.3 – Layout of experiment test rig

The rig can be divided in three main sections: the driving unit, the driven unit and the brake unit. The driven unit consists of an AC motor[12] (3-phase, 0.75 hp, 1440 rpm) adapted to a Kopp Variator[11]. The Variator provides an adjustable motor speed in the range between 200 and 1200 rpm. The output of the Variator is connected through an Oldham coupling to the input shaft[10]. The input shaft is supported by two roller bearings. The bearing housings are split in two halves to obtain easy access to the bearings. The pinion[9] is mounted on the shaft opposite to the coupling.

The layout of the driven unit (output shaft[6]) is symmetrical to the input shaft. The unit is mounted on a movable base plate[5]. The base plate slides on a guide, which guarantees the alignment between the input and the output shaft. The movable base plate makes the change of the test gears easier and allows the usage of gears with different pitch diameter. The connection between the output shaft[6] and the brake unit consists of two universal couplings in a telescopic assembly[4]. Note that a key was placed on the output shaft[6]. This was used to give a reference signal, indicating the position of the gear at a given time.

Finally, the brake unit consists of the brake calliper[1], the brake shaft[2] and the brake disc[3]. A brake shaft has one end connected onto the brake disk, and the other end connected to the universal joint on the opposite end to the brake disc. The bearings of the brake shaft are mounted in the same manner as on the other two shafts. The brake calliper is mounted on a bracket, and is operated by compressed air. The brake pressure can be adjusted by means of a needle valve (and pressure gauge) between the pneumatic circuit and the brake calliper.

3.2.2. Instrumentation

The instrumentation used to acquire the vibration data from the test rig described above consisted of: an accelerometer, a stress-wave sensor (SWS), a proximity sensor, charge amplifiers, a spectrum analyser, and a data-acquisition card software. The diagram below illustrates how these devices were interfaced.

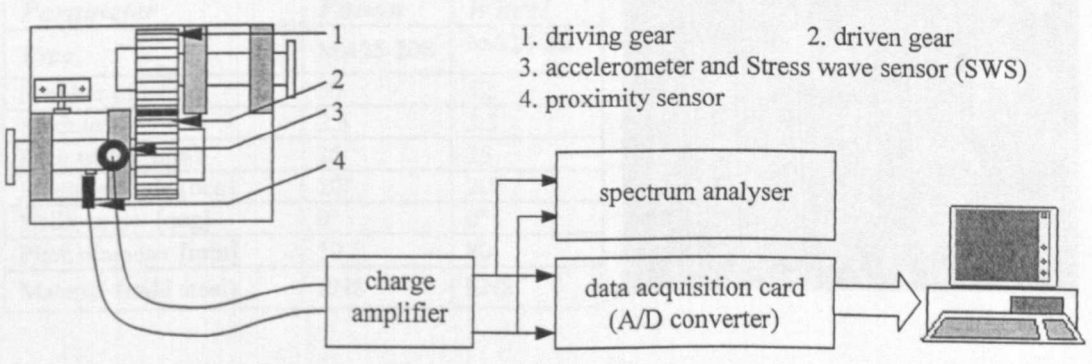


Figure 3.4 – Layout of instrumentation used in test rig

As it can be seen, the vibration signature (captured by the accelerometer, and the stress wave sensor) and the revolution trigger signal (captured by the proximity

sensor) were fed into a charge amplifier. The amplified signals were used as inputs to a spectrum analyser. This allowed an on-line visual inspection of the meshing frequency, to ensure that the rotational speed of the gears was constant.

The amplified signals were also fed into an analogue-to-digital converter. The digital signal of the vibration signature and the trigger signal were stored in a personal computer. The digitised signals were used by all the signal processing methods studied in this research.

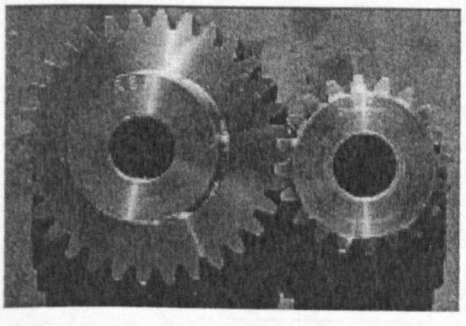
3.3. Test Gears and Simulated Faults

The test gears used in the experiments were manufactured by Davall Gears. A full description for these gears can be found in appendix 1, hence only a brief description will be included here.

In all seven mild steel unhardened spur gears were used. These were non-hardened and non-undercut. The reason for this selection is related to the difficulties in implementing faults (fatigue cracks) in hardened gears. For example for a hardened gear to show worn effect would take a much longer time than non-hardened gears. The gears were manufactured to the following standards: DIN 3962, DIN 3963 and DIN 867. The table below gives the main technical characteristics of the gears used.

Table 3.1 – Characteristics of Test Gears

Parameter	*Pinion*	*Wheel*
Type	MA25-20S	MA25-32S
Number of teeth	20	32
Module	2.5	2.5
Face width [mm]	25	25
Pressure angle [deg]	20°	20°
Helix angle [deg]	0°	0°
Pitch diameter [mm]	50	80
Material (mild steel)	EN8	EN8



The seven gears were used as follows:

- 2 gears used as normal reference gears, obtained by recording the vibration data after 5 minutes of gear usage (under a rotational frequency of 5Hz - or 300rpm - and brake torque of 20Nm)

- 1 gear used as a reference for a **brand new** gear, obtained by recording the vibration data soon after it was first used.
- 1 gear used as a worn out gear, obtained by using the gear (under a rotational frequency of 5Hz – or 300 rpm - and brake torque of 20Nm) until worn out effects (pitting and scuffing) could be seen.
- 3 gears with simulated fatigue cracks with different severity were used as the faulty gears. A description of how these faults were implemented is included below.

3.3.1. Fault Implementation

The faults were introduced into three of the test gears. The overall effect of a crack is a reduction of stiffness and mass of a gear tooth. The crack was simulated by cutting into the root of a tooth with a thin disc cutter. The disc cutter had a diameter of 55mm and a thickness of 0.3mm. The cut aimed at replicating a crack at the critical tooth section which is growing along the critical stress line. In practice, due to access restrictions caused by the adjacent teeth, the crack could be started at the critical section, but its direction deviated from that of the critical stress line [107].

For simulating the growth of a crack and comparing fault severity, the cuts were introduced into three different gears. The cuts started at one gear face and progressed in three incremental steps across the face width of the gears. The final, being the most severe cut. The cuts introduced, and their geometry, are shown below:

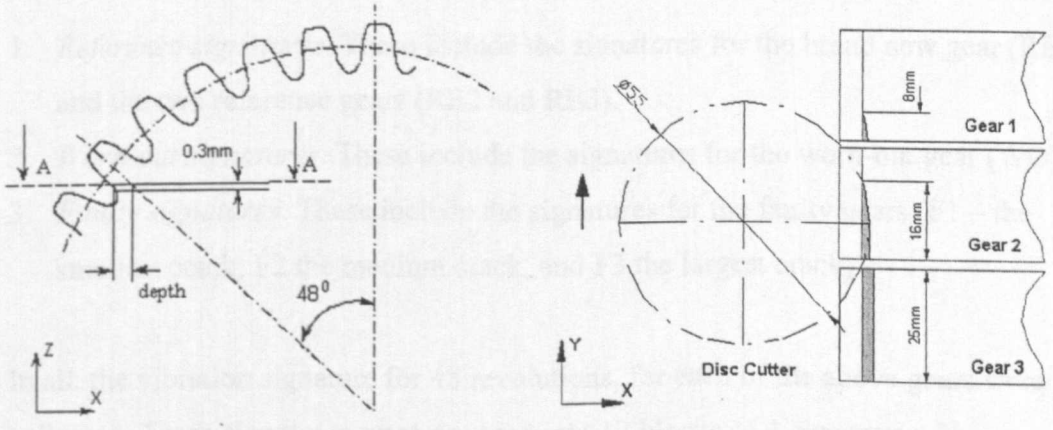


Figure 3.5 – Schematic diagram of spur gear used in the experiments

The schematic of the 32-tooth spur gear, onto which the faults were introduced, is shown above. The disc cutter access restriction would only allow a minimum cut angle of 42°, from the tooth centre line, compared to the critical stress line, which lies at approximately 40°. The schematic on the right, shows an A-A plain view of the faulty sections of the test gears. Gear 1 has the smallest cut (Fault 1) and gear 3 has the largest cut (Fault 3), which extends across the whole gear face width. The table and figures below summarises the cut geometry, and display the implemented faults.

Table 3.2 – Gear cut geometry

Gear Condition	Cut geometry			
	Depth (mm)	Width(mm)	Thickness(mm)	Angle (°)
F1	0.8	8	0.35	40
F2	1.6	16	0.35	40
F3	2.4	25	0.35	40

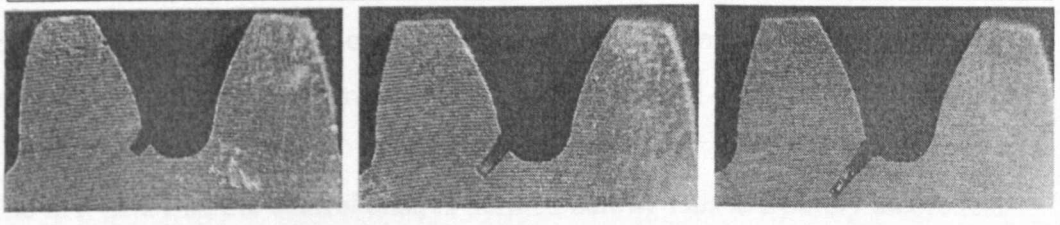


Figure 3.6 – Implemented fatigue cracks

3.4. Test procedure and Data labelling description.

The test rig described in the previous section was used to collect vibration signatures for the gear under test. The signatures were grouped into 3 classes:

1. *Reference signatures*. These include the signatures for the brand new gear (RE1) and the two reference gears (RE2 and RE3).
2. *Worn-out signatures*. These include the signatures for the worn-out gear (WO).
3. *Faulty signatures*. These include the signatures for the faulty gears (F1 – the smallest crack, F2 the medium crack, and F3 the largest crack)

In all, the vibration signature for 48 revolutions, for each of the above gears were collected. These signatures were grouped into 12 blocks of 4 signatures. The following names were given for these blocks:

- Brand new gear: RE1B1 to RE1B12
- Reference gears: RE2B1 to RE2B12 and RE3B1 to RE3B12
- Worn-out gear: WO1B1 to WO1B12
- Faulty gears: F1B1 to F1B12 (for smallest fault), F2B1 to F2B12 (for medium fault), and F3B1 to F3B12 (for largest fault)

These labels are used to identify to which gear specific vibration signatures relate. Also in chapter 7 the stress wave sensor data is used. This is differentiated by the identifier '*s*' after the gear condition (e.g. RE1sB1 to RE1sB12 for stress wave sensor data for a brand new gear).

3.5. Simulation data

This data set was generated numerically to expand on the results obtained from the analysis of the experimental data sets. Some of the files simulated here also served as the control data input, as its contents are very well known. This allows for a judgement related to the correctness and performance of each individual method.

The first data file, the control data, is defined below. This is a purely theoretical signal. It serves as a basis for the assessment of the accuracy of each method studied. In equation 3.1 '*t*' is the time (in milliseconds, ranging from 0 to 1023, hence the sampling rate is 1KHz) and θ is a constant (2π) that converts the sample number to the corresponding angular position.

$$x(t) = 2Sin(40\theta t) + Sin(70\theta t)[1 + Sin(3\theta t)] + 2Sin(100\theta t + Sin(3\theta t)) + Sin(130\theta t + Sin(3\theta t))[1 + 0.2Sin(3\theta t)] + fault \qquad \text{Eq. 3.1}$$

where:

t	time in milliseconds	$0 \square t < 1024$
θ	conversion factor	$\theta = 2\pi$
fault	can be one of the many simulated fault shapes shown in Table 3.3	

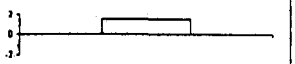
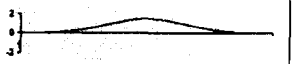
This signal has 5 components, namely:

- A pure sin wave at 40 Hz
- An amplitude modulated sine wave at 70Hz (modulation freq. 3Hz)
- An frequency modulated sine wave at 100Hz (modulation freq. 3Hz)

- A frequency and amplitude modulated sine wave at 130Hz (frequency and amplitude modulation at 3Hz)
- *Fault.* This takes any of the shapes of the transient functions shown below.

In this research all these numerical faults were tested, but only the results for the Gaussian window (fault no. 3) were included.

Table 3.3 – Numerical Faults

Fault		Equation	Time domain representation
N°	Description		
1	Square window	$\begin{cases} \text{for } 0.1+T_0+\frac{\sigma}{2} > t > 0.1+T_0-\frac{\sigma}{2} & f_1(t)=G \\ \text{for all other } t & f_1(t)=0 \end{cases}$	
2	Sine + Square window	$\begin{cases} \text{for } 0.1+T_0+\frac{\sigma}{2} > t > 0.1+T_0-\frac{\sigma}{2} & f_2(t)=G\sin(2\pi ft) \\ \text{for all other } t & f_2(t)=0 \end{cases}$	
3	Gaussian window	$f_3(t) = Ge^{(-(\sigma^2*(t-0.1-T_0)^2)/1024^2)}$	
4	Sine + Gaussian window	$f_4(t) = Ge^{(-(\sigma^2*(t-0.1-T_0)^2)/1024^2)}Sin(2\pi ft)$	
5	Exponential decay	$\begin{cases} \text{for } t < 0.1+T_0-\sigma & f_5(t)=0 \\ \text{for all other } t & f_5(t)=Ge^{(-(t-0.1-T+\sigma)/0.5\sigma)} \end{cases}$	
6	Sine + exp. Decay	$\begin{cases} \text{for } t < 0.1+T_0-\sigma & f_5(t)=0 \\ \text{for all other } t & f_5(t)=Ge^{(-(t-0.1-T+\sigma)/0.5\sigma)}Sin(2\pi ft) \end{cases}$	

The width, amplitude (and frequency, for fault 2,4 and 6) of these faults were controlled by the variables: G (gain), σ (width) and f (frequency).

From the above functions simulated time-series were generated. Those contained 2^{10} samples (as this was the length of the experimentally recorded time-series). Those signals were grouped into two classes. The first represented a vibration signature from a good condition gear (i.e. with no transients), and the second represented a faulty condition gear (i.e with the Gaussian pulse transient simulating a possible fault).

3.6. *Vibration signature samples*

The figures below show samples of typical vibration signatures for a brand new, a good condition and a faulty (F3 – largest fatigue crack) gear. As it can be seen from these time domain plots it is difficult, if not impossible, to differentiate between a good and a faulty condition gear

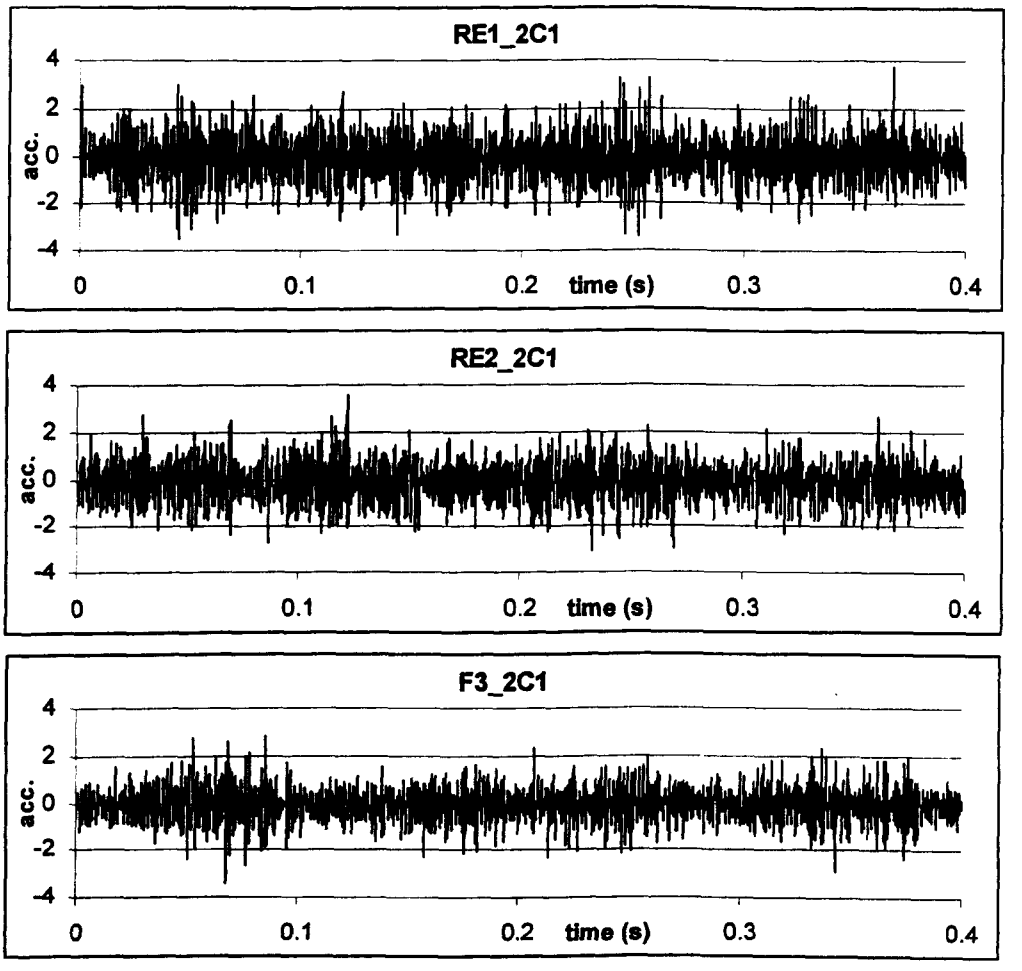


Figure 3.7 – Typical vibration signatures

Chapter 4

Condition monitoring based on statistical methods

In this chapter two main approaches to statistical condition monitoring are discussed. The first is based on a pure moment analysis of vibration signatures. The second approach is based on statistical comparisons of vibration signatures. Although the methods discussed can only lead to simple estimates of the condition of the device under observation, they are also very fast and have proved to be very effective, especially for giving insight about the signal to be analysed.

The results obtained when processing the experimental data files described in the previous chapter will be given at the end of each section. Note that the tests were performed for blocks of four cycles (signatures) of the rotating gear.

4.1 Moment Analysis

In this section a moment analysis of the collected data is performed. This analysis is crucial to establish the data's integrity and validity. It is assumed that moments of the vibration signature from the gearbox should not change unless the mechanical condition of the gearbox changes.

Moment analysis is centred around the measures of central tendency of a distribution. Namely: mean, variance, standard deviation, absolute deviation, skew and kurtosis.

A mathematical definition and a full description for each of these parameters follow.

4.1.1. Mean

This is a well know quantity, and for a discrete series it is defined as (N is the number of samples in the series):

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{Eq. 4.1}$$

The mean value of a distribution estimates the value, around which, central clustering occurs. For the collected data this value to be very close to zero, after all the mechanical equipment being analysed is fixed at a given position, and an average acceleration would indicate overall movement of the equipment.

4.1.2. Variance

This is defined as:

$$Var(x_1 \ldots x_N) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad \text{Eq. 4.2}$$

The variance value of a distribution estimates the 'width' of the distribution. Indicating the average variability of the values around the mean value. It gives a good indication of the spread of the data around its mean value.

4.1.3. Standard Deviation

This is defined as:

$$\sigma(x_1 \ldots x_n) = \sqrt{Var(x_1 \ldots x_N)} \qquad \text{Eq. 4.3}$$

The standard deviation is another measure of the variability of the data series around the central value. It estimates the mean squared deviation of x from its mean value. This quantity also gives an indication of the spread of the data around its central clustering value (mean).

4.1.4. Skew

This is defined as:

$$Skew(x_1 \ldots x_N) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3 \qquad \text{Eq. 4.4}$$

The skew number (also called third moment) characterises the degree of asymmetry of a distribution around its mean value. It is a non-dimensional quantity. A positive value of skew signifies a distribution with an asymmetric tail extending out towards more positive values of x. A negative skew signifies an asymmetric tail extending out towards negative values of x.

This estimator must be used with caution, as most sets of N measured values are likely to give a non-zero skew number. As a rule of thumb meaningfully relevant skewness occurs when the skew value exceeds the threshold $\sqrt{(15/N)}$.

4.1.5. Kurtosis

This is defined as:

$$Kurt(x_1 \ldots x_N) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\frac{x_i - \bar{x}}{\sigma} \right]^4 \right\} - 3 \qquad \text{Eq. 4.5}$$

The Kurtosis number (also called fourth moment) characterises the relative peakedness or flatness of a distribution when compared to the normal distribution. A positive Kurtosis number, signifies a distribution more peaked than the normal distribution (leptokurtic). A negative Kurtosis, signifies a flatter distribution than the normal distribution (platykurtic).

4.1.6. Crest Factor

This quantity also aims to establish a measure of the overall shape of the signal. Two versions for this measure can be seen in existing literature. These are defined below:

$$CF_{peak} = \frac{\max(|x_i|)}{X_{rms}} \quad \text{or} \quad CF_{peak-to-peak} = \frac{\max(x_i) - \min(x_i)}{X_{rms}} \qquad \text{Eq. 4.6, Eq. 4.7}$$

where:

$$X_{rms} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i^2\right)}{N}} \qquad \text{Eq. 4.8}$$

For very long vibration signatures both versions are valid, and this measure aims at establishing how high is the peak of the signal compared to the signal's *rms* value.

4.1.7. Form Factor

This is defined as:

$$FF = \frac{X_{rms}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \qquad \text{Eq. 4.9}$$

The form factor is another measure which aims to describe the waveform of the signal. It compares the signal's *rms* value to its mean absolute values.

4.1.8. Results of moment analysis of data files

In this section the results for the moment analysis of the vibration signatures is presented, both in graphical and tabular form. These results refer to the experimental data collected at the Condition Monitoring Laboratory at Hertfordshire University. The data collection procedure and file description has been included in chapter 3.

From Figure 4.1 to Figure 4.8, if is seen that moment analysis alone is not robust enough to accurately estimate the condition of the gear under observation. Also it is very clear that the only statistical measure to provide some indication about the presence of a fault is:

- **Vibration signature variance**. This tends to decrease if the fault (fatigue crack) is present. However it must be noted that a simple visual inspection of the variance value is not statistically valid. In fact, in order to assess if two variances significantly differ one must recur to the F-test, which is discussed in section 4.2.1.
- **Vibration signature standard deviation**. Again this measure decreases with the presence of a fault.
- **Vibration signature RMS value**. Again this measure decreases with the presence of faults.
- **Vibration Kurtosis value**. This measure clearly indicates the presence of a fault in its very early stages. This is clearly shown in Figure 4.4, the kurtosis value for F1 is very close to the kurtosis values of the reference signal, while the kurtosis value for F2 is distinctively different. Note that, as the fault advances (F3) the Kurtosis values decreases again, failing to indicate the presence of a fault.

The changes on these statistical measures can be mainly attributed to a reduction in the tooth stiffness due to the presence of the fatigue crack. However it must be noted that these changes although visible were very small. The remaining statistical measures did not show any visible trends.

The next pages contain the results first in graphical and later in tabular format for the collected vibration signatures. The charts on the left show the results for 12 blocks of 4 cycles (i.e. 48 gear revolutions). The figures on the right show the average of all these blocks.

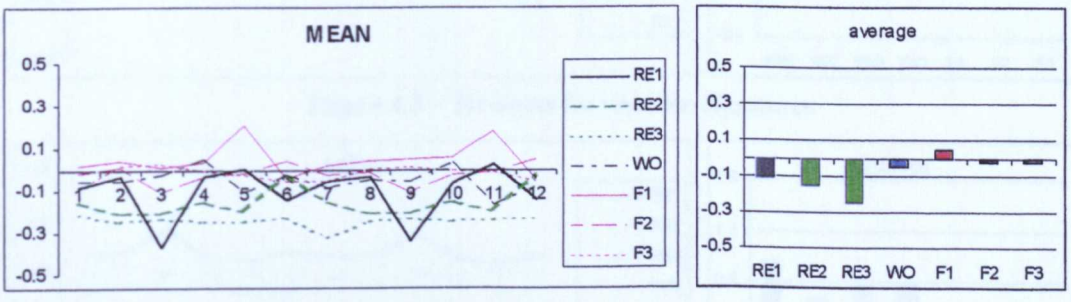


Figure 4.1 – Mean values for vibration signatures

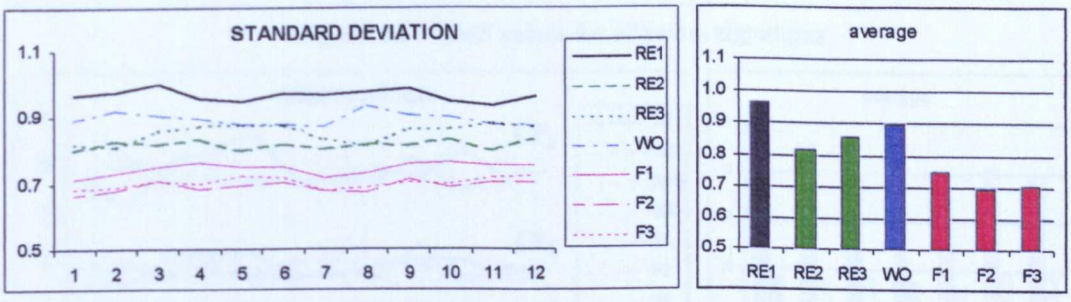


Figure 4.2 – Standard deviation for vibration signatures.

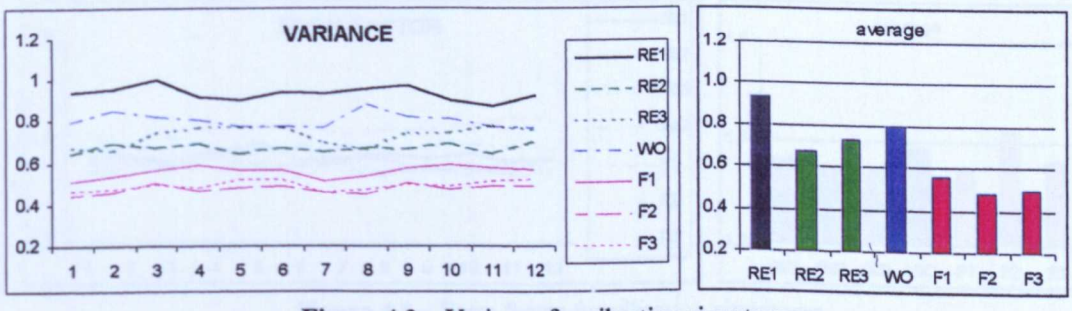


Figure 4.3 – Variance for vibration signatures

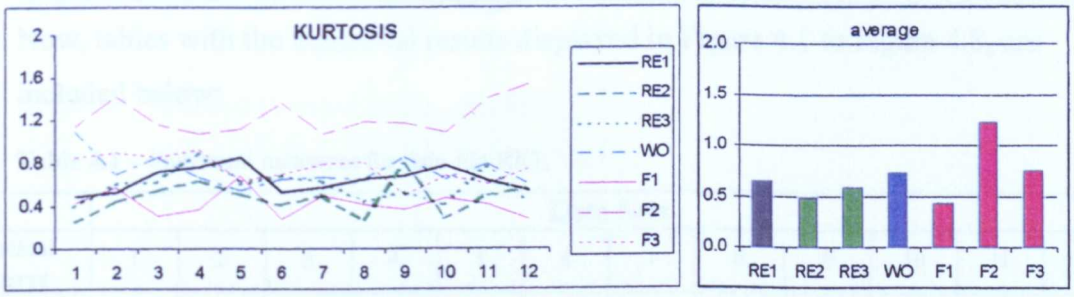


Figure 4.4 – Kurtosis for vibration signatures

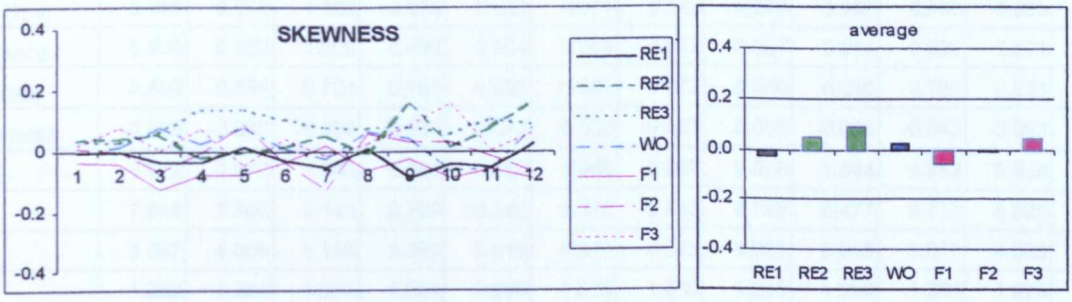


Figure 4.5 – Skewness for vibration signatures

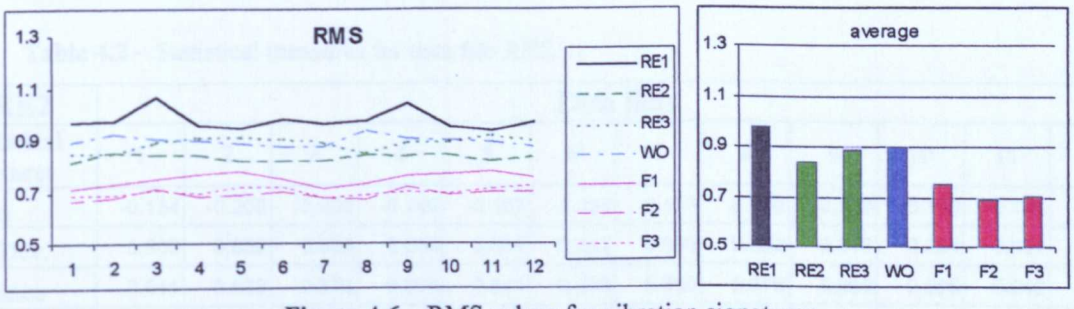


Figure 4.6 – RMS values for vibration signatures

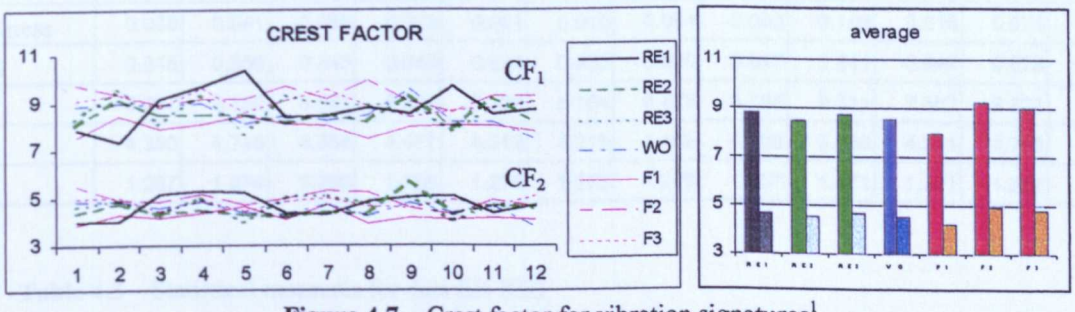


Figure 4.7 – Crest factor for vibration signatures[1]

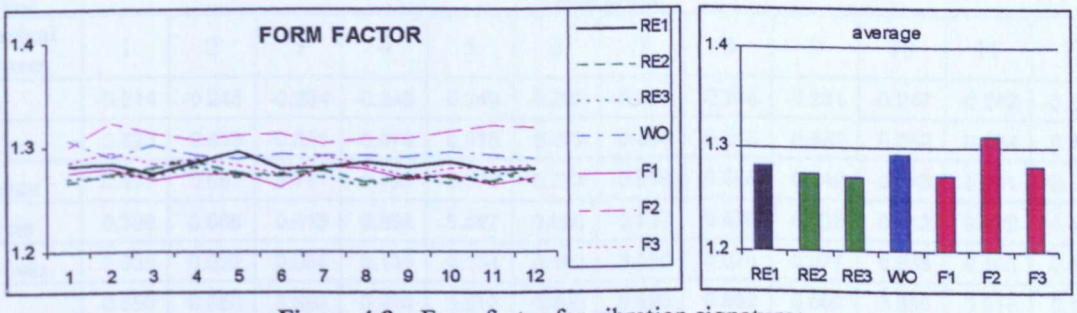


Figure 4.8 – Form factor for vibration signatures

[1] CF1=CFpeak-to-peak and CF2=CFpeak on **Eq. 4.6, Eq. 4.7**

Now, tables with the numerical results displayed in Figure 4.1 to Figure 4.8, are included below:

Table 4.1 – Statistical measures for data file RE1.

RE1	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.083	-0.031	-0.365	-0.034	0.001	-0.143	-0.055	-0.030	-0.341	-0.064	0.025	-0.142
Std. Dev.	0.969	0.976	1.002	0.959	0.951	0.971	0.965	0.978	0.987	0.950	0.933	0.962
Variance	0.939	0.953	1.005	0.920	0.904	0.944	0.932	0.957	0.974	0.903	0.871	0.925
Kurtosis	0.489	0.534	0.724	0.761	0.853	0.533	0.572	0.648	0.680	0.786	0.628	0.546
Skewness	-0.008	-0.002	-0.039	-0.037	0.017	-0.033	-0.052	0.010	-0.088	-0.043	-0.053	0.025
RMS	0.972	0.977	1.066	0.960	0.950	0.982	0.967	0.979	1.044	0.952	0.934	0.972
CF_1	7.845	7.348	9.148	9.709	10.346	8.375	8.463	8.790	8.477	9.712	8.525	8.682
CF_2	3.957	4.069	5.155	5.282	5.219	4.377	4.343	4.869	5.015	5.077	4.383	4.723
FF	1.282	1.284	1.274	1.284	1.296	1.279	1.280	1.285	1.280	1.286	1.278	1.279

Table 4.2 – Statistical measures for data file RE2.

