Condition Monitoring of Gas Turbines using Acoustic Emissions

Shahram Shahkar

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science at

Concordia University

Montréal, Québec, Canada

April 2018

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Shahram Shahkar

Entitled:

Condition Monitoring of Gas Turbines using Acoustic Emissions

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. M. Kahrizi

_____ Examiner, External
Dr. Y. Zhang To the Program

_____ Examiner
Dr. M.O. Ahmad

_____ Supervisor
Dr. K. Khorasani

Approved by: _____
Dr. W. E. Lynch, Chair
Department of Electrical and Computer Engineering

_____20_____ _____
Dr. Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Acoustic emission (AE) technology has recently found its way in condition monitoring of rotary equipment due to its advantage of earlier detection of defects and anomalies in comparison to vibration analysis. However, there has been very little industrial application of AE signals for condition monitoring of safety-critical equipment if any, partly due to the difficulty in processing, interpreting and classifying the acquired data in a highly reliable fashion. The motivation in this thesis was to develop a methodology for inferring health related information in a gas turbine without intruding the engine.

Our work has targeted a broad class of rotary equipment known as cyclostationary processes, therefore, instead of analyzing particular AE samples of gas turbines we have tried to build a mathematical framework that would suit any arbitrary machine complying certain conditions. The result of our work mainly encompasses a feature extraction technique that eliminates the random effects associated with a gas turbine AE signal, and a hypothetical testing method for classification of AE signals with any desirable level of certainty, subject to a set of assumptions and conditions.

We have validated our methodologies and derivations using actual real-life gas turbine AE signals, and compared our solutions with some of the techniques published in the literature.

Contents

1 Introduction **16**

1.1 Condition Monitoring of Gas Turbines 18

1.2 Acoustic Emission Technology 23

1.3 Problem Statement and Contributions 29

1.4 Data Collection for this Thesis 34

2 Background Information **39**

2.1 Multiresolution analysis 43

2.2 Wavelet Packet Decomposition and Hilbert-Huang Transform (WPD-HHT) 57

2.3 Largest Lyapunov Exponent (LLE) 74

2.4 Approximate Entropy (ApEn) 85

2.5 Conclusion 89

3 Characteristics of a Gas Turbine AE Signal **93**

3.1 Cyclostationarity 94

3.2 Wide-Sense Stationarity 98
3.3 Mean Square Periodicity 101
3.4 Ergodicity 109
3.5 Convergence 114
3.6 Summary 119

4 Condition Space and Hypothesis Testing 123

4.1 Experiments 127
4.2 Hypothesis Testing and Feature Space 137
4.3 Numerical Validations 147
4.4 Summary 151

5 Machine Learning and Autonomous Condition Spacing 154

5.1 Practical Feature Extraction Methods 156
5.2 Condition Spacing 165
5.3 Machine Learning and Challenges (Future Work) 177
5.3.1 Vigilance Vector 182
5.3.2 Non-Homogenous Grids and Training Time 184
5.4 Summary 186

6 Summary and Conclusion 188

6.1 Summary of the Thesis 188

6.2 Future Direction of Research 196

Bibliography **198**

List of Figures

2.1 Decomposition of a signal into lower and higher subbands. 51

2.2 Decomposition of a signal into subbands. 53

2.3 Percentage energy of wavelet coefficients for $x_1(t)$. The greyscale bar on the right shows the magnitude of the percentage energy corresponding to each shade. 54

2.4 Percentage energy of wavelet coefficients for $x_2(t)$. 54

2.5 Autocorrelation of the discrete version of AE signals $x_1(t)$ and $x_2(t)$ for 1000 time samples (refer to page 32 for data sampling). 55

2.6 Percentage energy of wavelet coefficients for a gas turbine in a thermally transient state. 57

2.7 IMF1 to IMF3 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively. . . 64

2.8 IMF4 to IMF6 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively. . . 64

2.9 IMF7 to IMF9 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively. . . 65

2.10 IMF10 to IMF12 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively. . . 65

2.11 IMF13 to IMF15 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively. . . 66

2.12 Instantaneous frequency of IMF-1 and IMF-2 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds 67

2.13 Instantaneous frequency of IMF-3 and IMF-4 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds 67

2.14 Instantaneous frequency of IMF-5 and IMF-6 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds 68

2.15 Instantaneous frequency of IMF-7 and IMF-8 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds 68

2.16 Instantaneous frequency of IMF-9 and IMF-10 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds 69

2.17 WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-1. . 72

2.18 WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-2. . 72

2.19 WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-3. . 73

2.20 WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-4. . 73

2.21 λ_1 (equation (2.58)) of $x_1(t)$ 83

2.22 λ_1 (equation (2.58)) of $x_2(t)$ 83

2.23 λ_1 (equation (2.58)) of $[x_1(t), x_2(t)]$ 84

3.1 Time signal and autocorrelation function of three gas turbine AE Samples with different conditions. 96

3.2 Variance of frequency components $|X_k|^2$ of the first subplot of Figure 3.1. 106

3.3 Fourier transform of the third subplot of Figure 3.1. 107

3.4 Fourier transform of the last subplot of Figure 3.1. 108

3.5 Variations of the Fourier coefficients for a gas turbines AE signal. 116

4.1 Fourier transform of (a) 113 samples of a gas turbine AE signal under unchanging operating conditions (note the scattering), (b) Same 113 samples divided into 11 vectors, each averaged over 10 subsamples, (c) Same 113 samples divided into 4 vectors, each averaged over 28 subsamples, and (d) Average of all the 113 samples. It follows that as m is increased in equation (4.1), the features are converging more. 129

4.2 Feature vector estimate $\widehat{\mathbb{X}}$ of AE samples belonging to a pair of gas turbines (i.e., GT No. 1 and No. 2). For each gas turbine we have taken two vectors in different days with different signal lengths, under different climatic conditions. Plots belonging to the same machine are similar whereas plots of different machines are not. 130

4.3 Feature estimates (i.e., $\widehat{\mathbb{X}}_k$) of AE Samples taken from inside the Gearbox compartment of GT-2. Comparison of the powerful sources of energy with Fig. 4.2 indicates that the acoustic source at 2920Hz in Fig. 4.2 was belonging to a malfunction in the gearbox. 132

4.4 The first 1000 features of GT-1 AE samples when (a) GT-1 is operating under thermally stable condition (as in Example 2), (b) When GT-1 was in warm-up condition, i.e., 4 hours after the start-up, and (c) When GT-1 was in a cold state, i.e., 2 hours after the start-up. 135

4.5 Condition spaces of the gas turbines in the 2^{nd} and 3^{rd} subplots of Fig. 3.1, i.e., $\mathscr{R}_{\text{Perfectly Commissioned}}$ (depicted by '$\star$') and $\mathscr{R}_{\text{Poorly Commissioned}}$ (depicted by 'o') respectively, for a two-dimensional feature space $\mathbb{R}^2_{[0,1]}$ with $m = 1$. Each point represents a feature vector estimate $\widehat{\mathbb{X}}_i = [\widehat{\mathbb{X}}_i(1), \widehat{\mathbb{X}}_i(2)]^T$ for $i = \{1, 2, ..., 100\}$. 148

4.6 Condition spaces for the example illustrated in Figure 4.5 with a varying m. 149

4.7 Condition spaces for the experiments of Section 4.1. Assume that we have four conditions to monitor, namely a perfectly stable condition denoted by $\mathscr{C}_{\text{stable}}$ (represented by stars), an ageing effect denoted by $\mathscr{C}_{\text{ageing}}$ (represented by circles), a gearbox problem denoted by $\mathscr{C}_{\text{gearbox}}$ (represented by squares), and a rotor imbalance denoted by $\mathscr{C}_{\text{start-up}}$ (represented by diamonds). For each condition the border is illustrated with a circle. In case a feature vector estimate falls inside a circle it can be inferred that the machine is in the corresponding condition. 150

5.1 Feature extraction block diagram. Commands illustrated by solid arrows to the integrators are used to reset the integrating units to zero, every time the switch is opened. 157

5.2 Filtering a signal $x(n)$ into 8 symmetric sub-bands $x_1(n)$ to $x_8(n)$ using the wavelet decomposition/reconstruction blocks (as illustrated in Fig. 5.3).159

5.3 Multi-rate sub-band filtering block. 160

5.4 Wavelet FIR filters 161

5.5 The left hand side plot is the input to the WDR block ($x(n) = \delta(n)$) as shown in Fig. 5.3. The right hand side plot is the output of the block $h(n) = x_R(n)$. 162

5.6 The time-domain output of FIR wavelet filters (X_{DH}, X_{DL}, X_H, X_L shown in Fig. 5.3) to an impulse input $x(n) = \delta(n)$, and corresponding frequency domain responses (i.e., Fourier Transforms $\mathscr{F}(X_{DH}), \mathscr{F}(X_{DL}), \mathscr{F}(X_H), \mathscr{F}(X_L)$). The x-axis of the frequency response plots represent normalized frequency and the y-axis represents the gain. 162

5.7 Frequency response of the 8-Channel filter of Fig. 5.2. In this figure $H_i = \mathscr{F}(x_i(n))$ (i.e., the Fourier transform of x_i), when the filter input $x(n) = \delta(n)$. The x-axis in all plots represents normalized frequency and the y-axis represents the gain. 163

5.8 Frequency response of the sum of all channels (i.e., $\sum_{j=1}^{8} H_j$) of the multi-rate filter of Fig. 5.2. 163

5.9 Wavelet FIR Filters of the order 150, as an alternative to the multi-rate filtering architecture proposed in Fig. 5.2, from high frequency to low frequency. The plots are x_1 to x_8 in Fig. 5.2 (upper left side plot to the lower right side plot) when the input in $x(n) = \delta(n)$ 164

5.10 ART-1 network architecture. Rectangles represent fields where Short Term Memory (STM) patterns are stored. Semi-circles represent adaptive filter pathways, and arrows indicate paths which are not adaptive. Filled circles represent gain control nuclei, which sum input signals. Output paths are not specified in the sense that at any given time a uniform signal is sent to all nodes in a receptor field. Gain Control F_1 and F_2 coordinates STM processing with input presentation rate [78]. 181

Nomenclature

$\mathscr{C}_{x(t)}$	Machine condition for $x(t)$.
$C_i^m(r)$	Similarity of a segment of a signal with length i to the entire signal with length m.
C_x	Covariance function of $x(t)$.
$c_i(n)$	i^{th} level coefficient of wavelet transform.
$d_i(n)$	i^{th} level detail of wavelet transform.
$E_{f(n)}$	Energy of the discrete function $f(n)$.
$E[\cdot]$	Expected value of a random variable.
$\widehat{E\left[\lvert X_k\rvert^2\right]}$	Average of m samples of $\lvert X_k\rvert^2$.
F_s	Sampling frequency.
$\hat{f}(t)$	Hilbert transform of $f(t)$.
$\mathscr{F}(x(t))$	Fourier transform of $x(t)$.
j	The unit imaginary number $(-1)^{1/2}$.
λ_1	Largest Lyapunov exponent.
m	Number of samples for estimating the feature vector.
$m_X(t)$	Mean of the stochastic process X at point t.
μ^+	Upper condition space border.
μ^-	Lower condition space border.
N_C	Number of desirable conditions in a condition monitoring application.

N	Dimension of the feature vector.
$\mathcal{N}$	Gaussian distribution.
$\phi_{i,k}(t)$	Scaling function with scale i and translation k.
$\psi_{i,k}(t)$	Wavelet function with scale i and translation k.
$p_{\widehat{\mathbb{X}}\|\mathscr{C}_{x(t)}}$	Probability distribution function of $\widehat{\mathbb{X}}$ under condition $\mathscr{C}_{x(t)}$.
$\Phi(\cdot)$	Cumulative distribution function for the Gaussian distribution.
P_{FA}	Probability of false-alarm rate.
$P[\mathscr{C}_i\|\mathscr{C}_j]$	Probability of detecting condition $\mathscr{C}_i$ when the machine is in condition $\mathscr{C}_j$.
$p(k)$	Discrete real valued low-pass filter.
$p'(k)$	Discrete quadrature mirrored low-pass filter.
$q(k)$	Discrete real valued high-pass filter.
$q'(k)$	Discrete quadrature mirrored high-pass filter.
$R_x(t_1, t_2)$	Autocorrelation function of $x(t)$.
$\bar{R}_{[x(t)]_{T_x}}(\tau)$	Mean of $R_{[x(t)]_{T_x}}(\tau)$ samples.
$\mathbb{R}$	The set of real values.
$\mathbb{R}^N_{[0,1]}$	N dimensional feature space.
$\mathscr{R}_{\mathscr{C}_{x(t)}}$	Condition space associated to $\mathscr{C}_{x(t)}$.
ρ_{x_1,x_2}	Correlation coefficient between x_1 and x_2.
$\sigma^2_{x(t)}$	Variance of a random process $x(t)$.
$\dot{s}(t)$	Partial differentiation of $s(t)$ with respect to t.

T	Period of the AE cyclostationary signal $x(t)$.
T_x	Sampling duration of the AE signal $x(t)$.
T_p	Duration in which the machine condition of $x(t)$ remains constant.
$W_f(i,k)$	Wavelet coefficients of function f.
$X(\omega)$	Fourier transform of $x(t)$.
$x^{(i)}_{\text{IMF}}(t)$	i^{th} level intrinsic mode function of $x(t)$.
$\big[x(t)\big]_{T_x}$	A random segment in $x(t)$ with length T_x.
$\langle x(t)\rangle_{T_x}$	Time average of a random process $x(t)$ for the duration T_x.
X_k	k^{th} Fourier coefficient of the discrete signal $[x(n)]_{T_x}$.
$\lvert X_k\rvert^2$	Energy of the k^{th} subband of $[x(n)]_{T_x}$.
$x^*(t)$	Complex conjugate of $x(t)$.
$\mathbb{X}_k$	k^{th} element of the feature vector of $x(t)$.
$\mathbb{X}$	Feature vector of $x(t)$.
$\widehat{\mathbb{X}}\big\vert\mathscr{C}_{x(t)}$	Feature vector estimate of $x(t)$ taken for condition $\mathscr{C}_{x(t)}$.

Chapter 1

Introduction

Condition monitoring (CM) as defined by Wikipedia [60], is the process of monitoring physical parameters of a rotary machinery in order to identify a significant change beyond normal operating circumstances, which is indicative of a developing fault, or a shortening normal lifespan. Condition monitoring techniques discussed in this thesis however deal with estimating the health condition of a rotary equipment through (continuously) analyzing acoustic emissions (AE) of the respective system under service, through a set of microphones that are mounted exterior but in a close proximity to the equipment emitting noise. Although acoustic signals emitted from rotary equipment are generally of a random nature, in the meantime exhibit the average behaviour of the machine, just like the body temperature can indicate the onset of an infection or a virus in the human body. In the above analogy one can realize that the body temperature, changing

randomly and instantaneously in effect of numerous biological interactions, still project symptoms of an illness when measured on an average manner, whereas instantaneous measurements (individually) entail very little (and often unreliable) information, if any. Therefore, in this thesis we will introduce another general notion for *machine condition*. We will later see that the scope of derivations and results of this thesis encompasses mechanical equipment of a *cyclic* nature only. Therefore, we make this notion explicit below.

Definition 1.1: *The average behaviour of an acoustic signal emitted from a cyclic mechanical equipment, under a fairly constant set of inputs, is hereinafter regarded as a 'machine condition', or simply a 'condition' and is denoted by* $\mathscr{C}_{x(t)}$.

Definition 1.2: *Definition 1.2: Quantitative characteristics of an AE signal that uniquely represent machine conditions are referred to as features. A feature vector is a vector that comprises a set of features.*

We will define a feature vector in Chapter 4 and will show that it can uniquely represent a machine condition based on the Definition 1.1 (through Theorem 3.1), and also distinguish anomalies in the system with respect to a benchmark, being the average behaviour of that system under the 'healthy' state.

1.1 Condition Monitoring of Gas Turbines

According to [76], guidelines for the production of condition monitoring systems applied to gas turbine engines have been available for over thirty years. Engine monitoring is performed using either 'on-line' systems, mounted within the engine, that perform analysis of engine data, or 'off-line' systems to which engine data is recorded and processed later.

On-line condition monitoring systems for flight applications are subject to strict certification requirements from international aviation authorities. The cost of producing systems to these required standards is very high, primarily due to the technical challenge of suppressing the large numbers of false alarms generated by many condition monitoring algorithms. Methods currently used in on-line condition monitoring are relatively simple due to these constraints.

Off-line condition monitoring systems are limited by the amount of data being recorded. Typically, data storage and transfer costs are high, which constrains the quality and quantity of data available to commercial offline monitoring systems. Standard methods of gas turbine CM involves comparison of recorded engine parameters with

fixed operational limits set by engine manufacturers, or simple univariate analyzes in which trends in engine data are monitored in order to detect sudden shifts in operational behaviour, often indicators of potential failure. Current research on the condition monitoring of aircraft engines is focussed on providing more sophisticated methods of reliably determining engine condition and resulting maintenance recommendations, within the constraints of on-line and off-line monitoring systems.

Expert systems aim to capture the knowledge of domain experts typically from engine manufacturers, in order to diagnose faults from engine data and recommend appropriate maintenance actions [76]. The accuracy of diagnosis may be improved through the construction of ever larger databases of fault information for gas turbine engines [56]. Case-based reasoning (CBR) uses the intuition that solutions to new problems are likely to be similar to previously encountered problems of a similar nature [28]. In this approach, a user friendly software environment is utilized to capture expert knowledge and manage fault diagnoses. This can then be used to provide coordinated construction of expert-system databases for monitoring engines in civilian aircraft between many different sites, using a framework for distributed Internet based data analysis.

CBR has also been applied to the management of the maintenance process for gas turbine engines in military aircraft. However, the CBR approach is limited by the

variation in operational behaviour between classes of aircraft engine, and even variation between engines of the same class, which can render the use of general fault databases ineffective when diagnosing engine specific faults. To avoid costs associated with re-constructing expert system databases for use with individual engines, the automated generation of the logical rules in the expert-system databases using 'soft computing' techniques has been proposed [76]. 'Fuzzy' expert systems consisting of fuzzy logic rules for characterizing engine specific faults have been combined with data mining techniques in which rules are formulated from trends in the value of measured engine parameters [16]. Performance parameter measurements are divided into segments modelled using linear regression, and described using clauses of fuzzy logic. Rules for diagnosing engine faults are formed using these fuzzy descriptions of each segment.

Within a conventional expert system, neural network methods have been used to match features detected in performance parameter signals with diagnosis rules [30]. These methods have also been applied to fault diagnosis of space shuttle engines, and to military armour. Here, diagnoses are obtained by combining sets of rules (comparing sensor output to fixed operational thresholds) and the outputs of a neural network classifying engine operation into one of several states (idle, full power, etc.). The combination of expert system databases from multiple engines of the same class has been attempted, aiming to generate systems that might generalize to other engines of that class.

Thermodynamical models of gas turbine engines have been used for condition monitoring by comparison of a model derived from actual engine data (usually temperatures and pressures) with a model constructed from the theoretical relationship between those engine data [32]. The residual error between the two models may be used to identify possible engine failures. Such methods combine a theoretical approach to thermodynamical model construction with knowledge derived from engine testing. Models derived using finite element analysis have been used in the condition monitoring of helicopter engines [24].

Linear flow models have been constructed from fuel-flow data and applied to condition monitoring of the engine combustion system [43]. Frequency-domain approaches to modelling engine parameters have been used in fuel-flow monitoring, exhaust gas monitoring, and more general diagnosis of engine faults [57]. These models have been combined with autoregressive moving average (ARMA) models for fuel flow monitoring, and for analysis of engines used in industrial plants [37].

Determining engine condition by comparison of the engine data with a model of normal behaviour may also be achieved by using 'soft computing' methods of model creation, such as neural networks, or using statistical pattern recognition techniques.

These methods offer the advantage that engine condition may be assessed without detailed theoretical system models.

These data-driven approaches can be used to construct models of engine behaviour directly from engine data, in contrast to rule-based approaches that encode expert knowledge, and may be adapted for use with individual engine units, providing engine-specific condition monitoring. However, these methods can be sensitive to the quality and quantity of example data from which models of engine operation are constructed.

Neural networks have been used to identify abnormalities in the operation of gas turbine engines within industrial plants, in space shuttle engines, and in engines used by heavy ships [44]. The support vector machine (SVM) method, a statistical pattern recognition approach, has been used to identify examples of 'abnormal' engine vibration behaviour from a 'normal' training set in aircraft engine production tests, and in water-pump gas turbine engines [53].

These one-class classifiers learn a model of normality from a training set of normal examples, and identify patterns that fall outside that class as 'abnormal', in comparison with multi-class classifiers, which require approximately equal numbers of examples for

each class. Other methods of constructing models of normal engine behaviour within engine production tests using data-driven techniques have been successfully demonstrated using data from military aircraft engines [10]. Automated extraction of patterns describing normal and abnormal gas turbine engine operating conditions with engine vibration data have been investigated, within a general fuzzy-neural analysis framework [41].

1.2 Acoustic Emission Technology

'Acoustic Emission (AE) is defined as the range of phenomena that results in the generation of structure-borne and fluid-borne (liquid, gas) propagating waves due to the rapid release of energy from localized sources within and/or, on the surface of a material' [81].

Acoustic Emission (AE) signals are recognized as a complementary measure for detecting incipient faults in rotary machinery due to their containment of sources of potential fault energy, in a sensitive way. However, determination of the potential sources of faults cannot be easily realized due to non-linearity and non-stationarity of the AE signals. Industrial condition monitoring techniques of gas turbine engines are mostly based on vibration and thermal analysis of certain parts in conjunction with simultaneous measurements of the system performance.