RE2	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.154	-0.200	-0.205	-0.149	-0.197	-0.028	-0.153	-0.202	-0.200	-0.150	-0.196	-0.028
Std. Dev.	0.800	0.832	0.820	0.835	0.801	0.822	0.806	0.822	0.817	0.834	0.802	0.838
Variance	0.641	0.693	0.673	0.698	0.641	0.676	0.650	0.676	0.668	0.695	0.642	0.702
Kurtosis	0.243	0.448	0.540	0.637	0.521	0.408	0.500	0.267	0.772	0.281	0.494	0.570
Skewness	0.036	0.041	0.059	-0.023	0.001	0.016	0.054	-0.003	0.149	0.016	0.071	0.159
RMS	0.815	0.856	0.845	0.848	0.824	0.822	0.820	0.847	0.841	0.847	0.825	0.838
CF_1	8.077	9.150	8.501	8.540	8.529	8.194	8.476	8.155	9.311	7.862	9.221	8.158
CF_2	4.350	4.746	4.354	4.477	4.312	4.216	4.452	4.302	5.689	4.311	4.735	4.650
FF	1.267	1.274	1.280	1.286	1.276	1.273	1.279	1.267	1.271	1.271	1.267	1.280

Table 4.3 – Statistical measures for data file RE3

RE3	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.214	-0.248	-0.234	-0.248	-0.249	-0.227	-0.325	-0.248	-0.231	-0.247	-0.242	-0.230
Std. Dev.	0.823	0.813	0.861	0.876	0.878	0.879	0.836	0.816	0.862	0.862	0.884	0.871
Variance	0.677	0.661	0.741	0.767	0.771	0.773	0.698	0.666	0.742	0.743	0.781	0.759
Kurtosis	0.399	0.606	0.613	0.895	0.597	0.680	0.637	0.438	0.532	0.412	0.772	0.434
Skewness	0.039	0.032	0.064	0.135	0.134	0.103	0.076	0.074	0.077	0.055	0.105	0.091
RMS	0.850	0.850	0.892	0.910	0.912	0.908	0.896	0.852	0.892	0.896	0.916	0.901
CF_1	8.295	9.636	8.113	9.390	7.641	9.273	8.965	8.878	8.912	7.768	8.784	9.104
CF_2	4.630	4.916	4.281	5.033	4.013	5.024	5.087	4.800	5.124	4.031	4.712	4.574
FF	1.270	1.273	1.273	1.281	1.271	1.274	1.271	1.264	1.273	1.265	1.278	1.266

Table 4.4 – Statistical measures for data file WO

WO	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.053	-0.058	-0.025	0.051	-0.108	-0.023	-0.086	-0.046	-0.055	0.036	-0.164	-0.077
Std. Dev.	0.893	0.921	0.904	0.898	0.879	0.878	0.873	0.936	0.904	0.899	0.880	0.874
Variance	0.797	0.848	0.818	0.806	0.772	0.771	0.761	0.876	0.818	0.809	0.774	0.765
Kurtosis	1.103	0.686	0.855	0.665	0.540	0.650	0.667	0.654	0.848	0.654	0.801	0.603
Skewness	-0.006	0.015	0.078	0.025	0.079	0.012	-0.026	0.060	0.039	-0.062	-0.059	0.080
RMS	0.894	0.923	0.905	0.899	0.885	0.879	0.877	0.937	0.906	0.900	0.895	0.878
CF_1	8.848	8.985	8.928	8.948	8.140	8.448	8.254	8.433	9.562	8.213	8.184	7.444
CF_2	4.806	4.858	4.588	4.811	4.379	4.302	4.734	4.264	4.797	4.383	4.637	3.805
FF	1.306	1.290	1.301	1.290	1.287	1.292	1.290	1.291	1.290	1.294	1.291	1.287

Table 4.5 – Statistical measures for data file F1

F1	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	0.016	0.041	0.014	0.035	0.212	-0.043	0.015	0.043	0.050	0.053	0.184	-0.050
Std. Dev.	0.714	0.732	0.751	0.767	0.752	0.751	0.717	0.732	0.760	0.767	0.757	0.753
Variance	0.510	0.537	0.564	0.588	0.566	0.565	0.514	0.535	0.578	0.588	0.573	0.567
Kurtosis	0.430	0.575	0.310	0.362	0.684	0.286	0.501	0.415	0.362	0.478	0.419	0.282
Skewness	-0.036	-0.043	-0.122	-0.080	-0.076	-0.037	0.021	-0.062	-0.025	-0.090	-0.058	-0.089
RMS	0.714	0.734	0.751	0.767	0.781	0.753	0.717	0.733	0.762	0.768	0.779	0.755
CF_1	7.517	8.427	7.909	8.095	8.423	8.097	8.036	7.436	7.962	8.057	8.024	7.778
CF_2	3.854	4.327	4.176	4.353	4.624	4.100	4.699	3.904	4.154	4.117	4.213	4.090
FF	1.2754	1.2786	1.2685	1.2741	1.2746	1.2620	1.2815	1.2790	1.2675	1.2727	1.2642	1.2704

Table 4.6 – Statistical measures for data file F2

F2	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.020	0.009	-0.104	-0.035	-0.020	0.041	-0.027	-0.010	-0.108	-0.024	-0.015	0.046
Std. Dev.	0.666	0.680	0.710	0.686	0.694	0.701	0.678	0.673	0.712	0.689	0.703	0.701
Variance	0.444	0.463	0.504	0.470	0.482	0.491	0.460	0.453	0.507	0.474	0.494	0.492
Kurtosis	1.141	1.433	1.155	1.071	1.121	1.370	1.067	1.190	1.179	1.091	1.409	1.545
Skewness	0.009	-0.001	-0.068	-0.005	0.090	-0.037	-0.126	0.055	0.032	0.032	-0.015	-0.069
RMS	0.666	0.680	0.717	0.686	0.694	0.702	0.679	0.673	0.720	0.689	0.702	0.703
CF_1	9.777	9.388	8.934	9.166	9.188	9.684	9.165	9.927	9.058	8.550	9.301	9.311
CF_2	5.512	4.924	4.835	5.181	4.700	4.995	5.309	5.210	4.543	4.310	4.880	4.669
FF	1.300	1.323	1.312	1.307	1.311	1.313	1.303	1.308	1.307	1.314	1.320	1.317

Table 4.7 – Statistical measures for data file F3

F3	Data Sets											
Statistical measures	1	2	3	4	5	6	7	8	9	10	11	12
Mean	-0.095	0.021	0.023	-0.004	-0.023	-0.053	-0.057	0.006	0.014	-0.007	-0.011	-0.053
Std. Dev.	0.682	0.690	0.708	0.697	0.722	0.719	0.679	0.689	0.704	0.700	0.715	0.723
Variance	0.466	0.475	0.501	0.486	0.522	0.517	0.461	0.475	0.496	0.490	0.511	0.522
Kurtosis	0.857	0.895	0.857	0.930	0.638	0.687	0.798	0.805	0.501	0.677	0.597	0.756
Skewness	0.021	0.077	0.003	0.075	-0.049	0.044	-0.018	0.027	0.085	0.157	0.006	0.064
RMS	0.689	0.690	0.708	0.697	0.722	0.721	0.681	0.689	0.704	0.700	0.715	0.724
CF_1	9.099	9.344	8.741	9.597	8.542	9.090	9.560	8.845	8.374	8.636	8.782	9.059
CF_2	4.975	5.040	4.426	4.890	4.420	4.836	4.959	4.649	4.575	4.897	4.644	5.123
FF	1.285	1.295	1.283	1.289	1.282	1.275	1.291	1.282	1.271	1.278	1.281	1.282

4.2. Statistical Comparison

In the previous section we have demonstrated how the visual inspection of simple statistical measures can be used to evaluate the properties of a time series. Now we will extend those concepts, and use more sophisticated statistical tools aiming to compare the difference between two time series. Note that although tools like regression analysis, and its counterparts, are very useful for modelling time series, they are not particularly well suited to perform a direct comparison between two series.

This section concentrates on methods, which are able to perform this direct comparison. Three tests are presented here:

- F-Test to assess if two distributions have the same variance,
- Kolmogorov-Smirnov KS test to assess if two distributions are equal,
- autocorrelation.

Finally, the last section of this chapter presents a summary of the results obtained by processing the experimental data with the methods described here.

4.2.1. F-test : Assessing if two distributions have the same variance

In order to observe whether two sets of data have the same variance, one needs to recur to the *F*-test. The F-Test returns a statistic value *'f'*, which is then fed into the F-distribution, so a similarity probability is returned. Again the closer this probability is to 1, the more confidence one has that the variances are equal. This statistic measure *f* is defined as:

$$f = \frac{\text{var}(x_A)}{\text{var}(x_B)} \qquad \text{for} \qquad \text{var}(x_A) > \text{var}(x_B) \qquad \textbf{Eq. 4.10}$$

$$f = \frac{\text{var}(x_B)}{\text{var}(x_A)} \qquad \text{for} \qquad \text{var}(x_A) < \text{var}(x_B) \qquad \textbf{Eq. 4.11}$$

A full definition of the F-distribution can be found in [108] and hence will not be repeated here.

4.2.2. Kolmogorov-Sminov (KS) test: Assessing if two distributions are equal

The F-test is extremely effective in analysing the similarity between the variance from two vibration signatures. Now, the need for a more general comparison is obvious. There are two standard statistical comparison methods.

The first is the Chi-Square test. This test is particularly useful when analysing binned data (i.e. data that have been grouped into specified ranges). However, since the collected vibration signature is not binned, this standard distribution comparison method is not applicable.

The second method is known as the Kolmogorov-Smirnov (KS) test. This testing procedure is classified as a frequency test of the degree of agreement between distributions. This method is widely used to test generated random numbers against specific distributions. In this research it is suggested that the KS test can also be used to compare vibration signatures, and estimate the condition of a mechanical device.

The KS test works on the null hypotheses that the CDF (cumulative density function) of a target distribution, denoted by $F(x)$, is statistically similar to the CDF of the reference distribution, $R(x)$. Hence, it is possible to compare two vibration signatures, and assess if both have the same CDF. Note that the usage of this test for condition monitoring assumes that the fault is strong enough to cause a variation in the CDF of the original vibration signature.

From the two CDFs, $F(x)$ and $R(x)$, a statistic distance D can be calculated. This distance is defined as the maximum absolute difference between $F(x)$ and $R(x)$. Mathematically this is represented by:

$$D = \max_{-\infty<x<\infty} |F(x) - R(x)| \qquad \textbf{Eq. 4.12}$$

The diagram below illustrates how this distance D is found. It uses the experimental data files RE2 (vibration signature for a reference gear) and F3 (vibration signature for the gear with the largest fatigue crack). Firstly the CDF for the vibration signatures are plotted (this is shown by lines RE2 and F3). From these plots, a distance d can be calculated by obtaining the module of the subtraction between the two cumulative probability functions, this is shown by line d. The maximum value of d is then used as D. In the example below the maximum value for d is D=0.048 occurring at x=-0.54.

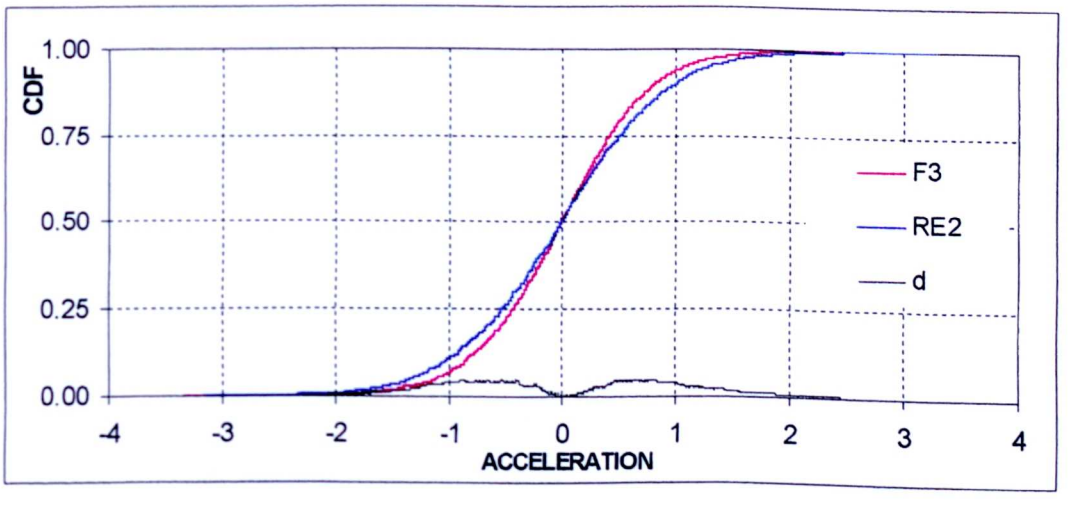


Figure 4.9 – Cumulative probability distribution function

As it can be seen when comparing a signature from a reference gear with that from a faulty gear, a reasonably large value for D is obtained. This value can be converted into a similarity probability using the Kolmogorov-Smirnov probability distribution function Q_{ks}. This is defined as:

$$prob(D) = Q_{KS}\left(\left[\sqrt{N_e} + 0.12 + \frac{0.11}{\sqrt{N_e}}\right]D\right), \qquad \text{Eq. 4.13}$$

$$N_e = \frac{N_1 N_2}{N_1 + N_2}. \qquad \text{Eq. 4.14}$$

Where N_e is the effective number of data points and is calculated according to Eq. 4.14. N_1 is the number of points in the first data set and N_2 is the number of points on the second data set. The Kolmogorov-Smirnoff distribution function is defined as:

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty}(-1)^{j-1} e^{-2j^2\lambda^2} \qquad \text{Eq. 4.15}$$

This is a monotonic function with limiting values of:

$$Q_{KS}(\lambda) = 1 \quad \text{As} \quad \lambda \to 0$$
$$Q_{KS}(\lambda) = 0 \quad \text{As} \quad \lambda \to \infty \qquad \text{Eq. 4.16}$$

Note that the choice of the statistic distance (maximum, rms, mean, etc) between the CDFs greatly affects the sensitivity of this method.

4.2.3. Autocorrelation

The autocorrelation function of a time series is used as a guide to provide some insight into the probability model that generated the time series. The autocorrelation coefficients are a measure of the correlation between observations at different distances apart. This series of coefficients can be displayed as a correlogram.

In the correlogram the distribution's autocorrelation coefficients r_k are plotted against the lag k. Visual inspection of the correlogram allows a quick insight into the type of time series (i.e random, short-term correlation, alternating, non-stationary and/or seasonal) being analysed.

The autocorrelation coefficients for a time series can be approximated by Eq. 4.17. A full derivation can be found in [9]. In this equation the autocorrelation coefficient (with significant values above the ($\pm 2/\sqrt{N}$) threshold), r_k, at lag k, is given by:

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \qquad \text{Eq. 4.17}$$

4.2.4. Results of Statistical Comparison

In order to evaluate the performance of the methods described above, the experimentally collected data files were processed. The results, together with a discussion of each method's performance are presented in the following sub-sections.

For the F-test and the KS- test the results will be in form of a 'similarity probability'. This is a statistical measure, which attempts to evaluate how similar two distributions are (100% for equal distributions, and 0% for different distributions).

4.2.4.1. Result for F-test: equal variance assessment

This test aimed at establishing whether the difference in the variance value of two distinct data sets could be suitably used as a simple condition monitoring technique. The results are tabulated below. Note that the basis of comparison was the vibration signature for a gear in good condition (data file RE2), hence high similarity values indicate that the two signatures are both from a good gear, while low values indicate that one of the signals is not from a good condition gear.

The results in each cell related to the comparison of the signatures included in the cell's respective row and column label. An overall averaged result of the individual comparisons is given in the eighth column.

Table 4.8 – Results for the F test.

Comparison							AVERAGE
VERSUS	RE2C1	RE2C2	RE2C3	RE2C4	RE2C5	RE2C6	
RE1C1	0	0	0	0	0	0	
RE1C2	0	0	0	0	0	0	
RE1C3	0	0	0	0	0	0	
RE1C4	0	0	0	0	0	0	
RE1C5	0	0	0	0	0	0	**RE1xRE2**
RE1C6	0	0	0	0	0	0	**0.0**
RE2C7	65	4	26	2	63	21	
RE2C8	9	43	88	31	8	98	
RE2C9	19	24	81	16	18	71	
RE2C10	1	92	30	90	1	36	
RE2C11	93	2	14	1	91	11	**RE2xRE2**
RE2C12	0	89	18	87	0	23	**37.9**
RE3C1	8	45	86	32	8	96	
RE3C2	30	14	59	9	29	50	
RE3C3	0	0	0	0	0	0	
RE3C4	0	0	0	0	0	0	
RE3C5	0	0	0	0	0	0	**RE2xRE3**
RE3C6	0	0	0	0	0	0	**13.0**
WO1C1	0	0	0	0	0	0	
WO1C2	0	0	0	0	0	0	
WO1C3	0	0	0	0	0	0	
WO1C4	0	0	0	0	0	0	
WO1C5	0	0	0	0	0	0	**RE2xWO**
WO1C6	0	0	0	0	0	0	**0.0**
F1C1	0	0	0	0	0	0	
F1C2	0	0	0	0	0	0	
F1C3	0	0	0	0	0	0	
F1C4	1	0	0	0	1	0	
F1C5	0	0	0	0	0	0	**RE2xF1**
F1C6	0	0	0	0	0	0	**0.0**
F2C1	0	0	0	0	0	0	
F2C2	0	0	0	0	0	0	
F2C3	0	0	0	0	0	0	
F2C4	0	0	0	0	0	0	
F2C5	0	0	0	0	0	0	**RE2xF2**
F2C6	0	0	0	0	0	0	**0.0**
F3C1	0	0	0	0	0	0	
F3C2	0	0	0	0	0	0	
F3C3	0	0	0	0	0	0	
F3C4	0	0	0	0	0	0	
F3C5	0	0	0	0	0	0	**RE2xF3**
F3C6	0	0	0	0	0	0	**0.0**

As it can be seen there the difference in the variance value of the different data files is statistically significant. This shows that the variance of a vibration signature can be used as a tool to indicate the presence of faults. Now, it must be noted that this method does not classify the reference files accordingly, as RE2 x RE3 gives a very low similarity value (13%). This suggests that this measure is not suitable for applications where transfer of knowledge amongst gear (devices) is required.

4.2.4.2. Results for the KS Test: equal distribution assessment

In this test, all data files were once again compared against the data file RE2. The table below shows the files compared and the results obtained. These results are displayed in the form of a similarity probability. This is the probability that two vibration signatures, come from the same statistical distribution function (i.e. the signatures are similar). So if the similarity probability is close to 100% then it follows that the two signatures compared are similar, on the other hand if the similarity probability is close to 0% then it follows that the two signatures are different.

Although this method is very simple, it is also very effective. The results clearly show that when two reference signals are compared the similarity probability is greater than when comparing a good signal with a faulty signal. Table 4.9 shows these results.

Table 4.9 – Results for the KS test.

COMPARISON							AVERAGE
VERSUS	RE2C1	RE2C2	RE2C3	RE2C4	RE2C5	RE2C6	
RE1C1	0	0	0	0	0	0	
RE1C2	0	0	0	0	0	0	
RE1C3	0	0	0	0	0	0	
RE1C4	0	1	0	1	0	0	
RE1C5	0	4	1	2	0	2	**RE1xRE2**
RE1C6	0	0	0	0	0	0	**0.3**
RE2C7	92	29	98	55	94	64	
RE2C8	75	77	70	87	23	97	
RE2C9	84	68	68	72	29	85	
RE2C10	42	81	29	40	14	66	
RE2C11	95	68	87	70	38	81	**RE2xRE2**
RE2C12	77	31	45	59	17	61	**63.0**
RE3C1	64	77	53	85	30	66	
RE3C2	97	52	75	64	66	79	
RE3C3	30	48	14	22	2	62	
RE3C4	28	81	20	37	5	28	
RE3C5	14	26	10	15	4	6	**RE2xRE3**
RE3C6	18	55	7	24	5	11	**38.3**
WO1C1	48	20	31	53	28	55	
WO1C2	2	15	3	6	1	6	
WO1C3	14	43	20	31	10	40	
WO1C4	12	18	5	15	2	18	
WO1C5	33	12	22	26	24	55	**RE2xWO**
WO1C6	28	87	45	32	23	43	**25.7**
F1C1	1	0	1	0	1	0	
F1C2	3	0	2	0	2	0	
F1C3	19	1	14	10	18	4	
F1C4	28	4	28	20	34	18	
F1C5	17	2	7	2	17	5	**RE2xF1**
F1C6	31	2	18	14	37	7	**10.2**
F2C1	0	0	0	0	0	0	
F2C2	0	0	0	0	0	0	
F2C3	0	0	0	0	0	0	
F2C4	0	0	0	0	0	0	
F2C5	0	0	0	0	0	0	**RE2xF2**
F2C6	0	0	0	0	0	0	**0.0**
F3C1	0	0	0	0	0	0	
F3C2	0	0	0	0	0	0	
F3C3	0	0	0	0	0	0	
F3C4	0	0	0	0	0	0	
F3C5	1	0	1	0	0	0	**RE2xF3**
F3C6	1	0	1	0	1	0	**0.1**

Note that as expected the similarity level drops as the fatigue crack advances. Also it was observed that the similarity level between a brand new gear (RE1) and a gear at operating condition (RE2 and RE3) was very low. This can be attributed to the fact that the brand new gear still has to adjust itself to the mesh of the system. In fact a similar behaviour is seen in most failure rate charts which follow a "bath-tub" shape.

As it can be seen this method identifies the presence of a fault very effectively, even for faults in its early stages (F1). A possible extension of this method would be in applying it to identify the age of the fault (fault development).

4.2.4.3. Results for autocorrelation

The correlograms for the experimentally collected data files give a very brief insight into the probability model that generated the vibration signature (time series). As it can be seen in the correlograms below, the time series seems to have cyclic period of 0.2s (5Hz), this is the rotational frequency of the gear. Note, also, that the correlograms for the data file RE1 (brand new gear) does not show this cyclic behaviour at all, suggesting that consecutive vibration cycles for this gear possesses different vibration properties. This phenomenon was also observed by the KS test, and is related to the fact that brand new gears have to adjust itself to the mesh.

On all the graphs below the lag time (in seconds) is shown on the x-axis and the autocorrelation coefficient is on the y-axis. Note that a coefficient of 1 means a perfect match. This value was only achieved at lag 0 (this can be seen in the last plot showing a detail of the correlogram for the worn out gear). Noise is the main reason why the autocorrelation coefficients at lag 0.2 and 0.4 are less than 1. Other reasons which also contribute (in a smaller scale) to this are: fluctuations in rotational speed of gear, fluctuations in the sampling rate of the A/D card and fluctuations of the braking torque applied to the gear.

Another point that must be noted is that the cyclic behaviour seems to increase for fatigue cracks at the very early stages. This is also expected as the periodical vibration induced by the fault can only contribute to the cyclic behaviour of the system. From the plots it can also be observed that for larger fatigue cracks the

periodical behaviour seems to be reduced, suggesting that larger cracks do not tend to induce repetitive deterministic vibrations.

The correlograms for the experimentally collected data for the shaved fault is shown below.

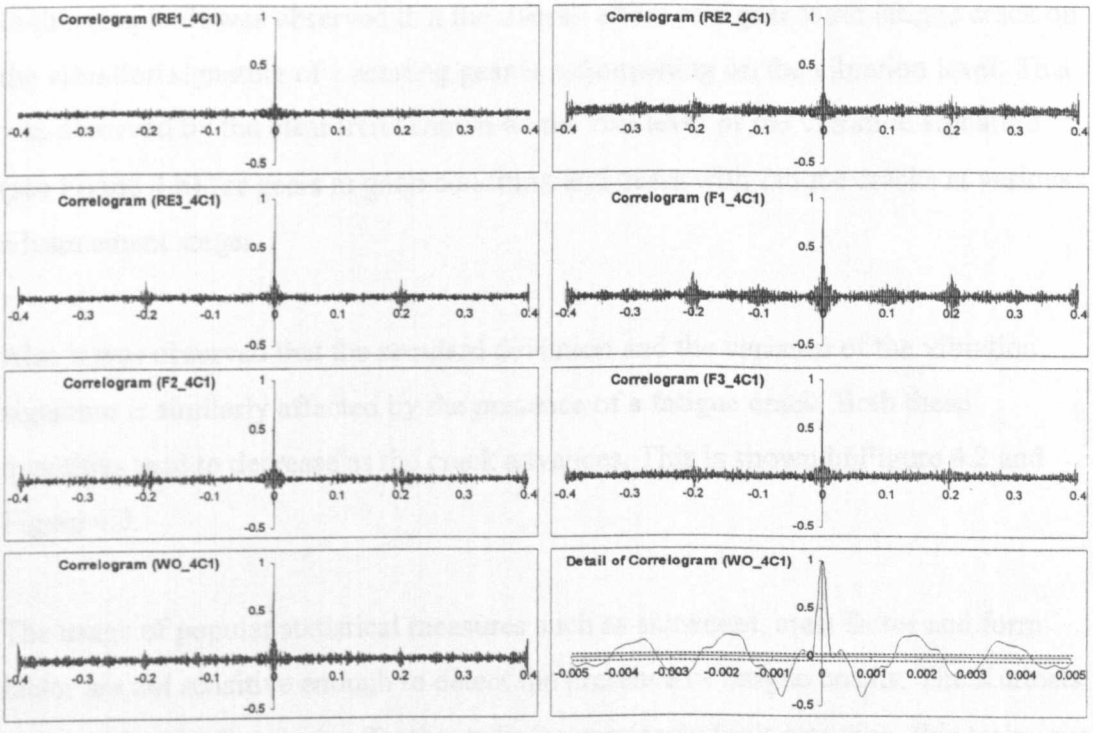


Figure 4.10 – Sample correlogram for data files

In the detailed plot for the worn-out gear (Figure 4.10), the threshold for significant values is also shown as the two lines parallel to the x-axis. It can be seen that the coefficients at lags around ±0.2 and ±0.4s are statistically significant.

This method is not effective for early fault detection. Still as it can be seen in the correlograms that when the fault becomes more intense the cyclic behaviour is also intensified. This behaviour is expected and it can be related to the fact that the fault acts as another cyclic input to the system. Hence it is adding to the cyclic properties of the vibration signature (time series).

Finally, a close look the detailed correlogram plot (Figure 4.10) also leads to the observation of the predominant frequency in the time series, which is usually the

meshing frequency or one of its harmonics. In the detailed plot for the worn-out gear the dominant frequency seems to be approximately 500Hz (period of 2ms). This is close to the second meshing harmonic (480Hz).

4.3. Summary of results

In this chapter it was observed that the overall effect of a gear tooth fatigue crack on the vibration signature of a rotating gear is a dampening on the vibration level. This was observed by the clear trend shown on the *rms* level of the vibration signature (see Figure 4.6) for gears in good condition and gears with fatigue cracks at various advancement stages.

Also it was observed that the standard deviation and the variance of the vibration signature is similarly affected by the presence of a fatigue crack. Both these quantities tend to decrease as the crack advances. This is shown in Figure 4.2 and Figure 4.3.

The usage of popular statistical measures such as skewness, crest factor and form factor are not sensitive enough to detect the presence of fatigue cracks. The Kurtosis value of the distribution is effective only for very early fault detection, this technique only detected fault (F2), and did not detect fault (F3). This suggests that these techniques must be used in conjunction with the other techniques here described for robust condition monitoring based on statistical analysis alone.

This chapter also introduces a novel technique, the KS test, to the field of condition monitoring based on time domain statistical analysis. The results obtained here showed that this test is able to identify the presence of a fatigue crack on a gear tooth in its very early stages. This test is very sensitive to the noise present in the signal (as the comparison of similar signatures, lead to a similarity probability of 65% in contrast to the expected 100%), but this does not pose a major problem. The sensitivity of the method can be easily adjusted by modifying the calculation of the distance D. In this research D was assumed to be the maximum absolute difference between the two cumulative distribution functions being compared. However tests

should be carried out using other definitions (such as mean absolute difference, or root mean square absolute difference, etc) for the calculation of D.

Finally in the last section, the usage of correlograms was analysed. It was shown that this technique identified the presence of a fatigue cracks in its very early stages. This was shown by the greater correlation coefficients at lag periods of 2ms (1 revolution of gear under analysis). However this technique is not suitable for signatures with a low noise to signal ratio.

Chapter 5

Frequency and time-frequency approach to condition monitoring.

In this chapter a full theoretical description of the frequency and time-frequency techniques included in this research (namely, spectral analysis, cepstral analysis, spectrogram and Wigner - and its variants - distributions) is given. Results obtained when processing the numerical and experimental data with these techniques are also included.

Condition monitoring systems based on frequency and time frequency distributions rely heavily on Fourier Integrals. This is a powerful method for extracting spectral information from time domain signals. It leads the way to many different industrial condition monitoring techniques. A full command of the Fourier transform properties must be attained for a full understanding of these frequency and time-frequency distributions.

The Fourier theorem states that any single periodic function, $x(t)$ with period P, can be represented by a series of orthogonal functions (Sines and Cosines) with different harmonic frequencies. The Fourier Transform is defined as:

$$X(f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt . \qquad \textbf{Eq. 5.1}$$

Hence from the above equation it is possible to convert a time series into the frequency domain, allowing for the observation of the frequency components in a time series. Note that if the Fourier series is to represent a function $x(t)$, then $x(t)$ must meet the following conditions (*Dirichlet Conditions*, [23]), so that its Fourier series converge to $X(f)$, the frequency components of $x(t)$:

- The function $x(t)$ must be defined and single-valued
- The function $x(t)$ must be continuous or have a finite number of finite discontinuities within a periodic interval
- The function $x(t)$ and $x'(t)$ must be piecewise continuous in the periodic interval

For vibration signatures, and most practical cases, these conditions are met, leading to the accurate (and fairly quick) description of a time series in frequency domain. Note that the frequency resolution achieved by this method is inversely proportional to the period of the time-series, and the frequency ranges from 0 to the Nyquist frequency $F_{Nyq}=F_{sampling}/2$

5.1. *Spectral Analysis*

The spectral analysis was the first tool developed for the frequency analysis of a given signal. It relies only on the Fourier Transform. Today there are many algorithms available to perform the Discrete Fourier Transform (DFT) of a signal. In this research the FFT algorithm was chosen. The following equation shows the digitised version of the Fourier integral.

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)\, e^{-j\frac{2\pi kn}{N}} \qquad \textbf{Eq. 5.2}$$

The vibration signatures analysed in this research all have 4096 (4 cycles per block) samples and a sampling frequency of 5.12 kHz. Hence the frequency resolution is 1.25 Hz, and the frequency ranges from 0Hz to 2.56 kHz.

5.1.1. Results of spectral analysis of experimental data

In this section the FFT was used for the spectral analysis of the experimental data (described in chapter 3). The plots in Figure 5.1 show the frequency components of these signals. The graph title refers to the origin of the data (i.e. gear condition). The plots on the left relate to the raw data with no pre-processing, the ones on the right relate to the time domain average of the vibration signature. This average is made from 12 blocks (4 cycles on each block) of data. The time average plots were included for comparison purposes, and as it can be seen their overall shape is similar to the shape of the spectrums of the raw vibration signatures.

As it can be seen from the plots in Figure 5.1, the mesh frequency and its first three harmonics can be clearly identified for all the gear conditions, both in the spectrum of the raw data and the time averaged data. The plots show no distinguishable frequency components over 800 Hz, except low level noise. Note that as it would be expected the noise level for the time averaged signal is much lower than that for the raw vibration data.

The dominant peaks occurred at the second harmonic (320 Hz) and the third harmonic (480Hz). Also the region around 600 Hz shows high signal intensity, these peaks do not relate to a meshing harmonic and could be attributed to structural resonance. This phenomenon appears in all spectra making its interpretation harder. Finally, a high side band activity in the spectrums can be recognised, with the reference (RE2 and RE3) and fault 1 (F1) data showing the lowest activity. To detect possible trends, comparisons of the signal levels and side band activity for the fundamental and the first two harmonics were carried out.

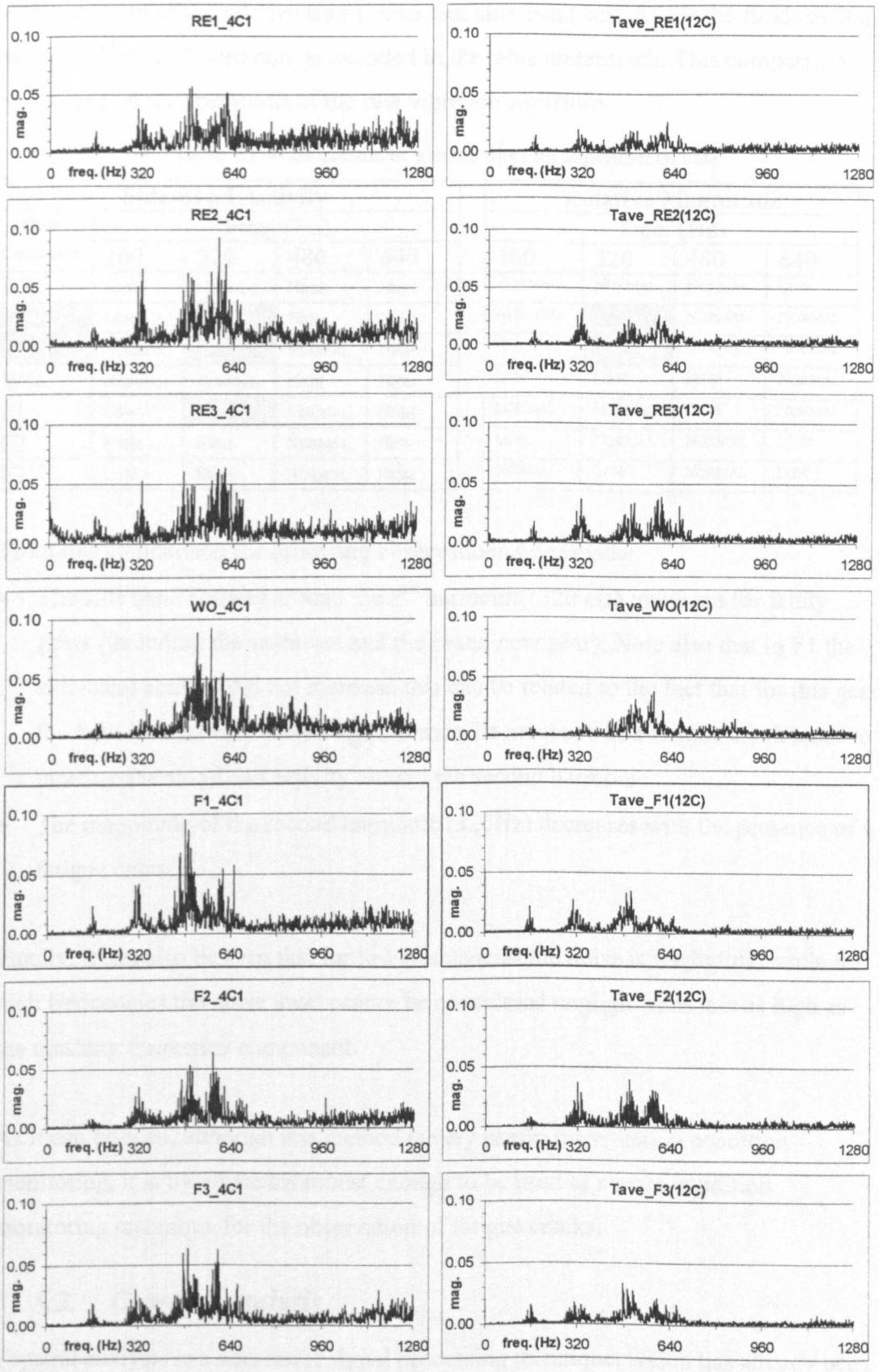


Figure 5.1 – Spectral analysis of experimental data

A comparison of the relative magnitudes and side-band activity for the fundamental and the 2nd and 3rd harmonic is included in the table underneath. This comparison was based on the spectrums of the raw vibration signature.

Table 5.1 – Comparison of spectral plots for experimental data

Side-band Activity				
Gear Condition	Freq. (Hz)			
	160	320	480	640
RE1	Low	Normal	High	High
RE2	Low	Low	High	High
RE3	Normal	Low	Normal	High
WO	Normal	Normal	High	High
F1	Low	Low	Normal	High
F2	High	High	Normal	High
F3	Low	High	Normal	High

Relative Magnitude			
Freq. (Hz)			
160	320	480	640
Normal	Normal	Normal	Low
Normal	High	Normal	Normal
Normal	High	Normal	Normal
Low	Low	High	Normal
normal	Normal	High	Normal
low	Normal	Normal	Low
Normal	Low	Normal	Low

From this comparison the following observations were made:

- The side band activity around the 2nd harmonic (320 Hz) increases for faulty gears (including the worn-out and the brand new gear). Note also that in F1 the side-band activity did not increase, this can be related to the fact that for this gear the fault is in its very early stages. Hence the presence of a fatigue crack tends to increase the side-band activity around the second harmonic.
- The magnitude of the second harmonic (320Hz) decreases with the presence of a fatigue crack.

Finally, it can also be seen that for low frequencies the noise is negligible, while for high frequencies the noise level cannot be considered negligible, as it is as high as the meshing frequency component.

As it can be seen, although this method is very useful for vibration condition monitoring, it is by no means robust enough to be used as a main condition monitoring technique for the observation of fatigue cracks.

5.2. Cepstral analysis

Cepstral analysis is a non-linear signal processing technique, which has already been successfully used in fields such as echo detection [29], speech analysis [31] and

others. It has also been used for time series analysis and condition monitoring systems.

Cepstral analysis also relies heavily on Fourier Integrals. There are two main variations on signal cepstral analysis. These are the 'complex cepstrum' and the 'real cepstrum'. The complex cepstrum *c(t)* of a signal *x(t)* is defined as the inverse Fourier transform of the logarithm of the Fourier transform of an input signal *x(t)*. Generally this requires the evaluation of the phase as a continuous function of frequency. Alternatively, the real cepstrum can be computed by using the logarithm of the magnitude instead of the complex logarithm of the Fourier transformed signal.

In this research only the real cepstrum is included. Furthermore, it is suggested that the real cepstrum reconstruction is effective for identifying the presence and strength of the transients present in the time series. This trait leads to an effective identification of faults in vibration signatures.

5.2.1. Real cepstrum

A formal definition of the real cepstrum and its inverse is given by [30, chapter 10]:

$$c(n) = \mathrm{IDFT}\left(\log\left|\mathrm{DFT}(x(n))\right|\right), \qquad \textbf{Eq. 5.3}$$

$$\hat{x}(n) = \mathrm{IDFT}\left(\exp\left|\mathrm{DFT}(c(n))\right|\right) . \qquad \textbf{Eq. 5.4}$$

Where *c(n)* is the real cepstrum of the input signal *x(n)*; and $\hat{x}(n)$ is the inverse real cepstrum, or cepstral reconstruction, of *c(n)*. DFT is the discrete Fourier transform and IDFT is the inverse discrete Fourier transform.

The output of the cepstral analysis is a signal in the cepstral domain. This is commonly shown as a plot, similar to a spectral plot, but with quefrency in the x-axis instead of the Frequency. The unit of quefrency is $[\mathrm{Hz}]^{-1}$. The graphs below illustrate the performance of the cepstral analysis method.

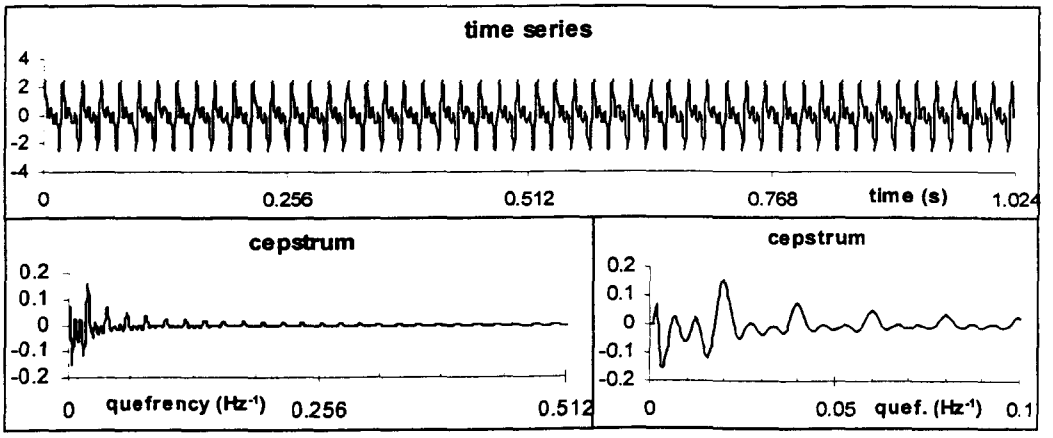


Figure 5.2 – Cepstral analysis of Sin(50t)+Sin(100t)+Sin(150t)

The time data is shown on the top figure. It consists of three sine waves (frequencies of 50 Hz, 100Hz, 150Hz). The sampling rate was set to 1000Hz and there are 1024 samples in the series. The lower right figure shows the whole cepstral plot with quefrencies ranging from $0.002Hz^{-1}$ (500Hz) and $0.512Hz^{-1}$ (approx. 2Hz). Note that because of the symmetry of the Fourier transform of a real signal, only quefrencies from 0 to $0.512\ Hz^{-1}$ are shown. Finally, the third figure is a zoom of the second figure. The third figure shows quite well two fundamental characteristics of the cepstral analysis.

The first point to be noted is that the amplitude of quefrencies smaller that 0.002s are zero. This is a result of the sampling theorem, which states that the sampling rate must be at least twice the highest frequency in the signal. Remember that the sampling rate was set to 1000Hz, hence it follows that the largest observable frequency should be 500Hz (i.e. quefrency of 0.002s).

The second point to be noted is that the dominant peak has a quefrency of 0.02s. This corresponds to a frequency of 50Hz, which is the fundamental frequency on the input signal. It is also important to note that no dominant peaks are present at 0.01s (100Hz) and at 0.007s (150Hz), as the cepstral analysis "pushes" these frequencies to their fundamental value (50Hz).

As it can be seen cepstrum analysis is an extremely powerful tool to observe the presence of harmonics in a signal. It is also very useful for observing transient components, as these tend to influence the initial values along the quefrency axis.

This suggests that the cepstrum transform of a time signal (with transient components) will show values, with a large magnitude (positive or negative), in the first samples along the time axis. While the cepstrum transform of a signal (without transient components) will show values with small magnitudes on the first samples along the quefrency axis. The concentration of information in the first samples of the quefrency axis is a result of the phase shift during the DFT and IDFT operations. This property of the cepstrum is illustrated in the figure below, showing how five different transients are reconstructed. The notation for these charts is included overleaf.

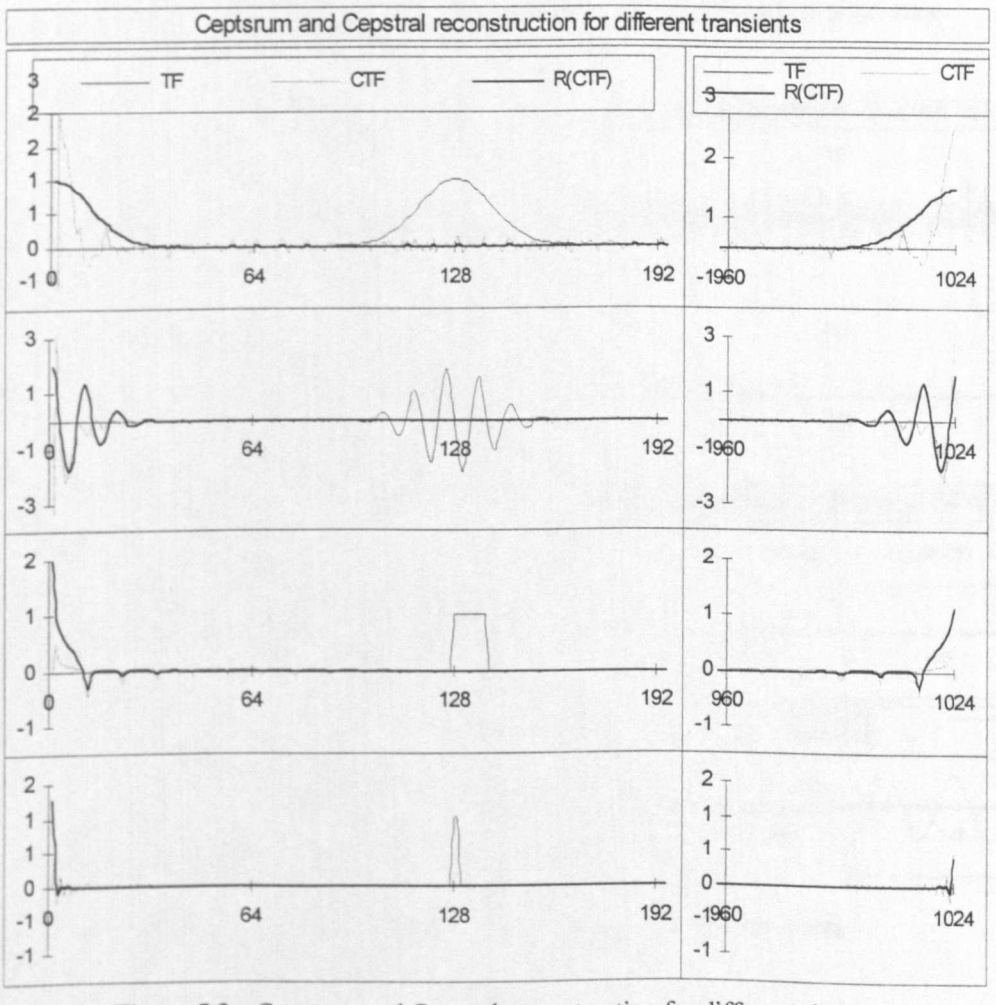


Figure 5.3 – Cepstrum and Cepstral reconstruction for different signals

The notation in the above diagram is as follow:

- TF denotes the time domain function;
- CTF denotes the cepstrum of TF
- R(CTF) denotes the cepstral reconstruction of CTF

As it can be seen the linear piecewise transients (the Gaussian pulse on first row, and the sine enveloped Gaussian pulse on second row) are reconstructed back to their original shape however these were shifted from t=0.128s to t=0. In contrast the two square pulses (wide pulse on third row, and very narrow pulse on the last row) are not reconstructed back to their original shape, still these are shifted to t=0s and increase the initial reading on the quefrency axis.

The charts below show how the shift property behaves when transients are embedded into non-transient signals. This simulates a fault in a real vibration signature.

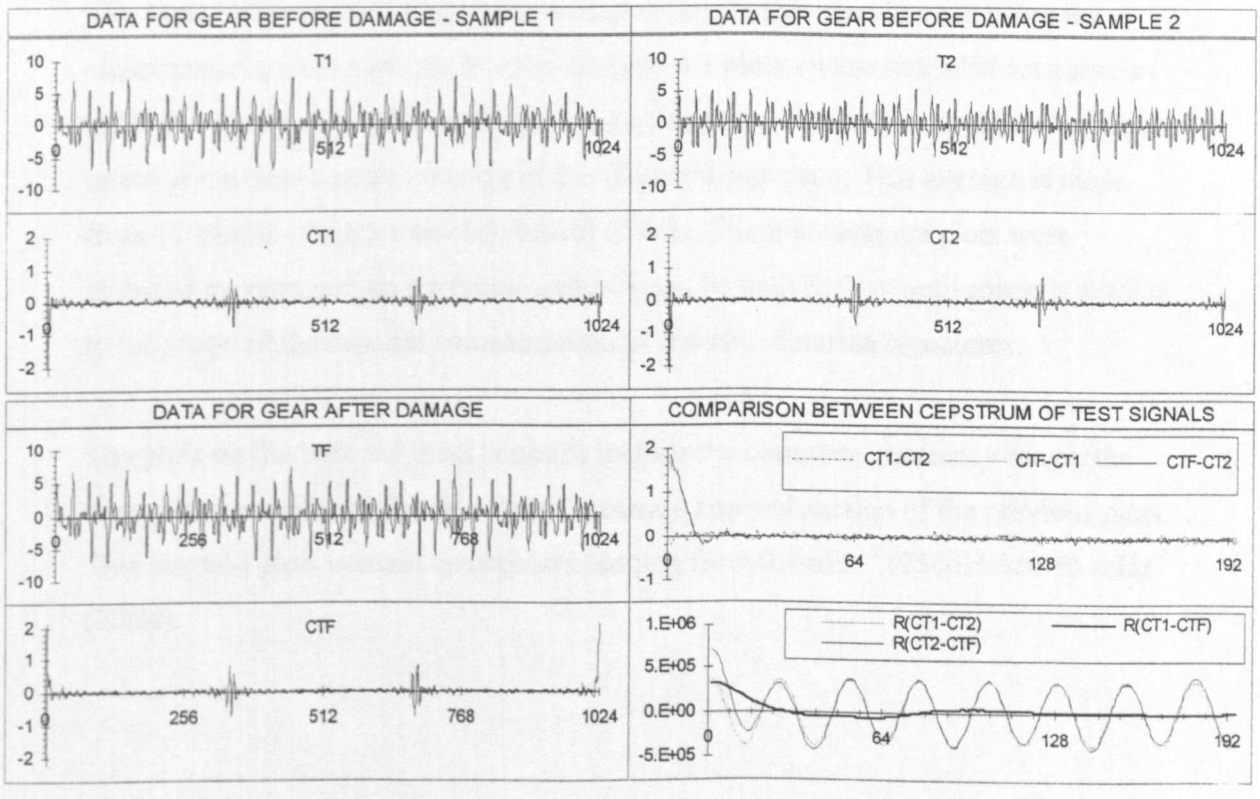


Figure 5.4 – Cepstral analysis of numerically simulated data

5.2.2. Results of cepstral analysis of experimental data

In this section cepstral analysis coupled with cepstral reconstruction was used, as this provided the best indication for the 'quefrencies' in the experimental vibration signatures. The chart below illustrates this choice.

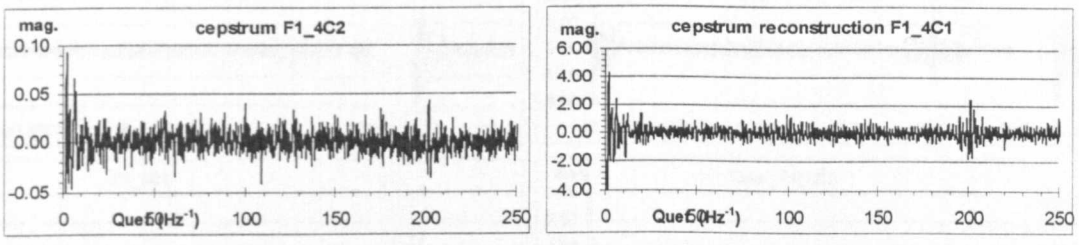


Figure 5.5 – Comparing the cepstral and the cepstral reconstruction plot

As it can be seen the sharpness of the peaks on the plots for the cepstral reconstruction is greater than on the plots for the simple cepstrum.

The plots below show the quefrency components for the experimental vibration signatures. Again as with the Fourier analysis the plots on the two leftmost columns relate to the raw data with no pre-processing, the ones on the two rightmost columns relate to the time domain average of the vibration signature. This average is made from 12 blocks (4 cycles on each block) of data. The time average plots were included for comparison purposes, and as it can be seen their overall shape is similar to the shape of the cepstral reconstruction of the raw vibration signatures.

The plots on the first and third columns include the complete quefrency range; the ones in the second and fourth column contain a zoomed section of the previous plots. This zoomed plots contain quefrencies ranging from 0.4mHz^{-1} (2500Hz) to 50 mHz^{-1} (20Hz).

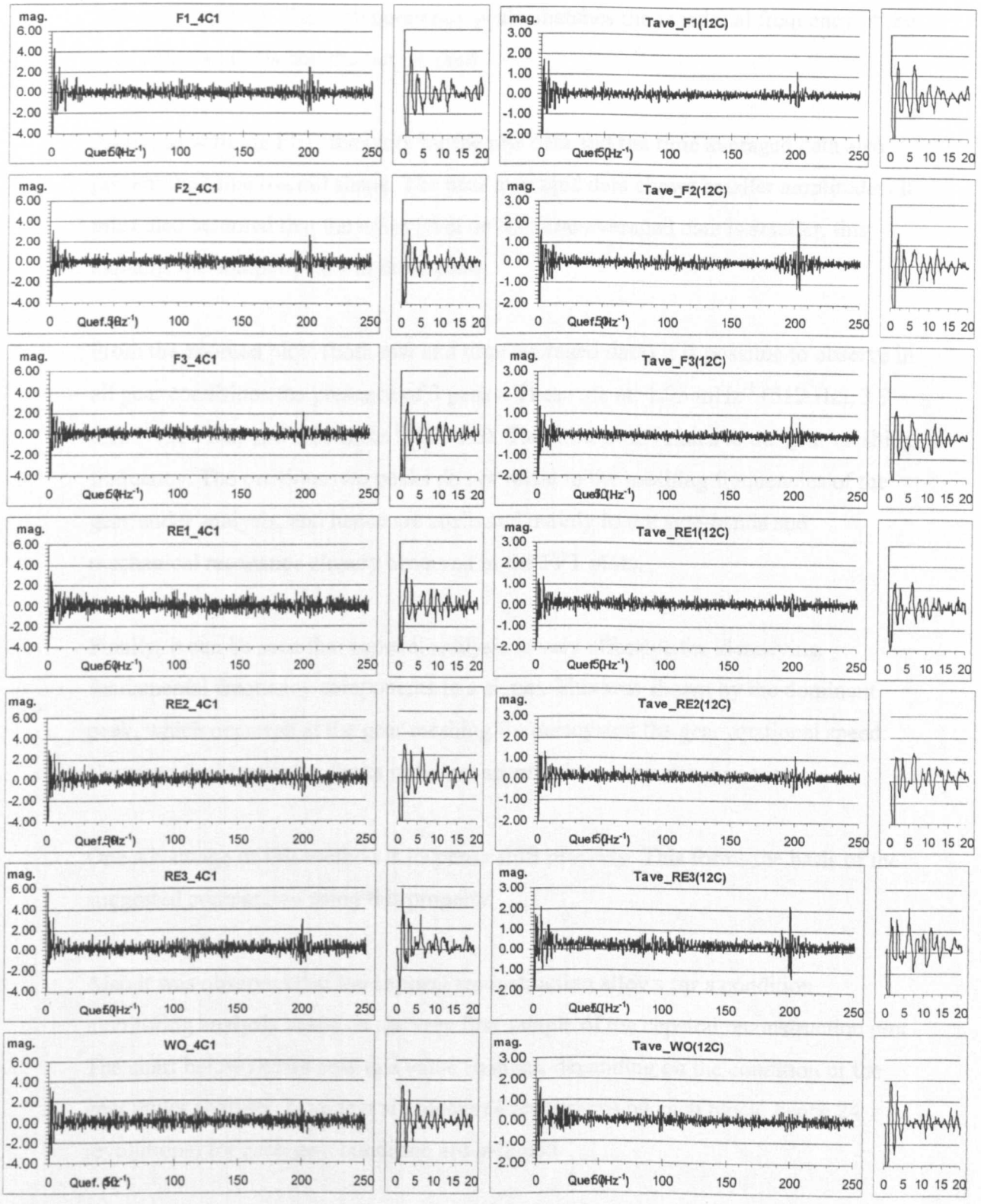


Figure 5.6 – Cepstral reconstruction plots for experimental data

From these plots it can be clearly seen that the vibration signatures all seem to present peaks at low quefrencies (high frequencies) and a peak at a higher quefrency,

200 mHz^{-1}, (5Hz). The high quefrency peaks matches the rotational frequency of the gear under analysis and the output shaft.

Again, as with the FFT, the plots for the raw data and the time averaged data also present the same overall shape. The time averaged data shows smaller amplitudes. It must also be noted that the noise level on the time-averaged data is smaller, this explains the sharper peaks in those plots.

From the zoomed plots (both raw and time averaged data) it is possible to observe in all gear conditions the presence of 3 peaks. These are at: 1.95 mHz^{-1} (512 Hz), 3.9 mHz^{-1} (256 Hz) and 6.25 mHz^{-1} (160 Hz). The third peak relates to the gear meshing frequency. The previous two peaks do not relate to the meshing frequencies of the gear under analysis, and hence are attributed mainly to the side-bands and mechanical resonance already observed in the FFT plots.

Finally, it can be seen that cepstral analysis is very effective for identifying fundamental frequency components in a signal. This was shown by the dominant peak, which occurred at the gear meshing frequency and the gear rotational speed. However for identifying faults it did not seem to be a very effective tool.

One advantage of this method is its phase shift property. This forms the basis of the suggested comparison using this property.

Also it was observed that the cepstral reconstruction allows for a condition monitoring analysis based on the very first sample of the cepstral reconstruction plot. The chart below shows how this value changes, depending on the condition of the gear under analysis. Note that six blocks (4 revolutions on each block, hence 24 revolutions) for each gear condition are included.

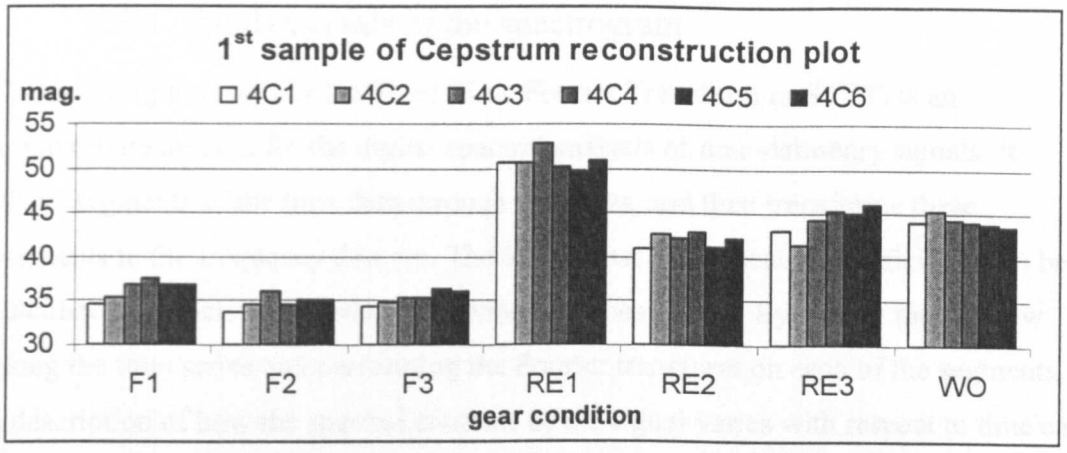


Figure 5.7 – Magnitude of first sample of cepstral reconstruction plots for experimental data

As it can be seen the magnitude of the very first sample of the cepstral reconstruction plot is very affected by the gear condition. Note that this phenomenon is also observable, to a lesser extent, on the cepstrum plots.

It is believed that the reason behind this behaviour is related to the phase shift, which is intrinsic to the real cepstral analysis of time series. So, as the phases of the transients are removed, when performing the cepstrum reconstruction, these transients are shifted to the beginning of the quefrency axis, affecting mainly the initial reading.

These results could be an explanation to the very high accuracy obtained by the artificial neural networks in [109], which were used as a pattern recognition tool for the condition monitoring of a similar gear train, with cepstral analysis as a pre-processing tool.

5.3. Spectrogram

The spectrogram was one of the first time-frequency distributions. It was first used in the 1940's to analyse human speech [42]. Since then subsequent developments turned this technique into a powerful tool for time-frequency analysis of stationary and non-stationary signals.

5.3.1. Fundamentals of the spectrogram

The spectrogram (also called Short Time Fourier Transform or STFT) is an approximate method for the digital spectral analysis of non-stationary signals. It views segments of the time data through windows, and then transforms these segments to the frequency domain. The idea is that local spectral coefficients can be calculated for each time t, where the window is positioned. By sliding the window along the time series and performing the Fourier transform on each of the segments, a description of how the spectral contents of the signal varies with respect to time can be obtained.

The most commonly used window function is the Gaussian function, as this is the only function that possesses similar shapes in the time and frequency domain. Therefore, once the Fourier transform (FT) is performed no unwanted side lobes (or ripples) are introduced in the spectrogram. Other functions, which have also been used for the windowing process, are the Hamming, Hanning, rectangular and triangular windows [48].

5.3.2. Theoretical background

The concept behind the spectrogram is simple. In order to analyse how the spectral contents of the signal vary with time, a window is positioned on each particular time and the FT of the windowed signal is performed. This process is repeated for each instant of time of the signal. So if $x(\tau)$ is the time data series to be analysed and $h(\tau)$ is the window function used in the STFT. The windowed signal x_w is:

$$x_w(t,\tau) = x(\tau).h(\tau - t), \qquad \textbf{Eq. 5.5}$$

where: t is the time instant where the window is centred and τ represents time.

Now performing the Fourier transform on the windowed signal x_w, the following definition is obtained:

$$X(t,f) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x_w(t,\tau)e^{-i\omega t}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(\tau).h(\tau - t)e^{-i\omega t}dt\,. \qquad \textbf{Eq. 5.6}$$

The spectrogram is defined as the power spectral density (PSD) of $X(t,f)$ [53]. This is given by:

$$S_w = |X(t,f)|^2. \qquad \textbf{Eq. 5.7}$$

Note that if $h(t)=1$ then this method is equivalent to the ordinary Fourier Transform of the signal, giving, as an output, the signal spectrum for all window positions.

The choice of the window function will greatly affect the performance of the spectrogram. This research only focuses on the Gaussian window, as it is the only window that does not add unwanted sidelobes to the spectrogram. The Gaussian function and its FT are defined as:

$$h(t) = ce^{-\sigma^2 t^2}, \qquad \textbf{Eq. 5.8}$$

$$H(f) = \frac{c\sqrt{\pi}}{\sigma} e^{-(\pi/\sigma)^2 f^2}, \qquad \textbf{Eq. 5.9}$$

where c and σ are both positive constants.

The constant 'c' controls the gain of the window, affecting the local sensitivity of the window function. The constant σ controls the width of the window function, setting the time/frequency resolution. A variable called window width was introduced to facilitate the process of choosing the window. This variable is used to determine the number of samples that will be enhanced by the window function.

The diagram below illustrates the meaning of the variable window width. It shows a Gaussian window centred at sample 256 and with window width=10. As it can be seen, the window enhances only 10 samples (from 251 to 261). The remaining samples (0 to 250 and 262 to 1024) are suppressed as the magnitude of the window is smaller than unity.

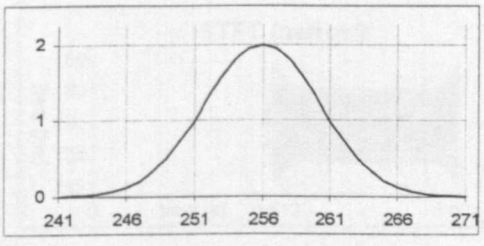


Figure 5.8 – Gaussian Window

The window peak and width should be selected to enhance the local patterns in the signal. In order to achieve this the window peak and window width must be set so that it

covers the period during which the pattern occurs [47]. This is one of the difficulties associated with this method, as one needs previous knowledge about the pattern being searched so that the window width can be set appropriately.

Note that this approach has one major drawback. It can never attain good frequency and time resolution simultaneously. If the width of the window is T, its frequency bandwidth will be 1/T. Therefore, when σ is set to be large (narrow window width and high time resolution) it follows that the frequency resolution will be very poor (as $1/\sigma$ will be large). On the other hand when $1/\sigma$ is set to be small (wide window and high frequency resolution) then σ will have to be large leading to poor time resolution. Therefore the two ideal requirements - a short time window and a narrow frequency bandwidth - are impossible to achieve.

The three plots in Figure 5.9 illustrate graphically this time-frequency resolution trade-off. The input signal is shown on the top figure (a pure sine wave with frequency of 100Hz starting at time 0.512 seconds). The middle figure is the spectrogram map for the input signal obtained when using a wide window (width of 300 samples i.e. 300ms). The lower figure is the spectrogram map obtained when using a narrow window (width of 65 samples i.e. 65ms).

As it can be seen one cannot achieve good frequency and time resolution simultaneously. In the middle figure a good frequency resolution is obtained, however time resolution is a problem. In the lower figure a good time resolution is obtained however now it is frequency resolution that poses as a problem.

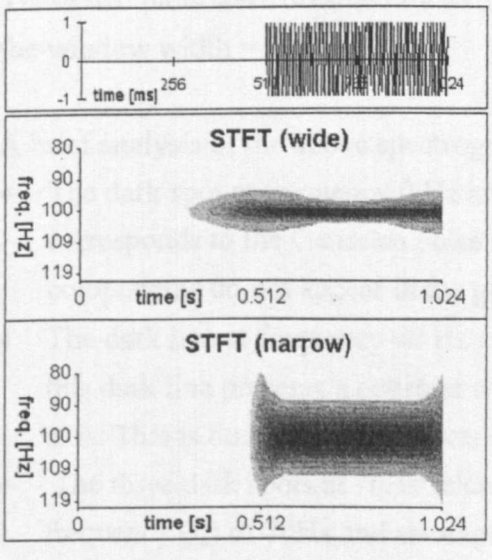


Figure 5.9 – TF resolution trade off

Unfortunately, with the spectrogram, there is no way around this time-frequency resolution compromise. Hence before the window width is set tests, must be performed to evaluate the performance of windows with different widths.

A discrete version of Eq. 5.6 can be implemented in a program which will then perform the STFT of a given time series. The discrete version of Eq. 5.6, follows:

$$X(t,f) = \frac{1}{2\pi}\frac{1}{N}\sum_{n=0}^{N} x(n)\, h(n-t)\, e^{-(2\pi Kt)/N} \quad \text{Eq. 5.10}$$

5.3.3. Numerical Example

Before analysing the experimental data a preliminary analysis on the numerically simulated data generated as described in chapter 3 was performed. A spectrogram routine written in C++ was used to process the two sets of data (i.e. good gear and the faulty gear). The results can be seen below. The x-axis represents the frequency, and the y-axis represents gear position. Note that the time series represents one full gear revolution. Therefore 360° is equivalent to 1.024 seconds.

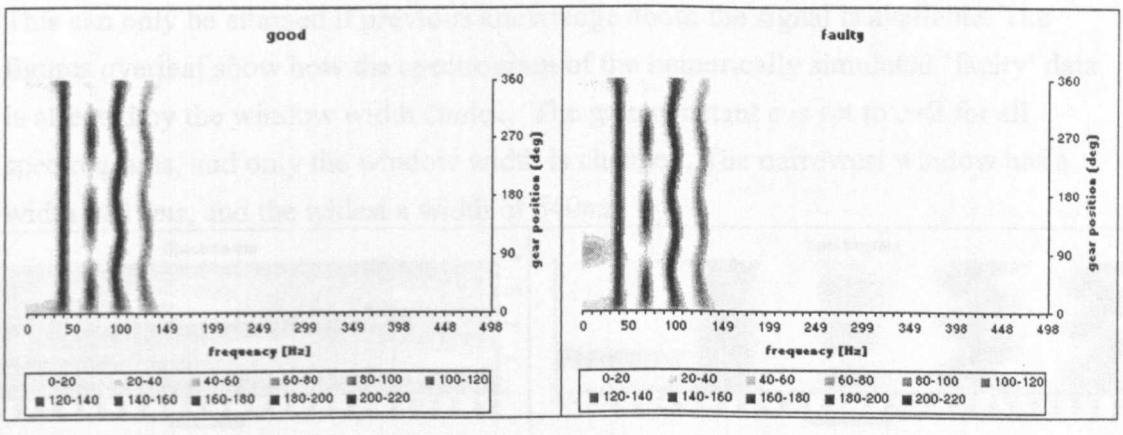


Figure 5.10 – Spectrograms for numerical data

The graph on the left shows the spectrogram for the 'good' condition gear signal, while the graph on the right shows the spectrogram for the 'faulty' condition signal. These two plots were obtained by setting the window function constants to c=2.0 and the window width = 100ms.

A brief analysis of the above spectrogram shows its main properties:

- The dark spot at frequency 0 Hz and angular position 90° on the 'faulty' plot corresponds to the Gaussian pulse (i.e. simulated fault). Note that this components do not appear in the good gear condition plot (on the left hand side).
- The dark line at frequency 40 Hz corresponds to the pure sine wave. Note that this dark line presents a constant width except at the end and the start of the time axis. This is due to the end effects in the Fourier Transform.
- The three dark spots at 70Hz relate to the amplitude modulated sine, with carrier frequency (f_c) of 70Hz and an amplitude modulation frequency (f_a) of 3 Hz.

- The dark '*s*' shaped curve at 100 Hz corresponds to the frequency modulated sin wave, with f_c=100Hz and a frequency modulation frequency (f_f) of 3Hz.
- The three dark '*s*' shaped spots appearing at 130 Hz correspond to the amplitude and frequency modulated sine wave, with f_c=130Hz, f_a=3Hz and f_f=3Hz.

As it can be seen the spectrogram is an effective signal processing method for the time-frequency analysis of dynamic systems. It has been shown that it can be used to detect a wide range of signals. Also in the case of amplitude and frequency modulated signals it is possible to obtain the carrier and the modulation frequency.