Vibration sensors are supplied and equipped on the bearings and/or casing, and on the gearbox in the form of proximity sensors, velocity meters and accelerometers, depending on the required range of spectrum of the vibration frequencies. The fluid temperature is periodically measured through thermocouples at certain locations, such as the compressor inlet/outlet as well as turbine wheel spaces and at the exhaust. Meanwhile, pressure transducers process fluid pressure where collective processing of the parameters yield decisions about the overall health of the machine.

However, condition monitoring is counted as an invasive and disturbing process by machine manufacturers, due to the added complexity, and the inherent intrusion of additional sensors in machine components. It would therefore be highly desirable and advantageous if one could devise a condition monitoring system through an external and lateral unit and outside the main engine.

Acoustic emission signals are remarkable candidates especially if the signals could be monitored through microphones or sensors located outside the engine compartment. There are publications already about the usage of AE sensors mounted inside gas turbine engines, aiming early fault detection of turbine shaft cracks, blade rubs, or bearing deteriorations, as introduced in [11].

Douglas *et al.* [1] have reported on potential use of AE signals for detection of major rotary components on steam turbines by Armor *et al.* [2]. Sato [3] proposed a diagnostic system for a steam turbine that can detect rotor-stator rubbing and bearing faults, from AE acquired at the bearing housings, and Mba *et al.* [4] concluded that a defect at the shaft-seal tip of a steam turbine can be detected by an AE transducer mounted on the bearing housing, also observed a relationship between AE signals and load on a steam turbine unit, and indicated that a rotor passing through its critical speed can be identified. Zuluaga-Giraldo1 *et al.* [5] were among the first who documented a correlation between the vibration/load and AE activities on a power generation steam turbine application.

Douglas *et al.* [1] later repeated similar experiences on a laboratory scale gas turbine and identified blade passing frequencies, and also demonstrated speed/load correlations with AE activities through sensors located along the gas flow path. Nashed *et al.* [6] combined statistical time domain and frequency domain features and revealed correlations between speed/load (and implicitly the fuel flow rate) and the AE signal intensity on a laboratory scale gas turbine with the impeller locked or removed.

These experiments indicate that in turbomachinery with relatively high rotating speeds, AE signals contain potential information about early fault stages of different

components. Moreover, there have been significant efforts to evoke appropriate features from AE signals that can correctly represent the health state of a machine under different operating conditions. Purarjomandlangrudi *et al.* [7] categorized principal time-domain statistical parameters, as root mean square (RMS), skewness and Kurtosis, and concluded that a pickup in these quantities can indicate a damage in wind turbines. He and Wu [8] suggested that wavelet packet technique can extract the pulse signal representing fault source characteristics, and could obtain effective failure information from the AE signals in steam turbine rotor cracks. When the wavelet packet technique is applied to the actual acoustic emission testing system, it will greatly improve the relevant technical indicators such as the rate of fault coverage, fault diagnosis and would reduce false alarming rates.

Farhat and Al-Taleb [9] conducted a study on industrial gas turbines in a power generation project and investigated the combustion stability that is supplied with lean fuel, and reported the power spectrum, auto correlation and the acoustic wave shape produced by the thermo-acoustic instability as the most effective features for analyzing the *'humming'* effects. Concerning the feature extraction techniques that are used in health monitoring using AE, Caesarendra *et al.* [11] published a comprehensive collection of available methodologies for low speed rotary components, and broadly categorized the techniques into two classes, namely (1) AE hit parameters (which is broadly referred to

as the occurring of a particular parameter in a signal, such as a peak or RMS beyond a certain amplitude), and (2) AE signal processing feature extraction methods, and pattern recognition. Among the AE hit parameters, amplitude and RMS monitoring of the AE signal has found many applications [13].

In some applications the AE energy (the integral of the rectified voltage signal over the duration of the AE hit) was used as a feature [14]. 'Counts' referred to as the number of detected AE signal excursions over a predetermined threshold of the AE signal, and 'Events' termed as the phenomenon which releases elastic energy into the material, which then propagates as an elastic wave, are other forms of features that are used in the literature. Moreover, the reference [17] defined the Average Signal Level (ASL) as a measure of the continuously varying and 'averaged' amplitude of the AE signal, and finally the AE Burst, a qualitative description of the discrete signal related to an individual emission event that occurs within the material were also considered in the literature as AE features.

In the field of signal processing, AE features were mainly considered to be Auto-regression (AR) coefficients [18], [19]. Reference [20] introduced the Energy Index (EI) technique that was used in detecting masked AE signatures associated with the loss of mechanical integrity in bearings and investigated the efficiency and applicability of the

technique at low signal-to-noise ratios. It was shown that the EI technique is effective in detecting AE burst that is buried in random noise.

He *et al.* studied feature extraction of acoustic signals based on Complex Morlet Wavelet [21]. Other feature extraction techniques that are referred to in the literature are the Short-time energy function, the short-time average zero crossing rate and median smoothing [22]. The peak ratio (PR) [23], the zero-inflated poisson (ZIP) regression [25], Mahalanobis distance [26] and the ratio of AE mean and the AE standard deviation [27].

The aforementioned techniques have unfortunately not addressed the AE condition monitoring of a gas turbine engine in its entirety (in contrast to component-based approaches), firstly because the lack of a universal feature extraction measure, and secondly due to the lack of a criterion for detecting a condition change. The next chapter will examine some of the nonlinear and nonstationary feature extraction methods. We will investigate the applicability of these methods for feature extraction from gas turbine AE samples, and will show the shortcoming of each method. Then in the succeeding chapters we will introduce a remedy to the problem. Before that let us explicitly represent the problem statement, and the contributions of this thesis.

1.3 Problem Statement and Contributions

Machine condition monitoring consists of three key steps: (1) data acquisition, (2) feature extraction, and (3) condition identification. It is therefore a procedure of mapping the information that is obtained in the measurement space (AE signals) to the features in the *feature space* (to be precisely defined in Chapter 4), and finally to machine conditions in the *condition space* (to be precisely defined in Chapter 4) [87]. Hence, it is essentially a problem of pattern recognition and classification (often referred to as clustering analysis).

Cluster analysis, as a multivariate statistical analysis method, is a statistical classification approach that groups signals into different (condition) categories on the basis of the similarity of the characteristics (or features) they possess. It seeks to minimize within-group variance and maximize the between-group variance [15]. The result of the cluster analysis is a number of heterogeneous groups (i.e., groups of samples with non-similar features), with homogeneous contents (i.e., contents with similar features).

Since in this work AE signal represents the machine condition, changes in the features associated with an AE signal also represent changes in the machine condition. From these changes, experts can make deductions about the underlying conditions of a machine, where the ultimate goal is to replace the human experts with computerized

systems. However, AE signals from the rotating and reciprocating machinery can be *non-stationary* (to be formally defined subsequently), implying that 'the shorter the time waveform is used to calculate statistical parameters, the greater the likelihood that the parameters do not accurately describe the underlying signal' [82]. Even an average of the parameters calculated from a large ensemble of waveforms may not be particularly useful, as it does not describe the spread of a parameters' values. Hence, the calculated average will also depend on the number of waveforms in the ensemble and the duration of each [82]. Consequently, the random nature of the AE signals entails reliability concerns to the decisions that are rendered through standard AE CM techniques.

As a solution to non-stationarity of AE signals, researchers have devised methods to quantify the non-stationarity associated with an AE signal and calculate the confidence limits in addition to notions of their ensemble averages [83]. Changes in the underlying machine condition are therefore observed as changes not only in the mean parameters, but also their confidence intervals. However, the problem still lies in the fact that in order to deduce safe decisions pertaining to a machine condition a large number of samples would be required to compute an ensemble average, which is generally not always feasible or practical.

Moreover, interpretation of the confidence limits depend on the associated parameter. A wide confidence interval for continuous signal features (e.g. root mean square (RMS), energy, etc.) implies that the continuous AE level is non-stationary (back-ground signal level is fluctuating significantly). A similar confidence interval for burst features (e.g. maximum amplitude) can indicate that, either few hits are occurring per unit time (and thus the underlying signal is relatively stable), or there is significant variability in the hits themselves [82]. Consequently, multiple parameters may need to be considered collectively to make accurate decisions (e.g. average amplitude, amplitude confidence interval, and hit count). Therefore, another challenge is that one has to assign appropriate weights to individual parameters in order to attain reliable features of a signal [87].

Therefore, more advanced techniques are widely being used in various condition monitoring fields to manage non-stationary signals [84]- [86]. Unfortunately, these methods are (generally) computationally intensive and the interpretation process is difficult to automate (as will be shown in the next chapter).

Neural networks can also be used to find patterns and correlations in the data that may not be otherwise detectable due to non-stationarity [82]. However, a large and representative volume of data is required to train the network adequately. Furthermore,

some researchers believe that making sense of neural network outputs is not as straightforward as it may sometimes seem [82].

The most widely quoted benefit of AE monitoring in the literature is its apparent high signal-to-noise ratio (SNR) [82]. It is true that typical process machine noise in the audible frequency range has limited impact on its AE signal. However, noise refers to more than just extraneous sound. It is widely defined as 'any unwanted part of the signal (or dataset)' and therefore in order to ensure valid CM conclusions, noise needs to be eliminated (or quantified) at the time of data collection or during post-processing. A variety of techniques have been developed over the past 20 years to de-noise condition monitoring data, with varying degrees of success when applied to AE signals. This is in part due to what constitutes noise and what is regarded as genuine data changes from one application to the next. Therefore, it is impossible to develop a single set of processing techniques that will be applicable in all situations [88].

In view of the above discussions, the main contributions of this thesis can be listed as follows:

1. There are numerous successful feature extraction techniques that are used for AE condition monitoring of certain applications. However, variations in the available techniques imply that no universal solution exists that could ensure successful

results for a wide range of applications [89], even under a specified set of conditions [88], [11]. Although, the feature extraction technique introduced in this thesis may have been successfully applied in some particular cases in the past [15], this work attempts to stipulate conditions (Assumption 3.2) under which our proposed methodology can guarantee to produce optimum results (Theorem 3.1), regardless of the application.

2. This work underlines conditions under which a gas turbine AE signal of sufficient length can be used to deduce properties of an ensemble of samples. It further shows that the features to be obtained based on the proposed procedures derived in this thesis could be free of all random effects (if certain conditions are met). This could be considered as a remedy to deal with a certain class of non-stationary signals, as addressed in [82].

3. Contrasting with the feature extraction techniques that are available in the literature, not much research has been done in clustering analysis [11]. The available CM methods still challenge to render a seamless procedure to evoke and classify features with a desirable level of confidence [15], [88]. Although our work encompasses certain conditions to be met, we have proposed a classification system with a definable level of certainty, (virtually) capable of classifying machine conditions with practically no error.

Not withstanding the above, the proposed methodologies have certain shortcomings that

is left for a future work, as addressed in Section 6.2.

1.4 Data Collection for this Thesis

This thesis is a data driven based methodology for CM and its contributions are resulted from real life gas turbine AE data processing. The main idea of the thesis was to develop a procedure for condition monitoring of gas turbine engines without adding additional instruments to the gas turbine. It was also intended that mobile devices be used for recording AE signals and processing the data, so that CM could be realized through portable data processing modules such as mobile phones, fully online.

Therefore, we have used the microphones embedded in ordinary mobile phones for digitizing and recording AE data, using a sampling rate of 44,100 Samples/Second . The personnel taking the recordings had no prior training, nor did they follow any specific procedure for taking the recordings. We had asked them to stand in the vicinity of the gas turbines when recording the data pointing the microphone toward the engine, so that ambient noise would be less interfering.

Every AE recording has taken 3 to 7 seconds time, when the condition of the gas turbine could be assumed fairly constant. The geometric location of the microphone was also selected randomly in order to assure that the data processing method is not

sensitive to any particular location, except that we prohibited positions close to miscellaneous sources of loud noise.

In order to assure a reliable CM process we have chosen identical gas turbines, in brand, type, and size for all our AE samples. The author had access to several power generating facilities where identical Hitachi H-25 Gas Turbine and Generators had been installed and put into commercial operation. The gas turbines had been commissioned on different dates, hence in different age and had a differing operational history, as explained below.

In total 7 Hitachi H-25 gas turbines were pinpointed from two sites as listed below:

- **4 H-25 Gas Turbines (i.e.,** $4 \times 30MW$**)** belonging to a power generating facility in west of Iran, commissioned between 2013 to 2014

- **2 H-25 Gas Turbines (i.e.,** $2 \times 30MW$**)** belonging to an industrial power generating facility in Tehran, the capital of Iran, commissioned in 2012

- **1 H-25 Gas Turbines (i.e.,** $1 \times 30MW$**)** belonging to the same industrial power generating facility in Tehran, the capital of Iran, commissioned in 2015

All 7 gas turbines were simple cycle units with open exhaust, consisting of a 3 stage gas turbine with a 17 stage compressor unit, coupled to a generator (made by Brush HMA, Netherland) through a gearbox and a hydraulic coupling. All 7 gas turbines were fired

by natural gas and connected to the main electricity grid of Iran.

The gas turbines located in west Iran were set on droop mode, loaded according to the demand controlled by the dispatch control center of Iran, whereas the industrial units (also set on droop) were controlled through the local control center inside the industry. All units were mostly operating part load.

AE data were collected in a 4 month period between November 2015 and March 2016. In order to collect the data, people had to be dispatched to the site with prior coordination with respected authorities that often took weeks of paperwork. At the time of recording representatives from the Owner had to attend and the data were then handed over for inspection before being released. Therefore, the data collection process was expensive and very time consuming, so the author could not manage to take numerous samples for the thesis.

In total 17 audio files were emailed from Iran to the author who was in Montreal in 2015, that were recorded from the gas turbines listed above. We intended to capture as much conditions as possible, therefore we have taken audio signals of gas turbines with different aging factors, although almost all were brand new engines. Also we have tried to emulate conditions that were impossible to capture at that time. For instance,

we have taken AE recordings when a particular gas turbine was put into service after a long rest. Such data were taken when the unit had not reached thermal stability and the rotor still had mild bow due to the standing weight resulting the resting period. Obviously as the rotor warms up, the bowing effect disappears due to the high temperature of the main shaft and the very high centrifugal forces resulting from high rotational speed (around 7200 rpm). Hence we were able to emulate a mild unbalance condition, artificially.

We were also very lucky to capture a temporary problem in the gear box of a unit installed in west Iran site. This was a very mild anomaly caused by the build up of debris and foreign particles in the teeth of the gearbox. It is recommendable that the hydraulic system including the pipings interconnecting the fin fan cooler to the gas turbine are flushed with oil after the welding is complete, and then complete draining of the oil and refilling the oil tank after internal cleanup with new oil. This is a very expensive burden for the contractors due to the price and large quantity of oil required. The contractors often neglect the recommendations and tend to filter the oil after commissioning and reusing the first lot of oil for operation.

This had caused particles that could not be perfectly filtered out move and stuck in the labyrinth between the gear teeth and cause very mild vibrations in the gearbox,

entirely beyond the range of embedded vibration sensors, or the thermocouples measuring the oil temperature flowing outside the gearbox. This problem was first identified by the AE condition monitoring technique developed in this thesis and was rectified by changing of the oil.

Apart from the gearbox effect and the emulated unbalance, we were able to monitor the ageing condition by comparing the AE of gas turbines having different ages. In all these cases we have tried to maintain equivalent operating conditions to capture the ageing factor only. However insufficient amount of data has not allowed us to extend our results beyond Chapter 5 and we are hopeful that this research could be progressed by access to other sources of real data in the future.

Chapter 2

Background Information

The core part of the condition monitoring process encompasses deploying methods and techniques that can best extract unique characteristics and features of the particular condition of the machine embedded in the acoustic signal. Indeed it is these characteristics and features that distinguish different machine conditions, and allow the condition monitoring process to take place. In our application feature extraction involves reducing the amount of resources required to describe the often large set of temporal samples obtained from the acoustic waveforms, and characterize their oneness among a set of given waveforms. We have briefly named the most common of these methodologies in Chapter 1, used in the domain of feature engineering, without deeply investigating their details. In this chapter, we will contemplate more advanced methods that are applied in condition monitoring feature extraction, and will investigate the effectiveness of each

for the problem in hand, i.e., gas turbine acoustic emissions. Before that let us look into some important characteristics of the feature extraction method we apply, and the notion of feature itself.

The desirable feature should have the lowest possible dimension without sacrificing important information contained in the signal. This is an essential attribute for an appropriate feature vector so that further processing would not entail huge computational cost. Additionally, a vector of large dimension might involve redundant (unnecessary) information, overlooking essences that makes differences between separate AE samples. In fact, we envisage the set of minimal quantities that evoke all the differences between given signals, in the most discriminative fashion.

An important characteristic of AE signals that is often addressed in the literature is non-stationarity. By definition, a *stationary* signal is one that its properties do not change in the course of the signal [65]. In stochastic theory, a stationary signal in its strict sense is one whose statistical properties are invariant to a shift in the time axis. And in wide sense it is a process whose second-order statistics depend only on the time difference.

Generally for a stationary process it is possible to study the behaviour of the

process through its spectral analysis. A gas turbine AE signal as it will be seen in a while, tends to be a short time *cyclostationary* stochastic process. A stochastic process X is said to be cyclostationary if

$$m_X(t + mT) = m_X(t) \tag{2.1}$$

and

$$R_X(t_1 + mT, t_2 + m'T) = R_X(t_1, t_2) \tag{2.2}$$

where m_X and R_x denote the mean and autocorrelation function of a process X respectively, and T is a cycle which the process repeats, and m and m' are real integers. This can physically make sense if we consider that the mechanical and thermodynamic processes taking place in a gas turbine repeat themselves every time the rotor travels a complete cycle. As such, we expect T being the periodicity in equations (2.1) and (2.2) to also be the time period of the rotor speed, which implies the process X being an *ergodic* process .

In probability theory, an ergodic dynamical system is one that, broadly speaking, has the same behaviour averaged over time as averaged over the space of all the system's states in its phase space [60]. Now considering the process X being a cyclostationary ergodic process, we can assume that taking two subsamples of equal length from $x(t)$ at time instants t_1 and $t_2 = t_1 + mT$, is statistically equivalent to taking two different AE samples $x_1(t)$ and $x_2(t)$ from the same process X. These two important conditions,

i.e., ergodicity and cyclostationarity are the two basic properties on which this thesis is based upon. We will however show in the subsequent chapters that ergodicity can be resulted from cyclostationarity under certain conditions of AE data acquisition.

Another characteristic of gas turbine AE signals is what is often referred as *non-linearity*. Unlike linear dynamical systems where the system has constant natural frequencies and vibrates at the frequency of an externally applied harmonic excitation, in a nonlinear system there are amplitude-dependant natural frequencies called 'internal resonances' at which the system may vibrate, completely different from an externally applied harmonic excitation [69]. The technique of Fourier Transform has been dominated in the field of signal analysis because of its prowess and simplicity. However, there are some crucial restrictions on the use of Fourier transform for feature extraction of nonlinear signals.

The signal generated by the inspected machine must be linear and temporally stationary; otherwise, the resulting Fourier spectrum will have little physical sense. Unfortunately, in machine condition monitoring, the signals to be analyzed are often non-stationary. The resulting frequency components are not always consistent as the patterns of the acquired AE signals often change with time. Therefore, Fourier Transform cannot fulfil the requirements of condition monitoring, particularly in real applications [63].

Hence, in this chapter we will first focus on some prominent signal processing techniques that are applicable to nonlinear and nonstationary signals, and see how well they can perform feature extraction of gas turbine AE signals. Then in subsequent chapters we will see that wide sense stationary samples of AE signals can be obtained, based on the assumption of cyclostationarity, that are indeed linear and ergodic.

2.1 Multiresolution analysis

Acoustic emission is the transient elastic waves within a material, caused by the rapid release of localized stress energy. An event source is the phenomenon which releases elastic energy into the material, which then propagates as an elastic wave. Rapid stress-releasing events generate a spectrum of stress waves starting at 0 Hz, and typically falling off at several MHz [60]. Therefore, we can assume that event sources are detectable through the frequencies their stress waves propagate at, and a particular phenomenon can be potentially identifiable by decomposing the AE signal into frequency levels (subbands), and looking for signatures of that particular phenomenon in its associated subband.