Unfortunately, this method relies heavily on the correct choice for the window width. This can only be attained if previous knowledge about the signal is available. The figures overleaf show how the spectrogram of the numerically simulated 'faulty' data is affected by the window width choice. The gain constant c is set to c=2 for all spectrograms, and only the window width is changed. The narrowest window has a width of 15*ms*, and the widest a width of 540*ms*.

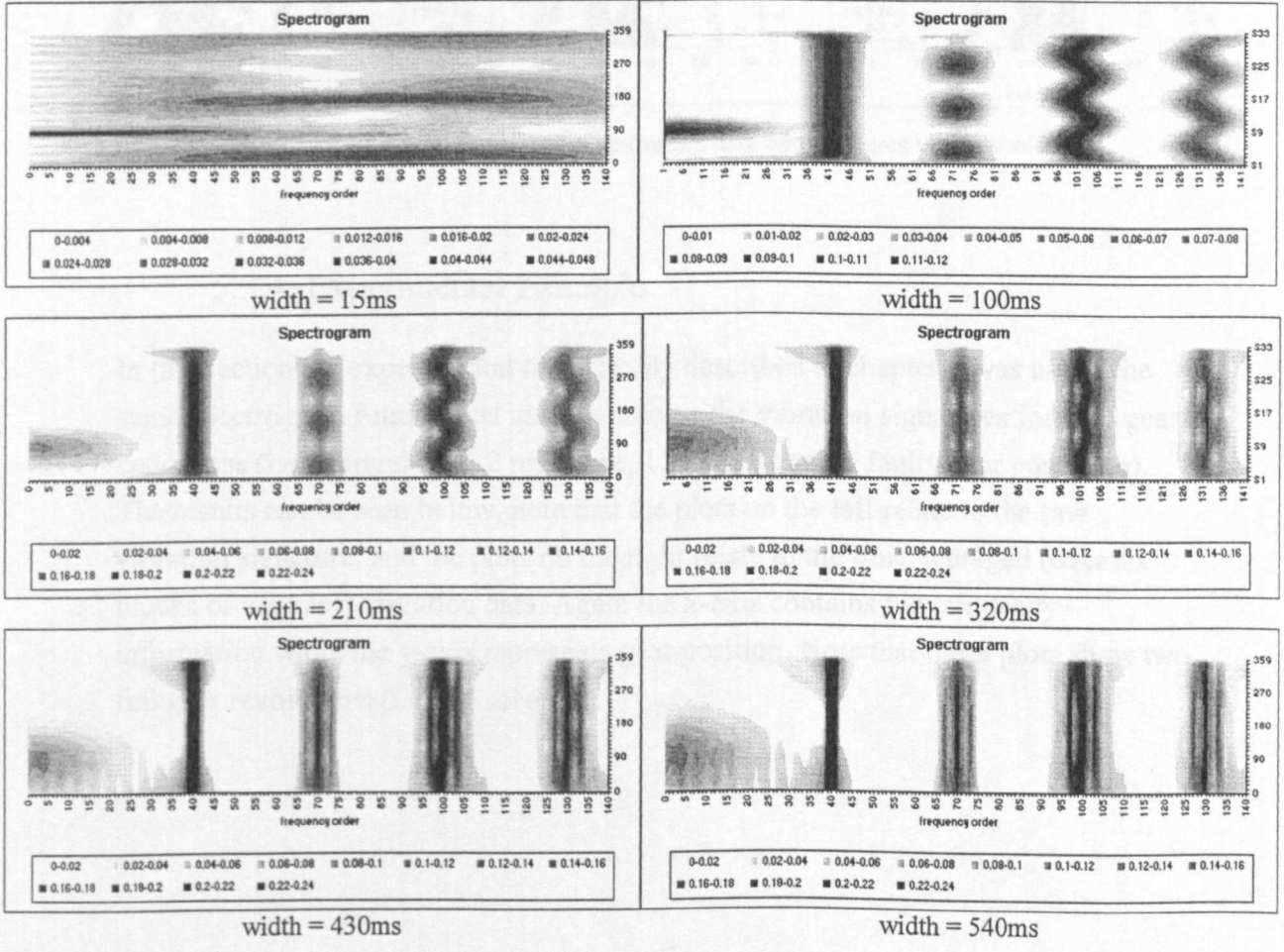


Figure 5.11 – Spectrogram of numerical data with different window width.

As it can be seen the window width choice greatly affects the performance of this method. For width=15ms one can hardly observe the presence of any of the signal components; and for width=100ms very little frequency resolution. In contrast for width=320, 430 and 540ms the three highest frequency components are shown with sidebands as result of the window width choice.

For this reason in the analysis of the experimental data, tests for different window widths were performed. From this it was observed that the ideal width for the analysis of the collected vibration signatures was 10ms (51.2 samples). This is illustrated by the Figure 5.12 and Figure 5.13.

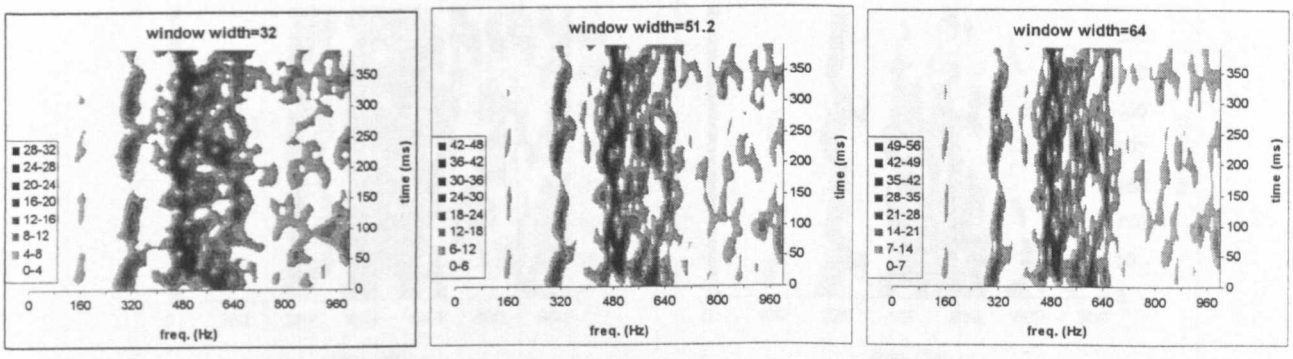


Figure 5.12 – Spectrogram of experimental data with different window width.

5.3.4. Experimental Example

In this section the experimental data already described in chapter 3 was used. The same spectrogram routine was used to process the vibration signatures for all 7 gear conditions (i.e. 1 brand new, 2 reference, 1 worn out and 3 faulty gear condition). The results can be seen below, note that the plots on the left relate to the raw vibration signature, and the plots on the right relate to the time averaged (over six blocks of 4 cycles) vibration data. Again the x-axis contains the frequency information while the y-axis represents gear position. Note that these plots show two full gear revolutions (i.e. 0.4 seconds).

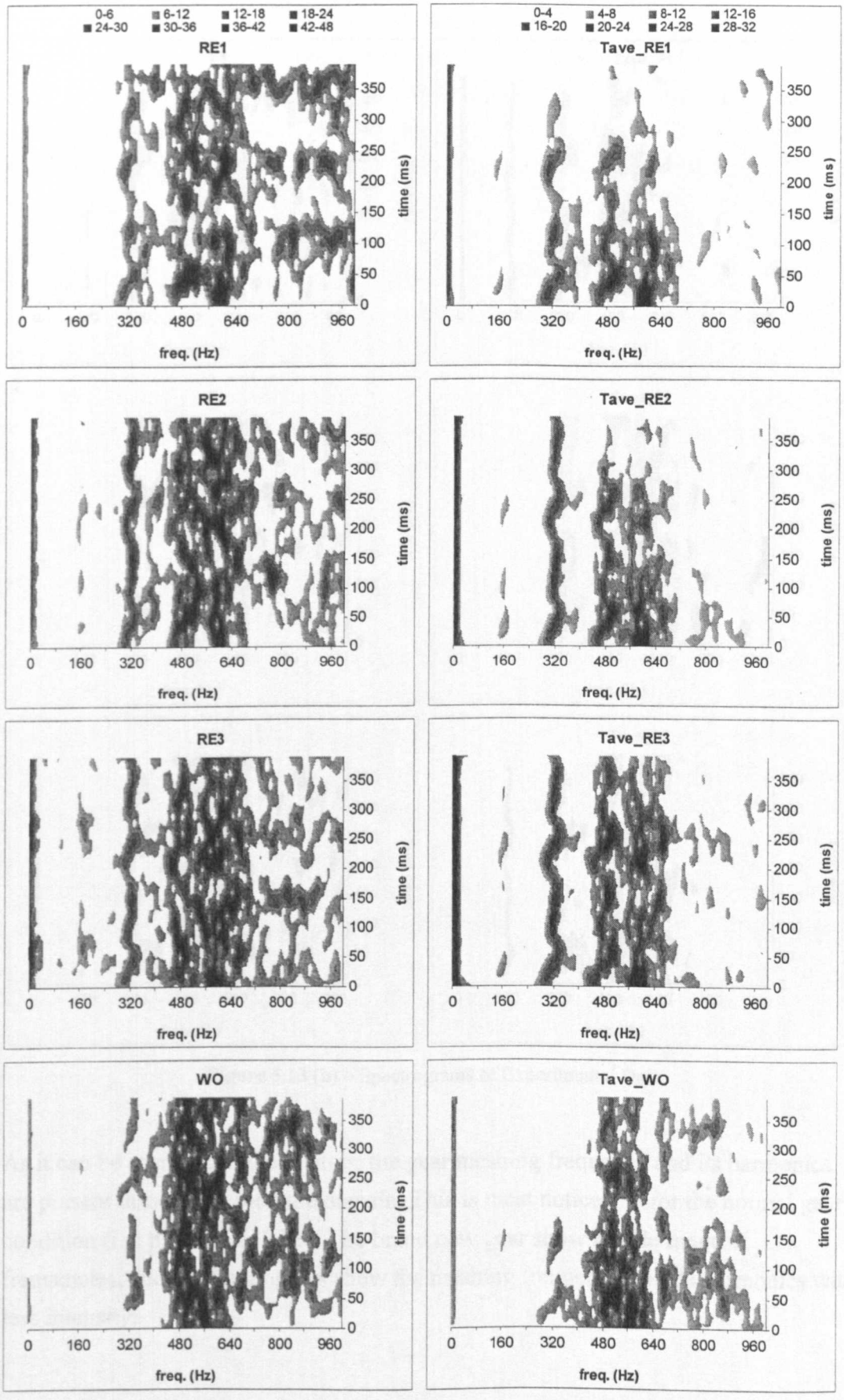


Figure 5.13 (a) – Spectrograms of experimental data

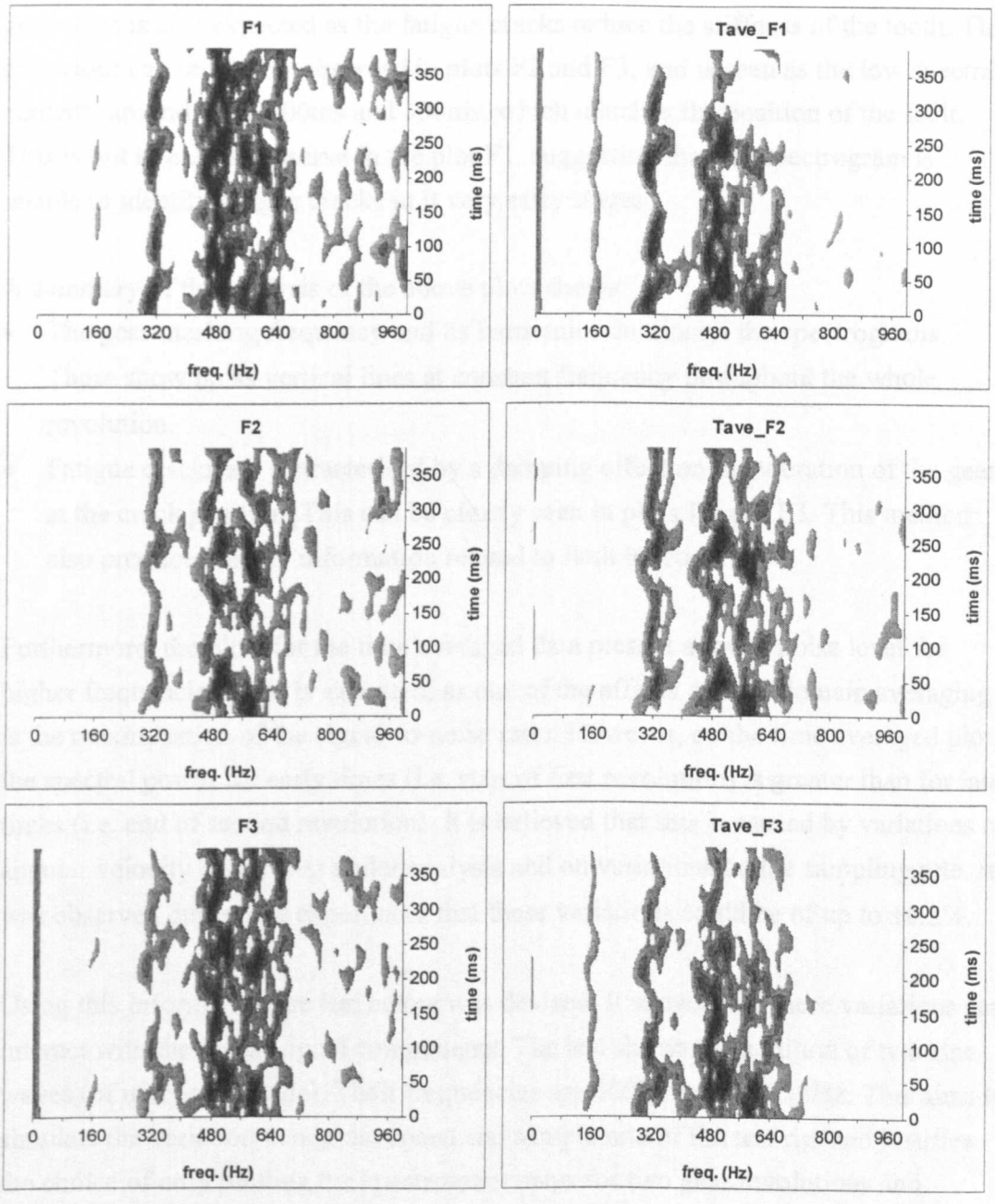


Figure 5.13 (b) – Spectrograms of Experimental Data

As it can be seen in the above plots, the gear meshing frequency and its harmonics are present throughout the time domain. This is most noticeable for the normal gear condition (i.e. plots RE2, RE3). The brand new gear show a wide range of frequencies, and the faulty gears show the meshing frequency and its harmonics with less intensity.

Also the fatigue crack has a definite damping effect on the overall vibration of the gear. This is also expected as the fatigue cracks reduce the stiffness of the tooth. This behaviour can be clearly observed in plots F2 and F3, and is seen as the low spectral contents around time, 100ms and 300ms, which matches the position of the fault. This is not as easy to observe in the plot F1, suggesting that the spectrogram is unable to identify fatigue cracks in it very early stages.

A summary of the analysis of the above plots shows:

- The gear meshing frequency and its harmonics dominates the spectrograms. These show up as vertical lines at constant frequency throughout the whole revolution.
- Fatigue cracks are characterised by a damping effect on the vibration of the gear at the crack position. This can be clearly seen in plots F2 and F3. This method also produced useful information related to fault location.

Furthermore, the plots for the time averaged data present a lower noise level for higher frequencies. This is expected, as one of the effects of time domain averaging is the maximisation of the signal-to-noise ratio. However, on the time averaged plots, the spectral power for early times (i.e. start of first revolution) is greater than for later times (i.e. end of second revolution). It is believed that this is caused by variations on angular velocity of the gear under analysis and on variations on the sampling rate. It was observed during the experiment that these variations could be of up to ±0.2%.

Using this information the test below was devised. It shows how these variations can interact with the actual signal components. The test shows the addition of two sine waves (of unity amplitude). Their frequencies are 600Hz and 601.18Hz. This aims to simulate the variation in angular speed and sample rate of the test rig, and justifies the choice of only plotting the spectrogram maps for two gear revolutions and frequencies up to 1000 Hz.

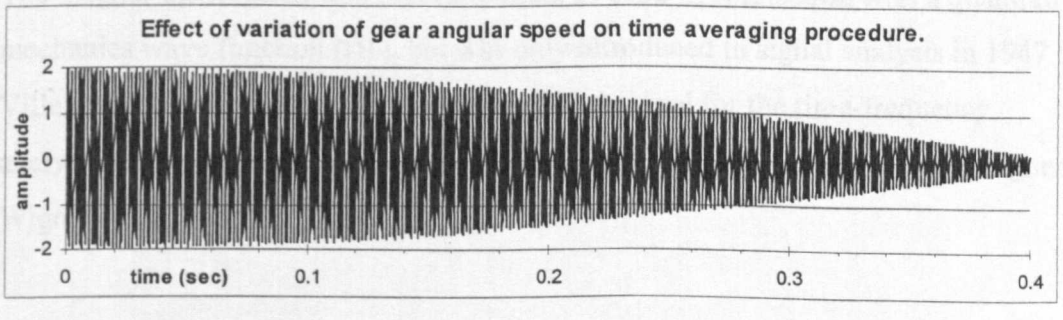


Figure 5.14 – Effect of variation of gear angular speed and/or sample rate on test rig.

This is equivalent to a 0.2% change in the angular speed and/or sample rate. As it can be seen after two cycles (i.e. 0.4s) these signal components are nearly 180° out of phase, causing destructive interference. This leads to the conclusion that the time domain averaging must be used with extreme care, as it is directly affected by: the signal length; the accuracy of the position measuring transducer (which tracks the gear position), and; the accuracy (and constancy) of the angular speed of the gear in question.

5.3.5. Summary of Spectrogram performance

It has been shown that the spectrogram is an effective method for industrial condition monitoring diagnostics if previous knowledge about the signal to be analysed is available. Although it can also be used to analyse signals with no previous knowledge, it is not recommended as the visual analysis of the spectrogram might lead to erroneous conclusions. A brief summary of the advantages and disadvantages of this method is shown below.

Advantages

- The spectrogram of a signal can be seen as the distribution of the signal energy not only in the frequency domain (like a spectrum) but also in the time domain.
- The method is very effective in analysing stationary signals.
- It is a very simple method, which can also be used to analyse non-stationary signals.

Disadvantages

- This method does not allow for simultaneous high time and frequency resolution.
- Previous knowledge about what type of patterns one is investigating is essential for setting the window width correctly. If the window width is not set correctly then the pattern being searched might not be found.

5.4. Wigner distribution (and its variants)

The Wigner distribution was first developed in 1932 in connection with a quantum mechanics wave function [110], but was only introduced to signal analysis in 1947 by Ville [54].This distribution is yet another effective tool for the time-frequency analysis of signals. There are several variations on the Wigner distribution (Wigner, Wigner-Ville, Pseudo-Wigner, etc.).

This research focuses only on the Wigner Distribution (WD), Wigner-Ville Distribution (WVD), and the Pseudo Wigner Distribution (PWD). These three distributions are very similar in form and very often mislabelled. A formal definition for these distributions is included in the next sections.

5.4.1. Description of fundamental aspects of the Wigner distribution (and its variants)

The Wigner distribution (WD) produces a plot (similar to spectrograms) showing how the spectral content of a signal varies with time. Like spectrograms this distribution also uses the Fourier Transform Algorithm. However it differs from the spectrogram as it does not suffer from the same time-frequency resolution compromise, leading to a better time resolution than the spectrogram [111]. This is illustrated by calculating the Pseudo-Wigner Distribution (PWD) of a time signal. In this calculation a Gaussian window (similar to the spectrogram window) was used.

The input signal is shown on the top figure (a pure sine wave with frequency of 100Hz). The middle figure show the PWD calculated with a wide window (window width of 300 ms), while the bottom figure shows the PWD for a narrow window (window width of 65ms). As it can be seen the window width only affects the frequency resolution, it does not affect the time resolution.

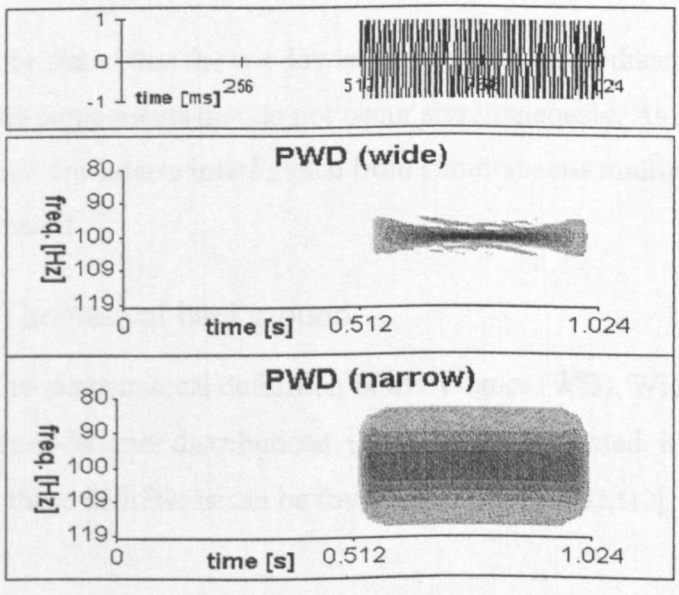


Figure 5.15 – PWD Time-Frequency resolution

Unfortunately this method suffers from a different problem than the spectrogram: cross-term interference from different signal components. The cross-term interference can be minimised in many ways. These are: by means of performing the Wigner distribution of the analytic signal (Wigner-Ville distribution) or by means of a sliding window function (Pseudo-Wigner distribution).

In the Wigner-Ville distribution the analytic signal, which does not contain any negative frequency, is used instead of the real signal. This eliminates the cross-term interference between positive and negative frequencies. Note that this is not sufficient to eliminate the cross-term interference arising from multiple components in a signal.

In the Pseudo-Wigner distribution a window function is used to place different weights on the time segments being analysed, this will reduce even further the cross term interference from the signals multiple components that occur at different times. The PWD can have different functions as its window function. The most common are Gaussian, Hamming, Hanning, and Kaiser. It is important to note that even though the PWD presents less cross-term interference than the WD and the WVD, it no longer has some of the desirable properties of the WD and WVD. The time frequency marginals (i.e. energy in time domain = energy in frequency domain) are no longer valid [59].

Finally, it must be stated that the windowing procedure only reduces the interference between multiple components that do not occur simultaneously. As it will be shown in the next section cross-term interference from simultaneous multiple components cannot be eliminated.

5.4.2. Theoretical background

In this section the mathematical definition of the Wigner (WD), Wigner-Ville (WVD) and Pseudo-Wigner distributions (PWVD) are presented. More detailed explanation for these definitions can be found in references [53,112].

The Wigner distribution of a real signal *x(t)* is defined as:

$$WD(t,f) = \int_{-\infty}^{\infty} x\left(t+\tau/2\right)x\left(t-\tau/2\right)e^{-2\pi f\tau i}d\tau . \qquad \text{Eq. 5.11}$$

The Wigner-Ville distribution of a real signal *x(t)* is defined as:

$$WD(t,f) = \int_{-\infty}^{\infty} \tilde{x}\left(t+\tau/2\right)\tilde{x}^{*}\left(t-\tau/2\right)e^{-2\pi f\tau i}d\tau . \qquad \text{Eq. 5.12}$$

Where $\tilde{x}$ is the analytic signal of x and is found by [61]:

$$\tilde{x}(t) = x(t) + \frac{i}{2\pi}\int_{-\infty}^{\infty}\frac{x(\eta)}{t-\eta}\,\mathrm{d}\eta . \qquad \text{Eq. 5.13}$$

Alternatively, the analytic signal can also be found in frequency domain using the Fourier transform *X(t)* of the real signal *x(t)* then zeroing the negative frequencies and doubling the positive frequencies as follows:

$$\tilde{X}(t) = \begin{cases} 2.X(t) & f>0 \quad (i.e.\ 0<t<N/2) \\ X(t) & f=0 \quad (i.e.\ t=0,N/2) \\ 0 & f<0 \quad (i.e.\ N/2<t<N) \end{cases} \qquad \text{Eq. 5.14}$$

Note that the above defined distribution is also named Wigner distribution by many authors, as it is already widely used in place of the original WD due to its better performance. Finally the Pseudo-Wigner distribution is defined as:

$$PWD(t,\omega) = 2\int_{-\infty}^{\infty}\tilde{x}(t+\tau)\tilde{x}^{*}(t-\tau)h(\tau)h^{*}(-\tau)e^{-4i\omega t}\,\mathrm{d}\tau , \qquad \text{Eq. 5.15}$$

where the asterisk denotes complex conjugate and tilde denotes analytic signal.

Note that the window function *h(t)* can be any filtering function. The most commonly used is the Gaussian function but other known window functions (such as Hamming, Hanning, Kaiser, etc.) can also be used. A full listing of the most common window functions and their respective properties can be found in [48].

Figure 5.16 illustrate how the PWVD (or PWD) differs from the WVD. These plots were obtained by processing two different time series. The time series that generated these plots are also shown above the Wigner plots. The plots on the left are for a time

series consisting of one sine wave (constant frequency 40 Hz) from 0 to 342 ms, then noise from 342 to 684 ms, and a second sine wave (constant frequency 80Hz) from 684 to 1024ms. The plots on the right were obtained by processing a time series composed of the summation of two sine waves (of frequencies 40 and 80Hz).

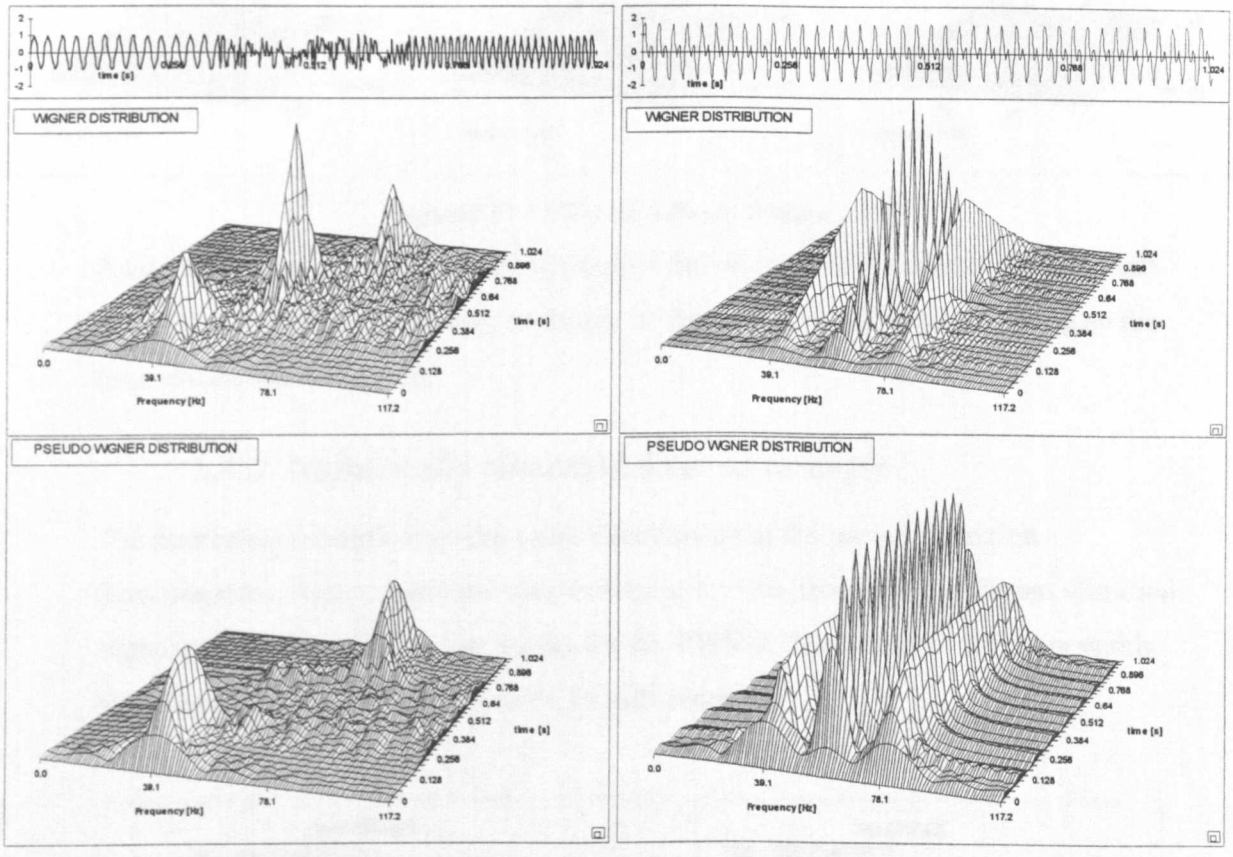


Figure 5.16 – PWD and WVD for different signals

As it can be seen the pseudo Wigner distribution is very effective at cancelling the cross-term interference from signal components that occur at different times, however for simultaneously occurring components it is not so effective. Still the PWVD seems to present a simpler output than the original Wigner-Ville distribution. For this reason, only the PWVD will be used to analyse the numerical and experimental data. This agrees with the current practice in practical applications [60].

Note that the window size also affects the output of the PWVD (especially the frequency resolution). As the window size is increased the PWVD results tend to be like those obtained from the WVD (i.e. high frequency resolution and high presence of cross-terms). This is expected from theory. The plots below show how the window

size affects the PWVD. The title of each plot indicates the window variance in milliseconds.

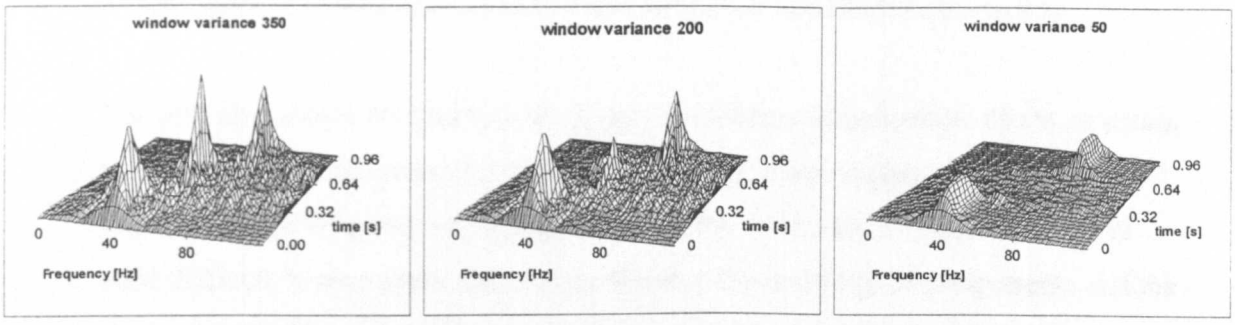


Figure 5.17 – PWD for different window width

As it can be seen by using a narrow window the cross-term (middle peak in the left plot) interference is eliminated, as shown in the right plot. However in doing so the time resolution is reduced.

5.4.3. Numerically simulated data: an example

The numerical example uses the same time data as in the previous section (spectrogram). Again, a routine was written in C++ to process the different vibration signatures. Underneath are the results for the PWVD. The optimum window width has been found experimentally to be 70 milliseconds.

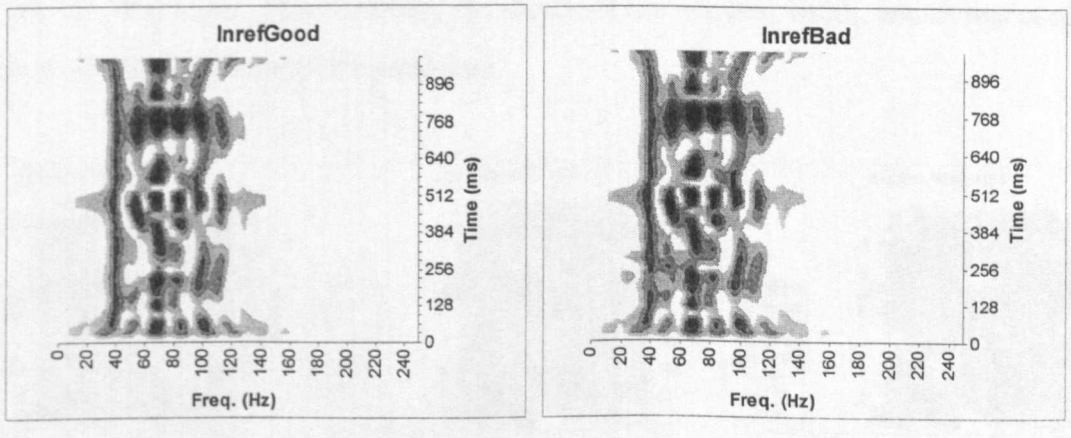


Figure 5.18 – PWD for numerical data

As it the can be seen the PWD still shows the frequencies present in time domain series. This is shown by:

- the dark line at 40Hz (sine wave component);

- the dark spots at 70Hz (the amplitude modulated sines wave);
- the curved spots 100Hz (the frequency modulated sine wave), and;
- the spots at 130Kz (the frequency and amplitude modulated sine wave).

The plot also shows the multiple frequency component interference, this is as strong as the frequency components themselves. Note that these appear in-between vibration signal frequency components (i.e. 55Hz, 85Hz and 115Hz). In fact it is quite difficult to accurately distinguish between the main signal components and the cross term interference. However, the fault can still be easily identified. It appears at position 256ms and 24 Hz. It is important to note that in this method window size also plays an important role. Finally, from this test it is possible to observe the effect of the inevitable cross term interference from simultaneous cross term components.

5.4.4. Experimental Example

The data for this example was obtained as described in chapter 3. The window width selected was: width=100 samples (19.5ms). The figures below show the PWD plots for the faulty condition gear (F1) with different window widths. The title on each plot indicates the window width as a number of samples (W_s), this can be converted to time by: $W_t=W_s/F_s$ – where F_s is the sampling frequency (5120 Hz); hence for Ws=50, $W_t\approx$ 10ms. This illustrates the choice of the window width, which was used in processing all the experimental data.

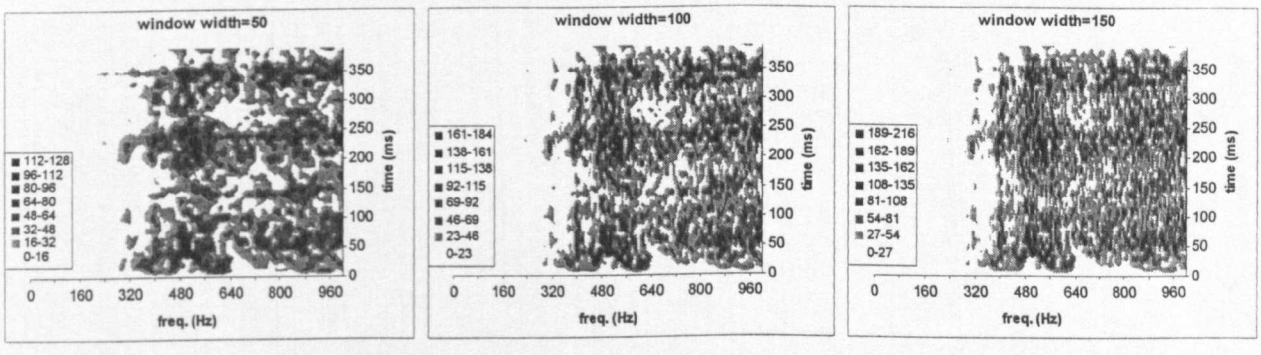


Figure 5.19 – PWD of experimental data with different windows

From the above plots it can be seen that only the frequency resolution is affected by the different window widths. The dominant peaks on the first plot (width=50

samples) are wider in the frequency domain than those in the third plot (width=150 samples). Also note that, for visualisation reasons these plots have different z-axis scales.

The plots for the experimental data are shown overleaf. These follow the same pattern as with the previous condition monitoring methods (i.e. plots on the left refer to raw vibration signature and plots on the right refer to time averaged signals).

It must be emphasised that, as with the spectrogram plots, the plots for the raw data show two gear revolutions and all have the same z-axis scale. This allows for a true comparison to be made between the raw vibration data plots for the different gear conditions. This is also true for the time averaged vibration signature. Note, however, that the z-axis scale for the time averaged vibration signature is different to the z-axis scale for the raw data. This adjustment is needed to allow for the difference in the power from these signals. To avoid repetition the z-axis scale was included only in the top plot.

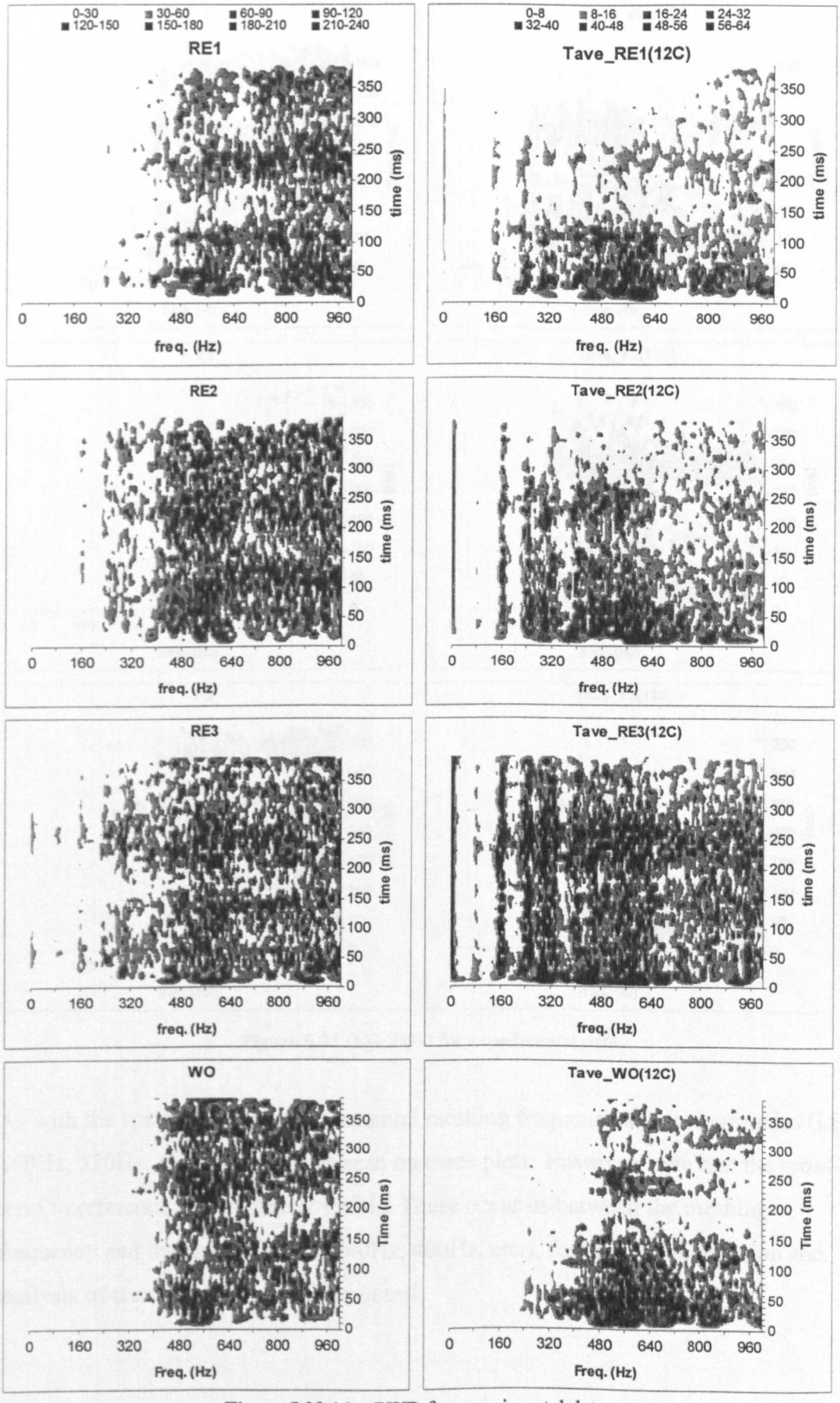


Figure 5.20 (a) – PWD for experimental data

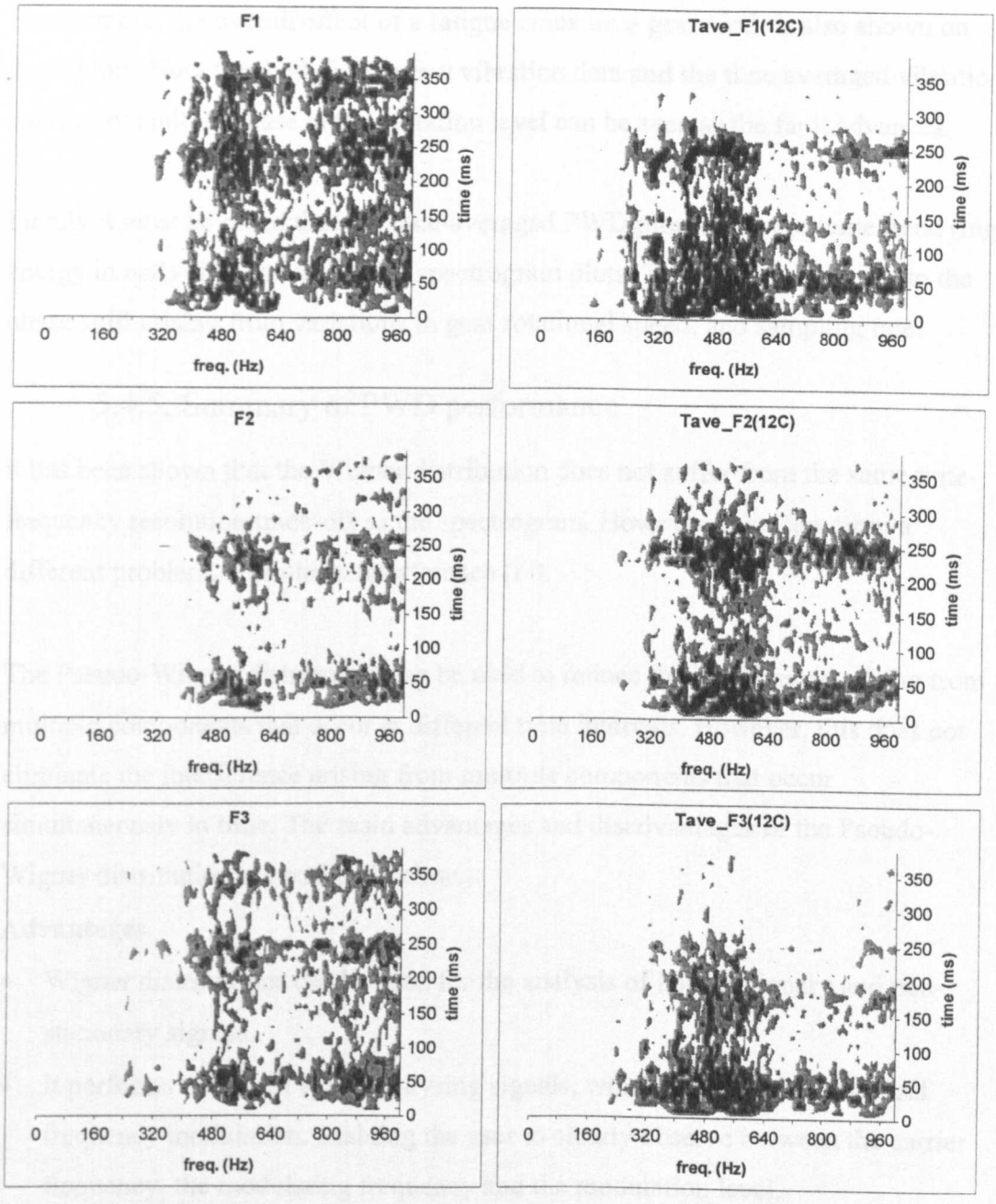


Figure 5.21 (b)– PWD for experimental data

As with the spectrogram, the fundamental meshing frequency and its harmonics (i.e. 160Hz, 320Hz, etc.) can easily be seen on these plots. However, note that the cross-term interference is also clearly visible. These occur in-between the meshing frequency and its harmonics (i.e. 240Hz, 400Hz, etc.), forcing the observation and analysis of these plots to be a difficult task.

Furthermore, the overall effect of a fatigue crack on a gear tooth is also shown on these plots. Note that both for the raw vibration data and the time averaged vibration data, an overall decrease in the vibration level can be seen as the fault advances.

Finally it must be noted that the time-averaged PWD plots show the same trend (high energy in early time values) as the spectrogram plots. This can be attributed to the phase shift arising from variations in gear rotational speed, and sampling rate.

5.4.5. Summary of PWD performance

It has been shown that the Wigner distribution does not suffer from the same time-frequency resolution trade-off as the spectrogram. However, it suffers from a different problem, cross-term interference [14].

The Pseudo-Wigner distribution can be used to reduce the interference arising from multiple components that occur at different time intervals. However, this does not eliminate the interference arising from multiple components that occur simultaneously in time. The main advantages and disadvantages of the Pseudo-Wigner distribution are outlined below.

Advantages

- Wigner distributions can be used for the analysis of both stationary and non-stationary signals.
- It performs very well when analysing signals, which contain amplitude and frequency modulation, enabling the user to clearly observe between the carrier frequency, the modulating frequency and the modulation level.
- Wigner distributions can filter out the white noise or random interference in the signal to be processed. This filtering can not only be done in the frequency domain but also in the time domain.

Disadvantages

- The WD and PWD are not very effective at processing signals with multiple components occurring simultaneously. The cross term interference that is generated in these cases can be very intense. This leads to complex time-frequency plots, which might lead to misinterpretation.

Chapter 6

Condition monitoring systems based on wavelets.

In the last decade, wavelets have received attention from a number of researchers from a wide range of fields. Wavelet analysis is a very versatile mathematical technique, which breaks down a signal into its constituent parts. Therefore, it can be used as a basis for representing functions with a finite number of components. Furthermore by usage of the wavelet mean square maps it can also be used as a technique for time-scale (frequency) analysis of time series (including vibration signatures).

Wavelets can be seen as a family of functions composed by translations and dilations of a single function (the wavelet, or mother wavelet). The set of products of all the dilated and translated wavelets with an arbitrary function is called the wavelet transform. Generally the wavelet transform is very effective at describing simultaneously both the local and large-scale features of a signal.

This chapter introduces the basic concepts behind wavelets. It also describes the two main classes of wavelets: Orthogonal and Harmonic wavelets.

6.1. *Basic Concepts*

The concepts shown here are valid for any wavelet decomposition, it forms the basis of the wavelet algorithm. The main idea behind the wavelet transform is to decompose a signal *x(t)* into its wavelet components. These components are called levels and are numbered from -1 upwards. The addition of these wavelet levels leads back to the original signal.

For orthogonal wavelets the number of levels in the wavelet decomposition of a signal is directly related to the number of points on the signal being decomposed.

The actual shape of the decomposed components (levels) depends on the signal under analysis and on the analysing wavelet. The latter can be a family of functions, which are the translation and dilation of a unique-valued function ψ(t). This can be defined as:

$$WT(t,s) = \int_{-\infty}^{\infty} x(\tau)\sqrt{s}\psi\left(s(\tau - t)\right)\mathrm{d}\tau . \qquad \textbf{Eq. 6.1}$$

Where $\psi(t)$ is the wavelet, t is the time and s is a scaling factor. From this it follows that a wavelet family can be generated by:

$$\left(\sqrt{s}\psi\left(s(\tau - t)\right)\right) \qquad \text{for} \quad (t,s) \in \Re . \qquad \textbf{Eq. 6.2}$$

And the reconstruction of the signal $x(t)$ can be obtained from:

$$x(t) = \sum_{t}\sum_{s} WT(t,s)\sqrt{s}\psi\left(s(\tau - t)\right). \qquad \textbf{Eq. 6.3}$$

6.2. *Orthogonal wavelets*

In this section the well-known orthogonal wavelets are introduced. These include: the Haar, the Daubechies Series and the Harmonic wavelets as defined in [70]. These wavelets share one main characteristic, the number of levels (L) resulting from the wavelet transform is related to the number of points in the data series to be transformed, This relation is described by: $L = n + 1$. Where L is the number of levels and 2^n is the number of elements in the time series.

The compact nature of the orthogonal (Haar, Daubechies and Harmonic wavelets) leads to many advantages in signal representation and reconstruction. Compact wavelets can decompose a signal into a minimum number of components (levels) very efficiently and with no redundancy. It is widely used as a means of signal representation and compression. Unfortunately, its compactness also leads to problems in time-frequency representations. Although the wavelet transform can be used to identify both the local and large-scale features of a signal, it can sometimes miss out features that happen to lie between the scales' (level) frequencies. Another handicap with orthogonal wavelets is the time translation variant [73]. Often, the same transients at different time may show up as a different pattern on the wavelet map.

6.3. *Theoretical background*

A comprehensive description of the theoretical background and properties of the orthogonal wavelets (Daubechies and Harmonic) can be found in [70,76,77] and only a short summary is repeated here.

Orthogonal wavelets arise from a very special case of the wavelet family. This special case arises when the condition described below is met:

$$\left(\sqrt{s}\psi\left(s\left(\tau-t\right)\right)\right) = \sqrt{2^j}\,y\left(2^j\left(t-2^{-j}n\right)\right) \qquad \text{for} \quad (j,n)\in Z\,. \quad \textbf{Eq. 6.4}$$

Hence the orthogonal wavelet can be defined as:

$$WT_o(t,s) = \int_{-\infty}^{\infty} x(t)\sqrt{2^j}\,\psi\left(2^j t - kT\right)\mathrm{d}t\,. \qquad \textbf{Eq. 6.5}$$

From which a signal can be reconstructed by:

$$x(t) = w_o + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} w_{2^j+k}\,\psi\left(2^j t - kT\right). \qquad \textbf{Eq. 6.6}$$

where:

$$w_o = \int_0^T x(t)\,\mathrm{d}t \quad \text{and} \quad w_{2^j+k} = \int_0^T x(t)\psi\left(2^j t - kT\right)\mathrm{d}t\,. \qquad \textbf{Eq. 6.7}$$

The results from this transform can be plotted as single vectors, (one vector for each level) or it can be plotted in the form of mean squares map. The latter can be easily obtained according to the algorithm presented in [70].

6.3.1. Illustrative Example

The concepts just described can be illustrated by the example below. It shows how the wavelet decomposition works by using the D4 wavelet (Daubechies Series with 4 coefficients) to analyse a test signal, shown in Figure 6.1. The frequency of the triangular wave is 50Hz, and it contains 2^7 (i.e. 128) points, sampled at 1000Hz. The wavelet transform of this signal is shown on the lower plot in Figure 6.1. This allows for the test signal to be decomposed into 8 levels. These are indicated by the vertical lines on wavelet plot of the test signal.

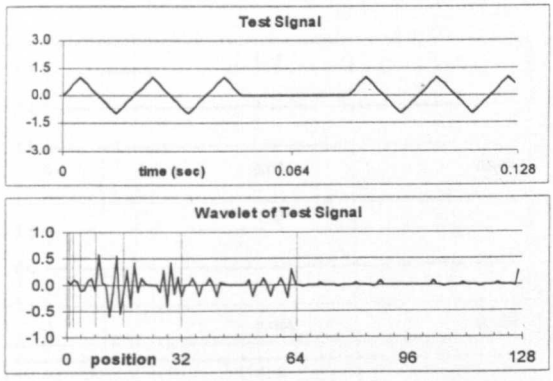


Figure 6.1 – Test Signal and its wavelet

The wavelet transform of the test signal is also a series with 2^7 samples. This series stores the signal information with respect to the different wavelet levels as shown in Table 6.1. Also the wavelet levels can easily be converted into frequency using the relation $2^n/T$ - where n is the wavelet level, and T is the length of time series. For the test signal described above T=0.128 seconds. In Table 6.1, in the last column the numbers in brackets give the equivalent frequencies (in Hz) for the wavelet levels of the test signal.

Table 6.1 – Identification of wavelet transform levels.

Level	Number of Wavelets	Elements position on transform	Frequency at Given Level (Hz)
-1	0	1	0/T (0)
0	1	2	1/T (8)
1	2	3 to 4	2/T (16)
2	4	5 to 8	4/T (31)
3	8	9 to 16	8/T (62)
4	16	17 to32	16/T (125)
5	32	33 to 64	32/T (250)
6	64	65 to 128	64/T (500)
n	2^n	$1+2^n$ to 2^{n+1}	2^n/T

After the wavelet analysis the following information related to the signal reconstructed levels can be obtained. The left column contains the eight-wavelet levels (from –1 to 6) of the test signal. The right column shows the progressive reconstruction of the signal, from its wavelet levels. This is obtained by adding the wavelet levels together. The titles on each chart in the left column refer to the specific wavelet level. In the right column, successive levels are added, leading back to the original signal in the lowest chart.

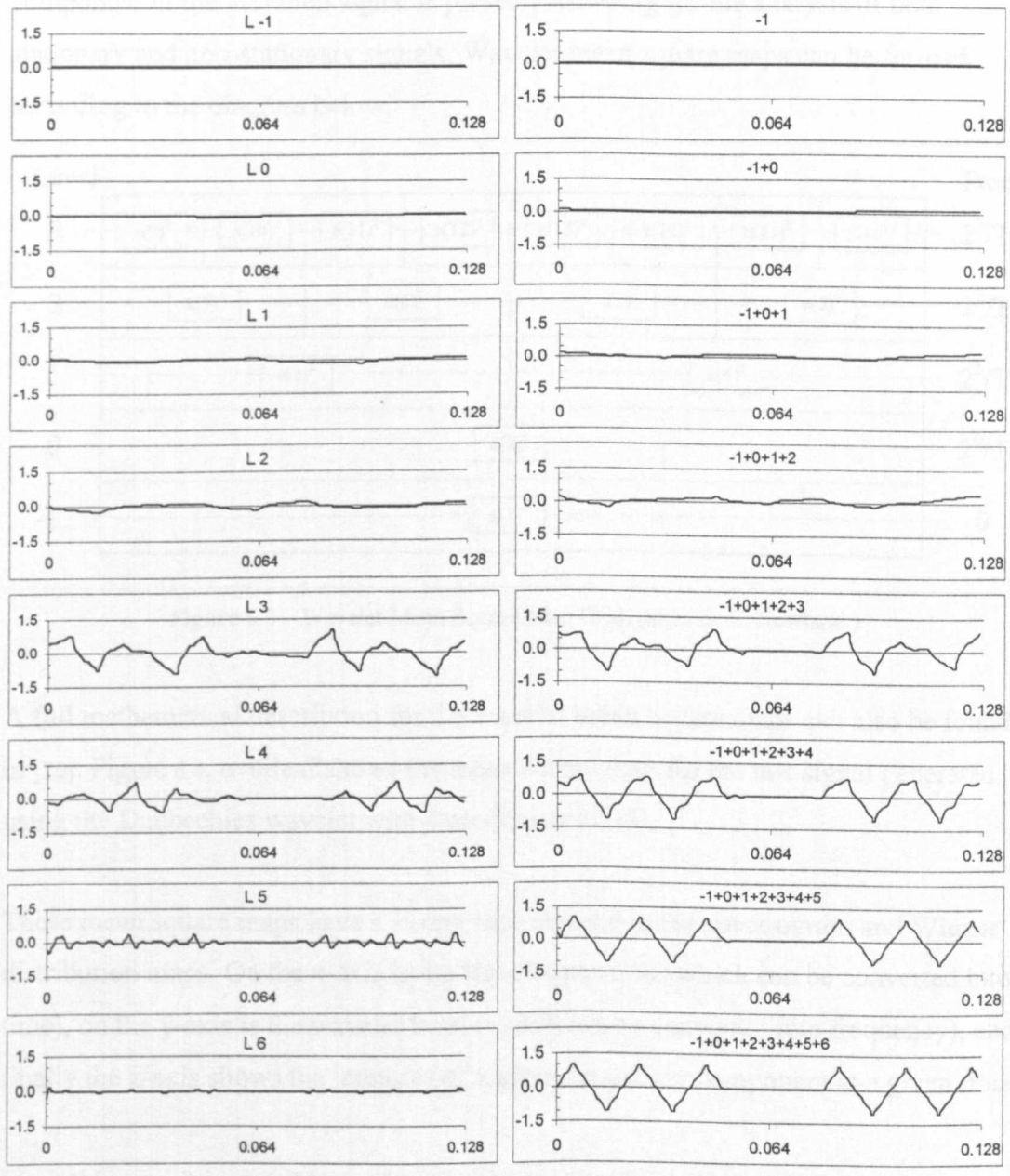


Figure 6.2 – Wavelet Analysis and Reconstruction of test signal.

A full mathematical description of the properties of the wavelet coefficients can be found in [70].

6.4. *Wavelet Mean Square Maps*

The wavelet levels coefficients can also be displayed in the form of a time-scale map (or time-frequency if the conversion described above is performed). This is used in vibration condition monitoring as a means to assess whether each frequency component in the vibration signal is present. Allowing for the analysis of both stationary and non-stationary signals. Wavelet mean square maps can be formed according to the diagram below.

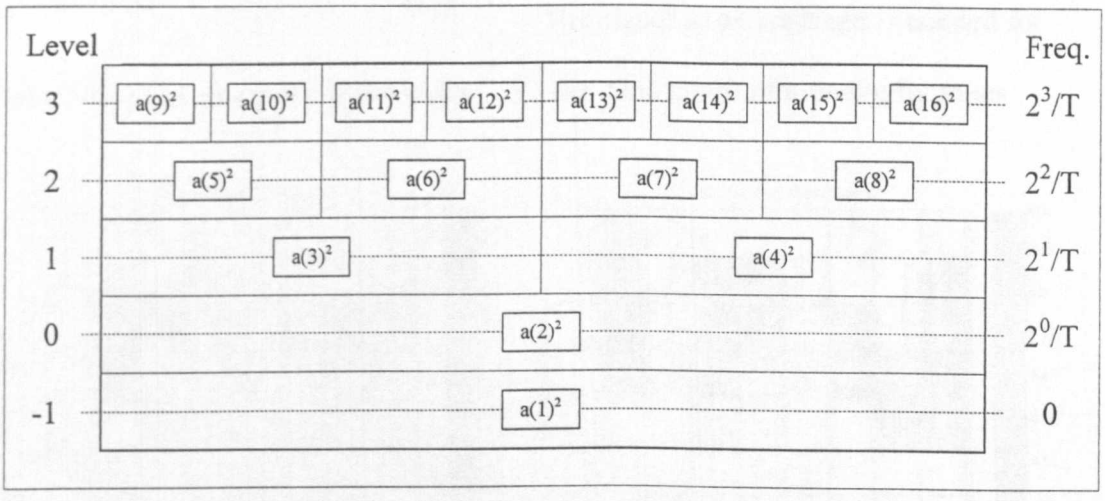


Figure 6.3 – Wavelet Mean Square Map Grid. (after D. E. Newland.)

A full mathematical description for the wavelet mean square maps can also be found in [70]. Figure 6.4, overleaf shows the mean square map for the test signal generated using the Daubechies wavelet with 4 coefficients (D4).

These mean square maps have a strong resemblance to the spectrogram and Wigner distribution maps. On the x-axis is the wavelet position (which can be converted into time), on the y-axis is the wavelet level (which can be converted into frequency), and finally the z-axis shows the intensity of a given frequency component at a given time.

From this map it is possible to observe that the signal is composed of a dominant frequency around 62Hz. Also between $t \approx 50$ms and $t \approx 0.80$ms seconds the signal is

zero (as no frequency component is present). This agrees well with the properties of the test signal except that the frequency of the triangular is f=50Hz. The poor frequency resolution on these maps is related to the small number of samples in the input signal. This affects the number of levels onto which the signal is decomposed.

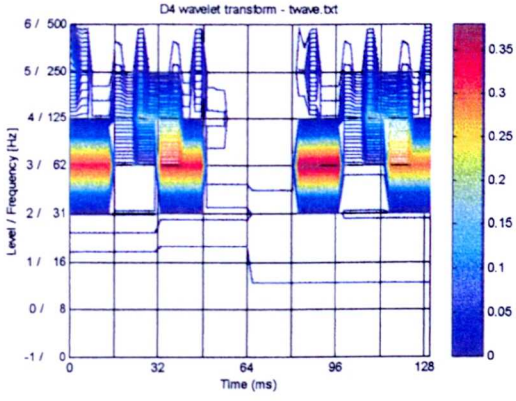


Figure 6.4 – Mean square map for test signal

Now it must also be observed that the time-frequency resolution of these maps is dependent on the shape of the mother wavelet used to decompose the signal. This is a handicap of this technique as it suggests that previous knowledge about the signal to be analysed is needed for the production of meaningful maps.

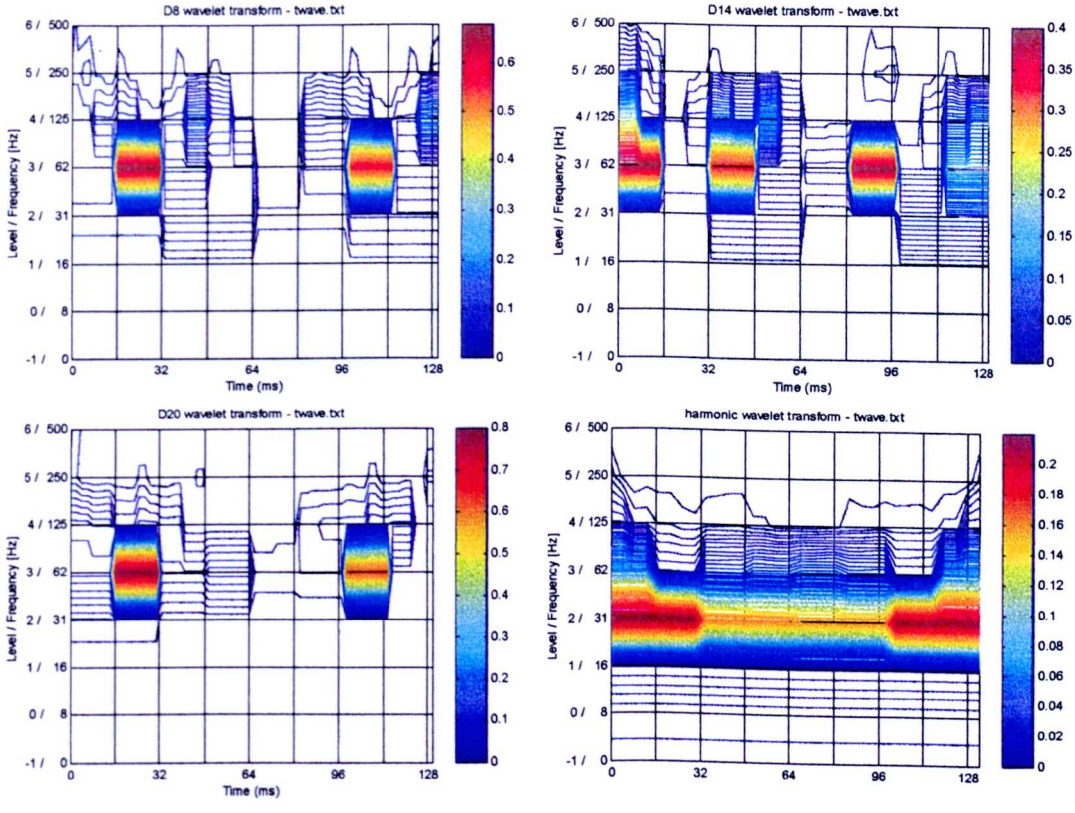


Figure 6.5 – Time-Frequency resolution of different wavelet families

6.5. *Wavelet Analysis*

To allow for a direct comparison between the wavelet decomposition technique and the previous time-frequency analysis techniques, only wavelet mean square maps and wavelet plots will be used throughout this chapter. Further analysis with wavelet level reconstruction could be performed for condition monitoring applications, however this will not be included in the scope of this research.

6.5.1. Numerical Example

Before analysing the experimental data a preliminary analysis on the numerically simulated data generated as described in chapter 3 was performed. A wavelet transformation routine written in MATLAB was used to process the two sets of data (i.e. good gear and the faulty gear). The results can be seen below. On these wavelet mean square maps the x-axis, which contains information about the wavelet position, has been converted to time. Also, the y-axis shows the level (and its respective frequency). Note that the time series represents one full gear revolution. Therefore 360° is equivalent to 1.024 seconds.

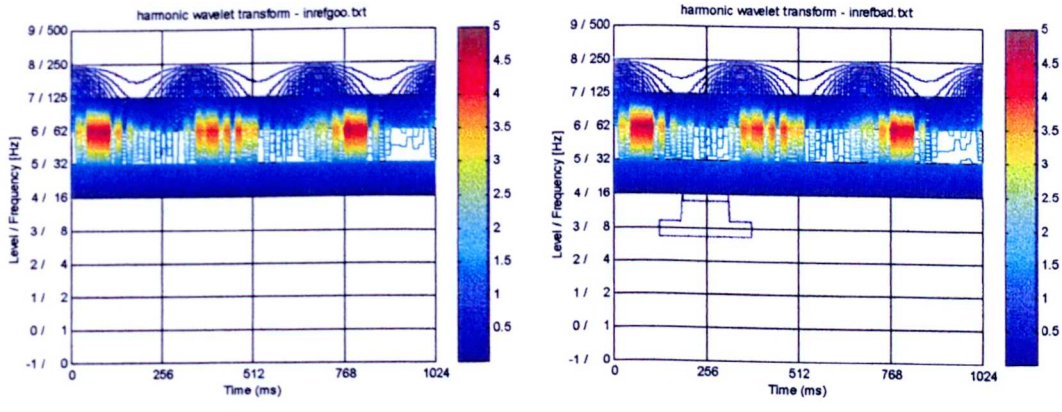


Figure 6.6 – Harmonic wavelet MSP for numerical data

These plots clearly show the simulated fault (occurring at t=256ms). The graph on the left shows the harmonic wavelet mean square map for the 'good' condition gear signal, and the left figure shows the same for map the 'faulty' condition signal. A brief analysis of the above maps shows its main properties:

- The blue and cyan contours ranging from 16Hz to 250Hz encompass the pure sine wave at 40Hz, the amplitude modulated sine wave (at 70Hz), the frequency

modulated sine wave (at 100Hz) and the amplitude and frequency modulated sine wave (at 130Hz). Note how the frequency modulation affected the upper shape of the contours, as compared to the straight contours found in the lower section of the map (due to the pure sine wave).

- The contours at frequency 8 Hz and 256ms (i.e. angular position 90°) on the 'faulty' plot corresponds to the Gaussian pulse (i.e. simulated fault). These do not appear in the good gear condition plot. These contours are at very low levels, indicating that this simulated fault is on the verge of not being identified by the wavelet map. This implies that the number of contour levels and their location plays an important role in the generation of the mean square maps.
- Finally, the three red peaks at frequency 62Hz correspond to the amplitude-modulated sine wave, with a carrier frequency (f_c) of 70Hz and an amplitude modulation frequency (f_a) of 3 Hz.

As it can be seen wavelets mean square map is an effective signal processing method for the time-frequency analysis of dynamic systems. It has been shown that it can be used to detect and display the time-frequency information of a wide range of signals. Unfortunately the frequency resolution of orthogonal wavelets are very dependent on the length of the signal to be analysed.

This suggests that non-orthogonal wavelets might have a better performance in the time-frequency analysis of vibration signals, than orthogonal wavelets. However these were not included in the scope of this research.

Furthermore, the previous tests suggests that in the same manner that spectrograms rely heavily on the correct choice of the window width, wavelet mean square maps rely heavily on the correct choice of the wavelet family. This can only be attained either experimentally, or if previous knowledge about the signal is available. The figures overleaf show how the wavelet mean square of the numerically generated 'good' data is affected by the wavelet family choice. Note that the z-axis (colours) have been adjusted for an optimal display for each plot.