The more levels the signal is broken into, the finer would become each subband, hence multiresolution analysis. This is in contrast to the Fourier analysis where the signal is decomposed into its frequency components as opposed to frequency subbands. It therefore transcends Fourier analysis due to its information packing ability that renders

lower dimensional feature vectors. It is worth noting that Fourier analysis precisely extracts frequency content of a signal only when it is analyzed infinitely in time. Taking shorter chunks of a signal and performing a Fourier transform inevitably invites noise and incorrect information. Specifically, let us suppose that the Fourier transform of a given time signal $x(t)$ is $X(\omega)$, that is

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt \tag{2.3}$$

In (2.3) the upper and lower bounds are stretched to infinity which accurately provides the frequency content of $x(t)$ in $X(\omega)$. However, focusing on a limited extent $[a, b]$ gives a revised version of the Fourier components of $x(t)$ as

$$X(\omega) = \int_{a}^{b} x(t)e^{-j\omega t}dt = \int_{-\infty}^{+\infty} w(t)x(t)e^{-j\omega t}dt \tag{2.4}$$

where $w(t)$ is the window function that is defined as

$$w(t) = \begin{cases} 1 & a \leq t \leq b \\ \\ 0 & elsewhere \end{cases} \tag{2.5}$$

Equation (2.4) is by definition the short term Fourier transform (STFT) of $x(t)$ and is related to the Fourier transform of $x(t)$ by:

$$w(t)x(t) \xrightarrow{\mathscr{F}} W(\omega) * X(\omega) = \frac{sin(\frac{\omega\tau}{2})}{(\frac{\omega}{2})}e^{j\omega\frac{\tau}{2}} * X(\omega) = \tau Sinc(\frac{\omega\tau}{2})e^{j\omega\frac{\tau}{2}} * X(\omega) \tag{2.6}$$

where $\tau = b - a$ and $*$ denotes the convolution operator. As $\tau \to \infty$, $w(t) \to 1$, and $W(\omega) \to \delta(\omega)$ and $w(t)x(t) \xrightarrow{\mathscr{F}} X(\omega)$. However, for a finite width of the window (τ), the

frequency spectrum of $x(t)$ would entail side-lobes as a result of the convolution operator, which distorts the true frequency content of the signal. If τ is selected sufficiently large enough, the side-lobes will stay apart and the frequency spectrum of $x(t)$ will be partly distorted and repeated through the side-lobes. However, narrowing the window width would cause a secondary phenomena that is referred to as time-aliasing where distortion takes place throughout the signal spectrum, and information would no longer be contained in the transformation.

One may therefore deduce that Fourier transform is an improper tool for analyzing local frequency information of a given signal. On the other hand, extending the window width results poor time resolution, and that the signal may no longer remain stationary. Due to the facts presented above there is a negative tendency applying Fourier analysis for feature extraction among researchers. 'Commonly employed spectral analysis technique such as FFT is based on the assumption of stationary signals [61], and is inherently unsuited for nonstationary, transient signal analysis' [62]. In this thesis we will revisit the use of STFT in feature extraction of gas turbine AE signals and will show that by selective sizing of the window width we can expect reasonable results.

However, from the past decade wavelet transform (WT) had become a fast evolving signal processing tool, handling transient nonstationary signals. The basis of wavelet

transforms is a set of *orthonormal*, oscillatory and decaying functions that are named *wavelets* $\psi_{s,\tau}(t)$. There can be special choices of the wavelet basis so that they are orthogonal to their own *translations* and *dilations* s, so that

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}) \tag{2.7}$$

and

$$\int \psi_{s,\tau}(t)\psi^*_{s',\tau'}(t)dt = \begin{cases} 1 & s = s', \tau = \tau' \\ 0 & otherwise \end{cases} \tag{2.8}$$

where $*$ denotes the complex conjugate operator. Orthonormality in (2.8) is only defined for discrete translations τ. If we assume $s = 2^i$ and $\tau = 2^i k$ where $i = \{0, 1, 2, ...\}$ and k is an integer, then the discrete wavelet ψ can be defined as

$$\psi_{i,k}(t) = \frac{1}{\sqrt{2^i}}\psi(\frac{t}{2^i} - k) \tag{2.9}$$

With the discrete wavelet basis, a continuous function $f(t)$ is decomposed into a sequence of wavelet coefficients

$$W_f(i,k) = \int f(t)\psi^*_{i,k}(t)dt =< f, \psi_{i,k} > \tag{2.10}$$

Daubechies [65] has proven that the necessary and sufficient condition for the stable reconstruction of $f(t)$ from its wavelet coefficients $W_f(i,k)$ is that the energy (i.e., the sum of square moduli) of $W_f(i,k)$ lies between two positive bounds, that is

$$A \parallel f \parallel^2 \leq \sum_{i,k} | < f, \psi_{i,k} > |^2 \leq B \parallel f \parallel^2 \tag{2.11}$$

where $\| f \|^2$ is the energy of $f(t)$ and A, B are positive bounded quantities, independent of $f(t)$. Hence,

$$f(t) = \sum_i \sum_k W_f(i,k)\psi_{i,k}(t) \tag{2.12}$$

if the wavelets satisfy the orthogonality conditions of equation (2.8). In the multiresolution analysis framework, the orthogonal wavelet transform is based on the scaling function $\phi(t)$, being a continuous and generally real valued function on the set of real numbers, usually normalized as

$$\int \phi(t)dt = 1 \tag{2.13}$$

The basic scaling function $\phi(t)$ as in the wavelet basis is shifted by discrete translation factors to construct an orthonormal basis at the same resolution level. That is,

$$\phi_{i,k}(t) = 2^{-i/2}\phi(2^{-i}t - k) \tag{2.14}$$

$$\psi_{i,k}(t) = 2^{-i/2}\psi(2^{-i}t - k) \tag{2.15}$$

As in the wavelet basis, an arbitrary function $f(t)$ can be expanded into a set of translated scaling functions. However, since the scaling function is a real valued function, we have

$$f(t) = \sum_k c_0(k)\phi(t - k) \tag{2.16}$$

The dilated $\phi_{i,k}(t)$ forms an orthonormal basis at each resolution level i by the discrete translations. The scaling function in two adjacent resolution levels satisfy the following

relation

$$\phi(t) = \sum_k p(k)\phi(2t-k) \tag{2.17}$$

where $p(k)$ is a discrete finite sequence, corresponding to a low-pass filter. Let us also assume that the wavelet basis $\psi(t)$ can also be expanded onto the orthonormal scaling function basis of $\phi(2t-k)$ as

$$\psi(t) = \sum_k q(k)\phi(2t-k) \tag{2.18}$$

whereas $q(k)$ is another discrete real valued finite sequence that corresponds to a high-pass discrete filter. Furthermore, we assume that the set of wavelet translates is orthogonal to the set of scaling function translates within the same resolution, that is

$$< \phi_{i,k}, \psi_{i,n} >= 2^{-i} \int \phi_i(t-k)\psi_i(t-n)dt = 0 \tag{2.19}$$

With the assumptions made above we are now able to decompose an arbitrary $f(t)$ into a sequence of wavelets and scaling functions

$$f(t) = \sum_n c_1(n)\phi_{1,n}(t) + \sum_n d_1(n)\psi_{1,n}(t) \tag{2.20}$$

where

$$c_1(n) =< f, \phi_{1,n} > \tag{2.21}$$

and

$$d_1(n) =< f, \psi_{1,n} > \tag{2.22}$$

where by substituting for $\phi_{1,n}$ from (2.17) we have

$$\begin{aligned} c_1(n) &= \int f(2t)2^{\frac{1}{2}}\phi(t-n)dt \\ &= \int f(2t+2n)2^{\frac{1}{2}}\phi(t)dt \\ &= \int f(2t+2n)2^{\frac{1}{2}}\sum_k p(k)\phi(2t-k)dt \\ &= 2^{\frac{1}{2}}\sum_k p(k)\int f(2t+2n)\phi(2t-k)dt \\ &= 2^{-\frac{1}{2}}\sum_k p(k)c_0(2n+k) \end{aligned}$$

and finally,

$$c_1(n) = 2^{-\frac{1}{2}}\sum_k p(k-2n)c_0(k) \tag{2.23}$$

Similarly,

$$d_1(n) = 2^{-\frac{1}{2}}\sum_k q(k-2n)c_0(k) \tag{2.24}$$

where $p(n-2k) = 2^{1/2} < \phi_{0,n}, \phi_{i,k} >$ and $q(n-2k) = 2^{1/2} < \phi_{0,n}, \psi_{i,k} >$. Generalizing equations (2.23) and (2.24) for the scale level i provides:

$$c_i(n) = 2^{-\frac{1}{2}}\sum_k p(k-2n)c_{i-1}(k) \tag{2.25}$$

$$d_i(n) = 2^{-\frac{1}{2}}\sum_k q(k-2n)c_{i-1}(k) \tag{2.26}$$

Equations (2.25) and (2.26) are the basis of multiresolution analysis. An arbitrary function $f(t)$ can be decomposed into M levels of resolutions, as

$$f(t) = \sum_k 2^{-\frac{M}{2}}c_M(k)\phi(2^{-M}t-k) + \sum_{i=1}^{M}\sum_k 2^{-\frac{i}{2}}d_i(k)\psi(2^{-i}t-k) \tag{2.27}$$

Substituting $f(k)$ (discrete time version of $f(t)$) for $c_0(k)$ in equations (2.23) and (2.24) provides the first level coefficients. Successive coefficients are then constructed through equations (2.25) and (2.26). Producing c_is and d_is can be demonstrated as shown in Figure 2.1. In this figure $M = 1$, $p'(n) = 2^{-1/2}\phi_{0,n}$, and $q'(n) = 2^{-1/2}\psi_{0,n}$. Also

$$p(n) = p'(L + 1 - n) \quad \text{and} \quad q(n) = q'(L + 1 - n)$$

where L is the filter length, and

$$q'(n) = (-1)^n p'(n)$$

$q'(n)$ and $p'(n)$ are referred to as *Quadrature Mirror* filters. Accordingly, an arbitrary signal $f(n)$ can be decomposed into its lower ($f_l(n)$), and higher ($f_h(n)$) subbands. Then $f_l(n)$ can in turn be decomposed into further low-low and high-low subbands by replacing $f(n)$ with $f_l(n)$ in Figure 2.1, and so forth, through recursive filtering.

With $M = 0$ in equation (2.27), $f(t)$ is decomposed into a set of translated scaling functions as in (2.16). If the function $f(t)$ is sampled at a rate producing N_s number of samples, the decomposition would result in the same number of translations N_s, and obviously the produced sequence of scaling translations (i.e., $\phi(n - k)$, $k = 0, 1, ..., N_s - 1$) would contain a bandwidth equal to the original sampled sequence $f(n)$, say f_s.

However, decomposing $f(n)$ into two levels of resolution (i.e., $M = 1$) would reduce (2.27) to (2.20) with a number of translations equal to $N_s/2$. However, this time $f(n)$

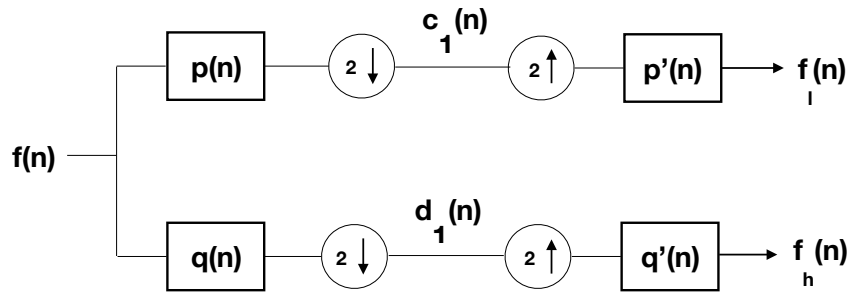

Figure 2.1: Decomposition of a signal into lower and higher subbands.

is decomposed into two subbands, each having half the original bandwidth, i.e., $f_s/2$. The low frequency component is composed of the translated scaling functions occupying $[0, f_s/2]$, and the high frequency component is composed of translated wavelets occupying $[f_s/2, f_s]$.

Signal decomposition into two levels increases the frequency resolution by two, since each component is now only half the original bandwidth. This gives us power to see frequency contents twice as separated (more finely tuned), however due to half number of time samples, we are now less able to locate the frequency components in time.

Increasing the resolution level will proportionally provide more resolution in frequency, and simultaneously less resolution in time. This is unfortunately the dilemma of wavelet multiresolution analysis, as we have to sacrifice the resolution of location in time for frequency, or resolution of frequency for location in time.

Figure 2.2 is an example of an AE signal decomposed into five subbands. The top left figure is the original signal which is decomposed into an approximation (top right hand figure), essentially being the low passed filtered version of the original signal. Successive details are shown in the second and third rows of the figure. The right hand side figure in the bottom is the highest frequency component of the signal. The sum of the approximation signals with their details represent the original signal.

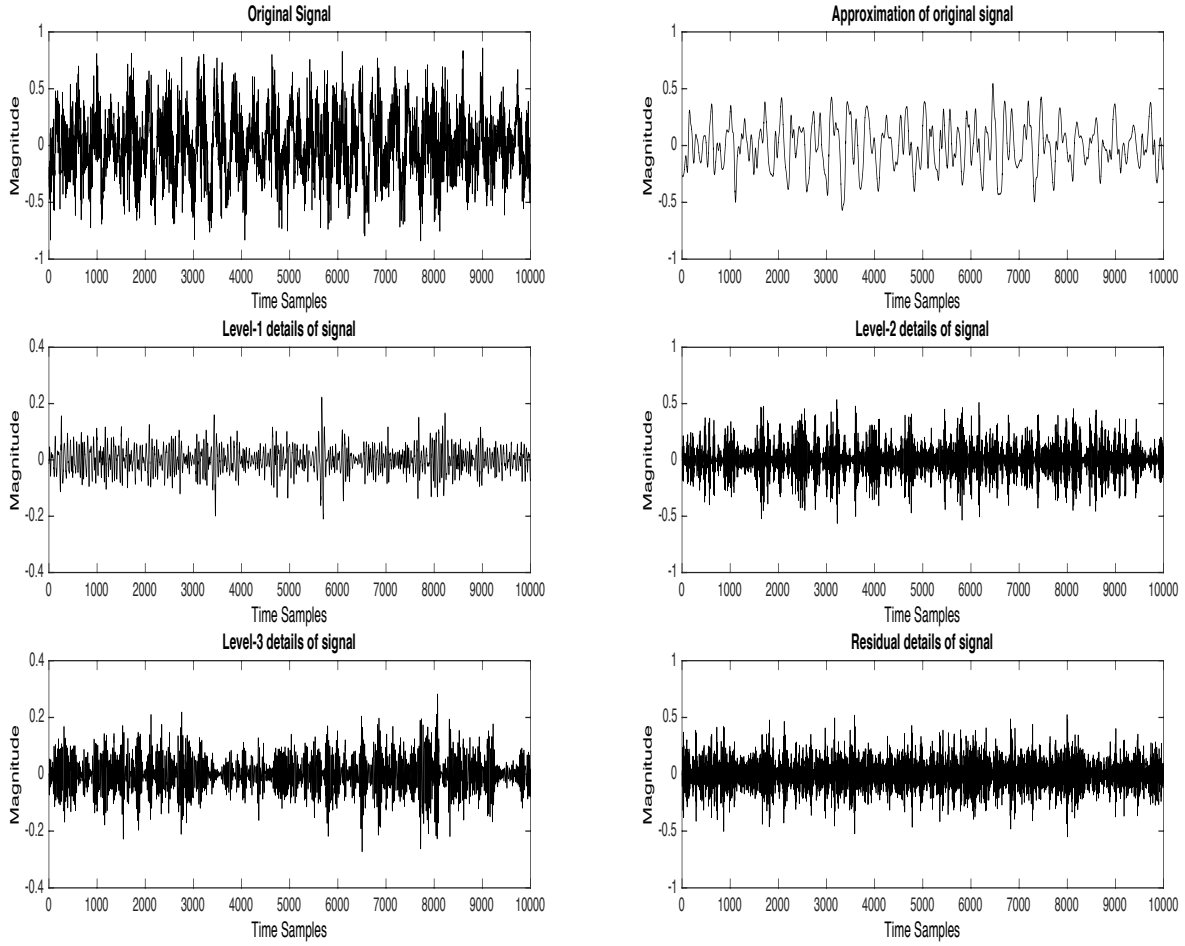

Figure 2.2: Decomposition of a signal into subbands.

Let us now compare two AE samples taken from a gas turbine in two different conditions installed in a power generating facility, as in Figure 2.3 and Figure 2.4. The first AE sample $x_1(t)$ belongs to a gas turbine operating in a healthy condition, whereas the second sample $x_2(t)$ is taken from the same gas turbine with a mild problem in

its gearbox (refer to page 37). In these two figures the AE signals $x_1(t)$ and $x_2(t)$

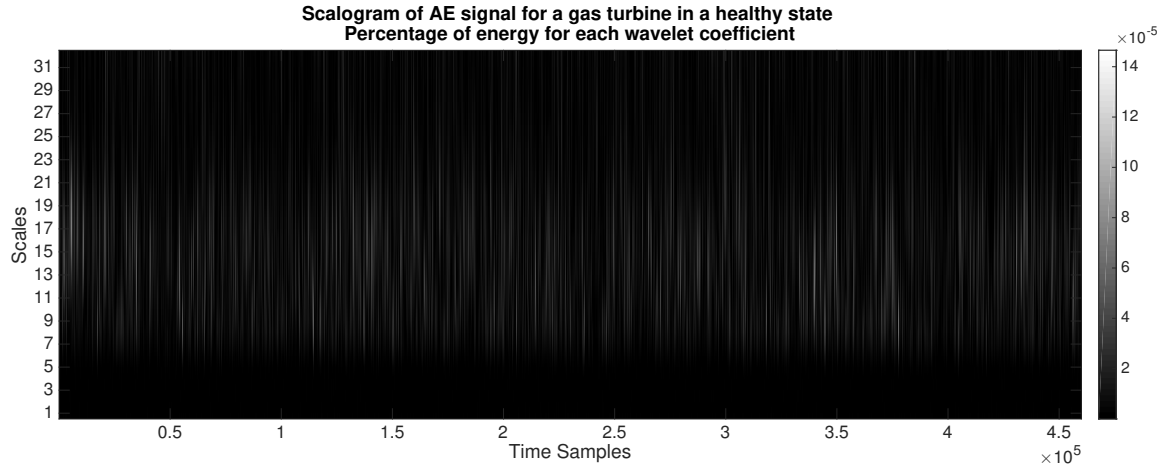

Figure 2.3: Percentage energy of wavelet coefficients for $x_1(t)$. The greyscale bar on the right shows the magnitude of the percentage energy corresponding to each shade.

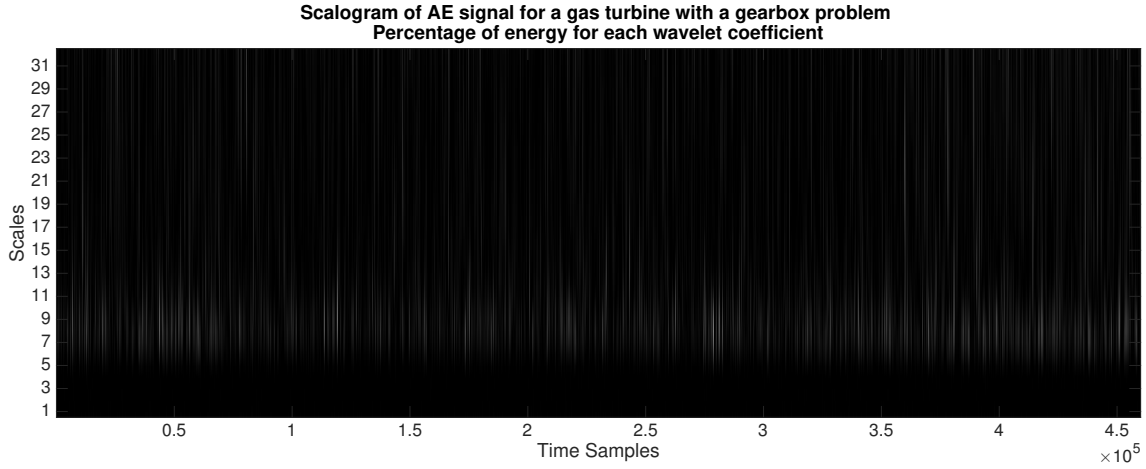

Figure 2.4: Percentage energy of wavelet coefficients for $x_2(t)$.

have been decomposed into wavelet coefficients $W_f(i,k)$, according to equation (2.12)

for $i = \{1, 2, ..., 31\}$, and the percentage energy of the corresponding wavelet coefficients have been plotted on a greyscale basis, where dark signatures are showing weaker components, and bright traces are showing (relatively) powerful components, in every instant in time. On the vertical axis the distribution of the energies are illustrated with respect to the scales, where lower scales correspond to higher frequencies. Therefore, white traces emerging at the bottom of the figure are indicating high frequency instantaneous components, whereas bright traces stretched in the middle of the vertical axis are indicating medium frequencies. By comparing Figures 2.3 and 2.4, it is difficult (if not impossible) to detect an anomaly in $x_2(t)$. This is while the anomaly in $x_2(t)$ is very evident if we look at the autocorrelation functions of the two signals as in Figure 2.5.

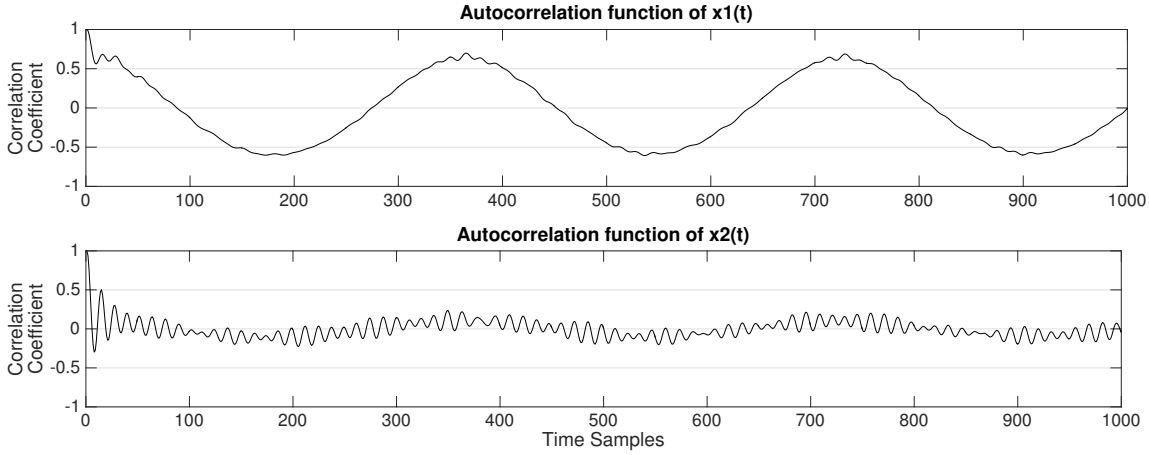

Figure 2.5: Autocorrelation of the discrete version of AE signals $x_1(t)$ and $x_2(t)$ for 1000 time samples (refer to page 34 for data sampling).

As mentioned before, gas turbine AE tends to a cyclostationary process. This can

be demonstrated in Figure 2.5, and specially in the first subplot, where the autocorrelation function $R_X(t)$ is plotted for a healthy gas turbine, and resembles a sinusoidal waveform. In the next chapter we will try to show that an additive noise to the autocorrelation function of an AE signal can indicate anomalies in the process, as long as the average behaviour of the signal is concerned. This is illustrated in the second subplot of Figure 2.5.

Another general characterization of nonlinear and nonstationary signal processing techniques, are their capability of extracting instantaneous features of the signal. For example the wavelet coefficients $W_f(i,k)$ (in Figure 2.6) are themselves functions of time k. This is in contrast with the autocorrelation function where the features are representing the signal in a mean sense.

In cases where the gas turbine condition excessively differs a normal condition, scalograms might be able to represent the distinctions. For example let us refer to Figure 2.6 where the percentage energy of wavelet coefficients are derived for a gas turbine that has started-up after a long rest, and has not achieved thermal equilibrium yet.

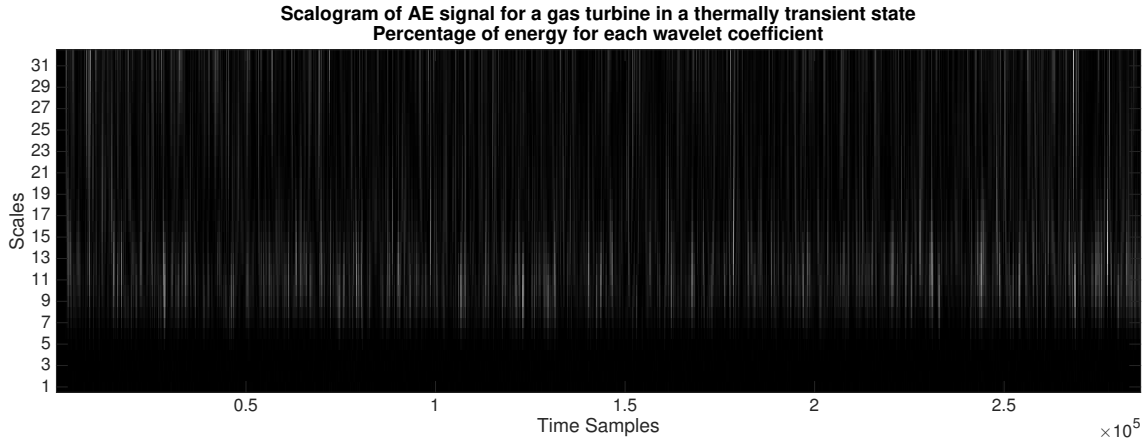

Figure 2.6: Percentage energy of wavelet coefficients for a gas turbine in a thermally transient state.

In these circumstances it might be possible to use machine learning solutions for autonomous classification of conditions. However, it is questionable if the instantaneous characterization of the features (wavelet coefficients) are holding any value for the classification purpose.

2.2 Wavelet Packet Decomposition and Hilbert-Huang Transform (WPD-HHT)

As we have seen wavelet transform is orthogonal (in the discrete form), local and adaptive. All these are vital for forming a basis to analyze nonlinear and nonstationary signals. The basic operation of wavelet transform involves the operations of dilation

and translation, which lead to a multiresolution analysis of the signal. Hence, it can extract both the time and frequency features of the inspected signal effectively. Although wavelet transform is capable of analyzing nonlinear and nonstationary signals and deemed suitable for machine condition monitoring, we have found deficiencies in demarcation of different machine conditions (as demonstrated in Figures 2.3 and 2.4).

The process of continuous wavelet transform for decomposing a signal into subbands is computationally intensive. On the other hand, the discrete wavelet transform has good computing efficiency. However, the resolution in frequency at a high-frequency range is poor, as discussed already. Due to the deficiencies of wavelet transform, a new type of time-frequency analysis called Hilbert-Huang transform (HHT) has been proposed for analyzing nonlinear and nonstationary signals [62]. The HHT is derived from the principals of empirical mode decomposition (EMD) and the Hilbert transform.

When applying the HHT, first, the EMD will decompose the acquired signal into a collection of intrinsic mode functions (IMF). The IMF is a kind of complete, adaptive and almost orthogonal representation for the analyzed signal. Since the IMF is almost monocomponent, it can determine all the instantaneous frequencies from the nonlinear or nonstationary signal. Second, the local energy of each instantaneous frequency can be derived through the Hilbert transform. Hence, the result is a HHT spectrum which

has an energy-frequency-time distribution of the signal. From the HHT spectrum, one can localize any event on its occurring time as well as its instantaneous frequency [62].

The definition of instantaneous frequency can be brought from Hilbert transform. The Hilbert transform of a signal $f(t)$ is defined as

$$\hat{f}(t) = \frac{P}{\pi} \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} d\tau = f(t) * \frac{1}{\pi t} \tag{2.28}$$

when the integral exists, and where P is the Cauchy's principal defined as

$$P = \lim_{\epsilon \to 0} [\int_{-\infty}^{\epsilon} \frac{dt}{\pi t} + \int_{\epsilon}^{+\infty} \frac{dt}{\pi t}]$$

The Cauchy integral is a figurative way to motivate the Hilbert transform. The complex view helps us to relate the Hilbert transform to something more concrete and understandable. Consider an integral in the complex z-plane on the form

$$\oint_{\Gamma} \frac{f(z)}{z-a}\, dz$$

If $f(z)$ is analytic and is a piecewise smooth closed contour, then the Cauchy integral theorem is applicable as

$$\oint_{\Gamma} \frac{f(z)}{z-a}\, dz = \begin{cases} 2\pi j f(a) & \text{if } a \text{ is inside } \Gamma \\ \\ 0 & \text{if } a \text{ is outside } \Gamma \end{cases} \tag{2.29}$$

If we define an arc γ_r as $z = a + re^{j\theta}$ where $\theta_1 \leq \theta \leq \theta_2$, then

$$\lim_{r \to 0^+} \int_{\gamma_r} f(z) dz = j(\theta_2 - \theta_1) Res(f(z))|_{z=a}$$

Now let us consider $f(z)$ is analytic on the upper half complex plane, and let Γ be defined as an infinitely large half circle encircling the upper half plane closed by the x-axis from $-\infty$ to $+\infty$, and let us assume a is a point on the x-axis. Cauchy's integral would then be applicable on Γ provided that an infinitely small half circle would encircle point a on the x-axis. Then

$$\oint_{\Gamma} \frac{f(z)}{z-a}\,dz = \lim_{R\to+\infty}\int_{R} f(z)dz + \int_{-\infty}^{a^{-}} f(z)dz + \lim_{r\to 0^{+}}\int_{\gamma_r} f(z)dz + \int_{a^{+}}^{+\infty} f(z)dz \quad (2.30)$$

and if $f(z)$ is so that

$$|f(z)| < \frac{C}{|z|}$$

then the left most term in the right hand side of equation (2.30) would vanish as $R \to \infty$, hence

$$P\int_{-\infty}^{+\infty} \frac{f(x)}{x-a}dx = j\pi f(a) \quad (2.31)$$

Also since

$$\frac{1}{\pi t} \xrightarrow{\mathscr{F}} -jsgn(\omega) \quad (2.32)$$

then

$$\hat{f}(t) = f(t) * \frac{1}{\pi t} \xrightarrow{\mathscr{F}} -jF(\omega)sgn(\omega) \quad (2.33)$$

The *analytical signal* of $f(t)$ is defined as

$$z(t) = f(t) + j\hat{f}(t) \quad (2.34)$$

and accordingly

$$z(t) \xrightarrow{\mathscr{F}} Z(\omega) = F(\omega) + F(\omega)sgn(\omega) = \begin{cases} 2F(\omega) & \omega > 0 \\ F(\omega) & \omega = 0 \\ 0 & \omega < 0 \end{cases} \tag{2.35}$$

Equation (2.35) is the basis for practical computation of the analytic signal, through the Fourier transform of the original signal. By computing the inverse Fourier transform of (2.35) one can deduce $z(t)$. As we can see the frequency and the power of $z(t)$ is identical to $f(t)$. Therefore, to approximate the instantaneous attributes of $f(t)$ one can instead compute the power and frequency of $z(t)$.

Let us consider $z(t) = a(t)e^{j\phi(t)}$, where

$$a(t) = [f(t)^2 + \hat{f}(t)^2]^{1/2}, \text{ and } \phi(t) = tan^{-1}[\frac{\hat{f}(t)}{f(t)}] \tag{2.36}$$

then $a(t)$ is the instantaneous amplitude of $f(t)$, which can reflect how the energy of the $f(t)$ varies with time, and $\phi(t)$ is the instantaneous phase of $f(t)$. The instantaneous frequency can now be defined as the time derivative of the instantaneous phase as [63],

$$\omega(t) = \frac{d\phi(t)}{dt} \tag{2.37}$$

It can be shown however that the instantaneous frequency derived in this manner provides incorrect results for signals with multiple components (i.e., signals composed of

multiple frequency components). This has been clearly demonstrated in [63]. In order to rectify this defect, the signal is first decomposed into almost monocomponent modes, named *Intrinsic Mode Functions* (IMFs) through a process referred to as the *Empirical Mode Decomposition* (EMD), and then the instantaneous frequency of each IMF is derived based on the Hilbert transform. According to Peng *et al.* [63], instantaneous frequency is physically meaningful only when the signal is symmetric with respect to the local zero mean, and have the same number of zero crossings and extrema.

That is, in an IMF function, the number of extrema and the number of zero crossings must either equal or differ at most by one in the entire data set, and the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at every point. However, the determined IMF may not satisfy these conditions precisely. Therefore, the resulting IMF is nearly a monocomponent signal, but not perfect. The detail process of eliciting IMFs is formulated in [66].

An upper and lower envelope of the signal is first prepared by connecting local maxima with a cubic spline (i.e., $e_u(t)$), and connecting the local minima with a cubic spline as a separate curve (i.e., $e_l(t)$). These two curves are embracing the signal through its entire course. Let $m_1(t)$ denote the mean

$$m_1(t) = \frac{e_u(t) + e_l(t)}{2} \tag{2.38}$$

the first IMF is determined through the difference between x and m_1, referred to as *sifting process*

$$x_{IMF}^{(1)}(t) = x(t) - m_1(t) \tag{2.39}$$

$x_{IMF}^{(1)}(t)$ should however possess the characteristics of the IMF, that is firstly, the number of its zero crossings should equate the number of extrema throughout the course of the signal, or at most differ by one. Secondly, $x_{IMF}^{(1)}(t)$ should be symmetric with respect to the time axis (i.e., the mean of its lower and upper envelopes to be zero). If $x_{IMF}^{(1)}(t)$ does not hold the conditions of an IMF, the same process is repeated replacing $x_{IMF}^{(1)}(t)$ for x, and repeating equations (2.38) and (2.39). This process is repeated until $x_{IMF}^{(1)}(t)$ is (close to) an IMF [66]. Then, the second IMF will be generated on the residual of $x(t)$. That is $x(t) \leftarrow [x(t) - x_{IMF}^{(1)}(t)]$, and going through the same process to determine $x_{IMF}^{(2)}(t)$. This algorithm is continued until the i^{th} IMF is derived, and the residual is no longer able to sift, or the residual is monotonic.

As an example the 15 intrinsic mode functions (IMF-1 to IMF-15) of the signals $x_1(t)$ and $x_2(t)$ have been derived for 0.1 seconds which corresponds to 5000 time samples in discrete mode, at a rate of 44,100 samples/sec. It is reminded that $x_1(t)$ and $x_2(t)$ corresponded to a healthy gas turbine AE and a gas turbine with some mild gearbox problem (refer to page 53 for details). These functions are plotted in Fig. 2.7 to Fig. 2.11.

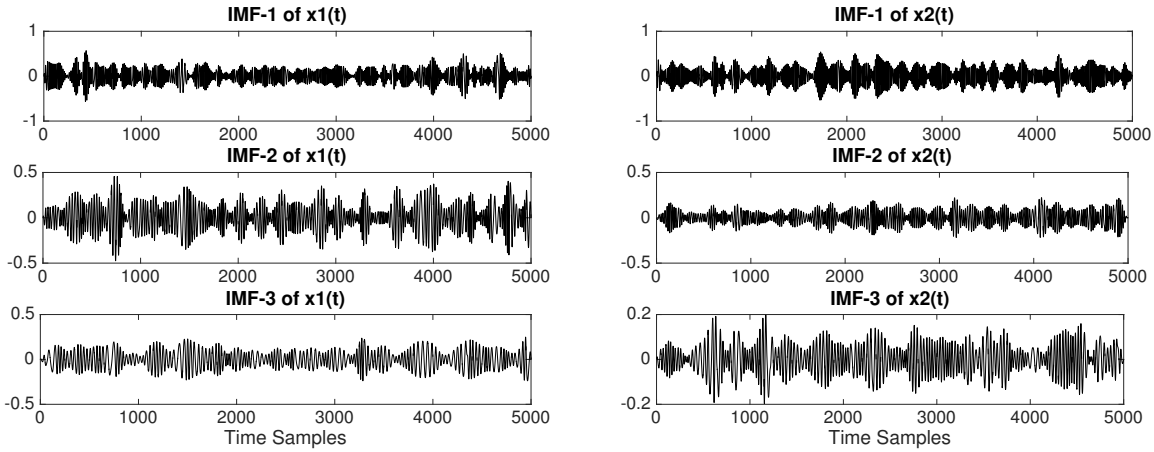

Figure 2.7: IMF1 to IMF3 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively.

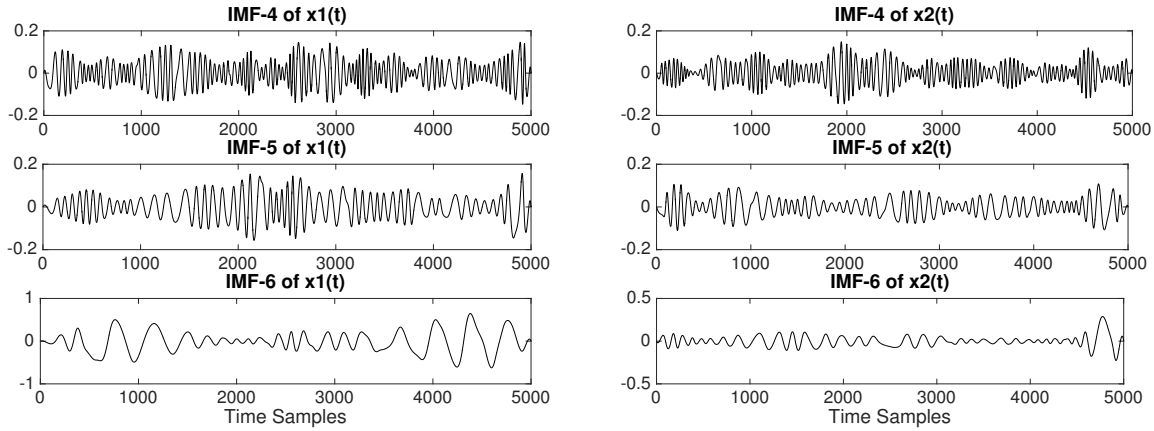

Figure 2.8: IMF4 to IMF6 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively.

The only apparent difference between the two signals seem to be in the relative energy of the 2^{nd} and 3^{rd} modes. However, the instantaneous frequency of the corresponding intrinsic modes of the signals do not appear to be very different.

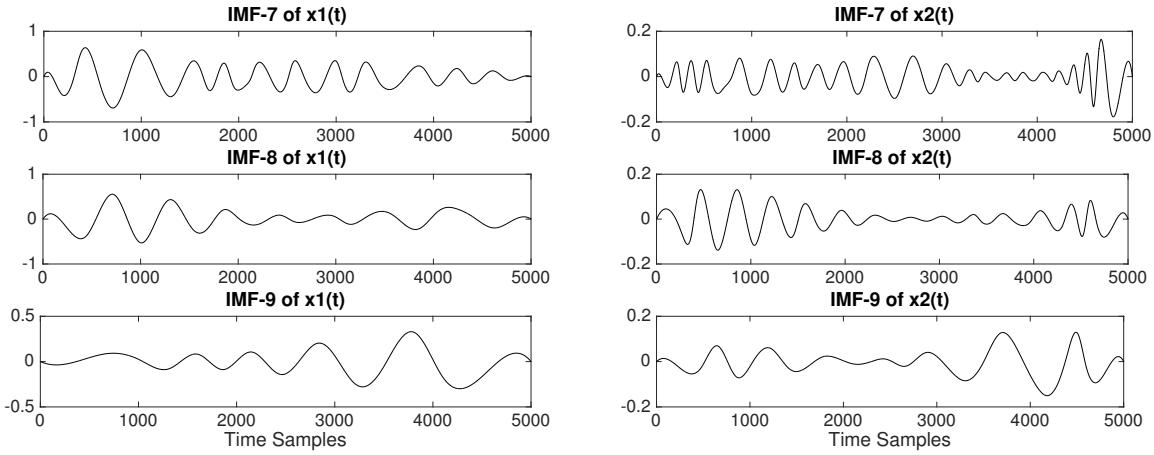

Figure 2.9: IMF7 to IMF9 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively.

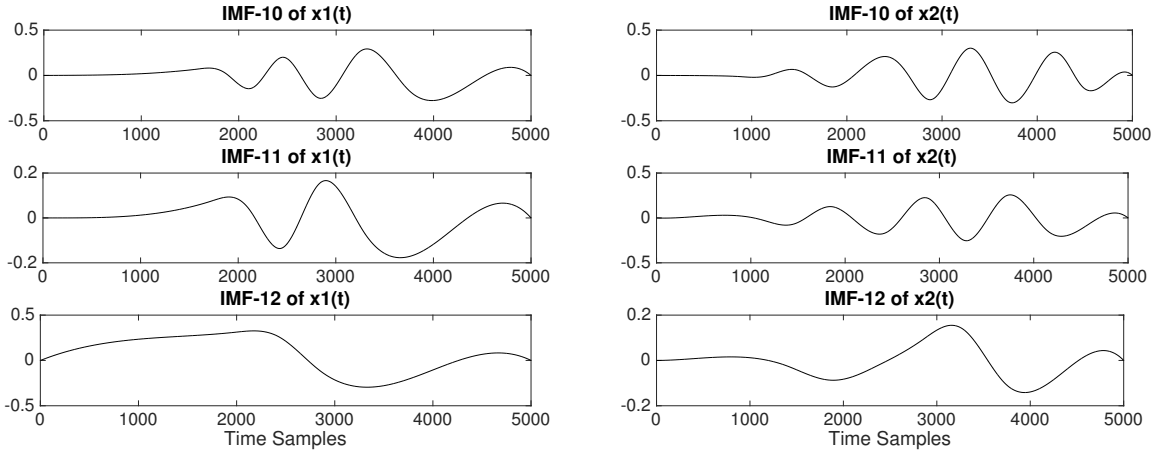

Figure 2.10: IMF10 to IMF12 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively.

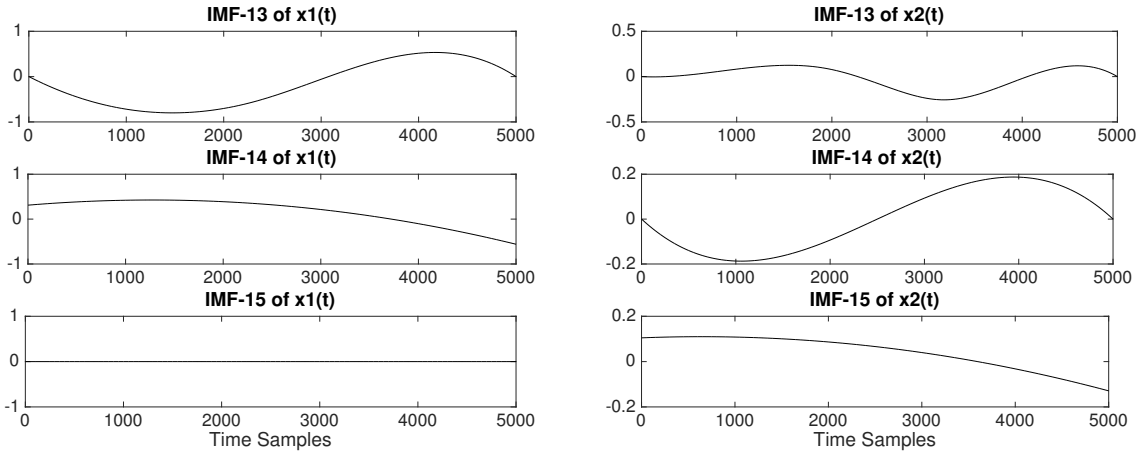

Figure 2.11: IMF13 to IMF15 of the signals $x_1(t)$ and $x_2(t)$ representing a healthy gas turbine and a gas turbine with a problem in the gearbox respectively.