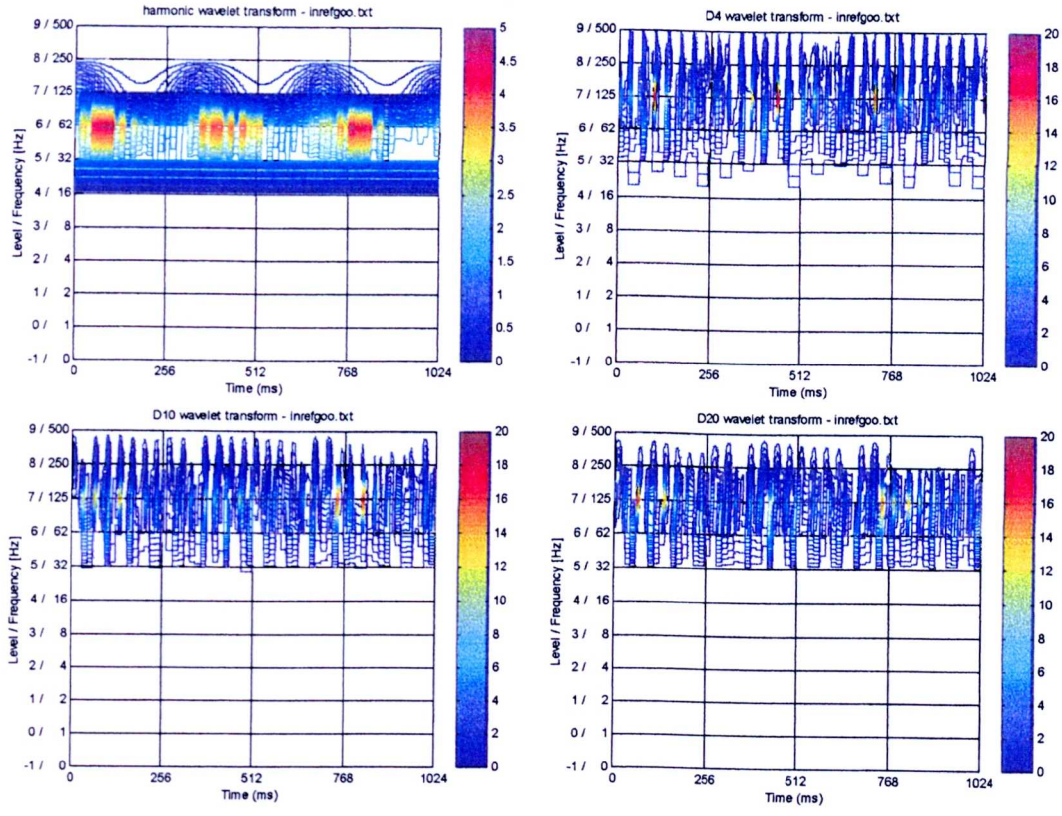


Figure 6.7 – Effect of mother wavelet choice on MSP of test signal

As it can be seen the harmonic wavelet maps contains more high level continuous regions. This simplifies its appearance, facilitating its visual inspection. This agrees with published results [72]. The other wavelet families also give good indication of the frequency components of the signal, but their maps are not so uniform.

As can be seen, for mean square map representations harmonic wavelets show better performance than the Daubechies series, as it produces more uniform maps. Still the latter wavelet family series is extremely useful for applications such as: acoustic analysis and signal compression. In fact for signal compression applications this series outperforms the harmonic wavelet in simplicity, as its wavelet coefficients are always real, while the harmonic wavelet coefficients are usually complex.

6.5.2. Experimental Example

Here the experimental data already (see chapter 3) was processed with the same harmonic wavelet routine used previously. This choice was based on the performance of the harmonic wavelet compared to the Daubechies wavelet.

The vibration signatures for all 7 gear conditions (i.e. 1 brand new, 2 reference, 1 worn out and 3 faulty gear condition) were analysed. The results can be seen below, note that the plots on the left relate to the raw vibration signature, and the plots on the right relate to the time averaged (over six blocks of 4 cycles) vibration data. All the plots contain two full revolutions (0.4s) of the gear under analysis. Again the x-axis contains the time information while the y-axis represents frequency.

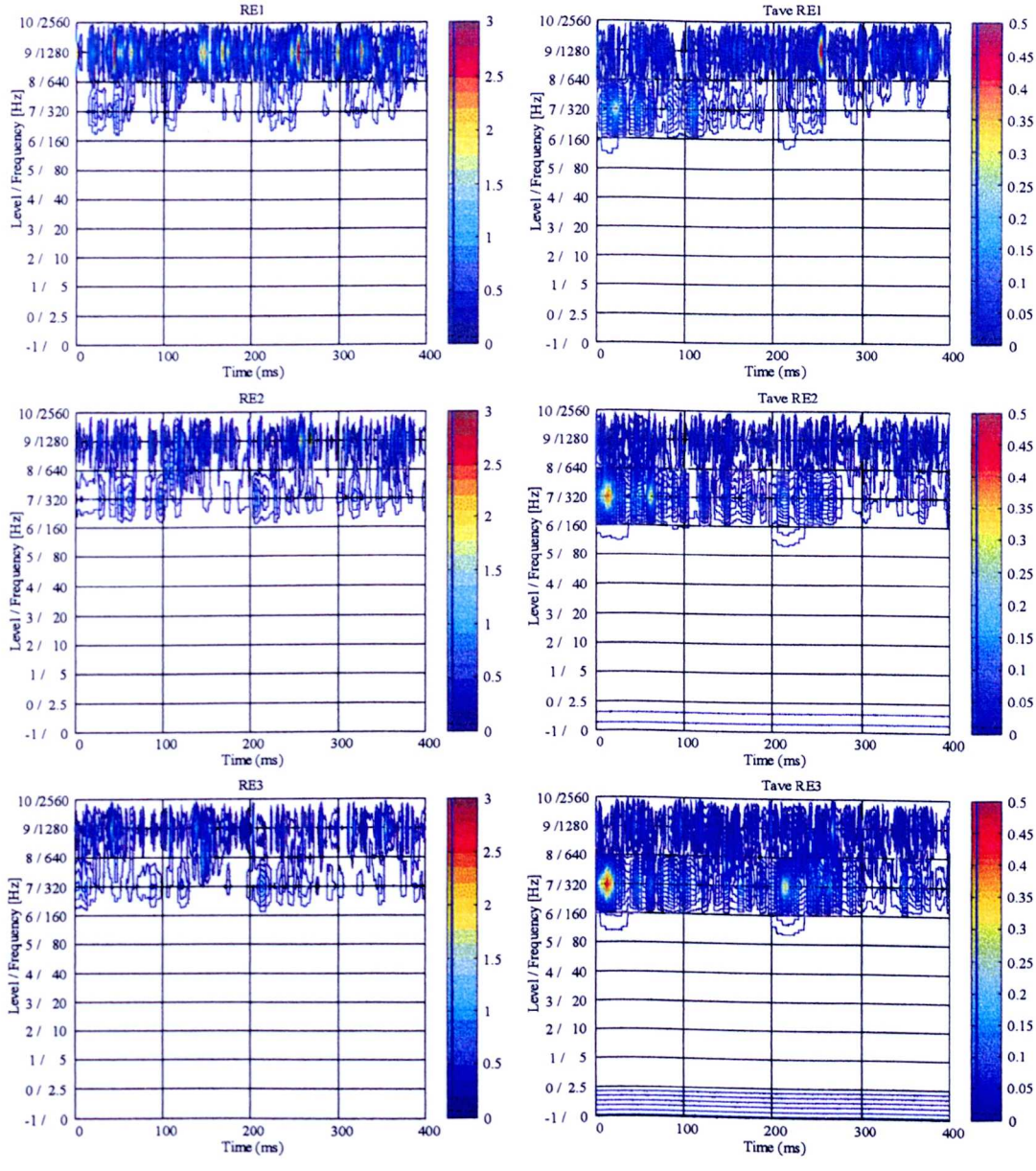


Figure 6.8 (a) – Wavelet mean square map for experimental vibration signatures

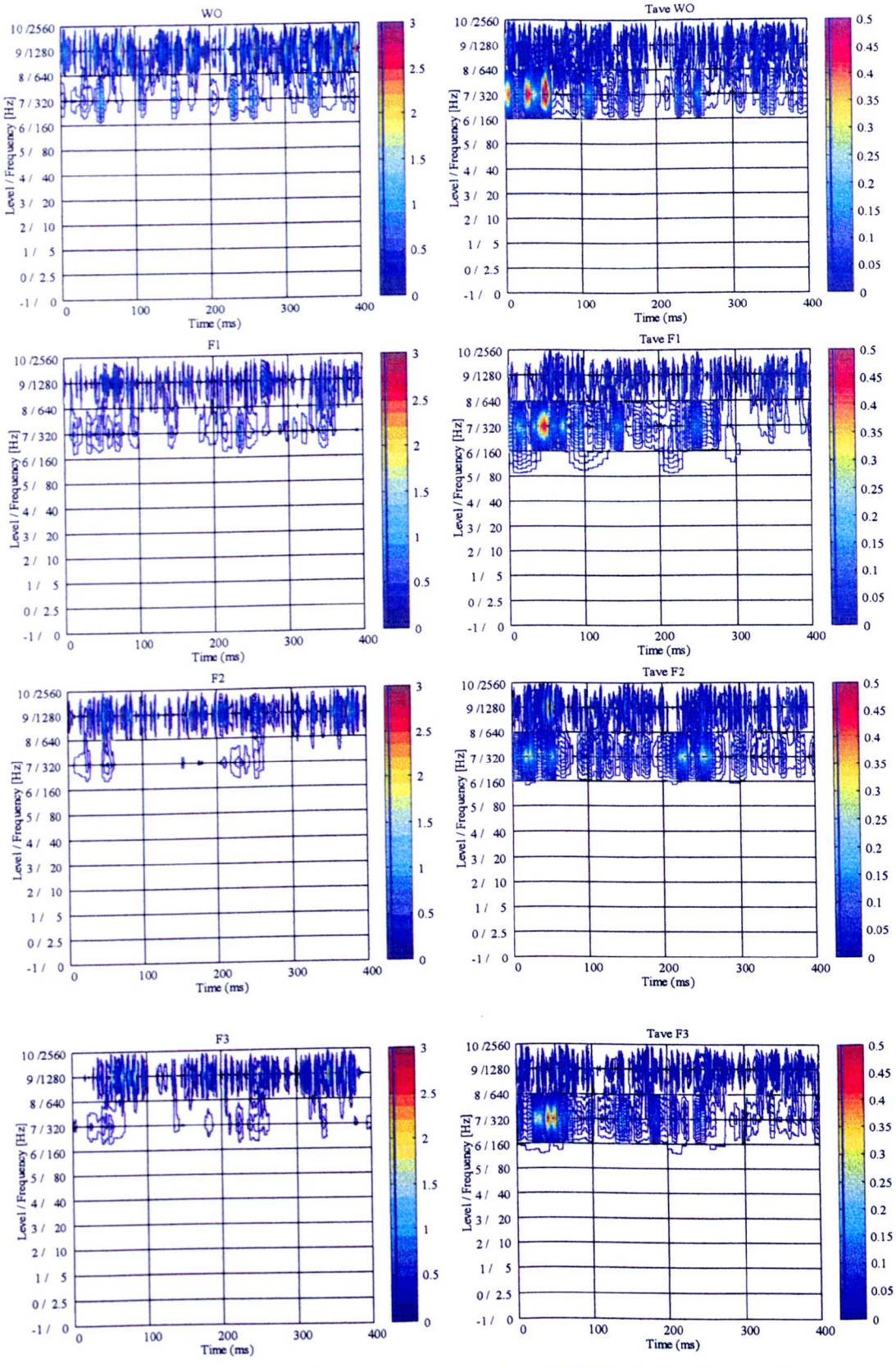


Figure 6.8 (b) – Wavelet mean square map for experimental vibration signatures

As it can be seem in the above plots, noise has a very strong effect on the maps for the raw vibration signatures. The strong peaks at the very high frequencies show this. The lower frequency appearing in the plots is 160Hz. This is the gear meshing frequency. These components are present throughout the time domain. This is most noticeable for the normal gear condition (i.e. plots RE2, RE3), and not so noticeable for the faulty condition gears (F1, F2 and F3). Also, the brand new gear shows the strongest high frequency components. This behaviour was also observed by the spectrogram analysis of these signals.

These plots also show the fatigue crack's damping effect on the overall vibration of the gear. This was also seen on the spectrogram analysis and is best observed in plots F2 and F3. Note, however that here this behaviour does not show so much intensity as seen in the spectrogram. These observations are stronger in the plots for the time-averaged data.

A summary of the analysis of the above plots shows:

- The gear meshing frequencies and its harmonics dominate the mean square maps.
- Fatigue cracks are characterised by a damping effect on the vibration of the gear at the crack position. This can be clearly seen in plots F2 and F3. This method produced useful information related to fault location.

Furthermore, as expected the plots for the time-averaged data present a lower noise level for higher frequencies. However, as with the previous methods, on the time averaged plots, the spectral power for early times (i.e. start of first revolution) is greater than for later times (i.e. end of second revolution). This behaviour has been discussed in chapter 5, and can be attributed to the inaccuracies on the rotational speed and the sampling rate of the data acquisition hardware.

6.5.3. Summary of wavelet mean square map performance

It has been shown that wavelet mean square map is an effective method for time-frequency decomposition, and can also be in industrial vibration condition monitoring and diagnostics. Although it can also be used to analyse signals with no previous knowledge, it is recommended that tests for different mother wavelets are to be carried out, as the mean square maps are directly related to the analysing mother wavelet. A brief summary of the advantages and disadvantages of this method follows.

Advantages

- The wavelet MSM of a signal can be seen as the distribution of the signal energy not only in the frequency domain (like a spectrum) but also in the time domain.
- It is very effective in analysing stationary and non-stationary signals; and in identifying both small and large scale features in signals.
- This method is much faster than spectrograms, and produces smaller maps indicating the time-frequency contents of a signal. This is of utmost importance in automated condition monitoring and diagnostics.

Disadvantages

- The frequency resolution of the MSM generated by orthogonal wavelets is dependent on the number of samples on the vibration signature. So for high frequency resolution a very long data series is needed.
- A heuristic test to evaluate the performance of the different mother wavelets is needed before choosing the wavelet family.

Chapter 7

Condition monitoring systems based on Non-linear methods.

The condition monitoring techniques, analysed so far (i.e. statistical measures, FFT, spectrogram, cepstrum, Wigner distribution and wavelet) are widely used for linear or quasi-linear systems analysis. Although it is possible to represent a non-linear system by dividing it into smaller linear (or quasi-linear) sections which are then analysed with the already well established theory for linear systems; it has been recognised that non-linear processes can only be analysed as a whole by non-linear methods. Hence the need for development and testing of non-linear condition monitoring techniques is real.

There are a number of non-linear system identification techniques that can be extended to vibration condition monitoring applications [113]. Some examples are:

- artificial neural networks, which have been widely used as pattern recognition tools to identify and differentiate vibration signatures from rotating devices with different and or multiple faults. This method has been very effective as a post-processing tool for vibration signatures, however it has not been useful for the analysis of the raw vibration signature [75];

- functional series methods (including Wiener and Volterra series), which allow the user to select the non-linearity order to be included in the system model [114], and;
- heuristic methods such as GMDH (Group Method Data Handling), parameter estimation methods and quadratic detectors [87, 116].

Some of these techniques have been suggested and tried in other areas of research, such as communication systems analysis and biological systems modelling. However out of these techniques only artificial neural networks (as a post-processing tool) have been applied to industrial condition monitoring applications. Leading to the conclusion that its performance is very dependent on the pre-processing technique used to extract the relevant information from the time series fed to the neural network.

This chapter introduces and performs a critical analysis of one specific non-linear technique, namely: the Volterra Series. The reason behind the choice of this method lies in the fact that, Volterra series has already been used with some success in the identification of non-linear biological signals, and this technique can be applied directly to the raw vibration data without additional pre or post-processing. This will be shown in the next section.

7.1. Volterra Series: Current Applications

In this section a short description of the fields, in which Volterra series have been applied, including remarks about the specific applications, is given. Note that this technique has not yet been introduced to vibration analysis.

7.1.1. Non-linear system modelling and identification

The identification and analysis of non-linear system plays a vital role in control theory. Most systems encountered in practice are non-linear. By restricting the operating range, these systems can be represented by a linear model. Unfortunately, the fact still remains; a non-linear system can only be adequately represented by a non-linear model.

From the above premise Koh applied Volterra filters[1] to model and predict the dynamic behaviour of moored vessels due to irregular sea waves. In [92] shows the use of Volterra filters for studying the non-linear drift oscillations of moored vessels when subjected to random sea waves has been shown.

Koh's results clearly show that the 2nd order Volterra filter is effective in modelling the low frequency drift oscillation of barges, also predicting the barge response. The latter result is very importance for the control and stabilisation of these systems.

7.1.2. Non-linear system analysis, and pattern detection and recognition

The Volterra series approach has been extensively used in the analysis of non-linear biological and physiological systems [103]. It has been found that its main advantage lies in its ability to model a system with no previous knowledge of the its structure. Also, Volterra kernel analysis lead to inferences of the model structure, as one is able to ascertain how past inputs might affect the future behaviour of the system.

In [115] the problem of detecting buried pipes is presented. Initially, an artificial neural network is used to discriminate between ground surface and actual pipe reflection from the return of a radar signal. Later the structure of the trained neural network (i.e. a network that correctly maps the site being surveyed) is compared to that of a Volterra series. It is shown that both the neural network and the Volterra series present the same structure. This suggests that these two systems have strong similarities and can share the same transfer function.

Bissessur showed that, by obtaining a mathematical formulation of the weights learnt by the artificial neural network and its nodal functions, it is possible to extract a set of Volterra kernels, formulating the Volterra series representation of a system, in this particular case, the pipe detection problem. These results are supported by those given in [116], which discusses the similarities and differences between neural networks and Volterra series.

[1] Volterra filters are non-linear filters with the Volterra series structure.

7.1.3. Discussion

The Volterra series is a well-established non-linear signal modelling technique, and it has been widely used in modelling non-linear biological and physiological systems. However, for no apparent reason, it has not yet been applied to the vibration analysis of non-linear mechanical systems.

The main advantage of the Volterra series approach is its ability to model the mathematical relation between input and output of a system, and this modelling does not require any previous knowledge of the system structure. Also, the analyses of the Volterra kernels enable one to make inferences about the underlying non-linear system. This knowledge can be used for fault identification in condition monitoring applications.

7.2. Volterra Series: Theoretical Background and Example.

The Volterra series emerged from studies of non-linear functionals of the form y(t)=F[x(t');t'<=t], and in [117] Volterra introduced the representation:

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2$$

$$+\ldots+\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} h_n(\tau_1,\tau_2,\ldots,\tau_n)x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)d\tau_1 d\tau_2\ldots d\tau_n \qquad \textbf{Eq. 7.1}$$

for n=1,2,3,...

Where: $x(t)$ and $y(t)$ are, respectively, the system's input and output at a given time t, and $h_n(\tau_1, \tau_2, \ldots, \tau_n)$ is the n^{th} order Volterra kernel.

The first application of the above functional to the study of non-linear systems was performed by Wiener [84], in characterising a system as a mapping between its input and output spaces. Wiener introduced another way of expressing the above equation:

$$y(t) = H_1[x(t)] + H_2[x(t)] + \ldots + H_n[x(t)] \qquad \textbf{Eq. 7.2}$$

where:

$$H_n[x(t)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_2) x(t-\tau_1) \ldots x(t-\tau_n) d\tau_1 \ldots d\tau_n \qquad \text{Eq. 7.3}$$

In this representation the H_n is called the nth order Volterra operator. Also, since most mechanical systems to be analysed are causal systems, it is common to replace the limits of integration shown above by 0 and ∞.

The solution of the identification problems based on the Volterra series requires the calculation of the Volterra kernel [118]. These can be calculated by different methods.

The method used in this research is known as the Lee-Schetzen (crosscorrelation technique) approach. This method is extensively explained in [96,102, 118], and only a brief explanation will be included here. The cross-correlation technique was proposed in 1965 by Schetzen [118], and relies on the assumption that the nth Volterra kernel is directly related to the nth order impulse response of the system. Also Schetzen observed that the kernel $h_1(\tau)$ is the system unit response, that $h_2(\tau_1, \tau_2)$ is the two dimensional impulse response of a second order system and that the same principle could be extended to higher order kernels.

In developing this technique Lee and Schetzen showed that the set of kernels h_i could be evaluated by using crosscorrelation techniques. Furthermore, they also showed that for a system S, with a white-Gaussian noise input $x(t)$ and a response $y(t)$, the kernels can then be calculated according to the schematic diagram in Figure 7.1 and Figure 7.2.

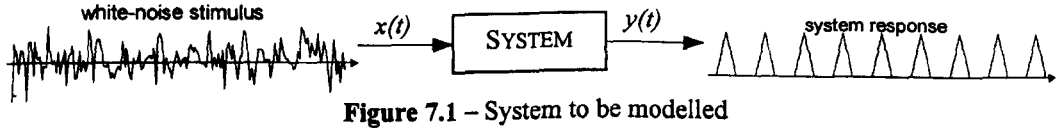


Figure 7.1 – System to be modelled

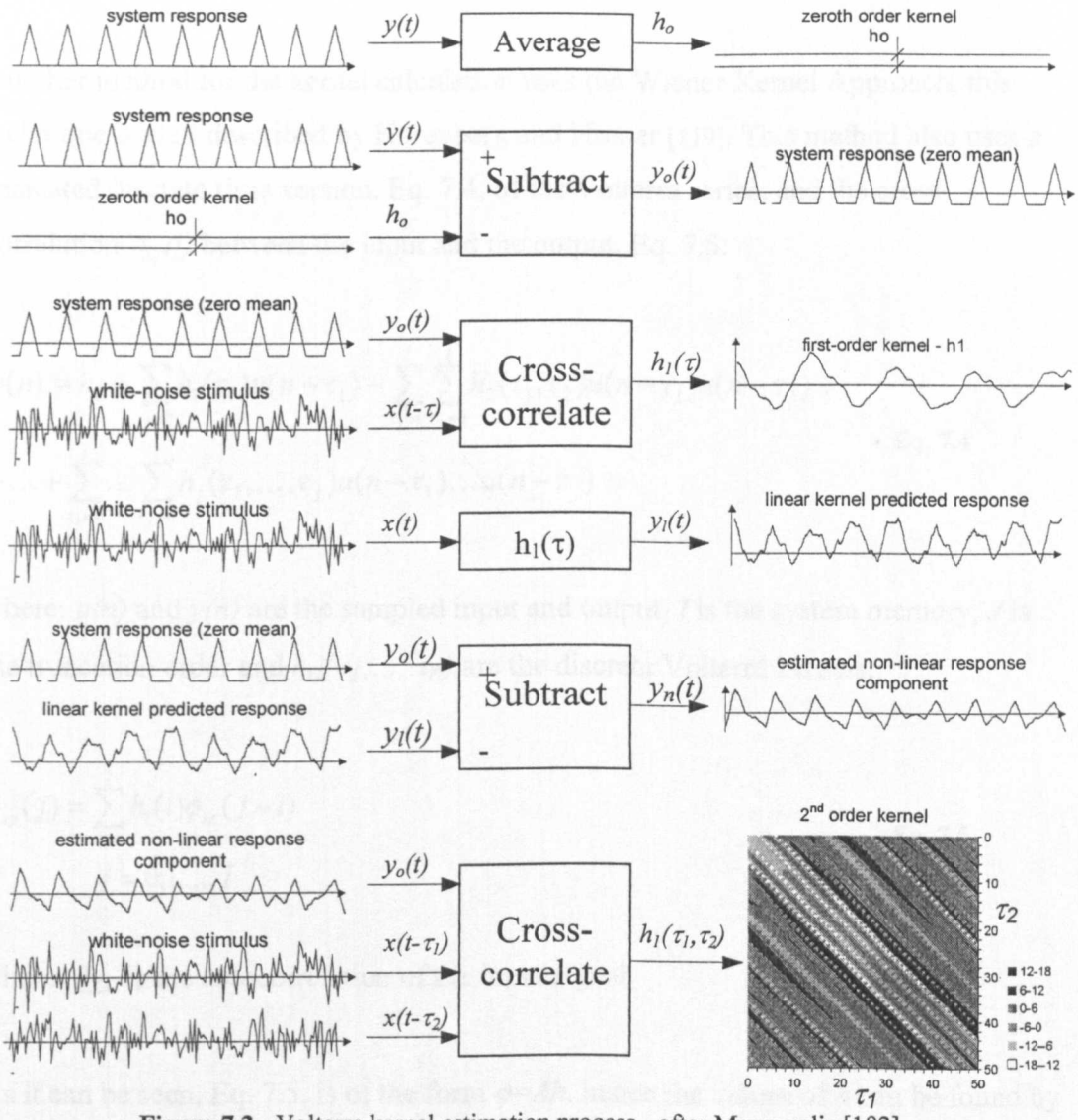


Figure 7.2 - Volterra kernel estimation process - after Marmarelis [102].

Note that in the above figure, the intermediate graphs shown are in fact what is obtained from the Volterra analysis of the stated non-linear system (Figure 7.1).

This approach to kernel estimation was selected due to its simplicity and also because it has several advantages over the Wiener approach. Firstly it estimates the kernels directly, giving an insight into the structure of the system under study. Secondly, this method is simpler as it does not involve Laguerre and Hermite transformations as required by the Wiener approach. Finally, this method requires

fewer computations than the Wiener approach, reducing the computational expense of this technique.

Another method for the kernel calculation uses the Wiener Kernel Approach, this technique is well described by Korenberg and Hunter [119]. This method also uses a truncated discrete time version, Eq. 7.4, of the Volterra series, and the cross-correlation $\phi_{xy}(j)$ between the input and the output, Eq. 7.5.

$$y(n) = h_0 + \sum_{\tau_1=0}^{I} h_1(\tau_1)u(n-\tau_1) + \sum_{\tau_1=0}^{I}\sum_{\tau_2=0}^{I} h_2(\tau_1,\tau_2)u(n-\tau_1)u(n-\tau_2) + \\ + \ldots + \sum_{\tau_1=0}^{I}\ldots\sum_{\tau_j=0}^{I} h_j(\tau_1,\ldots,\tau_j)u(n-\tau_1)\ldots u(n-\tau_j) \qquad \text{Eq. 7.4}$$

where: $u(n)$ and $y(n)$ are the sampled input and output, I is the system memory, J is the truncation order and $h_n(\tau_1,\ldots,\tau_n)$ are the discrete Volterra kernels.

$$\phi_{xy}(j) = \sum_{i=0}^{I} h_1(i)\phi_{xx}(j-i) \qquad \text{Eq. 7.5}$$
$$j = 0,1,\ldots,I$$

where ϕ_{xx} is the autocorrelation of the input signal.

As it can be seen, Eq. 7.5, is of the form $\phi=Ah$, hence the values of h can be found by inverting matrix A (i.e. $h=A^{-1}\phi$). Higher order kernels can be calculated with the same process, however considering the n^{th} order cross-correlation between the input and the output.

This technique also shows disadvantages. When it is used for higher orders it gives the Wiener kernels instead of the Volterra kernels [115].

Finally, more recently Korenberg introduced yet another kernel estimation method: The Fast Orthogonal algorithm [99]. This could be an alternative method to the cross-correlation technique used in this research.

7.2.1. Meaning of the Volterra kernels (theoretical example)

From the theoretical description in Figure 7.2 it should be obvious that the zeroth order kernel is only an indication of the DC level (i.e. average) of the signal being analysed. The meaning of the first and higher order kernels is not so obvious to visualise. In fact these kernels indicate the pattern in which the past values of the stimulus affect the present value of the system response.

"The n^{th} order Volterra kernel is the pattern of interaction among n pieces of the stimulus past with regard to the effect that this interaction has upon the system response" [102].

Hence, it can be understood that the first order kernels indicate how a stimulus at a given time in the past affect the present value. The figure below shows the first order kernel superimposed on the input signal. Since these series have very different amplitudes two y-axis were introduced. The y-axis on the left indicate the amplitude of the first order kernel, the axis on right indicate the amplitude of the system response.

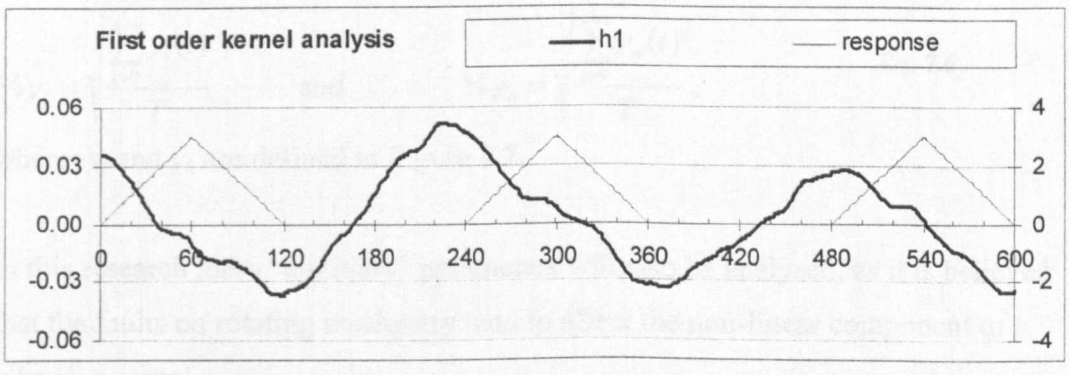


Figure 7.3 – Analysis of first order kernel

As it can be seen the first order kernel also has a cyclic (non-damped) behaviour with the same frequency of the system response. Also note that, from the analysis of the first order kernel it is obvious that the sample which has the greatest effect on the current sample lies at a lag of 240. While the samples at lag=120, have a strong negative effect. These lags were expected as the cyclic non-linearity has a period of 240 samples.

A similar interpretation can be used to analyse the second order kernel. However this is usually more complex, as it assumes that the present system response is dependent on two consecutive inputs occurring at respective lags. This means that a positive 2nd order kernel at position τ_1,τ_2 indicates that, the present input tends to be positive if a stimulus applied at τ_1 samples in the past is followed by a stimulus applied at τ_2 samples in the past. The 2nd order kernel will always be symmetrical about τ_1 and τ_2. Hence for practical purposes only half of the kernel matrix is plotted.

Finally the process of system identification using the Volterra kernel approach lies in the analysis of the shape of the first and second order kernels. This can usually be very complex. Other parameters can be used to allow for a comparison between different system responses. This includes the analysis of the first order kernel frequency (for under-damped cyclic systems), and the relative strength of linear and non-linear components in a given signal. Note that this relative strength, "$\%\, y_l$", "$\%\, y_n$" is calculated from the *rms* value of the estimated linear and non-linear signal components respectively. This is mathematically defined as:

$$\%y_l = \sqrt{\frac{\sum_{t=0}^{T} y_l(t)}{T}} \quad \text{and} \quad \%y_n = \sqrt{\frac{\sum_{t=0}^{T} y_n(t)}{T}}, \qquad \textbf{Eq. 7.6}$$

where: y_l and y_n are defined in Figure 7.2.

In this research these "auxiliary" parameters will also be analysed, as it is believed that the faults on rotating machinery tend to affect the non-linear component of a vibration signal.

7.3. Volterra kernel analysis

In this section the Volterra kernels for the numerical and experimental data (as described in chapter 3) are included. As already discussed previously, this analysis will be based on the visual observation of the first and second order kernels. Also the "auxiliary" measures will be used as a basic measure of comparing the kernels for the different time series.

7.3.1. Numerical Example

Before analysing the experimental data a preliminary analysis of the numerical data generated as described in chapter 3 was performed. A routine with the Lee-Schetzen kernel estimation algorithm written in Visual Basic was used to process the two sets of data (i.e. good gear and the faulty gear). The results can be seen below.

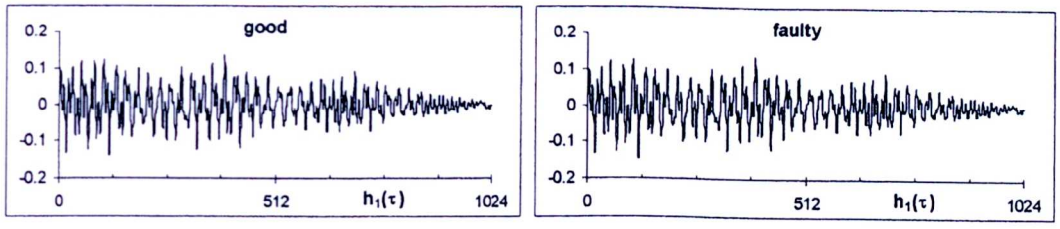


Figure 7.4 – First order kernel of numerical data

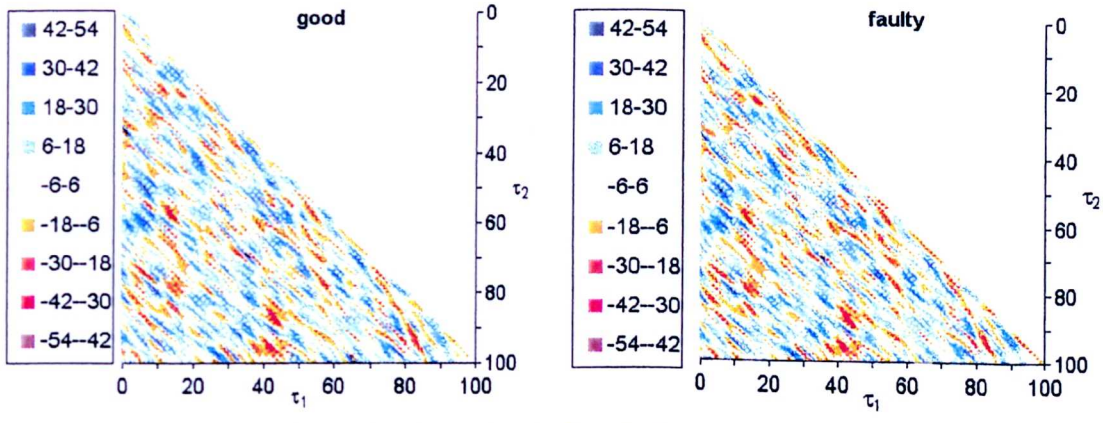


Figure 7.5 – Second order kernel of numerical data

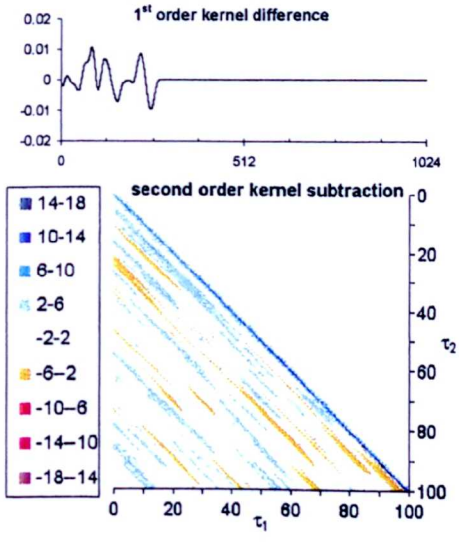


Figure 7.6 – Kernel subtraction

A simple visual inspection of these kernels does not allow for the faulty signal to be differentiated from the good signal. This is hardly surprising, since both signals are equal (except for the samples affected by the simulated fault – i.e. the Gaussian pulse). Therefore this method does not identify the simulated fault. Still it must be noted that these kernels are not equal. The figures on the left, show the subtraction of the first and second order kernels for the numerical data. As it can

be seen the two kernels are slightly different. Note how the scale of the subtracted signals is much smaller then the scales of the original kernels. Unfortunately this method of kernel comparison is not effective as most real vibration signatures will have embedded noise, therefore generating different minor variations to the kernel values.