The instantaneous frequency of IMF-1 to IMF-10 of the signals $x_1(t)$ and $x_2(t)$ is plotted in Fig. 2.12 to Fig. 2.16. As it is evident it is difficult to see noticeable differences between the corresponding figures, except some momentary and random sparks in some of the figures, such as IMF-6 of $x_2(t)$.

Additionally, the average frequency of IMF-7 and IMF-8 of $x_2(t)$ also seems slightly higher than those in $x_1(t)$. However, one has to pay attention to the fact that these derivations are taken from a very short interval of the AE signals (only 0.1 seconds), and there is no guarantee that the frequency patterns will remain consistent for some other samples, unless it can be shown that the statistical behaviour of $x_1(t)$ and $x_2(t)$ are constant throughout the course of the signals.

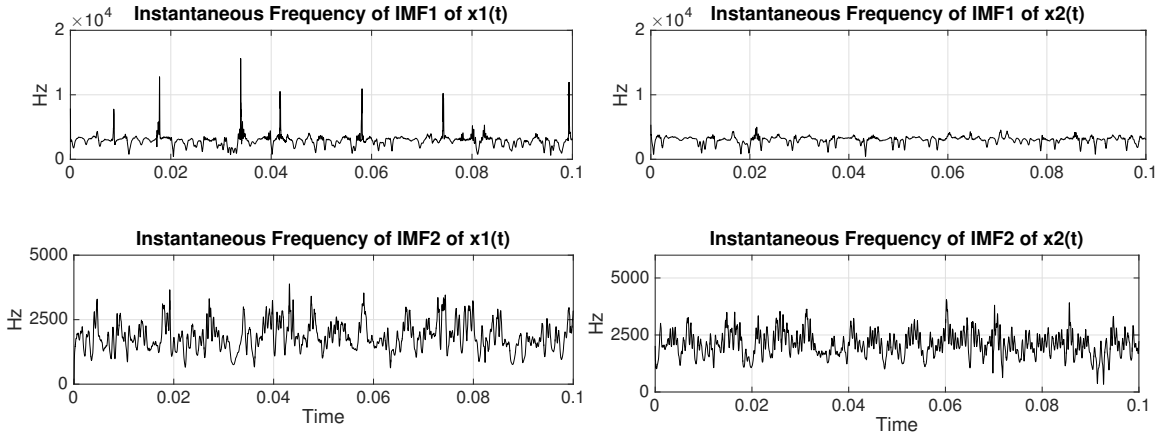

Figure 2.12: Instantaneous frequency of IMF-1 and IMF-2 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds

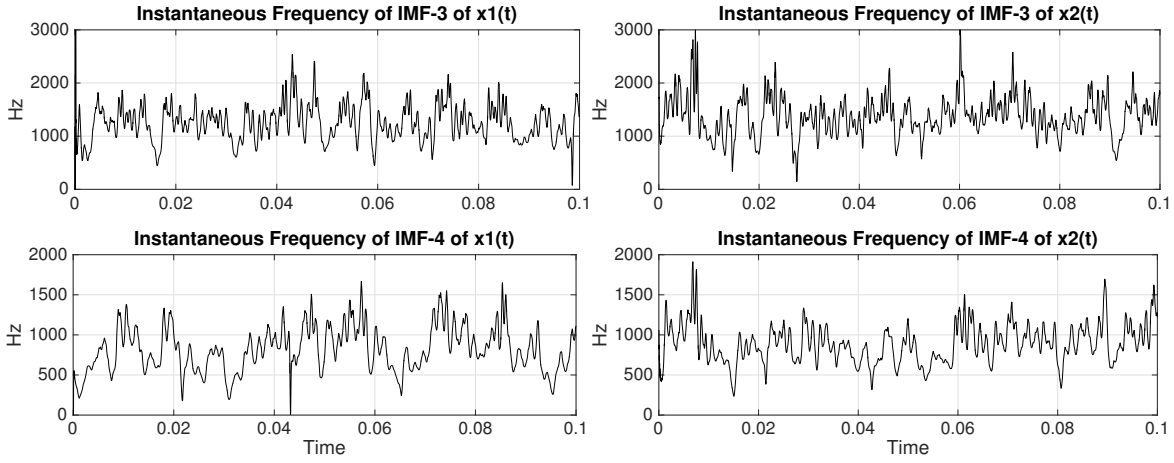

Figure 2.13: Instantaneous frequency of IMF-3 and IMF-4 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds

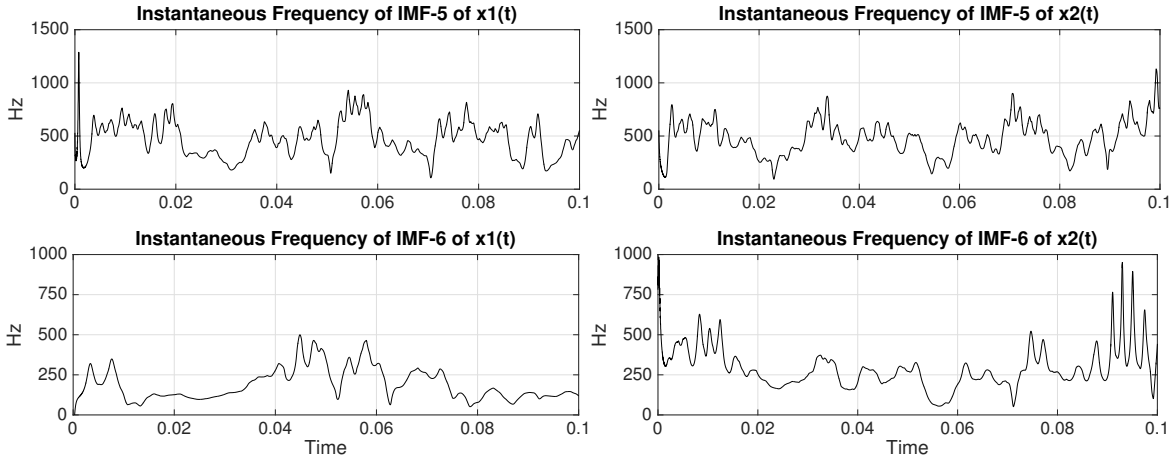

Figure 2.14: Instantaneous frequency of IMF-5 and IMF-6 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds

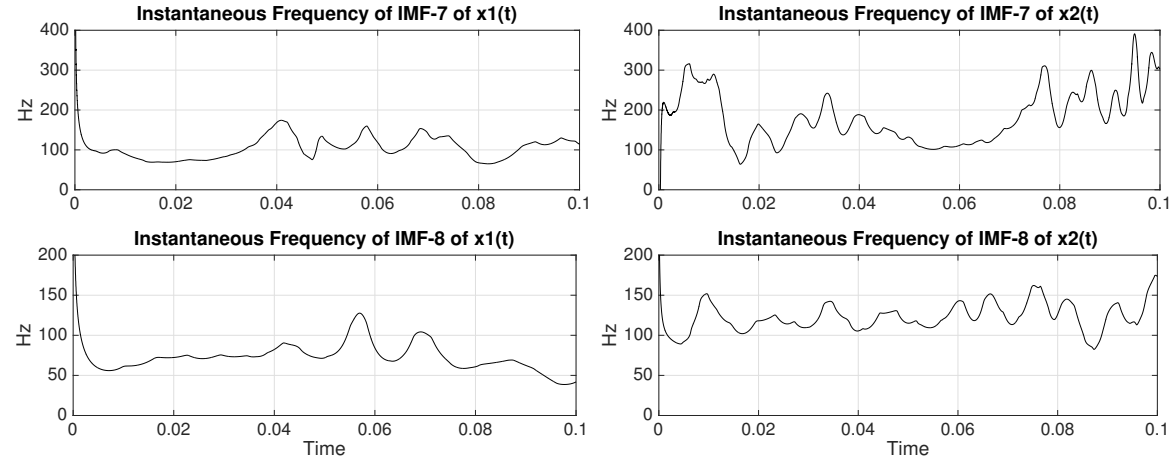

Figure 2.15: Instantaneous frequency of IMF-7 and IMF-8 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds

Researchers often criticize EMD for three major shortcomings [63]. These three major shortcomings include (1) the EMD generating undesirable IMFs at the low-frequency region, (2) the first obtained IMF may cover too wide a frequency range such that the property of monocomponent cannot be achieved, and (3) some signals

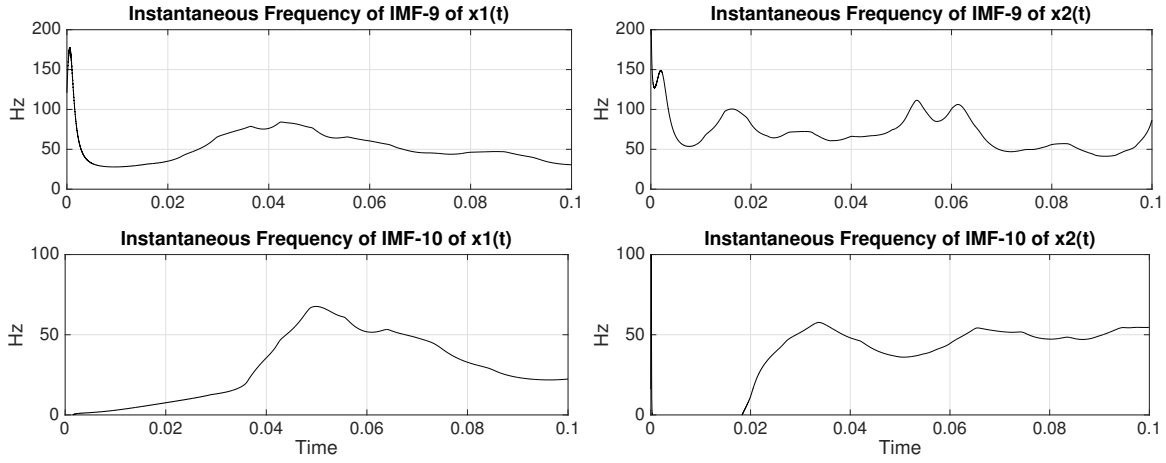

Figure 2.16: Instantaneous frequency of IMF-9 and IMF-10 of $x_1(t)$ and $x_2(t)$ for 0.1 seconds

that contain low-energy components are inseparable.

In order to rectify these issues a more sophisticated method was developed that is termed WPD-HHT which is based on wavelet packet decomposition of the signal first. After the signal is decomposed into certain subbands (refer to Figure 2.1), the contribution of each decomposed subband is screened. Let us assume a signal $f(n)$ is decomposed into k levels, that is,

$$f(n) = d_1(n) + d_2(n) + ... + d_k(n) + c_k(n) \quad \text{where} \quad n = 1, 2, ..., N$$

where d_is are the details and c_k is the approximation of the signal, and k denotes the lowest frequency subband and d_1 is the highest frequency component in the signal. The

normalized energy of the signal can be defined as

$$E_f = \frac{1}{N}\sum_{n=1}^{N} f^2(n) = \frac{1}{N}\sum_{n=1}^{N}[d_1^2(n) + d_2^2(n) + ... + d_k^2(n) + c_k^2(n)] \tag{2.40}$$

since $\sum_n d_i(n)d_j(n) = 0$ for all $i \neq j$ (due to orthogonality of the wavelet components), or,

$$E_f = E_{d_1} + E_{d_2} + ... + E_{d_k} + E_{c_k} \tag{2.41}$$

The decomposition for our application can be extended so that c_k has the highest correlation with the original signal among all possible integers for k. This usually happens when c_k is of equal frequency to the speed of the rotor. For all $i = 1, 2, ..., k$ where $E_{d_i}/E_f < 0.1$, the i^{th} detail is considered insignificant and lumped as a residual. In the next step, the significant components having a relative energy of more than 10% will go through the sifting process and for each subband component the IMFs will be computed.

Once again IMFs will be screened and insignificant ones will be excluded, and finally instantaneous frequency of the IMFs are derived. In this process the IMFs are closer to a monocomponent signal and therefore instantaneous frequency computation would be more meaningful and reliable.

For $x_1(t)$ and $x_2(t)$ wavelet filtering decomposed 8 level of details, and the screening resulted the following:

$$E_{d_1(x_1(t))} \approx 0, \text{ and } E_{d_1(x_2(t))} \approx 0 \quad \text{to be excluded}$$

$$E_{d_2(x_1(t))} \approx 0, \text{ and } E_{d_2(x_2(t))} \approx 0 \quad \text{to be excluded}$$

$$E_{d_3(x_1(t))} = 0.011, \text{ and } E_{d_3(x_2(t))} = 0.021$$

$$E_{d_4(x_1(t))} = 0.021, \text{ and } E_{d_4(x_2(t))} = 0.015$$

$$E_{d_5(x_1(t))} = 0.007, \text{ and } E_{d_5(x_2(t))} = 0.005 \quad \text{to be excluded}$$

$$E_{d_6(x_1(t))} = 0.005, \text{ and } E_{d_6(x_2(t))} = 0.002 \quad \text{to be excluded}$$

$$E_{d_7(x_1(t))} = 0.014, \text{ and } E_{d_7(x_2(t))} = 0.002 \quad d_7(x_2(t)) \text{ to be excluded}$$

$$E_{d_8(x_1(t))} = 0.111, \text{ and } E_{d_8(x_2(t))} = 0.006$$

$$E_{c_8(x_1(t))} = 0.002, \text{ and } E_{c_8(x_2(t))} \approx 0 \quad \text{to be excluded}$$

According to the procedure described above, we can carry out the sifting process for d_3, d_4, and d_8. The EMD of these signals after eliminating spurious low energy IMFs results in 21 final monocomponents, from which instantaneous frequencies could be evoked. As an example we have taken the 4^{th} detail level d_4 of $x_1(t)$ and $x_2(t)$. Among the high frequency components, IMF-1 to IMF-4 had substantial energies. Figure 2.17 to Figure 2.20 illustrate the WPD-HHT process of $x_1(t)$ and $x_2(t)$. Comparison between the two subplots in Figure 2.17 does not indicate any difference in $x_2(t)$. Now let us examine other IMFs.

In Figure 2.17 to Figure 2.20 it is evident that the average frequency of the sub-bands decrease from IMF-1 (the highest frequency component) to IMF-4 (the lowest

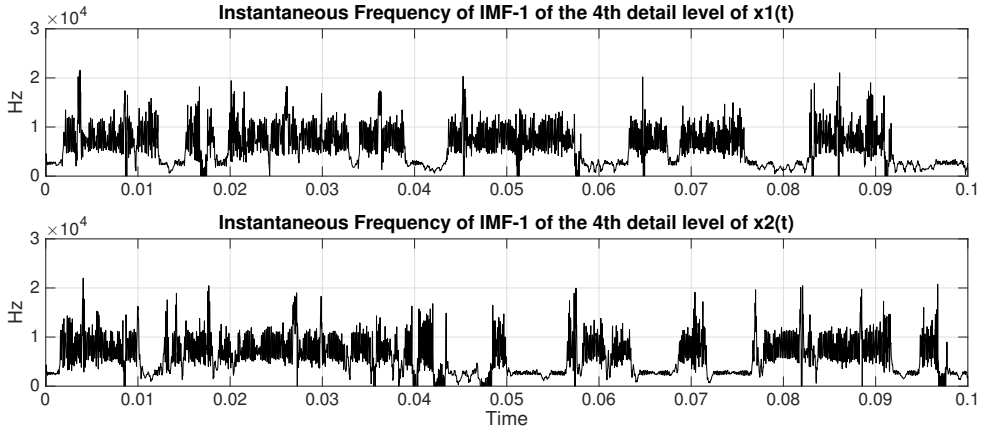

Figure 2.17: WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-1.

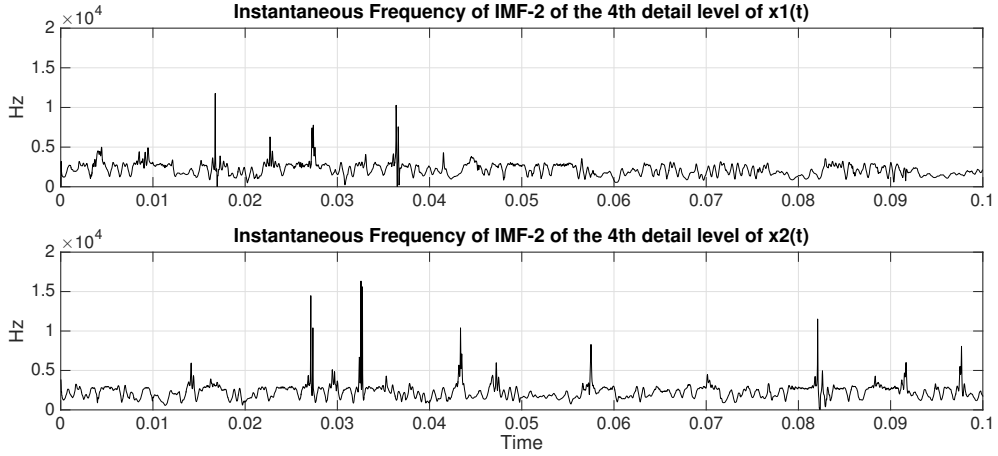

Figure 2.18: WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-2.

frequency component) for both $x_1(t)$ and $x_2(t)$. However there is no substantial distinction between the two subplots in each figure that could lead us to some anomaly in $x_2(t)$, except some random momentary bursts of high frequency sparks that sometimes happen in $x_2(t)$, and sometimes in $x_1(t)$.

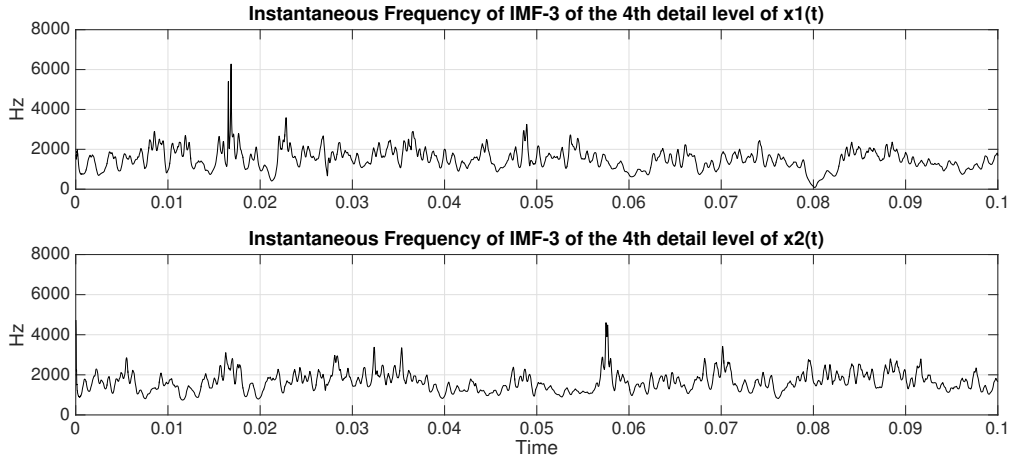

Figure 2.19: WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-3.

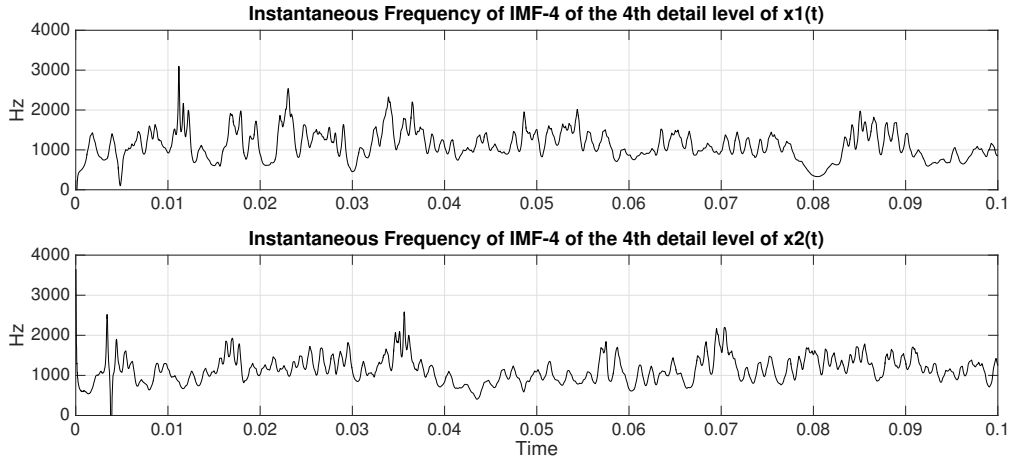

Figure 2.20: WPD-HHT Process of the 4^{th} detail levels of $x_1(t)$ and $x_2(t)$, for IMF-4.

Investigating the HHT of all other higher frequency detail levels (not shown here) also do not suggest any sign of anomaly in $x_2(t)$, concluding that WPD-HHT cannot be considered as a powerful feature extraction tool, as long as gas turbine AE signals are concerned.

In chapter 5 we will show that the per unit energy of the wavelet frequency sub-bands of a signal (denoted by $\mathbb{X}_k$), one could simply and perfectly distinguish average differences in a signal. The formal definition will be given by Definition 4.1, later on page 125.

2.3 Largest Lyapunov Exponent (LLE)

Detecting the presence of *deterministic chaos* in a dynamical system is an indication of non-predictable behaviour of the system dynamics. Deterministic chaos in contrast with noise is a characteristic behaviour of a dynamical system that is quantified by the divergence of trajectories of nearby initial conditions on an attracting manifold. When the attractor is chaotic the trajectories diverge on average at a rate characterized by the largest Lyapunov exponent [67]. Therefore, when the long-term behaviour of an initial condition that is specified with any uncertainty cannot be predicted; the system is characterized by chaos, and an attractor for a dissipative system with one or more positive Lyapunov exponents is said to be 'strange' or 'chaotic' [68].

In the previous sections we have seen that an operating condition in a gas turbine causes acoustic emissions that tend to be short-time cyclostationary (refer to page 41 for the definition). When conditions change, the process produces a new cyclostationary

time-series data that will retain its statistical attributes as long as the new operating conditions remain (as per the definition of condition on page 17, where we have defined a condition as the average behaviour of a signal with constant statistical parameters). However, change in the cyclostationary signal due to its non-predictive quality, may manifest as chaos which has been the basis of a nonlinear method for feature extraction, and has found its applications in very recent publications [70]. In this section we will first briefly explain the theoretical concept behind Lyapunov exponent spectrum, and then apply the method to our data and see its practicality for extracting features from acoustic emissions of a gas turbine.

Let us consider a discrete version of the AE signal, representing a certain operating condition of a gas turbine denoted by C_i through the sampling procedure discussed on page 34, denoted by $x(n)$. Ergodicity (as defined on page 41) enables one to consider $x_1(n)$ and $x_2(n)$, as two separate time samples of the same cyclostationary process from which $x(n)$ was obtained, where

$$x_1(n) = \begin{cases} x(n) & n_0 \leq n \leq n_1 \\ 0 & \text{otherwise} \end{cases} \tag{2.42}$$

and

$$x_2(n) = \begin{cases} x(n) & n_0 + mT \leq n \leq n_1 + mT \\ 0 & \text{otherwise} \end{cases} \tag{2.43}$$

and m is a positive integer and T the discrete period of the cyclostationary process (also assumed as the continuous period to avoid numerous notations). As per the definition of cyclostationarity one can assume equivalent statistical parameters (mean and auto-correlation) for $x_1(n)$ and $x_2(n)$.

Now let us define a time invariant dynamical system represented by a finite dimensional manifold, $s \in \mathcal{M} \subset \mathbb{R}^p$, characterizing a gas turbine operating condition so that

$$\dot{s}(t) = f(s(t), d(t), u(t)) \tag{2.44}$$

where $d \in \mathbb{R}^q$, and $\{d_j(t), j = 1, 2, ..., q\}$ is the set of all system uncertainties and external disturbances, $u \in \mathbb{R}^r$, $\{u_l(t), j = 1, 2, ..., r\}$ is the set of system inputs, and $f : \mathbb{R}^{p \times q \times r} \rightarrow \mathbb{R}^p$ a possibly nonlinear function denoting the evolution of the system states in time, and let us further assume that the system states $s = \{s_1, s_2, ..., s_p\}$ are measurable through the acoustic emissions of the gas turbine, so that

$$x(t) = h(s(t), v) \tag{2.45}$$

where $x(t)$ is the AE signal, and $v \in \mathbb{R}$ denotes the measurement noise, and $h : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ is another (nonlinear) mapping function causing random acoustic emissions $x(t)$ correspond to the particular operating condition C_i.

Equations (2.44) and (2.45) collectively define the cyclostationary process (presuming cyclostationarity as on page 74) represented by $x(t)$. By sampling $x(t)$ with a sampling rate T_s, the time-series $x(nT_s)$ (or simply $x(n)$) is produced, which subsequently produces a family of $x_i(n)$-s (such as $x_1(n)$ and $x_2(n)$ in equations (2.42) and (2.43) respectively). Additionally let us assume that $s = 0$ is a globally and *exponentially stable* equilibrium point for the system, which denotes the gas turbine 'trip'. As we will see later in Chapter 3, ergodicity and stationarity can be considered as valid assumptions for our application, whereas global exponential stability might be too restrictive.

However, for the time being let us assume all four conditions (ergodicity, stationarity, global exponential stability, and time invariance) to hold. In this case for the unexcited system, given bounded perturbations we have

$$\| d(t) \| \leq \delta_d \Longrightarrow \| s(t) \| \leq \delta_s, \quad \text{if} \quad \| s(0) \| \leq \delta_0 \tag{2.46}$$

for a $\delta_d, \delta_s \in \mathbb{R}$, and all $t \in \mathbb{R}^+$ [71]. For the system under excitation, if we combine the effect of the input $u(t)$ and the system perturbations $d(t)$ as $u_d(t)$, then

$$\| s(t) \| \leq \beta(\| s(t_0) \|, t - t_0) + \gamma(\sup_{t_0 \leq \tau \leq t} \| u_d(\tau) \|) \tag{2.47}$$

where β is a class $\mathcal{KL}$ function and γ is a class $\mathcal{K}$ function [72]. Given a deterministic excitation $u(t)$ to the system in (2.44), the system state $s(t)$ will still behave randomly solely due to the effect of the random disturbances $d(t)$. However, in absence of d and

v one can assume the system trajectory being a periodic orbit (due to cyclostationarity condition), since

$$x(t+mT) = x(t) \quad \Longrightarrow \quad h(s(t+mT),0) = h(s(t),0)$$

or

$$s(t+mT) = s(t) \text{ for all } t \text{ as long as the condition is unchanged} \tag{2.48}$$

where T is the period of the cyclostationary process x, and m is a positive integer. Equation (2.48) in turn requires that

$$\dot{s}(t+mT) = \dot{s}(t)$$

and since $f(s,v,u)$ was assumed to be time invariant,

$$f(s(t+mT),0,0) = f(s(t),0,0) \text{ for all } t, \text{ or } f_t = f_{t+mT} \tag{2.49}$$

Equation (2.49) implies that if $x(t)$ is assumed to be periodic (or cyclostationary in a broader sense) then the system f would be a nonlinear periodic system, if $d = 0$ and $v = 0$. Therefore, according to (2.47) with bounded disturbances, as $t \to \infty$ the system trajectory will be confined within a boundary around the origin, as long as the operating condition C_i is satisfied.

Now let us consider the closest point to $s(t)$ on a nearby trajectory $s(t+\Delta t)$, so that $T \leq \Delta t$. According to equation (2.47), as $t \to \infty$

$$\| s(t) \| - \| s(t+\Delta t) \| \leq \gamma(\sup_{t_0 \leq \tau \leq t} \| u_d(\tau) \|) - \gamma(\sup_{t_0 \leq \tau \leq t+\Delta t} \| u_d(\tau) \|) \leq \epsilon_1 \tag{2.50}$$

for some $\epsilon_1 \in \mathbb{R}^+$. It is evident from equation (2.50) that ϵ_1 is independent from t, as long as $t_0 << t$. Hence,

$$\| s(t) - s(t + \Delta t) \| \leq \delta_1 \tag{2.51}$$

for all $t \in \mathbb{R}^+ >> t_0$, as long as $u_d(t)$ remains bounded within a boundary $\| u_d(t) \| \leq \alpha$, for some $\delta_1, \alpha \in \mathbb{R}^+$, if the condition remains unchanged.

Equation (2.51) requires that the measurements made for $s(t)$ and $s(t + mT)$ do not diverge unboundedly with time, that is

$$\| x(t) - x(t + mT) \| \leq \delta_x \quad \text{for some } \delta_x \in \mathbb{R}^+ \tag{2.52}$$

Equation (2.52) is often reformulated in the literature as

$$d_x(t) = \int_0^{nT} \| x(t + \tau) - x(t + mT + \tau) \| \, d\tau \leq \delta_x \tag{2.53}$$

where $x(.)$ is taken during a non changing condition, and $n, m \in \mathbb{N}$, and $n >> m$. One can assume an exponential divergence (or convergence) for the separation of nearby orbits, i.e.,

$$d_x(t) = d_0 e^{\lambda_1 nT}$$

with

$$\lambda_1 = \frac{1}{nT} ln\big(\frac{d_x(t)}{d_0}\big) \tag{2.54}$$

where d_0 is the initial separation of some nearby orbits. λ_1 in equation (2.54) is referred to as the largest Lyapunov exponent (LLE) and is the basis for determination

of the divergence (or convergence) of nearby orbits. A positive LLE indicates diverging trajectories, whereas a negative LLE is an indication of contraction between trajectories.

We learned from equation (2.53) that the system trajectories cannot diverge endlessly for an unchanging condition, provided that the system disturbances and uncertainties remain bounded within a constant boundary, and in case exponential stability can be assumed for $s = 0$. Therefore, a persisting positive LLE might be indicative of a change in the conditions, or due to momentary chaotic perturbations in the system, or measurement noise. In other words, when the system condition is changing, one would expect to see momentary changes in $d_x(t)$. Note that, the method of largest Lyapunov exponent has been successfully applied in a number of recent fault detection applications [70].

An algorithm often used for numerical computation of the largest Lyapunov exponent in the literature (referred to as Rosenstein algorithm) is a bit different from the concept explained above because the period of the cyclostationary process T is not always constant. Therefore, the algorithm is structured in order to be implementable without *a priori* about the process.

Let us again assume that $x(n)$ (for $n = \{1, 2, 3, ..., N\}$) is a discrete version (refer to page 34) of $x(t)$ under some constant condition. The first step is to construct the

Phase Space Matrix (denoted by P_x) of the time-series $x(n)$ as below:

$$P_x = \begin{bmatrix} x(1) & x(1+J) & x(1+2J) & ... & x(1+(m-1)J) \\ x(2) & x(2+J) & x(2+2J) & ... & x(2+(m-1)J) \\ x(3) & x(3+J) & x(3+2J) & ... & x(3+(m-1)J) \\ & & ... & & \\ x(M) & x(M+J) & x(M+2J) & ... & x(M+(m-1)J) \end{bmatrix}_{M\times m} \tag{2.55}$$

where $M+(m-1)J = N$. Note that m and n are different parameters than those used in equations (2.52) and (2.53).

In equation (2.55) the re-arrangement parameter m is referred to as the *embedding dimension*, J is called the *delay*, and M is the number of *reconstruction vectors*, all being positive integers. For our particular application we can assume that $T = mJ$ that is the period of the cyclostationary process. That is equivalent to taking m equally spaced samples of a complete cycle starting at $x(i)$, $i \in \{1, 2, 3, ..., M\}$. In other words, we are making M reconstruction vectors, each representing a complete cycle with m points, starting at $x(i)$.

The second step is determining the *closest* cycle to the cycle starting at $x(i)$ (say $x(j)$), in a sense that:

$$\| x(i) - x(j) \| = \min_{j\neq i} \Big[\sum_{k=0}^{m-1} \big(x(i+kJ) - x(j+kJ)\big)^2 \Big]^{1/2} \tag{2.56}$$

where j can be any integer between 1 to M, except i, so that equation (2.56) is satisfied. Therefore, corresponding to every row of P_x there is an index representing the closest trajectory to that cycle.

The next step consists of computing the distance of the closest trajectory to the reference trajectory at point r, i.e.,

$$d_x(i) = \sum_{r=1}^{M} \parallel x(i+r) - x(j+r) \parallel \tag{2.57}$$

where j is the index of the closest cycle to the i^{th} cycle, as computed in equation (2.56). Finally the last step, which is an estimate for the largest Lyapunov exponent, i.e.,

$$\lambda_1 \approx \frac{1}{R} \sum_{i=1}^{R} ln(d_x(i)) \tag{2.58}$$

where R is an arbitrary number of iterations for the numerical estimation of the average LLE.

In Figures 2.21 and 2.22 the LLE of $x_1(t)$ and $x_2(t)$ is depicted for $N = 10,000$ points, where $x_1(t)$ and $x_2(t)$ are those explained on page 53 used in preceding examples. Figure 2.21 illustrates λ_1 (as in equation (2.58)) for $x_1(t)$ and shows that for $N = 10,000$ the closest orbits tend to diverge at a constant pace ($\lambda_1 \approx 0.75$) after lapse of some time.

Initially the distance between two closest random trajectories diverge and converge in a sinusoidal manner, which shows the cyclic effect of $x(t)$. However the mean

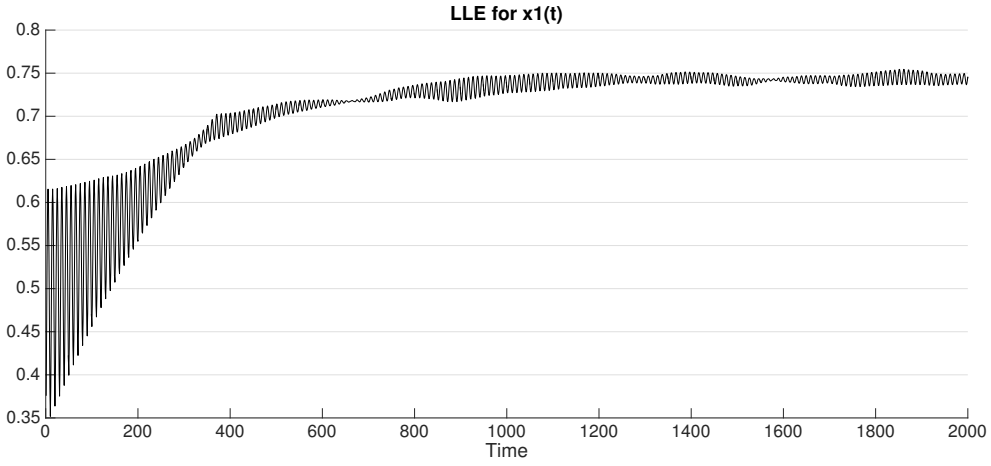

Figure 2.21: λ_1 (equation (2.58)) of $x_1(t)$

distance between the trajectories tend do diverge fast and after some time stabilize at a constant pace that is an indication of the chaos present in the signal. Figure 2.22

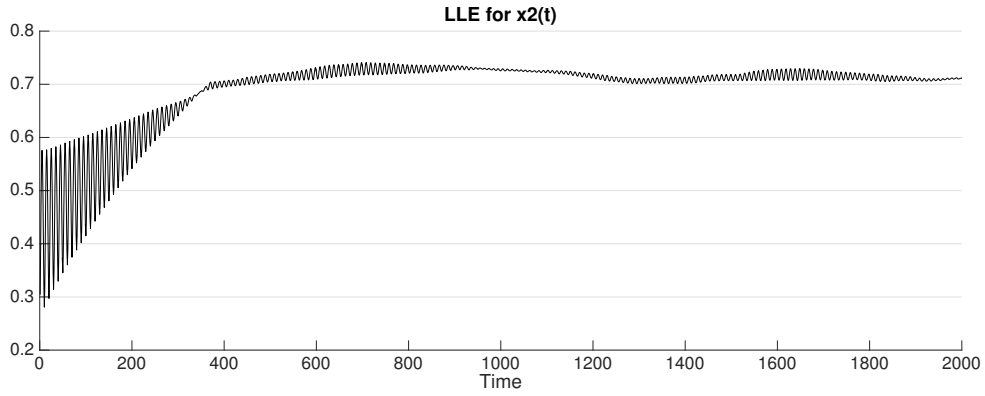

Figure 2.22: λ_1 (equation (2.58)) of $x_2(t)$

illustrates λ_1 of $x_2(t)$ which was expectedly more chaotic than $x_1(t)$, but actually shows a smaller λ_1 comparing $x_1(t)$. Finally in Figure 2.23 the LLE of a signal with a change in condition is depicted. In this figure we have generated a time sequence by merging

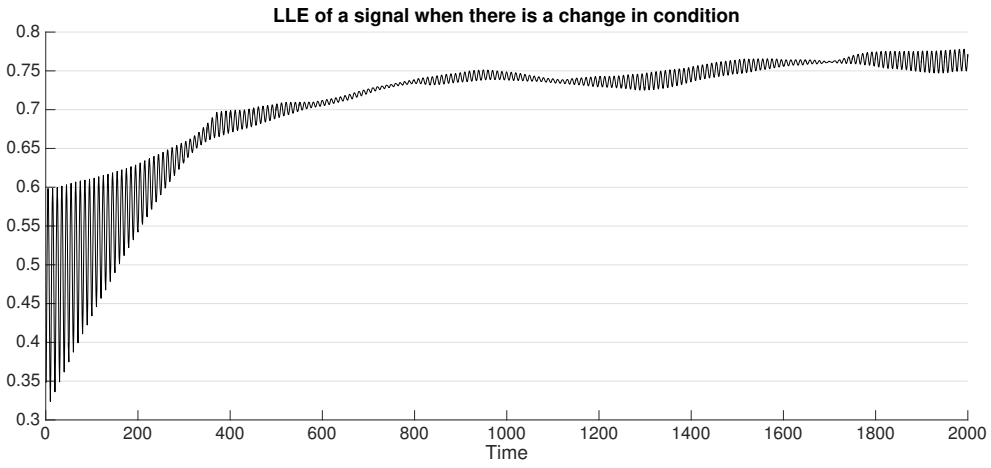

Figure 2.23: λ_1 (equation (2.58)) of $[x_1(t), x_2(t)]$

5000 points of $x_1(t)$ with 5000 points of $x_2(t)$ to emulate a sudden change in condition. We were expecting an abrupt change in λ_1 as a result of attaching signals with very different behaviours while this is not the case. However, the change in condition in Figure 2.23 shows a slightly bigger chaos than what can be seen in Figures 2.21 and 2.22.

However, since computation of LLE requires extensive computational resources it cannot be a promising technique for online condition monitoring. Additionally, as we can see this method cannot distinguish signals of longer periods, since the results are only comparable for signals of very short durations. As an alternative method to the LLE estimate discussed above, which is a measure of chaos in a non-linear signal, there are other techniques that are based on different definitions of 'distance' or 'separation' between vectors, other than that defined in equation (2.53). A commonly used statistical

test for 'regularity' in a time-sequence that has been recently used in many condition monitoring applications, is the measurement of Entropy which is the subject of our next section.

2.4 Approximate Entropy (ApEn)

As we stated in the previous sections, acoustic emissions of a gas turbine embed certain information about the conditions under which the machine is operating. We have demonstrated this through derivation of autocorrelation functions of the AE signal, and the periodic behaviour of which indicated cyclostationarity in the signal, hence *regularity*. Approximate entropy (ApEn) is a statistical technique that has found broad application recently for testing regularity, and even found its way into the field of fault detection and condition monitoring. The theoretical basis behind ApEn is very similar to what we have discussed for LLE, but with a different definition for distance between vectors.

Let us consider two discrete arbitrary signals $x(n)$ and $y(n)$ where $n = \{1, 2, ..., N\}$. Let us further take subsamples $x(i + n)$ and $y(j + n)$ of length m from each signal respectively, for some arbitrary i and j, so that $n = \{1, 2, ..., m - 1\}$ and $i + n \in [1, N]$ and $j + n \in [1, N]$ for all n.

The subsequences $x(i + n)$ and $y(j + n)$ and hereinafter referred to as vectors $x(i)$

and $y(j)$ respectively. We define the distance between $x(i)$ and $y(j)$ [73] as:

$$d(x(i), y(j)) = max_{k=1,2,...,m}(|x(i+k-1)-y(j+k-1)|), \quad \text{where} \quad i, j = 1, 2, ..., N-m+1 \tag{2.59}$$

In case we choose m to be approximately equal to the period of our cyclostationary process, then we can estimate the distance between any two cycles, as $d(x(i), x(j))$. Now let us consider that there are a total of $N - m + 1$ vectors (cycles) within a time sequence of length N. The idea is to first estimate the percentage of the vectors whose distances to a given vector $x(i)$ do not exceed a certain tolerance r, and name it the *similarity* of the time sequence to a vector $x(i)$, that is

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{j \neq i} \Theta[r - d(x(i), x(j))] \tag{2.60}$$

where

$$\Theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2.61}$$

In equation (2.60), $C_i^m(r)$ is the similarity of the time sequence x to vector $x(i)$ (which has a length m) with a tolerance r, where r can be $k\sigma_x$ for a positive factor of k, where σ_x is the standard deviation of x. By definition, ApEn is defined as

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{2.62}$$

where

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} ln[C_i^m(r)] \tag{2.63}$$

As stated above, ApEn is a measure of regularity or predictability in a time series. The presence of repetitive patterns of fluctuation in a time series renders it more predictable than a time series in which such patterns are absent. ApEn reflects the likelihood that similar patterns of observations will not be followed by additional similar observations. A time series containing many repetitive patterns has a relatively small ApEn; a less predictable process has a higher ApEn [60].

In our application as the process is cyclostationary, we can expect high regularity for a completely stable operating condition, whereas changes in operating conditions would expectedly cause transient unpredictability in the AE signal and accordingly increase in the signal ApEn.

In order to demonstrate the application of ApEn we have taken five different AE samples from three conditions, namely stable, transient, and thermally unstable conditions (as described on page 37) and computed ApEn for each sample. Results are given in Table 2.1. We expect all the five subsamples taken from same conditions to show similarity, i.e., same range of entropies. However, for example the 4^{th} sample of the transient condition shows similarity to the samples taken from the thermally stable condition. Same applies for the 3^{rd} sample of the unstable condition.

Sample No.	Thermally Stable	Transient Condition	unstable (Cold)
1^{st} Sample	6.1×10^{-4}	18×10^{-4}	24×10^{-4}
2^{nd} Sample	5.5×10^{-4}	19×10^{-4}	17×10^{-4}
3^{rd} Sample	4.4×10^{-4}	20×10^{-4}	6.6×10^{-4}
4^{th} Sample	7.9×10^{-4}	7.1×10^{-4}	14×10^{-4}
5^{th} Sample	7.1×10^{-4}	23×10^{-4}	21×10^{-4}

Table 2.1: ApEn results for five AE samples of a gas turbine running in different operating conditions.