Now the "auxiliary" kernel measures for the numerical data are presented below.

Table 7.1 – Volterra auxiliary parameters for numerical data

Auxiliary Parameters	**Time series**	
	Good	Faulty
h_1 *rms*	00.0309	00.0310
$\%\, y_l$	41.2	41.2
$\%\, y_n$	58.8	58.8

The main frequency component on the first order kernel is very close to the strongest frequency on the time series under analysis (i.e. 40Hz). Also the remaining parameters (h_1 *rms*, $\%y_l$ and $\%y_n$) of the two time series are also very similar. Therefore as it can be seen in this specific application the Volterra kernel analysis fails to identify the presence of the Gaussian pulse in the simulated time series.

7.3.2. Experimental example

Here the experimental data already described in chapter 3 was processed with the same kernel estimation routine used in the section 7.3.1. The vibration signatures for all 7 gear conditions (i.e. 1 brand new, 2 reference, 1 worn out and 3 faulty gear condition) were analysed. The results are shown below. As in the previous sections the charts on the left relate to the raw vibration data, and the charts on the right relate to the time averaged data. Note that these results were obtained using two full gear revolutions (2048 samples) as the system response. The Volterra model has a linear memory of 1024 samples (however only the first 200 first order kernels are displayed), and a second order memory of 100 samples.

These results are grouped in two blocks, the first block (Figure 7.7), show the first order kernels of the experimental data for all the gear conditions. The second block (Figure 7.8), shows the second order kernels for the same experimental data.

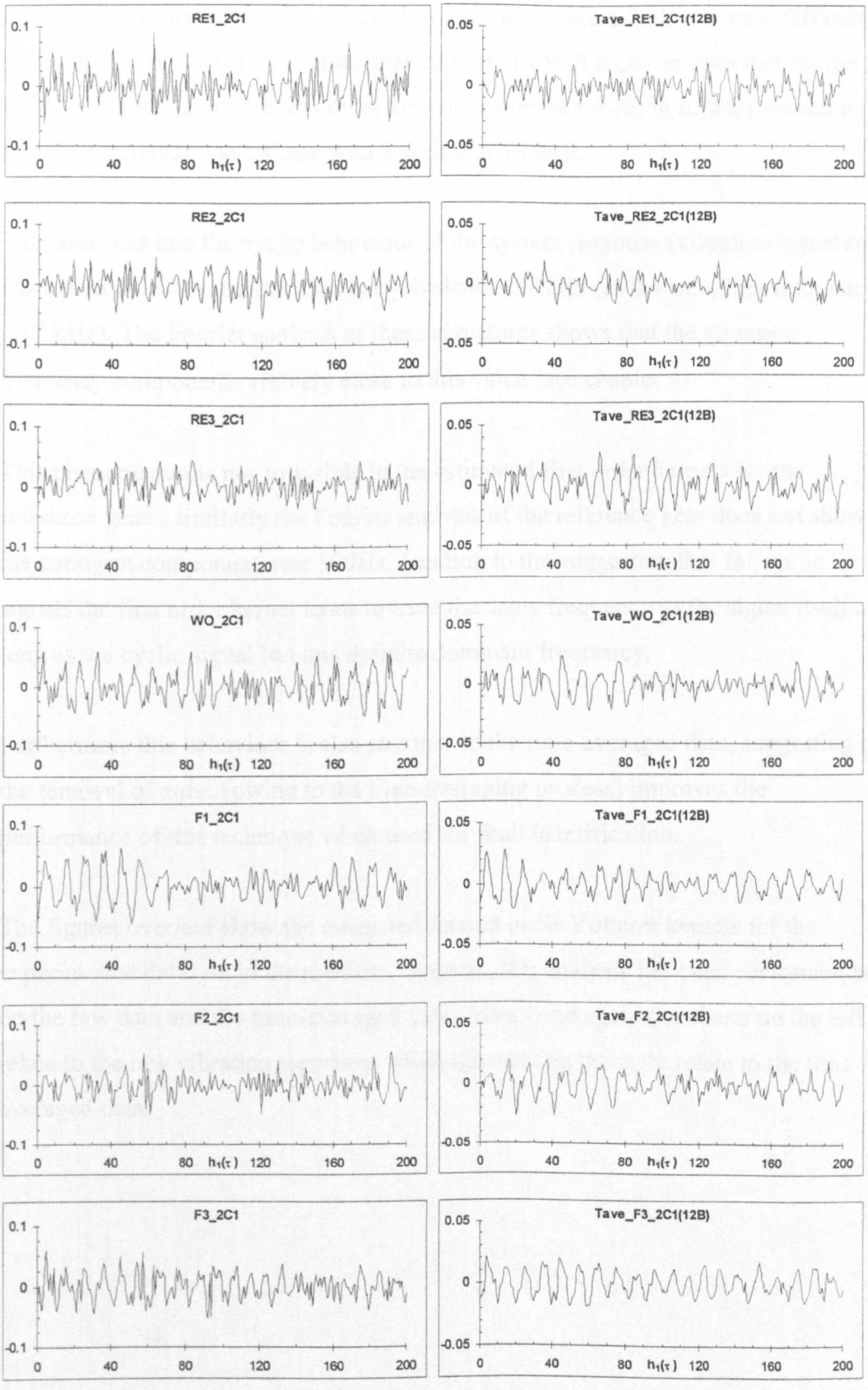


Figure 7.7 – First order kernel for experimental data

From a pure visual inspection of the above kernels (Figure 7.7), it is very difficult to differentiate between the different gear conditions. Still it can be seen that for the faulty condition gears the first order kernels (especially $h_1(0)$ to $h_1(40)$) present a definite triangular wave shape with a period of 10 lags.

This indicates that the cyclic behaviour of the system response (vibration signature) has a period of 10 samples. This is equivalent to 512Hz (as the sampling frequency is 5.12 kHz). The Fourier analysis of these signatures shows that the strongest frequency components are very close to this value (see chapter 4).

This phenomenon is not so visible in the estimated first order kernels for the reference gears, similarly the Fourier analysis of the reference gear does **not** show the strongest component near 500Hz. Leading to the suggestion that for cyclic signals the first order kernel tends to have the same frequency as the signal itself as long as the cyclic signal has one definite dominant frequency.

Furthermore this behaviour is also stronger in the time-averaged data, suggesting that the removal of noise (owing to the time-averaging process) improves the performance of this technique when used for fault identification.

The figures overleaf show the estimated second order Volterra kernels for the experimental data. As in the previous chapters, this analysis has been performed both on the raw data and the time-averaged data. Also, once again the charts on the left relate to the raw vibration signature, while the ones on the right relate to the time averaged data.

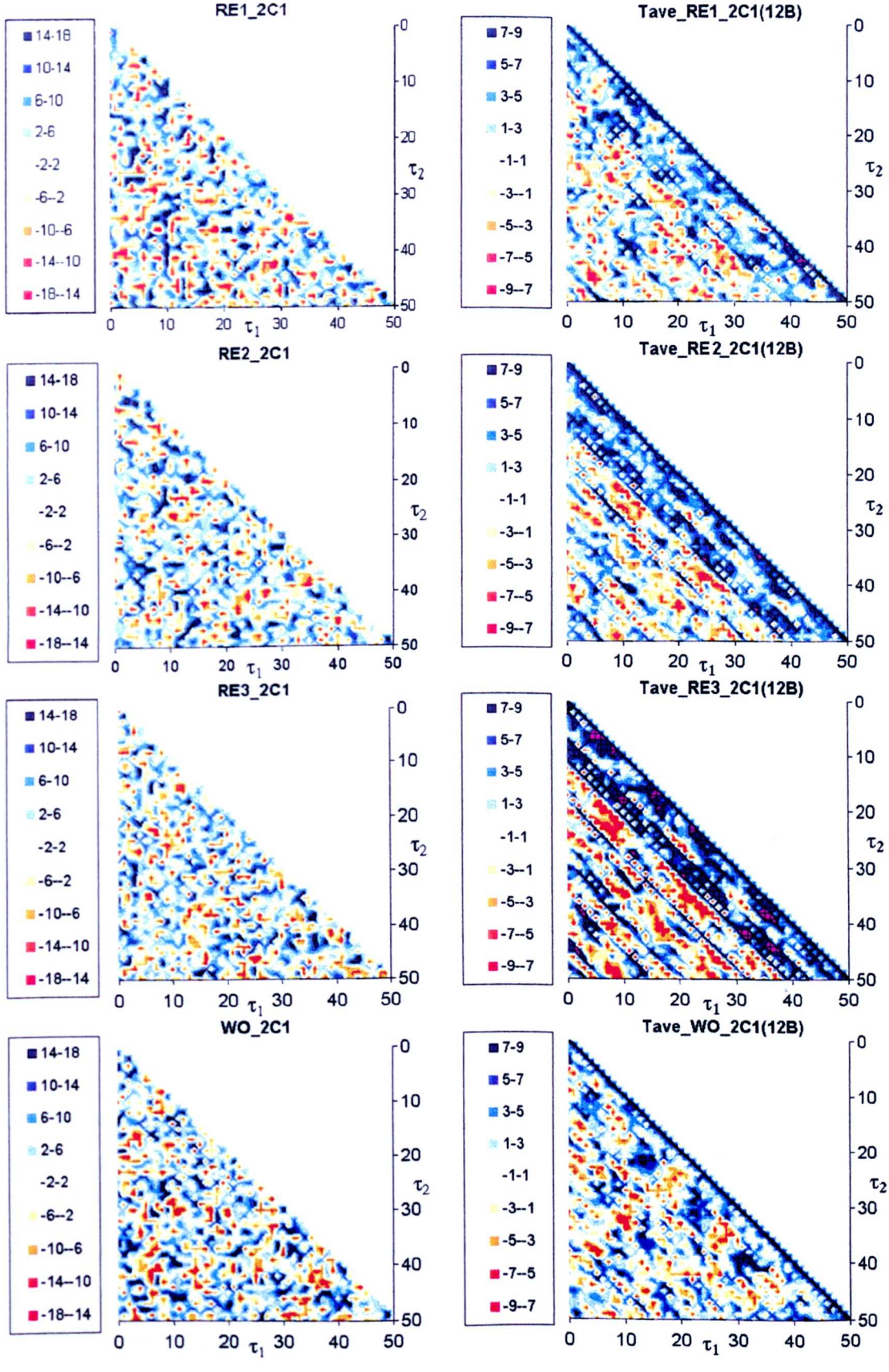


Figure 7.8 (a) – Second order kernels for experimental data

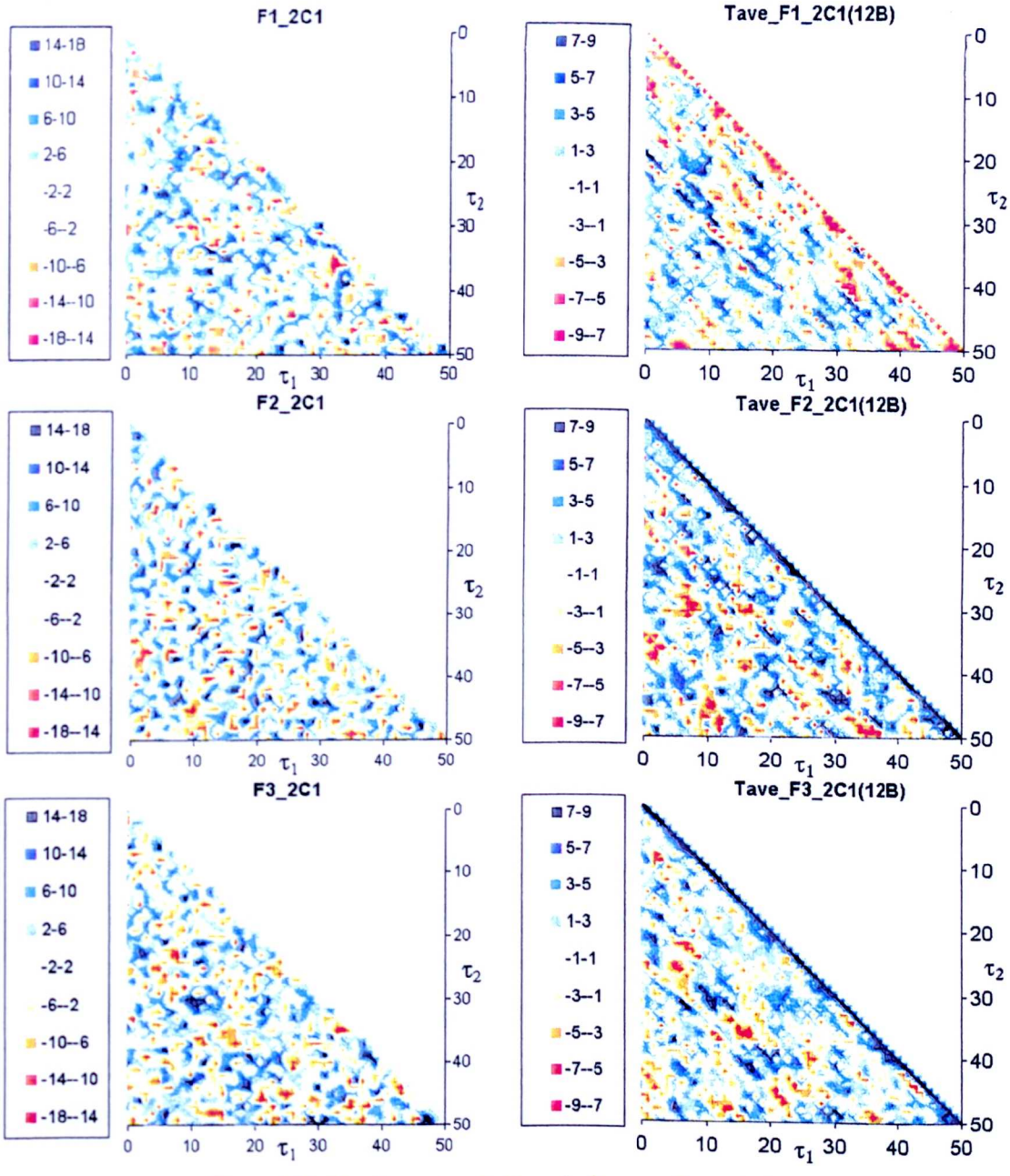


Figure 7.8 (b) – Second order kernels for experimental data

As discussed previously a full analysis of the second order kernels is a complex task, as the magnitude and both lags (τ_1 and τ_2) of the kernel must be considered. A visual inspection of these kernels (especially for the time-averaged data) can lead to a meaningful analysis and differentiation between the different gear conditions. It is clearly seen (in the time averaged plots) that as the fatigue crack develops the strong influence of periodical pulses (period of 10 samples) and respective lags of 10

samples (i.e. $|\tau_1-\tau_2|=10$) diminishes. This is also observed for pulses with very short respective lags (i.e. adjacent to the matrix diagonal).

This behaviour is not seen in the plots for the raw data. In fact on the raw data plots it is impossible to visualise any specific trends from a pure visual inspection of the kernel maps. This is attributed to the interference caused by the inherent noise present in the raw signal.

This hypothesis is testified in Section 7.3.3, when an additional test of the Volterra kernel approach is performed. This test uses the stress-wave signatures[2] (which is less susceptible to noise from other adjacent processes) and shows that the raw stress-wave data produces plots which are very similar to the time-averaged stress wave data. This test also shows the importance of the time domain averaging procedure for the effective use of the Volterra kernel approach to vibration condition monitoring.

Finally, the "auxiliary" kernel measures for the experimental raw data are included in Table 7.2. The first two columns of the table label the different vibration signatures (gear cycles) for all the gear conditions. For the raw data, the four auxiliary parameters (columns 2 to 4) are calculated for six cycles (each with two revolutions, 2C1 to 2C6) of vibration signatures. For the time averaged data only two cycles were calculated (2C1(12B) and 2C2(12B)). A measure (labelled *Ratio*) combining the relative strength of the linear to non-linear component was used. This is defined in the equation below:

$$Ratio = \frac{rms(\text{estimated linear response})}{rms(\text{estimated non - linear response})} = \frac{rms(y_l(t))}{rms(y_n(t))} \qquad \text{Eq. 7.7}$$

where: y_l and y_n are defined in Figure 7.2.

[2] This data was collected simultaneously with the acceleration data used throughout this research.

Table 7.2 – Volterra auxiliary parameters for experimental data

GEAR CONDITION		AUXILIARY PARAMETERS				
		h_1 rms	% Y_L	% Y_N	Ratio	$Ratio_{Tave}$
Brand New Gear (RE1)	2C1/2C1(12B)	0.0156	39.33	60.67	0.648	0.742
	2C2/2C2(12B)	0.0158	39.76	60.24	0.660	0.680
	2C3	0.0148	37.33	62.67	0.596	
	2C4	0.0156	38.22	61.78	0.619	
	2C5	0.0157	38.40	61.60	0.623	
	2C6	0.0161	39.68	60.32	0.658	
Normal Cond. 1 (RE2)	2C1/2C1(12B)	0.0123	37.04	62.96	0.588	0.761
	2C2/2C2(12B)	0.0113	36.68	63.32	0.579	0.618
	2C3	0.0122	36.88	63.12	0.584	
	2C4	0.0130	37.16	62.84	0.591	
	2C5	0.0125	37.43	62.57	0.598	
	2C6	0.0129	37.92	62.08	0.611	
Normal Cond. 2 (RE3)	2C1/2C1(12B)	0.0119	36.09	63.91	0.565	0.793
	2C2/2C2(12B)	0.0132	38.47	61.53	0.625	0.810
	2C3	0.0128	39.56	60.44	0.655	
	2C4	0.0128	38.07	61.93	0.615	
	2C5	0.0130	37.91	62.09	0.610	
	2C6	0.0131	37.85	62.15	0.609	
Worn-out (WO)	2C1/2C1(12B)	0.0153	42.62	57.38	0.743	0.709
	2C2/2C2(12B)	0.0165	45.04	54.96	0.820	0.533
	2C3	0.0136	36.53	63.47	0.576	
	2C4	0.0142	37.66	62.34	0.604	
	2C5	0.0153	43.71	56.29	0.777	
	2C6	0.0173	45.58	54.42	0.838	
Fatigue crack 1 (F1)	2C1/2C1(12B)	0.0099	35.27	64.73	0.545	0.616
	2C2/2C2(12B)	0.0095	32.39	67.61	0.479	0.579
	2C3	0.0129	44.05	55.95	0.787	
	2C4	0.0129	41.62	58.38	0.713	
	2C5	0.0122	40.03	59.97	0.667	
	2C6	0.0130	41.76	58.24	0.717	
Fatigue crack 2 (F2)	2C1/2C1(12B)	0.0108	39.49	60.51	0.653	0.679
	2C2/2C2(12B)	0.0105	38.15	61.85	0.617	0.631
	2C3	0.0100	36.93	63.07	0.586	
	2C4	0.0111	39.52	60.48	0.653	
	2C5	0.0111	38.76	61.24	0.633	
	2C6	0.0119	41.25	58.75	0.702	
Fatigue crack 3 (F3)	2C1/2C1(12B)	0.0117	40.69	59.31	0.686	0.623
	2C2/2C2(12B)	0.0112	42.08	57.92	0.727	0.677
	2C3	0.0112	38.96	61.04	0.638	
	2C4	0.0111	39.73	60.27	0.659	
	2C5	0.0124	42.07	57.93	0.726	
	2C6	0.0118	42.12	57.88	0.728	

A graphical representation of the results in Table 7.2 is included below. These results suggest that a possible use of kernel estimation for condition monitoring is by analysing how well the 1st order kernel models the vibration signal. This is based on the idea that different faults generates different non-linearities. This hypothesis is very plausible, however it still needs experimental investigation for full acceptance.

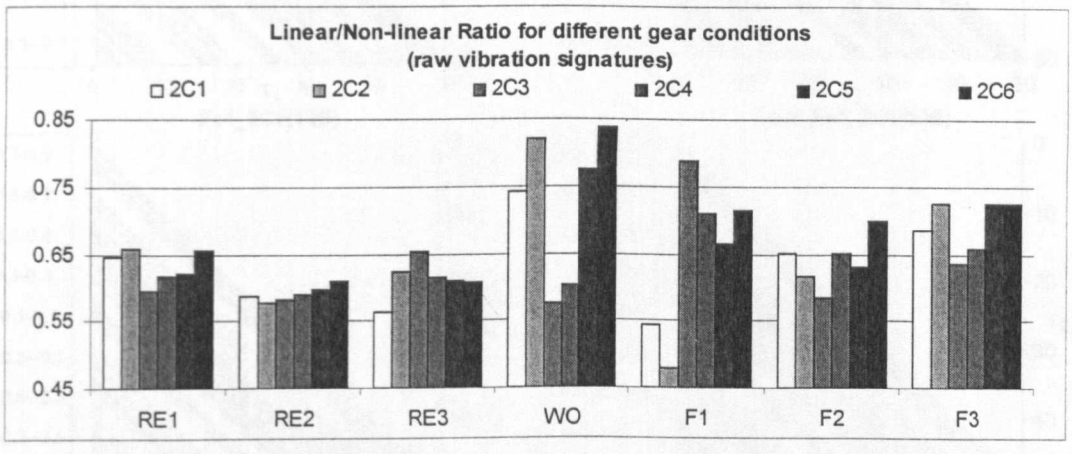


Figure 7.9 – Ratio of kernel estimated linear to non-linear signal content

In the present study, the y_l/y_n ratio for the reference gears is constantly around 0.6. In contrast the ratio for the worn-out and faulty condition gears are not. However, these plots indicate that fatigue crack tend to modify the y_l/y_n ratio. This could be due to its dampening effect, still to ascertain this fact for sure without further experimentation.

7.3.3. Volterra approach to CM using Stress wave signatures

Here, stress wave signatures (collected simultaneously with the vibration data) are used with the Volterra kernel approach to condition monitoring. This test testifies that, the noise inherent to the vibration signatures plays a major role on the 2nd order kernel estimation. Hence, if vibration data is to be used with the Volterra kernel approach to condition monitoring, then time averaging procedures (which reduces the random noise) are essential. In contrast, time averaging procedures are not needed for stress-wave signal, as these usually have a higher signal-to-noise ratio.

In this section only the second order kernels are calculated, as these are the kernels which were greatly affected by noise. These are shown overleaf, once again the left column gives the second order kernels for the raw stress-wave data, while the right column gives the results for the time averaged stress-wave data.

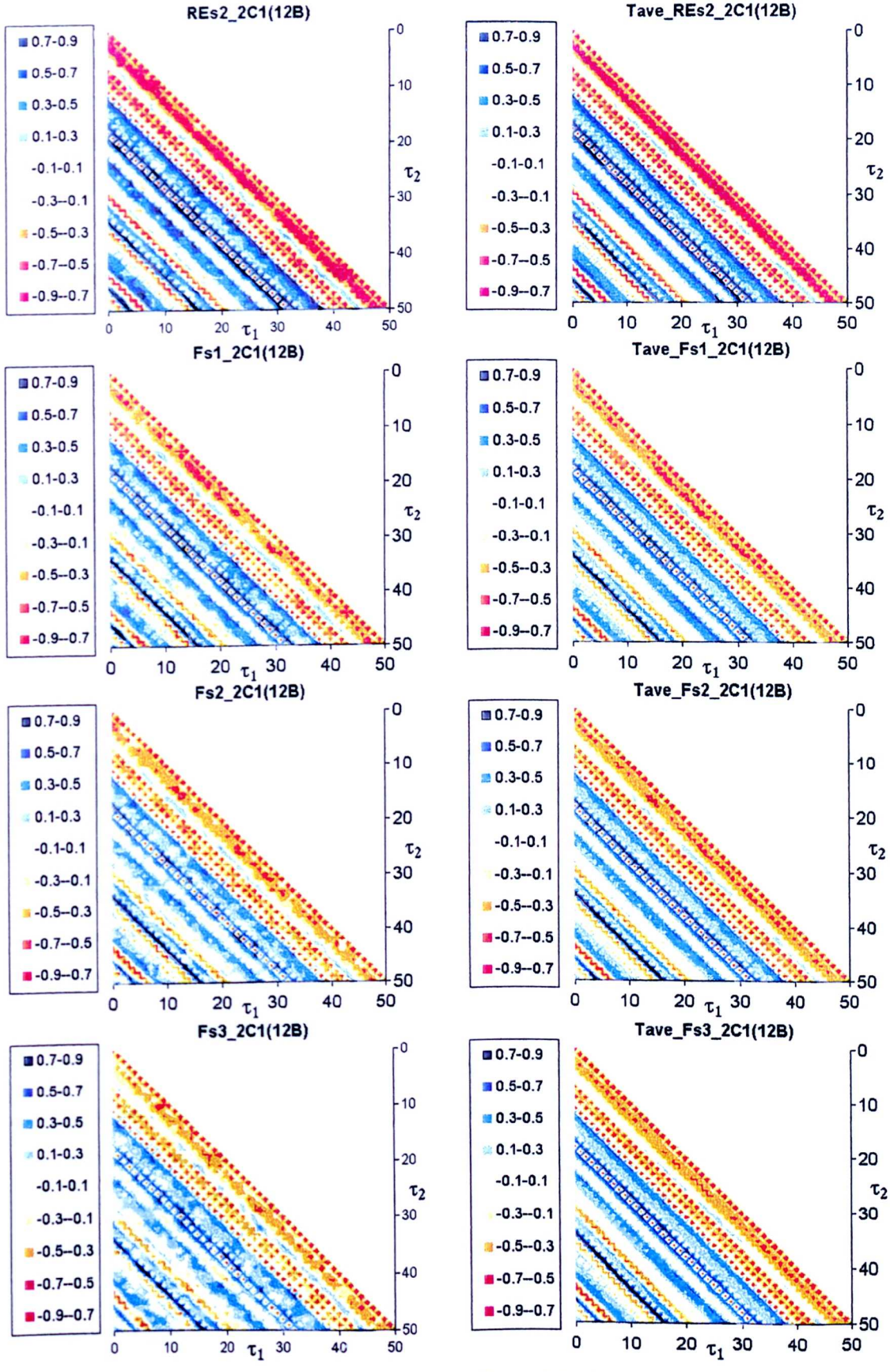


Figure 7.10 – Second order kernels for experimental data (stress wave signature)

As it can be seen now the plots for the time-averaged data are very similar to those for the raw data. Furthermore, the trends caused by the increasing fatigue crack are similar for both raw and time-averaged data. These trends are also similar to those observed in the second order kernels for the time-averaged vibration signatures.

This shows that the hypothesis formulated in the previous section is true and shows the importance of the time averaging procedure on noisy data signatures.

7.3.4. Summary of the performance of the Volterra approach to condition monitoring

It has been shown that Volterra kernel analysis can also be used for the identification of fatigue cracks on rotating spur gears. This technique has effectively identified the presence of the crack in two different sources of data: vibration signatures (after time-averaging procedure) and stress wave signatures. It must be observed that the performance of this technique is heavily dependent on the noise level in the collected data. In fact, this method failed to produce meaningful results for the raw vibration data, which presented the highest noise level, out of all the data sets analysed. Below is a summary of the advantages and disadvantages of this method:

Advantages

- Volterra kernels are able to model non-linear signals, hence allowing the analysis of truly non-linear systems.
- Previous knowledge of the signal properties is not required for the selection of the input parameters. This is a major advantage of this technique as most methods used in vibration condition monitoring require previous knowledge about the signal in order to select appropriate parameters (window function in TF methods, and the mother wavelet in time-scale decomposition) before the signal analysis.

Disadvantages

- This technique is extremely computationally expensive. Depending on the signal length and memory of the Volterra model, the processing can take several

minutes (on a typical PII 300 MHz, 64Mb RAM), making it impossible to apply this technique to online real-time condition monitoring. This problem will eventually be overcome as microprocessors become more and more powerful.

- This technique is not suitable for long cyclic signals with a low signal-to-noise ratio. So, for most real applications of these techniques to vibration signatures, a time-averaging procedure is essential to reduce noise influence on the time series.
- For under-damped signals very long memories are needed to accurately model the system.

Chapter 8

Discussion.

This chapter reviews and compares all the different condition monitoring covered by this research. Here four main sections are included, each relates to the different approaches to condition monitoring, namely: statistical, frequency and time-frequency, time-scale and non-linear approaches. These sections contain a summary of the results obtained in this research.

8.1 Statistical approach to CM.

This approach relies on well established statistical measures techniques such as moments, skewness and Kurtosis measures (section 4.1). These are the simplest approaches to vibration condition monitoring, and the results presented here show that these techniques are also suitable for the identification of tooth fatigue cracks on spur gears.

This research introduces and shows the effectiveness of one particular statistical test, namely: Kolmogorov-Smirnov test (section 4.2.2). This test, which has never been used in vibration condition monitoring applications, is based on the statistical comparison of two signals. This comparison return 1 if the *null hypotheses* (i.e. the two signals come from the same distribution, hence are statistically equal) is met, and returns 0 if the signals are not equal. These values can be easily converted into an easy to assimilate measure, which has been called "similarity probability".

The results obtained in this research clearly show that when comparing similar vibration signatures (i.e. reference with reference) the similarity is much greater than when comparing different signatures (i.e. reference with faulty). Furthermore, it was observed that as the fatigue crack increased, the expected decrease in the similarity probability was observed.

These results are of great importance as it shows that this technique can and should be used in vibration condition monitoring applications. Also it leads the way to the usage of this technique as a pattern classification tool. This would be based on the results here obtained and the theoretical contributions by Peacock [120] and Fasano and Franceschini [121] comparing this technique to time-frequency maps, time-scale maps and second order Volterra kernels.

It is believed that the performance of this technique is as effective as the performance of neural networks. The only difference being that this technique does demand the computational overhead linked to the training period of the Neural Network.

8.2. Frequency and Time-frequency approach to CM.

This approach is based on the assumption that different faults will generate different frequency components at certain times in the vibration signature. This approach decomposes the time domain signal onto a time-frequency map. This map forms the basis for the analysis of the system condition. Time-frequency maps have already been successfully used in a wide range of fields (including speech recognition, echo detection, gear and bearing failure).

This analysis can be done either visually, or if a feature extraction method is used, the maps can be fed to pattern recognition systems for automated condition monitoring. Consequently, these techniques can be used as a vibration signature pre-processing and post-processing tools.

8.2.1. Spectral analysis

Today, spectral analysis is widely used as a tool for a preliminary analysis of time series. This method applies the Fourier transform to a vibration signature, thereby identifying the dominant frequencies appearing in the spectral plot.

This tool is very effective in providing fundamental knowledge of the signal under analysis. However, it has been shown that this technique cannot be reliably used for early fatigue crack detection. In fact, from the visual observation of the spectral plots it is virtually impossible to ascertain for sure the condition of the gear (section 5.1).

Still, this technique is very powerful because of its simplicity. Furthermore, Fourier analysis forms the basis of many other time-frequency condition monitoring techniques.

8.2.2. Cepstral analysis

This method relies on the performance of the cepstrum technique on the time series. This can be seen as the spectrum of a spectrum. Cepstral analysis is very effective for the observation of major frequency components on a signal. It groups the harmonic contents on a single fundamental frequency (section 5.2).

Also it was observed in this research that cepstral reconstruction (an extension to the cepstral analysis) is a very effective method for identifying the relative strength of transients present in a signal. It is shown in this research that the observation of the first sample on the cepstral reconstruction plot is capable of clearly distinguishing between the different gear conditions tested. This suggests that this method is very suitable for pattern extraction from time-domain signals, as it not only indicates the fundamental frequency components, but also gives some indication to transient contents in the signal. The patterns extracted from this method can then be fed into automated pattern recognition techniques such as neural networks.

Unfortunately this technique has one major drawback. It does not allow for fault localisation, as all the signal components lose their phase information.

8.2.3. Spectrogram (STFT)

This technique is based on the well known Fast Fourier Transform algorithm. It selects portions of the time-domain signal and converts (these portions) to the frequency domain, giving the spectral content for each section (i.e. each time instant). The spectral information for the different time instants are plotted in the form of a 2-D map or a 3D surface, showing how the signal spectral contents change over time (section 5.3).

The main advantage of this technique lies in its simplicity and speed. Furthermore, spectrogram maps are of easy assimilation due to its uniform and smooth surface. Unfortunately, this technique shows two major disadvantages:

1) This method does not allow for a signal analysis with a simultaneous high time and high frequency resolution (time-frequency resolution trade off). Therefore, it is concluded that this technique is unsuitable for the accurate identification of both large and small-scale patterns in the signal.
2) The performance of this method is very much dependent on the correct choice of the window function (including shape and dimensions). The correct choice can only be made if additional information on the signal to be processed is available.

8.2.4. Wigner distribution

This technique is also based on the well known Fast Fourier Transform algorithm. Similar to spectrograms, this technique produces a 2D map (or 3D surface) showing the signal spectral content for different time instants (section 5.4).

The main advantage of this technique lies in the fact that it does not suffer from the time-frequency resolution trade-off. This is related to the method of selecting the "portions" of the time series described in section 5.4. Unfortunately, this technique shows one major disadvantage, which jeopardises the performance of this method:

The WD is a bilinear functional of a given waveform $f(t)$. Whenever $f(t)$ is a sum of various independent components, such as a combination of multiple frequencies and noise (this is the case for most vibration signals), then the WD introduces cross-

terms. These cross-terms can be seen as an artefact arising from the interference between the time series' multiple components. This interference can be reduced by the usage of windows that maximise the local signal components, and minimise the distant components. Unfortunately this only reduces the interference between non-simultaneous signal components. For simultaneous components the interference is inevitable.

A consequence of the cross terms interference is that the WD maps indicate the presence of non existent frequency components (ghost components). This reduces the clarity of the maps, making it very difficult to observe the real frequency components of the signal.

Still, this method is of invaluable need when a high time and frequency resolution is needed in the analysis of signals where multiple components do not occur simultaneously. One example is telephone tone dial detection.