Computational process for calculating ApEn is unfortunately too heavy for long time series, and therefore unaffordable for online implementation. Samples taken for Table 2.1 were 2000 points only, and insufficient to experiment actual transitional results, when a gas turbine is changing operating conditions from one to another. However, processing time series longer than 2000 are indeed tedious, and has not been investigated in this thesis.

Nevertheless, looking into the results of Table 2.1 shows that ApEn is on average capable of demarcation between a normal and abnormal operating condition, whilst it is still incapable of clearly separating transient condition versus cold, since the variation of the entropies of cold condition (6.6×10^{-4} to 24×10^{-4}) is almost equal to the variations

of entropies for the transient condition (7.1×10^{-4} to 23×10^{-4}).

For gas turbine condition monitoring we are interested in isolating various operating conditions, and it is essential to deploy a feature parameter that is capable of distinguishing different conditions accurately while being immune to system noise and temporary fluctuations. In conclusion, due to excessive computational process and insufficiency of accurate demarcation between different conditions, we still believe that ApEn cannot be a suitable feature parameter for gas turbine condition monitoring.

2.5 Conclusion

In this chapter, we have tried to examine some of the prominent available feature extraction methods used in fault detection and condition monitoring applications, which have specially attracted attention of the AE technology. Before going through the results it is worth noting that traditional transformation methods based on Fourier analysis are deemed to be inappropriate for nonstationary and/or nonlinear signals. Therefore, the main focus of available feature extraction methods for AE signals are generally based on nonstationary signal processing techniques, the first of which encompasses continuous time and discreet time wavelet decomposition with emphasize on multiresolutional analysis.

Continuous wavelet analysis, being difficult to implement numerically, has not drawn much attraction in the literature for fault detection and machine condition monitoring applications. We have seen that it is not capable of discriminating between different but close unstable operating conditions. Multiresolutional analysis on the other hand, with the ability to implement online through recursive filtering, has made a successful reputation in machine condition monitoring, especially in vibration analysis and AE technology. This chapter has reviewed the theoretical aspects of wavelet multiresolution analysis (refer to section 2.1) and has also applied this method for feature extraction of gas turbine AE signal in the latter part of the section. Due to low frequency resolution of wavelet coefficients in the higher end of the frequency spectrum, we have seen that wavelet analysis is unable to discriminate between different transient conditions.

Additionally, we have seen that the dimension of the feature vector is often very large (depending on the required number of resolution levels) and therefore necessitates sophisticated machine learning algorithms for interpretation of the features, i.e., wavelet coefficients in principal. Therefore, the historical challenge of making useful help from multiresolution analysis in condition monitoring applications had been in dealing with the dilemma of low-frequency versus low-time resolution. In this connection a more advanced feature extraction method was developed which uses Empirical Mode Decomposition (EMD) and Hilbert Transform, for instantaneous frequency analysis of a given

signal. While reviewing the theoretical aspects of Hilbert transform, we have highlighted that instantaneous frequency extraction through Hilbert transform is only meaningful when the subjected signal is close to a monocomponent signal.

In this connection, we have attempted to decompose the gas turbine AE signal through wavelets, and then decompose the dominant energy components further into Intrinsic Mode Functions (IMFs). Then in order to reduce the dimension of the feature vector, available publications suggest screening of the individual components, and removing non-dominant sub-components (energy-wise), before computing instantaneous frequency. Consequently, we have concluded that high-frequency sub-components are energy-wise so negligible that do not find chance for further process through the screening stage. Moreover, the methodology was found excessively involved computationally, and unaffordable for online monitoring purposes.

In addition to the frequency domain methods for feature extraction explained above, there are a number of purely statistical techniques in stochastic signal processing, which have found a lot of applications in machine condition monitoring very recently. In this chapter, we have reviewed the theory of Largest Lyapunov Exponent (LLE) in depth, measuring chaos of a signal and its relation with condition monitoring. We have also applied the theory and some computer algorithms in estimation of the LLE, and

have found that while the estimation of LLE is capable of correctly identifying 'boundedness' in system perturbations that give rise to the random behaviour of the AE signal, it is incapable of discriminating between different operating conditions.

We have finally re-defined the concept of 'distance' or 'separation' (used in ApEn theory) among a set of cyclic random signals, in contrast to the distance adopted in LLE, and used another statistical test of 'regularity' in a signal, referred to as Approximate Entropy (ApEn), trying to correlate it with condition monitoring of gas turbine AE. We have concluded certain shortcomings of the ApEn feature parameter in the gas turbine AE signals. Except multi-resolution analysis via wavelets, all other feature extraction techniques were unfortunately excessively involved computationally, and inappropriate for online processing.

In the next chapter we will try to rectify these shortcomings by introducing a customized feature function for the particular case of gas turbine condition monitoring of AE signals.

Chapter 3

Characteristics of a Gas Turbine AE Signal

In the previous chapter we have loosely used statistical terms such as stationarity, cyclostationarity, ergodicity, etc. and their relationship with feature extraction in a non-concrete fashion. We have pinpointed that conventional feature extraction methods are incapable of discrimination between certain gas turbine operating conditions, by analyzing samples of a gas turbine during transient thermal conditions, and have suggested that the autocorrelation function of the AE signal may provide better characteristics for condition monitoring. However, we have not addressed the fact that given a piecewise (time limited) AE sample, we are only able to compute an estimate of the autocorrelation function, for a particular duration in time, that cannot generally infer beyond the

duration of the sample. In this chapter we will try to build a mathematical framework for analysis of AE signals using piecewise segments and then try to extract characteristics that remain invariant regardless of the signal length. Then we will be able to infer about the conditions under which the machine is operating.

3.1 Cyclostationarity

Mechanical signals produced by rotating or reciprocating machines are examples of random processes which exhibit cyclic behaviour. For instance, a gas turbine can be depicted by the T-S (temparature-entropy) Brayton cycle, in which gas is first compressed and heated (ideally) under constant entropy in the compressor, and then ignited in the combustor during which its temperature and entropy rises excessively. The thermal energy so delivered to the gas flow is then converted into mechanical power through the turbine stages, and finally the exhaust cycle in which temperature and entropy returns to initial state. The four cycles described above exhibit a repetitive processes, although in each and every cycle there are influences from random and non-repetitive surrounding processes (hereinafter referred as *process noise*), that cause the random behaviour of the signal. For instance, in its simplest case, let us assume that a gas turbine AE signal to be composed of a set of cyclic random processes $x_p(t)$ with identical periods T_0, and a set of random uncertainties $x_r(t)$, that is

$$x(t) = x_p(t) + x_r(t) \tag{3.1}$$

so that

$$E\big[x_p(t_1)x_p(t_2)\big] = E\big[x_p(t_1 + mT_0)x_p(t_2 + nT_0)\big] \quad \text{for all } t_1,\ t_2 \text{ and integers } n, m$$

and,

$$E\big[x_p(t)\big] = E\big[x_p(t + nT_0)\big] \quad \text{for all } t \text{ and integer n}$$

and also,

$$E\Big[x_p(t)x_r(t)\Big] = E\Big[x_p(t)\Big]E\Big[x_r(t)\Big] \quad \text{for all } t$$

and if,

$$E\Big[x_r(t)\Big] = 0 \quad \text{for all } t$$

then

$$E\Big[x(t)\Big] = E\Big[x(t + T_0)\Big] \tag{3.2}$$

Also, we have

$$R_x(t,\tau) = E\Big[x(t)x(\tau)\Big] = E\Big[\Big(x_p(t) + x_r(t)\Big)\Big(x_p(\tau) + x_r(\tau)\Big)\Big]$$

$$= E\Big[x_p(t + nT_0)x_p(\tau + mT_0)\Big] + E\Big[x_p(t)\Big]E\Big[x_r(\tau)\Big] + E\Big[x_p(\tau)\Big]E\Big[x_r(t)\Big] + E\Big[x_r(t)x_r(\tau)\Big]$$

Therefore,

$$R_x(t,\tau) = R_{x_p}(t + nT_0, \tau + mT_0) + R_{x_r}(t,\tau) \tag{3.3}$$

If we assume $R_{x_r}(t,\tau) << R_{x_p}(t + nT_0, \tau + mT_0)$, then

$$R_x(t,\tau) \approx R_{x_p}(t + nT_0, \tau + mT_0) \tag{3.4}$$

A random process satisfying the conditions (3.2) and (3.4) is said to be a wide-sense cyclostationary random process. It has been shown through examples that signals of rotary mechanical equipment such as compressors, turbines, gears, bearings, etc. are generally remarkably well modelled as cyclostationary processes [33]. The cyclostationary family accepts all signals with hidden periodicities, either of the additive type (presence of tonal components) or multiplicative type (presence of periodic modulations) [60].

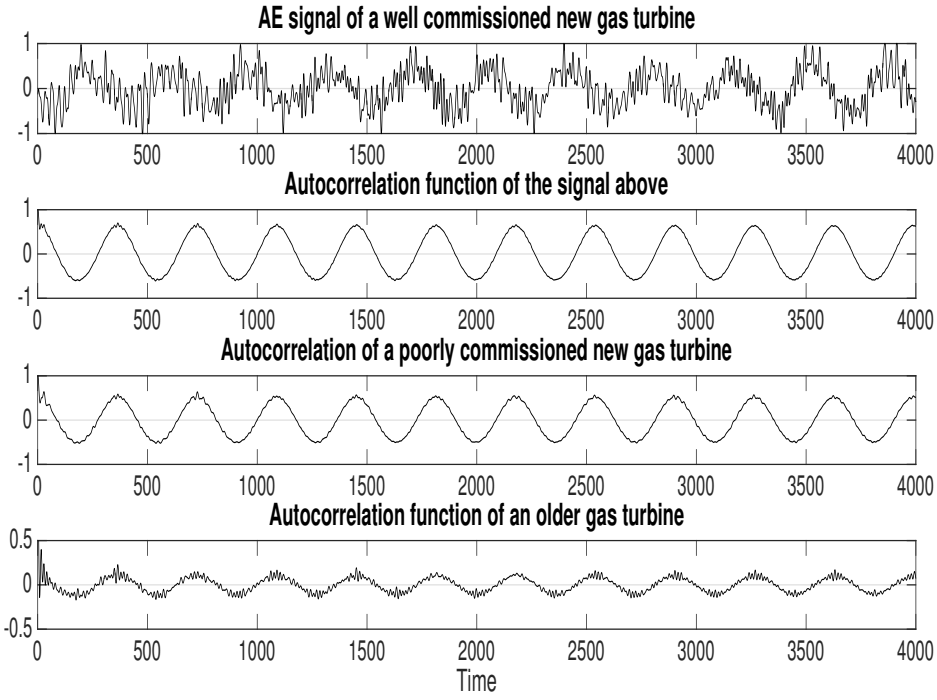

Figure 3.1: Time signal and autocorrelation function of three gas turbine AE Samples with different conditions.

Figure 3.1 shows the autocorrelation function of three different single-shaft gas

turbine AE samples. We can see an obvious sinusoidal waveform in all the three samples that is the R_{x_p} component in equation (3.3), added with a small noise R_{x_r}. In the first subplot the time sequence of a well commissioned gas turbine is depicted. The second subplot is the autocorrelation function of the first subplot, that is close to a pure sinusoidal waveform. A perfect sinusoidal autocorrelation is an example of an ideal cyclostationary process. The signal to noise ratio of the autocorrelation function may indicate the stability or predictiveness of the process. This can be more explained by looking into the last subplot of Figure 3.1, where the process noise of the old gas turbine is relatively larger than the newly commissioned engines. This is due to the fact that in an older engine there are more non-engineered processes involved (due to ageing), hence lowering the predictiveness of the process.

Interestingly, we can also see that in all the samples, the period of the autocorrelation function is exactly identical, despite the fact that these samples belong to different gas turbines and are taken at very different days. However, in all the three subplots, conditions (3.2) and (3.4) are evident. Hence, these are examples of a cyclostationary random process with its period equal to the inverse of the rotor speed. In this thesis, cyclostationarity is assumed as a pre-requisite for all the forthcoming derivations:

Assumption 3.1: *A gas turbine AE signal operating under a set of unchanging condition is assumed to be a cyclostationary random process.*

3.2 Wide-Sense Stationarity

By definition, a wide-sense stationary random process $x(t)$ satisfies the following conditions:

$$E\Big[x(t)\Big] = m_x \quad \text{for all } t \tag{3.5}$$

where we assume $m_x = 0$ in this thesis (as later specified in Assumption 3.2), and

$$C_x(t_1, t_2) = E\Big[\Big(x(t_1) - m_x\Big)\Big(x(t_2) - m_x\Big)\Big]$$

$$= E\Big[\Big(x\big(t_1 + (-t_2)\big) - m_x\Big)\Big(x\big(t_2 + (-t_2)\big) - m_x\Big)\Big] = C_x(t_1 - t_2, 0) \tag{3.6}$$

for all t_1, t_2, where C_x is the covariance function of $x(t)$, and $E[x(t)]$ is the expectation of $x(t)$. Equation (3.6) also obviously implies that $R_x(t_1, t_2) = R_x(t_1 - t_2)$, where R_x stands for the autocorrelation function of $x(t)$.

Practically speaking, if a random process such as $x(t)$ is not wide-sense stationary (WSS), its statistical parameters would not remain constant throughout the course of the time, and consequently its Fourier transform (for instance) would not hold any reliable physical meaning. Specifically, the Fourier transform of $R_x(t_1, t_2)$ has to be taken with respect to t_1 for a given t_2, or the other way around. However, in case of a wide-sense

stationary process, as per the Wiener-Khinchin theorem we have

$$R_x(t+\tau,t) = E\big[x(t+\tau)x(t)\big] = x(t) * x(-t)$$

and

$$\mathscr{F}_t\Big(R_x(t+\tau,t)\Big) = \mathscr{F}_t\Big(x(t) * x(-t)\Big) = X(\omega)X^*(\omega) = |X(\omega)|^2 \tag{3.7}$$

where in equation (3.7), $x(t) * x(-t)$ denotes convolution of $x(t)$ with $x(-t)$, and $X^*(\omega)$ denotes the complex conjugate of $X(\omega)$.

It follows that the Fourier transform conveys meaningful information. In case of cyclostationary random processes, R_x is periodic when $x(t)$ spans to infinity and $\mathscr{F}(R_x(\tau))$ returns Fourier coefficients of its cosine components, representing instantaneous frequencies of R_x, or the power spectral density of $x(t)$. However, in case of nonstationary processes, we are obliged to apply nonlinear signal processing methods (such as WPD, HHT, EMD, etc.) for instantaneous frequency determination of the process, as were done in Chapter 2.

Therefore, it is firstly important to assure that one is dealing with stationary signals when applying Fourier analysis. Secondly, when a cyclostationary process $x(t)$ ($t \in (-\infty, +\infty)$) with period T is sampled for a finite duration to produce $\big[x(t)\big]_{T_x}$,

where,

$$\big[x(t)\big]_{T_x} = \begin{cases} x(t), & t \in [0, T_x], \text{ where } T_x >> T \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

then $\big[x(t)\big]_{T_x}$ cannot be interchangeably used for $x(t)$, except through a careful procedure which requires another condition for $x(t)$ that is referred to as *ergodicity*. And finally, special care has to be taken to assure that the Fourier transform of $\big[x(t)\big]_{T_x}$ samples, converges into deterministic results to become useful. However, as discussed earlier, the first step is 'stationarizing' the original $x(t)$.

Property 3.1: *Assume that $x(t)$ is an i.i.d (independent and identically distributed) cyclostationary random process with period T and $E[x(t)] = 0$. Then $x(t + t_s)$ is Wide Sense Stationary (WSS), if $t_s \in [0, T]$ is a random variable with a uniform probability distribution.*

Proof. There are two conditions to be proven in order to show that $x(t + t_s)$ is Wide Sense Stationary (WSS), namely, (1) $E[x(t + t_s)] = 0$, and (2) $R_x(t + t_s, t + t_s + \tau)$ is a function of τ only.

In order to show $E[x(t + t_s)] = 0$, one should note that $x(t)$ is a random process and t_s is a random variable. Therefore, $E[x(t + t_s)]$ can be derived in two steps, first

with respect to t_s, and second with respect to $x(t)$, i.e.,

$$E[x(t+t_s)] = E_{t_s}\Big[E_x\big[x(t+t_s)\big]\big|t_s\Big]$$

Since t_s is assumed to be a random variable with uniform distribution in $[0,T]$, we have

$$E_{t_s}\Big[E_x\big[x(t+t_s)\big]\big|t_s\Big] = \frac{1}{T}\int_0^T E_x\big[x(t+t_s)\big]dt_s$$
$$= \frac{1}{T}\int_t^{t+T} E_x\big[x(\tau)\big]d\tau = 0$$

which proves the first condition. In a similar manner

$$R_x(t+t_s, t+t_s+\tau) = E\Big[x(t+t_s)x(t+t_s+\tau)\Big]$$
$$= E_{t_s}\Big[E_x\big[x(t+t_s)x(t+t_s+\tau)\big]\big|t_s\Big]$$
$$= \frac{1}{T}\int_0^T E_x\big[x(t+t_s)x(t+t_s+\tau)\big]dt_s$$
$$= \frac{1}{T}\int_t^{t+T} E_x\big[x(t')x(t'+\tau)\big]dt'$$

and due to cyclostationarity of $x(t)$

$$= \frac{1}{T}\int_0^T E_x\big[x(t')x(t'+\tau)\big]dt'$$

which is only a function of τ, therefore $x(t+t_s)$ is WSS. ■

3.3 Mean Square Periodicity

By definition a deterministic signal $x(t)$ is periodic (with the period T) if $x(t+T) = x(t)$. However, for a cyclic process $x(t)$ with a period T, $x(t+T)$ is equal to $x(t)$ in a *mean*

square sense, hence

$$E\big[\big(x(t+T)-x(t)\big)^2\big]=0 \tag{3.9}$$

It can be shown that when $x(t)$ is mean square periodic, it can be represented in a mean square sense by a Fourier series [74]:

$$x(t)=\sum_{k=-\infty}^{+\infty} X_k e^{j2\pi kt/T} \tag{3.10}$$

where the coefficients X_k are *orthogonal* (i.e., uncorrelated) random variables defined by

$$X_k=\frac{1}{T}\int_0^T x(t)e^{-j2\pi kt/T}dt \tag{3.11}$$

and

$$E\big[X_k X_n^*\big]=0, \quad \text{for } k\neq n. \tag{3.12}$$

Property 3.2: *If $x(t)$ is a cyclostationary process, it is also mean square periodic.*

Proof. Let us assume that $x(t)$ is a cyclostationary random process, then

$$R_x(0,0)=R_x(T,0)$$

However, due to the cyclostationarity condition, since $E[x(t+T)^2]=E[x(t)^2]=R_x(0,0)$, and since $R_x(T,0)=E[x(t+T)x(t)]$, we can write

$$E\Big[x(t+T)^2\Big]+E\Big[x(t)^2\Big]=E\Big[x(t+T)x(t)\Big]+E\Big[x(t+T)x(t)\Big]$$

or

$$E\Big[x(t+T)^2\Big]+E\Big[x(t)^2\Big]-2E\Big[x(t+T)x(t)\Big]=0$$

i.e.,

$$E\Big[\Big(x(t+T)-x(t)\Big)^2\Big]=0$$

This completes the proof. ■

Property 3.2 provides a number of useful results. For instance, let us take a another look at the autocorrelation function of a cyclostationary process $x(t)$ in a mean square sense:

$$\begin{aligned}
R_x(\tau) &= E\big[x(t+\tau)x(t)\big] \\
&= E\left[\Big(\sum_{k=-\infty}^{+\infty} X_k e^{j2\pi k(t+\tau)/T}\Big)\Big(\sum_{k'=-\infty}^{+\infty} X_{k'} e^{j2\pi k' t/T}\Big)\right] \\
&= \sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\big[X_k X_{k'}\big] e^{j2\pi kt/T} e^{j2\pi k' t/T} e^{j2\pi k\tau/T} \\
&= \sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\big[X_k X_{k'}\big] e^{j2\pi (k+k')t/T} e^{j2\pi k\tau/T} \qquad (3.13)
\end{aligned}$$

However $E[X_k X_{k'}]=0$ for all $k'\neq -k$, since $E[X_k X_{k'}^*]=0$ for all $k\neq k'$, and $X_k^*=X_{-k}$ (since $x(t)$ is a real signal). Accordingly all the terms in equation (3.13) vanish except for $k=-k'$, therefore equation (3.13) will reduce to

$$R_x(\tau)=\sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\big[X_k X_k^*\big] e^{j2\pi k\tau/T}$$

or

$$R_x(\tau)=\sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] e^{j2\pi k\tau/T} \qquad (3.14)$$

Equation (3.14) demonstrates that $E\Big[|X_k|^2\Big]$ can be used to perfectly reconstruct $R_x(\tau)$. We can obtain the variance of the process $x(t)$ directly from equation (3.14). If $E[x(t)]=$

$m_x = 0$,

$$VAR\Big[x(t)\Big] = \sigma_x^2 = E\Big[(x(t) - m_x)^2\Big]$$
$$= E\Big[x(t)^2\Big] = \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] \tag{3.15}$$

or,

$$\sigma_x^2 = \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] \tag{3.16}$$

Equation (3.16) indicates that the variance of the process $x(t)$ is equal to the sum of ensemble mean of its energy components. Equation (3.16) also implies that the variance of a cyclostationary process $x(t)$ is equal to the sum of variances of all its frequency components $\sigma^2_{|X_k|}$, since $\sigma^2_{|X_k|} = E[|X_k|^2]$. Therefore, if a change in the machine condition exhibits a new frequency component, the corresponding change in the variation of the signal is equal to the ratio of the fault energy to the total energy of the signal, which can be extremely small and undetectable. It therefore seems inappropriate to use the statistical parameters of an unprocessed signal for machine condition monitoring, due to small variation of these parameters in response to minor changes.