8.3. Time-scale approach (wavelet approach)

The wavelet analysis is a technique which extracts both frequency, phase and position information from a signal. It can be thought of as a Fourier analysis that uses other orthogonal functions (instead of sine and cosine) to decompose the signal. These orthogonal functions are called mother wavelets. Similar to the Fourier analysis this technique transforms the signal into one-dimensional vectors. These show how the intensity of the different wavelet scales change over time. Also, these vectors can be used to generate time-scale maps. These are very similar to time-frequency maps, in the sense that they can also be presented as a 2D map or a 3D surface, and more importantly, they contain the time-scale[3] information describing the signal under analysis.

In this research it was found (section 6.5) that this method is extremely effective as a pre-processing tool for automated pattern recognition system, since it is able to

[3] Note that each wavelet scale can be easily converted into frequency; this is shown in chapter 6.

concisely extract the most important features of a signal. However, the pure visual inspection of time-scale maps does not show the presence of fatigue cracks as well as some of the previous techniques here described. Still, it must be noted that the experienced engineer is able to recognise the faults in these maps.

The main disadvantage of this technique is related to the that fact the frequency resolution (i.e. number of scales) is governed by the number of samples of the signal under analysis. Also the conversion between scale to frequency is not a linear process. In fact the mean square maps tend to display the middle to high frequency components on its higher levels. This can be a problem for signals with low signal-to-noise ratio, as the middle to high frequency components can be overshadowed by the high frequency noise present in the signal.

This, opens the way to further work being carried out in this area, investigating how low-pass filters can be used to improve the performance of condition monitoring systems by means of wavelet mean square maps.

Furthermore, this research shows that the choice of the mother wavelet greatly affects the appearance of the time-scale map. This suggests that, as with time-frequency techniques, previous knowledge about the signal must available for the correct choice of mother wavelet.

8.4. Non-linear approach (Volterra Kernel)

This technique is based on the analysis of the estimated Volterra kernels of different signals. These kernels are estimated numerically, and many different algorithms exist for this purpose. This research introduced the use of Volterra series (for the first time) and showed the effectiveness of this new approach to condition monitoring, by calculating the zeroth, first and second order kernels for different vibration signatures (section 7.3).

The estimated kernels are presented as a single value (zeroth order kernel), as a 1D vector (the first order kernel) and as a 2D matrix (second order kernel). Again as in the previous approach only a visual inspection of these kernels was performed.

The experimental results showed that especially the second order kernels gave clear indication of the presence of fatigue cracks. This performance was cross-checked by using data sets collected with different transducers (namely an accelerometer and a stress wave sensor). This opens the way to further applications of this technique in the vibration condition monitoring and analysis.

The main advantage of this technique lies in the fact that it does not require any previous knowledge about the signal to be analysed. However, this technique also shows some drawbacks.

The major drawback of this technique is related to its high sensitivity to noise. Therefore it must be used with great caution for signals with low signal-to-noise ratios. This drawback is clearly seen when comparing the results from the raw vibration (accelerometer) signatures to those from the time-averaged vibration (accelerometer) signatures, and stress wave sensor. The second order kernels for the raw vibration data do not lead to the identification of the presence of a fatigue crack. In fact, it does not give any insight on the signal being analysed (hence condition of the gearbox). This is shown by the spurious peaks spread across the plots, which are attributed to the influence of noise present in the signal.

This technique has another minor drawback. The kernel estimation process is an extremely computationally expensive task. Therefore, any on-line application of this technique is of difficult implementation. Still, this does not pose as a major problem as the microprocessor power is ever increasing.

These two problems should be addressed in further works based on the ideas given here.

Chapter 9

Conclusions

This chapter concludes the ideas and findings of this research programme by summarising the main points of the topics here studied. The final section, further work, restates the questions opened by this research. These questions should be tackled in further research in this area.

This research reviewed today's most common techniques in wide used for vibration condition monitoring, namely:

- **Statistical approach**, including well known measures such as vibration signatures moments and statistical measures such as kurtosis and skewness.
- **Frequency and time-frequency approach,** including well known techniques such as FFT, Cepstrum, spectrogram and Wigner distribution.
- **Time-scale approach,** including the Daubechies wavelet series and the Harmonic wavelet.

In addition this research also introduces two new techniques which can be used in vibration condition monitoring applications, namely:

- **The Kolmogorov-Smirnov test.** This method belongs to the class of statistical methods for vibration condition monitoring.

The Volterra kernel approach. This approach belongs to the class of non-linear methods for vibration condition monitoring.

These techniques have not yet (to the best of the author's knowledge) been applied to the field of vibration condition monitoring, and experimental results have shown that these techniques are of invaluable importance to the condition monitoring engineer.

The performance of the two new techniques was tested and evaluated over the problem of early detection of tooth fatigue cracks on spur gears. The fatigue cracks were simulated by adding cuts, with an attack angle similar to those of real fatigue cracks, to the base of the tooth face. In all three cracks (cuts) of different depths were analysed. The results are presented in chapters 4 and 8. These clearly show the effectiveness of these methods, in fatigue crack identification. From the investigations performed the following conclusions were obtained.

1) The time averaging procedure is effective to magnify the signal-to-noise ratio of vibration signature. However, for high frequency component analysis, this method must be used with great care. In fact, it has been shown that minor variations in the rotational speed (or data acquisition sampling frequency) of the device under analysis, is sufficient to place the high frequency components in interference mode. This leads to an erroneous representation of the signal to be analysed, jeopardising the whole condition monitoring system.

2) The Kolmogorov-Smirnov test is of great importance for the comparison of signals. It has been shown that this test is able to accurately identify the presence of the fatigue faults in a time series. Also it is able to identify the advancement of the fatigue crack.

3) The well established frequency techniques for condition monitoring (spectral and cepstral analysis) are essential to give the condition monitoring engineer some preliminary knowledge about the signal under analysis. Furthermore, cepstral analysis in particular is extremely effective in providing information related to the transient content in the signal. This can be achieved by means of the cepstral reconstruction. Also this technique proved to be very effective for signal pattern extraction.

4) From the time-frequency (and time-scale) distributions the spectrogram showed the best performance for the detection of fatigue cracks. This is attributed to the uniformity of the maps generated by this technique. The maps generated from the WVD include too much cross-term interference (increasing the complexity of the map appearance), and the time-scale maps (from wavelet decomposition) present too little frequency resolution. In these maps most of the signal information is contained in two levels only.

5) Wavelet decomposition seems to be an ideal tool for signal pattern extraction, as it is able to fully represent a signal on a minimum number of patterns. This is of utmost importance when using wavelets as pattern extraction tools, for feeding automated pattern recognition systems (such as neural networks). Also the investigations here presented show that the choice of the mother wavelet greatly affect the appearance of the time-scale map. This suggests that for effective condition monitoring by means of wavelet mean squares maps, previous knowledge about the signal to be analysed is needed. This aids in the correct choice of the mother wavelet.

6) The Volterra kernel approach is another effective condition monitoring technique. It indicates the presence of fatigue cracks on spur gears. This method is not suitable for the analysis of signals with low signal-to-noise ratios, as the noise greatly influences the second order kernel (which gives the best indication about the fault presence).

Aiming to collect and summarise the results obtained in this research Table 9.1 was devised. The table compare the different methods studied in terms of their performance for the specific problem of fatigue crack detection.

Table 9.1 – Collection and summary of characteristics of the condition monitoring different method

Signal processing techniques	Fault detection capabality									Time domain localisation of fault			Computational expense/ complexity			Theoretical/ interpretation simplicity		
	crack			Crack growth			worn out											
	Low	Med	High	L	M	H	L	M	H	none	Low	High	L	M	H	L	M	H
Statistical moments	■			■			■			■			■			■		
F-test	■	■		■			■			■			■			■		
KS test			■			■			■	■			■			■		
Spectrum	■	■		■			■			■				■			■	
Cepstrum		■		■			■			■				■			■	
Spectrogram		■	■	■	■		■	■				■		■	■		■	■
Wigner distribution		■	■	■	■		■	■				■		■	■		■	■
Harmonic wavelet		■	■	■	■			■			■	■		■	■		■	■
Volterra kernel			■	■			■			■					■			■

The previous table summarises the main properties of the techniques studied. These related to:

- **Fault detection capability**, shows how the techniques are able to:
  1. identify presence of a crack,
  2. identify the different evolution stages of the crack, and;
  3. identify the symptoms from worn out gears.

- **Time domain localisation of fault**, relates to the capability of determing 'where' in the vibration signal the fault is present (phase information).

- **Computational expense/complexity**, relates to the difficulties involved in the algorithm coding and also the time required to process the vibration signals

- **Theoretical, interpretation simplicity**, relates to the simplicity of results obtained from the different techniques. This is of utmost importance in an industrial environment, as most of the condition monitoring programmes will be carried out on the factory floor.

Finally, it must be emphasised that the results in Table 9.1 are accurate for the specific problem of early fatigue crack identification. Therefore, it must be used only as a guiding tool for other condition monitoring applications.

The contributions of this research could be summarised as follows:

1) Introduction of two new techniques to the analysis of non-linear vibration signatures. Neither of these techniques requires any signal pre-processing, allowing for the direct use of the time-domain data.

2) A comprehensive comparative review of the performance of existing condition monitoring approaches to the problem of identifying early tooth fatigue cracks on spur gears.

3) A comprehensive comparison of the performance of the different condition monitoring approaches when using raw vibration signatures and time-averaged vibration signatures. This supports the previous claims that the time-averaging technique magnifies the signal-to-noise ratio of the signal. Also it is noted in this research that this technique must be used with great care for the analysis of high frequency components, as minor variations on the rotational speed or the gear, or on the sampling rate is sufficient to shift the time series. Consequently, high frequency components can enter in destructive interference mode, leading to an erroneous representation of the time series.

4) Importance of mother wavelet choice and how it affects the appearance of the time scale mean square maps.

9.1. Further Work

The work presented here aims at providing the vibration condition monitoring engineer with a wide range of tools that could be included in a vibration condition monitoring toolkit. These tools should be robust enough to identify a wide range of faults in mechanical systems. To enhance the toolbox presented, further work could be directed at the following issues.

1) **Usage of higher order Volterra kernels**: Investigations into the influence of the selected order for the Volterra series over the task of fault identification is urgently needed. This research has only calculated up to the second order kernels.

2) **Use of different statistical measures for the KS test**: The statistical measure used for the comparison of two signals with the KS test plays a major role in its performance. This research only used the maximum distance between cumulative distribution functions. However it is believed the by using different measures (such as mean, mean square or root mean square distances) the user is able to adjust the sensitivity of this method, tuning it to specific applications.

3) **Fault types**: The performance of the newly introduced techniques (KS test and Volterra Kernels) should be tested on other types of mechanical faults such as

blips, shaves, pitting and scoring, missing tooth, misalignment, unbalance, bearing defects, etc. This investigation would verify the true effectiveness and flexibility of these approaches.

4) **Pattern recognition**: Use of the 2 dimensional KS test for and automated comparison of time-frequency, time-scale and second order Volterra kernels obtained from different vibration signals. This, in effect should be able to accurately state whether 2D maps (such as spectrogram and others) representing vibration signatures are similar. Therefore, if a template with all the system conditions is known, this technique could be used to match the most likely condition of the system under analysis. This approach has one main advantage over neural networks, it does not require the vast amount of data needed for the network training.

Bibliography

1 B. K. N. Rao. Handbook of condition monitoring. Elsevier Advanced Technology. ISBN 1 85617 234 1. 1996.

2 T. M. Hunt. Condition Monitoring of Mechanical and Hydraulic Plant: A concise introduction and guide. Chapman & Hall. ISBN 0 412 70780 2. 1996.

3 J. H. Williams et al . Condition Based Maintenance and Machine Diagnostics. Chapman & Hall. ISBN 0 412 46500 0. 1994.

4 R. A. Cullacott. Mechanical Fault Diagnosis and Condition Monitoring, Chapman & Hall, London. ISBN 0-470-99095-3. 1977.

5 M. Neale and Associates. A Guide to the condition monitoring of machinery. ISBN 0 11 512126 9. 1979.

6 R. Patton, P. Frank, R. Clark. "Fault Diagnosis in Dynamic Systems. Theory and Applications. Prentice Hall Int. ASIN: 0133082636. 1989.

7 R. W. Thorne. PhD Thesis University of Illinois. 1987.

8 P. W. Hills. Vibration-based Condition Monitoring – the learning issue. Insight, V. 38 (8), pp. 576-579. 1996.

9 C. Chatfield. The analysis of Time-Series: An introduction. 4th ed. Chapman & Hall. ISBN 0412318202. 1989.

10 J. G. Proakis, C. M. Rader, F. Ling and C. L. Nikias. Advanced digital signal processing. Macmillan. ISBN 0 02 396841 9. 1992.

11 P. W. Hills. Production benefits from a vibration-based condition monitoring programme. Insight, vol. 38 (8), pp. 563-565. 1996.

12 T. I. El-Wardany et al. Tool condition monitoring in drilling using vibration signature analysis. Int. J. Mach. Tools Manufacture, vol. 36 (6), pp. 687-711. 1996.

13 G. Dalpiaz and A. Rivola. Condition monitoring and diagnostics in automatic machines: Comparison of vibration techniques. Mech. Systems and Signal Processing, vol. 11(1), pp. 53-73. 1997.

14 Q. Meng and L. Qu. Rotating Machinery fault diagnosis using wigner distribution. Mech. Systems and Signal Processing, vol. 5(3), Pp.155-166. 1997.

15 K. Zheng, D. J. Whitehouse. The application of the wigner distribution function to machine tool monitoring. Proc. Inst. Mech. Engineers, vol.206, pp. 249-264. 1992.

16 C. Pachaud et al. Crest factor and kurtosis contributions to identfy defects inducing periodical impulsive forces Mechanical systems and signal processing , vol. 11(6), 903-916. 1997.

17 M. N. M. Badi et al. Fault Classification of a model drive-line using time domain data. Proc. of COMADEM-96, pp. 43-51 Univ. of Sheffield, England. 1996.

18 B. X. Santiago, R. A. W. Elson and G. F. Gilmore. HST photometry of 47 Tuc and analysis of the stellar luminosity function in Milky Way clusters. Monthly Notice of the Royal Astronomical Society, vol.281(4), pp.1363-1374. 1996.

19 M. Akay, Y. M. Akay, P. Cheng and H. H. Szeto. Investigating the effects of opioid drugs on electrocortical activity using wavelet transform. Biological Cybernetics, vol.72(5), pp.431-437. 1995.

20 H. Arsham. A test sensitive to extreme hidden periodicities. Stochastic Hydrology and Hydraulics, vol.11(4), pp.323- 330. 1997.

21 F. Andrade, I Esat, M. Badi. "Gearbox Fault detection using Statistical methods,Time-Frequency methods (STFT and Wigner-Ville distribution)and Harmonic wavelet - A comparative study". 12th International Congress & Exhibition on Condition Monitoring and Diagnostic Engineering Management, Proc. of COMADEM '99, pp. 77-85. ISBN 1901892 13 1.

22 F. Andrade, I Esat, M. Badi. "Gearbox Fault detection using Statistical methods,Time-Frequency methods and Harmonic wavelet". Feature article. Condition Monitor, November 1999, n°. 155. pp. 10-13. ISSN 0268-8050.

23 K. A. Stroud. Further Engineering Mathematics. 2nd ed., Macmillan. ISBN 0 333 52610 4. 1990.

24 D. C. Champney. Fourier Transform and their Physical Applications.. New York: Academic Press. ISBN 1973.

25 E. O. Brignham. The Fast Fourier Transform. 1974. Englewood Cliffs, Nj: Prentice-Hall

26 D. F. Elliot and K. R. Rao. Fast Transforms: Algorithms, Analysis, Applications. NewYork: Academic Press. ISBN: 0122370805. 1983.

27 R. B. Randall. A new method of modelling gear faults. Trans. of ASME, J. of Mechanical Design, vol.104, pp 259-267. 1982

28 Digital Signal Processing Committee, IEEE ASSP Soc. Programs for Digital Signal Processing. Chapter 7. IEEE Press. ISBN: 0-87942-127-4.

29 B. P. Bogert, M. J. R. Healey, J. W. Tukey. The Quefrency Analysis of Time Series. Proc. Symp. Time Series Analysis, M. Rosenblatt, Ed., New York, John Wiley & Sons, pp. 209-243, 1963.

30 A. V. Oppenheim and R. W. Schafer, Digital Signal Processing. Chapter 10, Prentice-Hall Inc., Englewood Clifss, New Jersey, 1975.

31 A. V. Oppenheim and R. W. Schafer, "Homomorphic Analysis of Speech. IEEE Trans. Audio Electroacoust. Vol. AU-16(2), pp. 221-226, 1968.

32 J. Tribolet. Application of Homomorphic Filtering to seismic signal processing. Prentice-Hall Inc. Englewood Clifs, New Jersey, 1979

33 R. B. Randall. Advances in the application of cepstrum analysis to gearbox diagnostics. Proc, of the Int. Conference on Vibration in Rotating Machinery, ImechE, C276/80, pp 169-174, London, 1980.

34 R. B. Randall. Cepstrum analysis and gearbox fault diagnostics. Bruel & Kjar Application Note, pp. 233-280, 1982

35 P. Bradshaw and R. B. Randall. Early detection and diagnostics of machine faults on the trans-Alaska pipeline. 9th Biennial Conference on Mech. Vibration and Noise, ASME, 11-14 Sept, Dearborn, Michigan, USA. 1983.

36 A. A. Syed, J. D. Brown, M. J. Oliver and S. A. Hill. The cepstrum: a viable method for the removal of ground reflections. Journal of Sound and Vibrations, vol 71(2), pp 199-373. 1980.

37 P. Norton. Fundamentals of noise and vibration engineering analysis for engineers. Cambridge University Press, Cambridge. 1989

38 A. Bracciali, G. Cascini. Detection of corrugation and wheelflats of railway wheels using energy and cepstrum analysis of rail acceleration. Proc. of the IMechE - J. of Rail and Rapid Transit. Vol. 211(2) pt. F, pp. 109-116. 1997.

39 R. G. DeJong and J. E. Manning. Gear noise analysis using modern signal processing and numerical modelling techniques. SAE Technical paper 840478, 1984.

40 H. Tang, C. Jian-Zhong, W Yaunhui and C Zhang. The principle of cepstrum and its application in quantitative fault diagnostci of gears. Modal Analysis, Modelling, Diagnostics and Control-Analytical and experimental, ASME, DE-vol 38, pp. 141-144. 1991.

41 M. N. M. Badi, D. J. Dell and A. E. Fellows. Alternative methods of diagnosing gearbox faults. Profitable Cond. Mon. Conf. Stratford-upon-Avon, UK. 1992.

42 R. Koenig, H. K. Dunn and L. Y. Lacy. The sound spectrograph. J. Acoust. Soc. Am, vol. 18, pp.19-49. 1946.

43 A. Dziewonski, S. Bloch, and M. Landisman. A technique for the analysis of transient signals. Bull. Seism. Soc. Am., vol. 59, pp427-444. 1969.

44 A. V. Oppenheim. Speech spectrograms using the fast fourier transform. IEEE Spectrum, vol. 7 pp.57-62, 1970.

45 J. Flanagan. Speech analysis; synthesis and perception. New York, NY: Springer, 1972

46 K. Kodera, R. Gendrin and C. Villedary. Analysis of time-varying signals with small BT values. IEEE Trans. ASSP vol. 26, pp. 64-76, 1978.

47 W. Wang and P. McFadden. Early detection of gear failure by vibration analysis - I. Calculation of the time-frequency distribution. Mechanical Systems and Signal Processing, vol. 7(30), pp. 193-203. 1993.

48 F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. Proc. of the IEEE. Vol 66(1), pp. 51-83. 1978.

49 W. Wang and P. McFadden. Early detection of gear failure by vibration analysis – II. Interpretation of the time-frequency distribution using image-processing techniques. Mechanical Systems and Signal Processing, vol. 7(30), pp. 193-203. 1993.

50 G. Fazio and A. Molinaro. Detection of echoes using time frequency analysis techniques. IEEE Transactions on instrumentation and measurement, vol. 45(1). 1996.

51 W. Y. Kim, Y. C. Park, S. G. Kim. Discrimination of small earthquakes and artificial explosions in the Korean Peninsula using Pg/Lg ratios. Geophysical Journal International, Vol. 134(1), pp. 267-276. 1998.

52 B. E D. Kingsbury, N. Morgan, S. Greenberg. Robust speech recognition using the modulation spectrogram. Speech Communication. Vol.25, No.1-3, pp.117-132. 1998.

53 L. Cohen. Time-frequency distributions – A review. Proc. of the IEEE, vol 77(7). 1989

54 J. Ville. Theory and applications of the notion of complex signal. RAND Corporation Technic al Report T-92, Santa Monica, CA,1958. Translated from "Theorie et applications de la notion de signal analytique. Cables et transmission, vol. 2A, pp. 61/74, 1948" by I. Selin

55 G. S. Cunningham and W. J. Williams. Kernel decomposition of time-frequency distributions. IEEE Trans. on Signal Processing, vol 42(6), 1994.

56 T. A. C. Claasen and W. F. G. Mecklenbrauker. The wigner distribution – a tool for time-frequency signal analysis – Part I: continuous-time signals. Philips journal of research, vol 35 pp.217-250. 1980.

57 T. A. C. Claasen and W. F. G. Mecklenbrauker. The wigner distribution – a tool for time-frequency signal analysis – Part II: discrete-time signals. Philips journal of research, vol 35 pp.276-300. 1980.

58 T. A. C. Claasen and W. F. G. Mecklenbrauker. The wigner distribution – a tool for time-frequency signal analysis – Part III: Relations with other time-frequency signal transformations. Philips journal of research, vol 35 pp.372-389. 1980.

59 L. Cohen. Time frequency analysis. Prentice-Hall International (UK) Ltd, London. ISBN 0 13 594532 1. 1995.

60 W. J. Staszewski, K. Worden and G. R. Tomlinson. Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition. Mechanical systems and signal processing, vol. 11(5), pp. 673-692. 1997.

61 G. C. Gaunard and H. C. Strifors. Signal analysis by means of time-frequency (Wigner-Type) Distributions – Applications to sonar and radar echoes. Proc. of the IEEE, vol. 84(9). 1996.

62 W. Martin and K. Kruger-Alef. Application of the Wigner-Ville spectrum to the spectral analysis of a class of bio-acoustical signals blurred by noise. Acustica, vol. 61. Pp176-183. 1986.

63 W. Sweldens. Wavelets: What next?. Proc. of the IEEE, vol 84(4). pp.680-685.1996.

64 A. Grossmann and J. Morlet. Decomposition of Hardy function into square integrable wavelets of constant shapes. SIAM Journal of Mathematics, vol. 15(4), pp 723-736, 1984.

65 I. Daubechies. "Where do wavelets come from? - A personal view". Proceedings of the IEEE. vol. 84 (4), pp. 510-513. 1996.

66 D. F. Schomer, A. A. Elekes, J. D. Hazle, J. C. Huffman, S. K. Thompson, C. K. Chui, and W. A. Murphy. Introduction to wavelet-based compression of medical images. Radiographics, vol.18,(2), pp.469-481. 1998.

67 N. Kambhatla, S. Haykin, R. D. Dony. Image compression using KLT, wavelets and an adaptive mixture of principal components model. J. of VLSI Signal Processing Systems for Signal Image and Video Technology, vol.18(3), pp.287-296. 1998.

68 T. C. Bailey, T. Sapatinas, K. J. Powell, W. J Krzanowski. Signal detection in underwater sound using wavelets. J. of the Aerican Statistical Assoc., vol. 93(441), pp.73-83. 1998.

69 M. Jobert, C. Tismer, E. Poiseau, H. Schulz. Wavelets – A new tool in sleep biosignal analysis, J. of Sleep Research, vol.3 (4), pp.223-232. 1994.

70 D. E. Newland. "An introduction to random vibrations, spectral & wavelet analysis" (3rd edition). Longman Scientific & Technical. ISBN 0582 21584 6. 1994.

71 W. J. Staszewski and G. R. Tomlinson. "Application of the wavelet transform to a fault detection in a spur gear". Mech. Systems and Signal Processing, vol. 8 (3), pp. 289-307. 1994.

72 W. J. Wang and P. D. McFadden. "Application of orthogonal wavelets to early gear damage detection". Mech. Systems and Signal Processing, vol. 9 (5), pp. 497-507. 1995.

73 W. J. Wang and P. D. McFadden. "Application of wavelets to gearbox vibration signals for fault detection". Journal of Sound and Vibration, vol. 192(5), pp. 927-939. 1996

74 S. T. Lin and P. D. McFadden. "Gear vibration analysis by B-spline wavelet-based linear wavelet transform". Mech. Systems and Signal Processing, vol. 11(4), 603-609. 1997.

75 B. A. Paya, I. I. Esat, and M. N. M. Badi. "Artificial Neural network based fault diagnstics of rotating machinery using wavelet transforms as a preprocessor". Mech Systems and Signal Processing, vol. 11(5). Pp.751-765. 1997.

76 C. K. Chui. "An introduction to wavelets". Academic press, San Diego, ISBN: 0121745848 1992.

77 C. K. Chui. Wavelets: "A tutorial of theory and applications". Academic press, New York, ISBN: 0121745902. 1992.

78 W. J. Wang and P. D. McFadden. "Application of the wavelet transform to gearbox vibration analysis". American society of Mechanical Engineers, PD-Publication, PD-Vol. 52, pp. 13-20, ASME, New York, USA. 1993.

79 P. D. McFadden. "Application of the wavelet transform to early detection of gear failure by vibration analysis". Proceesing of Int. Conf. Of Condition Monitoring, University College of Swansea, Wales, 1994.

80 D. E. Newland. "Harmonic wavelet analysis". Proc. of the Royal Society of Lond. A 443, pp. 203-225, 1993.

81 S. A. Billings. "Identification of nonlinear systems – A survey". IEE Proceedings., vol. 127(6), pt. D, pp.272-285. 1980.

82 F. Andrade, I Esat. "Volterra Series for industrial condition monitoring and diagnostics?". IDPT Proc. of 3rd Biennial World Conference, vol 5, 1998. pp. 62-65. ISSN: 1090-9389.

83 M. Frechet. "Sur les fonctionnelles continues". Annales Scietifiques de l'Ecole Normal Superieure 27, pp. 193-210. 1910.

84 N. Wiener. "Response of a non-linear device to noise". Report V-165 Radiation Laboratory, MIT, Cambridge. April 6. US Dept. of commerce publication PB-58087.

85 N. Wiener. "Nonlinear problems in random theory". MIT Press, Cambridge, MA, 1958.

86 R. Beale and T Jackson. "Neural Computing. An introduction". Institute of Physics Publishing Ltd. ISBN 0 85274 262 2. 1990.

87 J. Fang. "Quadradtic Detectors for energy estimation". IEEE Trans. on Signal Processing, vol. 43(11), pp. 2582-2594. 1995.

88 J. H. Seinfeld. "Nonlinear estimation theory". Ind. Eng. Chem. vol. 62, pp. 32-42. 1970.

89 A. B. Gardiner. "Identification of process containing single-valued non-linearities". Int. Journal of Control, vol. 18, pp. 1029-1039. 1973.

90 G. Hung and L. Stark. "The kernel identification method- A review of theory, calculation, application and interpretation". Math Biosc., vol. 37, pp. 135-170. 1977.

91 J. F. Barret. "Bibliography on Volterra series, Hermite function expansions and related subjects". T. H. report 77-E-71. Eindhover University, 1979.

92 T. Koh and E. J. Powers. "Second-order Volterra filtering and its application to nonlinear system identification". IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-33 (6) pp. 1445-1455. 1985.

93 M. J. Korenberg and E. L. Morin. "Automatic discrimination of myoelectric signals via parallel cascade identification". Annals of Biomedical Eng., vol. 25, pp. 708-712. 1997.

94 M. J. Korenberg L. D. Paarmann. "Applications of fast orthogonal search: Time-series analysis and resolution of signals in noise". Annals of Biomedical Eng., vol. 17, pp. 219-231. 1989.

95 C. R. Arnold. and K. S. Narendra. "The Characterisation and identification of systems". Harvard University, technical report 471. 1965.

96 Y. W. Lee, and M. Schetzen. "Measurement of the Wiener kernels of a non-linear system by crosscorrelation". Int. Journal of Control. vol. 2, pp.237-254. 1965.

97 S. Y. Fakhouri. "Indetification of the Volterra kernels of nonlinear systems". IEE Proc. vol 127(6) pt. D. pp. 296-304. 1980.

98 S. Boyd, Y. S. tang, and L. O. Chua. "Measuring Volterra kernels". IEEE Trans. Circ. Sys. Vol. 30, pp. 571-577. 1983.

99 M. J. Korenberg. "Identifying non-linear difference equation and functional expansion representations: the fast orthogonal algorithm". Annals of Bio. Eng. Vol. 16, pp. 123-142. 1988

100 M. J. Korenberg, S. B. Bruder and P. J. McIlroy. "Exact orthogonal kernel estimation from finite data records: extending Wiener's identificatin of nonlinear systems". Annals of Biomedical Eng., vol. 16, pp. 201-214. 1988.

101 M. J. Korenberg. "Parallel cascade identification and kernel estimation for non-linear systems". Ann. Biomedical Eng. Vol. 19, pp. 429-455. 1991.

102 P. Z. Marmarelis and V. Z. Marmarelis. "Analysis of Physiological Systems. The white-noise approach". Plenum Press, New York. ISBN 0-306-31066-X. 1978.

103 M. J. Korenberg and I. W. Hunter. "The identification of nonlinear biological systems: Volterra kernel approaches". Annals of Biomedical Eng., vol 24, pp. 250-268. 1996.

104 W. J. Staszewski. "Gearbox vibration diagnostics. An overview". Proc. of COMADEM '96, pp. 95-111. 1996.

105 Alatass. "Experimental study of fault influence on vibration and noise measurements in a gear transmission mechanism". Int. J. gearing Conf. 1994, pp. 469-475. 1994.

106 D. Schönfeld. "Design and on-line condition monitoring of a model drive-line". University of Hertfordshire. Mechanical Engineering Department. BSc Final Year Project. 1995.

107 S. Engin . "Condition Monitoring of Rotating Machinery using wavelets as a pre-processor to artificial neural networks". University of Hertfordshire. Mechanical Engineering Department, PhD Thesis, 1988.

108 Von Mises. "Mathematical theory of probability and statistics." Academic Press, 1964

109 B. Paya. "Vibration condition monitoring and fault diagnostics of rotating machinery using artificial neural networks". Brunel university, Mechanical Engineering Dept, PhD Thesis. 1998.

110 E. P. Wigner. "On the quantum correction for thermodynamic equilibrium". Phys. Ver., vol. 40, pp. 749-759. 1932.

111 L. Jones and T. W. Parks. "A resolution comparison of several time-frequency representations". IEEE Trans. on Signal Proc., vol. 40 (2), pp. 413-420. 1992.

112 B. Boashash. "Note on the use of the Wigner distribution for time-frequency signal analysis". IEEE Trans. ASSP vol 36(9), pp. 1519-1521. 1988

113 F. Andrade and I. I. Esat. "A comparative study of signal processing techniques for monitoring and diagnostics". Proceedings of the American Society of Mechanical Engineers, ASME '97. Paper ref. 97-AA-93

114 J. C. P. Jones and S. A. Billings. "Describing functions, Volterra series, and the analysis of non-linear systems in the frequency domain". Int. Journal of Control, vol. 53(4), pp.871-887.

115 Y. Bissessur and R. Naguib. "Buried plant detection: A Volterra series modelling approach using artificial neural networks". Neural Networks, vol. 9(6), pp. 1045-1060. 1995.

116 G. Govind and P. A. Ramamoorthy. "Multi-layered neural networks and Volterra Series: the missing link". Proc. of the 90th IEEE Int. Conf. on Systems Engineering, pp. 633-636. 1990.

117 V. Volterra. "Theory of functionals and of integral and of integral and integro-differential equations". London: Backie & Sons Ltd.

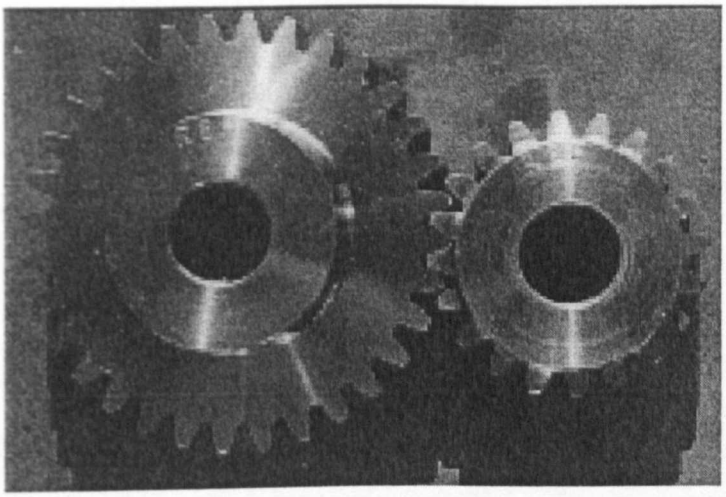
118 M. Schetzen, "Measurement of the Kernels of a Non-linear System of Finite order". Int. J. Control, vol. 1, pp. 251-263. 1965.

119 M. J. Korenberg and I Hunter. "The identification of non-linear biological systems: Wiener kernel approaches". Annals of Biomedical Eng., vol. 18, pp. 629-654, 1990.

120 J. A. Peacock. Monthly Notices of the Royal Astronomical Society, vol. 202, pp. 615-627. 1983.

121 G. Fasano, A. Franceschini.. Monthly Notices of the Royal Astronomical Society. vol. 225, pp 155-170. 1987.

Appendix 1 - Gear details



Parameter	*Pinion*	*Wheel*
Type	MA25-20S	MA25-32S
Number of teeth	20	32
Module	2.5	2.5
Face width [mm]	25	25
Pressure angle [deg]	20°	20°
Helix angle [deg]	0°	0°
Pitch diameter [mm]	50	80
Material - mild steel unhardened	EN8	EN8
Bore [mm]	90	90
Bore hole	H8	H8
Bore key [mm x mm]	6x6	6x6
General tolerance [μm]	±0.25	±0.25
Quality	"9gS	"9gS

These gears were manufactured to the following standards:

BS970.3:1991 Specification for wrought steels for mechanical and allied engineering purposes

DIN 3962-1 Tolerances for cylindrical gear teeth; Tolerances for deviations of individual parameters

DIN 3962-2 Tolerances for cylindrical gear teeth; Tolerances for tooth trace deviations.

DIN 3962-3 Tolerances for cylindrical gear teeth; Tolerances for pitch-span deviations.

DIN 3963 Tolerances for cylindrical gear teeth; Tolerances for working deviations.