Now let us assume that a change in the machine condition manifests certain monotonic frequency components. For instance the ageing effect depicted in Figure 3.1 is a result of a complicated process that in manifested through a set of monotonic frequency components superimposed on the main process. Specifically, a mechanical change in the gas turbine (if detectable) can be considered as an emergence of a new source of

energy. Therefore, detection of a change in the machine condition requires monitoring of the variance of the energy of all subcomponents constituting the signal according to equation (3.16). We have earlier presented through equation (3.10) that a mechanical signal of a rotary machine $x(t)$ (being a cyclostationary process), can be represented by a sum of monotonic frequency components with coefficients X_k, where these coefficients are themselves random variables. From equation (3.14), one can note:

$$E\Big[|X_k|^2\Big] = \int_{-\infty}^{+\infty} R_x(\tau) e^{-j2\pi k\tau/T} d\tau \tag{3.17}$$

The first subplot of Figure 3.1 shows a short fraction of the AE signal that is obtained from a gas turbine in stable operation (refer to Section 1.4 for details about the properties of the gas turbines used for collecting AE data, and the data acquisition method). The first plot displays the cyclic nature of the process (cyclostationarity). The second subplot is the autocorrelation function computed for 4000 points in time, which is also clearly periodic and close to a pure sinusoidal waveform. The frequency components of the autocorrelation function $R_x(\tau)$ are shown in Figure 3.2, representing the variance of the frequency coefficients of the original process $x(t)$. Peaks in Fig. 3.2 are each representing a *varying* energy source, that might be a cause of a change in the machine condition, an anomaly, or only a result of the random nature of the process. In order to detect changes in machine conditions and/or anomalies, one has to remove the randomness associated with the AE signal. We will elaborate this later in Section 3.5. Returning to Figure 3.2 the first two peaks occur at 121Hz and 242Hz which represent

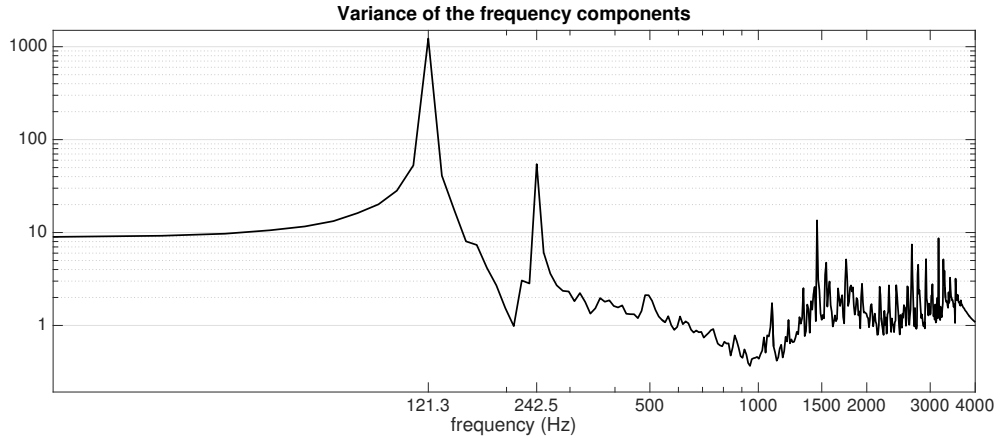

Figure 3.2: Variance of frequency components $|X_k|^2$ of the first subplot of Figure 3.1.

the rotor speed and its second harmonic, respectively.

The first peak at 121Hz seems perfectly normal because of the instantaneous power demand in the grid. The gas turbines are all coupled with the electrical grid through synchronous generators, which causes the turbine speed to change in response to the instantaneous power demand of the network. Accordingly we can expect variations in $|X_k|^2$ at 121Hz which seem normal. The second peak however at 242Hz might be an early indication of the rotor unbalanced weight or 'bowing'. There are also peaks at higher frequencies that cannot be explained and might be due to changes in the conditions, or even faults.

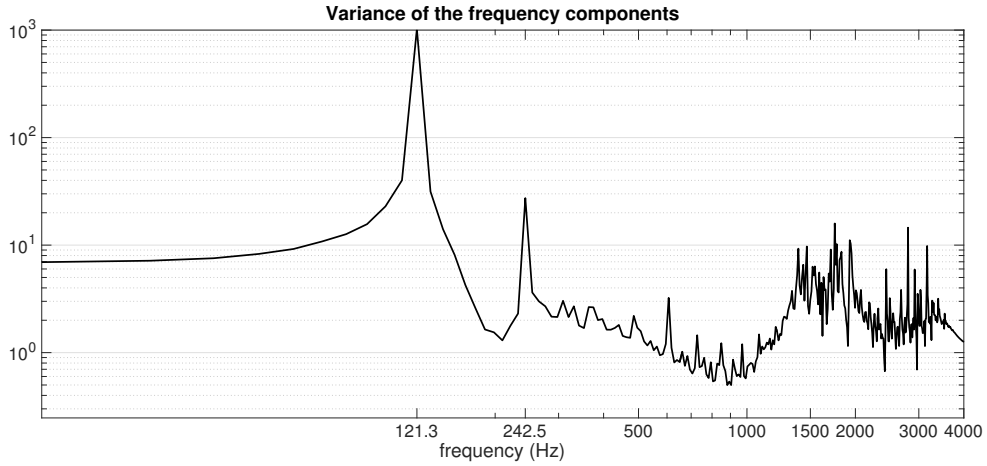

Figure 3.3: Fourier transform of the third subplot of Figure 3.1.

Figure 3.3 depicts the Fourier transform of the third subplot of Figure 3.1, indicating the variance of the magnitudes of the frequency components of an AE signal belonging to a new gas turbine that was not commissioned very professionally, as there are more powerful activities in the higher end of the frequency spectrum, which is an undesirable phenomenon due to the fact that more fuel energy is wasted in the process noise.

Figure 3.4 illustrates the Fourier transform of the last subplot of Figure 3.1, collected from an aged gas turbine of the same brand, make and size of Figures 3.2 and 3.3. It is interesting to note that the power of the second harmonic has diminished (possibly) due to a long uninterrupted operation of the gas turbine at full speed, that has gradually corrected the bowing effect of the shaft. However, unfortunately the power of the useful process (the 121Hz component) has also diminished, and more energy is wasted on the

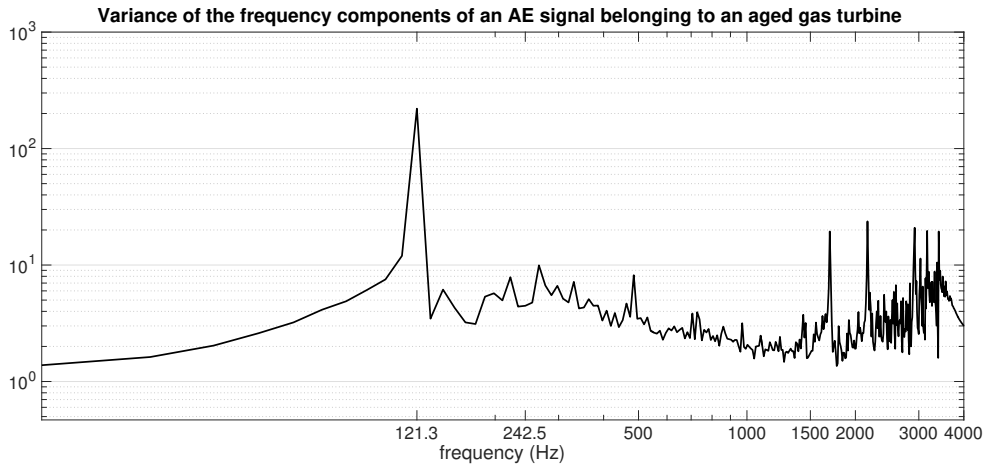

Figure 3.4: Fourier transform of the last subplot of Figure 3.1.

process noise above 500Hz, which is the expected effect of ageing. These derivations hint an important observation:

Observation: *Dominating variations in the energy sources of a rotary machine may discriminate machines and/or operating conditions, if one can demonstrate that these variations persist in location (i.e., in their frequencies) when the operating conditions are held constant.*

There are two fundamental aspects that need to be examined before we can finally verify this observation. (1) The autocorrelation function cannot be derived except when the process is *ergodic*. (2) There is no guarantee that if the experiment is repeated, the same results are obtained. Therefore, one cannot generalize the above observation,

except if one can prove the aforementioned statements. The focus of the subsequent sections will be to investigate these aspects.

3.4 Ergodicity

Ergodicity in a broad sense refers to the convergence of a process statistical measures *in time* into its *ensemble*'s statistical parameters. For instance, if the time average of a random process $x(t)$ is defined as:

$$\Big\langle x(t) \Big\rangle_{T_x} = \lim_{T_x \to \infty} \frac{1}{T_x} \int_{-T_x/2}^{T_x/2} x(t) dt$$

where T_x is a real positive quantity. *The process is then called mean ergodic, or* 1^{st} *order ergodic, if*

$$\lim_{T_x \to \infty} E\Big[\big\langle x(t) \big\rangle_{T_x} \Big] = E\Big[x(t) \Big] \quad \textit{and} \quad \lim_{T_x \to \infty} VAR\Big[\langle x(t) \rangle_{T_x} \Big] = 0 \tag{3.18}$$

In a similar way, the random process $x(t)$ *is said to be ergodic in autocorrelation function (ACF), or* 2^{nd} *order ergodic, if*

$$\lim_{T_x \to \infty} E\Big[\big\langle R_x(\tau) \big\rangle_{T_x} \Big] = R_x(\tau) \quad \text{and} \quad \lim_{T_x \to \infty} VAR\Big[\big\langle R_x(\tau) \big\rangle_{T_x} \Big] = 0 \tag{3.19}$$

where (by definition),

$$\big\langle R_x(\tau) \big\rangle_{T_x} = \frac{1}{T_x} \int_{-T_x/2}^{T_x/2} x(t) x^*(t-\tau) dt$$

Conditions (3.18) and (3.19) if satisfied, would allow us to take a long enough sample of an AE signal and discern about the process's mean and autocorrelation functions.

The following property would prove that a wide sense cyclostationary process can be considered ergodic.

Property 3.3: *Let $x(t)$ be a cyclostationary process and $E\big[x(t)\big] = 0$. If $x(t)$ is considered WSS (as per Property 3.1), then the random process $x(t)$ will be ergodic in the mean sense and in ACF, i.e., 1^{st} and 2^{nd} order ergodic.*

Proof. We can easily see that

$$\begin{aligned} E\Big[\big\langle x(t)\big\rangle_{T_x}\Big] &= E\Big[\frac{1}{T_x}\int_{-T_x/2}^{T_x/2} x(t)dt\Big] \\ &= \frac{1}{T_x}\int_{-T_x/2}^{T_x/2} E\Big[x(t)\Big]dt \\ &= \frac{E\Big[x(t)\Big]}{T_x}\int_{-T_x/2}^{T_x/2} dt \\ &= E\Big[x(t)\Big] = 0 \end{aligned}$$

Additionally,

$$\begin{aligned} \lim_{T_x\to\infty} E\Big[\big\langle x(t)\big\rangle_{T_x}^2\Big] &= \lim_{T_x\to\infty} E\Big[\frac{1}{T_x^2}\int_{-T_x/2}^{T_x/2}\int_{-T_x/2}^{T_x/2} x(t)x(t')dtdt'\Big] \\ &= \lim_{T_x\to\infty}\frac{1}{T^2}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} E\Big[x(t)x(t')\Big]dtdt' \end{aligned}$$

$$= \lim_{T_x \to \infty} \frac{1}{T_x^2} \int_{-T_x/2}^{T_x/2} \int_{-T_x/2}^{T_x/2} R_x(t - t') dt dt'$$

In this integral $t - t'$ spans from $-T_x$ to T_x. As the integrand is a function of only $t - t'$, the integrand is a constant function along the line $\tau = t - t'$. So the integral can be considered as the area under the curve $(1 - |\tau|/T_x)R_x(\tau)$, where $-T_x \leq \tau \leq T_x$, hence,

$$\lim_{T_x \to \infty} E\Big[\big\langle x(t) \big\rangle_{T_x}^2\Big] = \lim_{T_x \to \infty} \frac{1}{T_x} \int_{-T_x}^{T_x} \Big(1 - \frac{|\tau|}{T_x}\Big) R_x(\tau) d\tau \tag{3.20}$$

and as per equation (3.14), assuming that T is the period of the process (not to confuse that with T_x),

$$\lim_{T_x \to \infty} E\Big[\big\langle x(t) \big\rangle_{T_x}^2\Big] = \lim_{T_x \to \infty} \frac{1}{T_x} \int_{-T_x}^{T_x} \Big(1 - \frac{|\tau|}{T_x}\Big) \Big(\sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] e^{j2\pi k\tau/T} \Big) d\tau$$

$$= \lim_{T_x \to \infty} \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] \Big(\frac{1}{T_x} \int_{-T_x}^{T_x} \big(1 - \frac{|\tau|}{T_x}\big) e^{j2\pi k\tau/T} d\tau \Big)$$

However,

$$\int_{-T_x}^{T_x} \big(1 - \frac{|\tau|}{T_x}\big) e^{-j2\pi k\tau/T} d\tau = \mathscr{F}\big(s(t) * s(t)\big)$$

where $s(t) * s(t)$ denotes the convolution of $s(t)$ with itself, where,

$$s(t) = \begin{cases} \frac{1}{2T_x} & -T_x \leq t \leq +T_x \\ \\ 0 & \text{otherwise} \end{cases}$$

The Fourier transform of $s(t)$ can be given by:

$$\mathscr{F}\big(s(t)\big) = \frac{sin(2\pi k T_x/T)}{2\pi k T_x/T} = Sinc(2\pi k T_x/T)$$

therefore,

$$\frac{1}{T_x}\int_{-T_x}^{T_x}\big(1-\frac{|\tau|}{T_x}\big)e^{-j2\pi k\tau/T}d\tau=\frac{1}{T_x}Sinc^2(-2\pi kT_x/T)=\frac{1}{T_x}Sinc^2(2\pi kT_x/T)$$

and,

$$\lim_{T_x\to\infty}E\Big[\big\langle x(t)\big\rangle_{T_x}^2\Big]=\lim_{T_x\to\infty}\frac{1}{T_x}\sum_{k=-\infty}^{+\infty}E\Big[|X_k|^2\Big]Sinc^2(2\pi kT_x/T)=0 \tag{3.21}$$

or,

$$\lim_{T_x\to\infty}VAR\Big[\langle x(t)\rangle_{T_x}\Big]=0 \tag{3.22}$$

which proves mean-ergodicity of the process. On the other hand,

$$\lim_{T_x\to\infty}E\Big[\big\langle R_x(\tau)\big\rangle_{T_x}\Big]=\lim_{T_x\to\infty}E\Big[\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}x(t)x^*(t-\tau)dt\Big]$$

$$=\lim_{T_x\to\infty}\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}E\Big[x(t)x^*(t-\tau)\Big]dt$$

$$=\lim_{T_x\to\infty}\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}R_x(\tau)dt$$

that is,

$$\lim_{T_x\to\infty}E\Big[\big\langle R_x(\tau)\big\rangle_{T_x}\Big]=R_x(\tau) \tag{3.23}$$

which satisfies the first condition for ergodicity in ACF as per equation (3.19). However, from the mean square periodicity condition for $x(t)$ we also have,

$$\big\langle R_x(\tau)\big\rangle_{T_x}=\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}\Big[x(t)x^*(t-\tau)\Big]dt$$

$$=\frac{1}{T_x}\int_{-T_x/2}^{T_x/2}\Big[\sum_{k=-\infty}^{+\infty}X_ke^{j2\pi kt/T}\sum_{k'=-\infty}^{+\infty}X_{k'}^*e^{-j2\pi k't/T}e^{j2\pi k'\tau/T}\Big]dt$$

$$= \sum_{k=-\infty}^{+\infty} \sum_{k'=-\infty}^{+\infty} X_k X_{k'}^* \Big[\frac{1}{T_x} \int_{-T_x/2}^{T_x/2} e^{j2\pi(k-k')t/T} dt \Big] e^{j2\pi k'\tau/T}$$

$$= \sum_{k=-\infty}^{+\infty} X_k \sum_{k'=-\infty}^{+\infty} X_{k'}^* e^{j2\pi k'\tau/T} Sinc\Big(\pi(k-k')T_x/T\Big)$$

Therefore,

$$\Big\langle R_x(\tau) \Big\rangle_{T_x} = \sum_{k=-\infty}^{+\infty} X_k \Big[X_k^* e^{j2\pi k\tau/T} * Sinc(\pi k T_x/T) \Big] \tag{3.24}$$

where $*$ is the convolution operator. In equation (3.24), when $T_x >> T$, $Sinc(\pi k T_x/T) \rightarrow \delta_k$, and

$$\Big\langle R_x(\tau) \Big\rangle_{T_x} \rightarrow \sum_{k=-\infty}^{+\infty} |X_k|^2 e^{j2\pi k\tau/T} \tag{3.25}$$

and

$$E\Big[\Big\langle R_x(\tau) \Big\rangle_{T_x} \Big] \rightarrow \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2 \Big] e^{j2\pi k\tau/T} = R_x(\tau) \tag{3.26}$$

which proves the first condition for 2^{nd} order ergodicity in equation (3.19).

The last condition for 1^{st} and 2^{nd} order ergodicity to prove is

$$\lim_{T_x \rightarrow \infty} VAR\Big[\Big\langle R_x(\tau) \Big\rangle_{T_x} \Big] = 0$$

To prove this, it is sufficient to show that when $T_x >> T$,

$$E\Big[\Big\langle R_x(\tau) \Big\rangle_{T_x}^2 \Big] = R_x(\tau)^2$$

However, if $T_x >> T$

$$E\Big[\Big\langle R_x(\tau) \Big\rangle_{T_x}^2 \Big] = E\Big[\Big(\sum_{k=-\infty}^{+\infty} |X_k|^2 e^{j2\pi k\tau/T} \Big)^2 \Big]$$

$$= E\Big[\Big(\sum_{k=-\infty}^{+\infty} |X_k|^2 e^{j2\pi k\tau/T} \sum_{k'=-\infty}^{+\infty} |X_{k'}|^2 e^{j2\pi k'\tau/T}\Big)\Big]$$

$$= \sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\Big[|X_k|^2|X_{k'}|^2\Big] e^{j2\pi k\tau/T} e^{j2\pi k'\tau/T}$$

where $E\Big[|X_k|^2\Big]$ represents the average energy of the k^{th} monocomponent of $x(t)$. Assuming uncorrelated energy sources for $x(t)$, i.e., $E\Big[|X_k|^2|X_{k'}|^2\Big] = E\Big[|X_k|^2\Big]E\Big[|X_{k'}|^2\Big]$, we have,

$$\sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\Big[|X_k|^2|X_{k'}|^2\Big] e^{j2\pi k\tau/T} e^{j2\pi k'\tau/T} = \sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\Big[|X_k|^2\Big]E\Big[|X_{k'}|^2\Big] e^{j2\pi k\tau/T} e^{j2\pi k'\tau/T}$$

$$= R_x(\tau)^2$$

Hence,

$$VAR\Big[\big\langle R_x(\tau)\big\rangle_{T_x}\Big] = 0, \quad \text{when} \quad T_x >> T$$

which proves 1^{st} and 2^{nd} order ergodicity of $x(t)$.

■

3.5 Convergence

It was explained in the preceding sections that the Fourier analysis in general is incapable of evoking features of AE signals due to non stationarity. We have also shown that even by *stationarizing* the signal, convergence of Fourier coefficients is not simply achievable, which means that by taking multiple samples of a gas turbine AE signal,

Fourier coefficients of samples could be entirely different, even under non-varying conditions for $x(t)$. We can show again this by examining the variance of the coefficients X_ks.

$$\sigma^2_{X_k} = E[X_k X_k^*] - E[X_k]E[X_k^*] = E[|X_k|^2]$$

since we can assume that $E[X_k] = 0$ because,

$$\begin{aligned} E[X_k] &= E\Big[\int_{-\infty}^{\infty} x(t)e^{-j2\pi kt/T}dt\Big] \\ &= \int_{-\infty}^{\infty} E\big[x(t)\big]e^{-j2\pi kt/T}dt \\ &= E\big[x(t)\big]\int_{-\infty}^{\infty} e^{-j2\pi kt/T}dt = 0 \quad \text{for all} \quad k \neq 0 \end{aligned}$$

Therefore,

$$\sigma^2_{X_k} = E[|X_k|^2] = \mathscr{F}\big(R_x(\tau)\big)$$

as per equation (3.17). It follows that the Fourier coefficients X_k have a variance equal to the energy of its monocomponent. This implies that each and every time a recording is taken from an AE signal, a different Fourier coefficient is attained, which is the reason why Fourier analysis cannot be directly applied for the condition monitoring problem. Figure 3.5 illustrates Fourier coefficient magnitudes of 113 samples taken from a gas turbine AE signal. As can be seen, Fourier coefficients of individual samples vary randomly and differently.

The variance of the original sequence $x(t)$ can also be computed as follows:

$$\sigma^2_{x(t)} = E\big[x(t)^2\big] - E\big[x(t)\big]^2$$

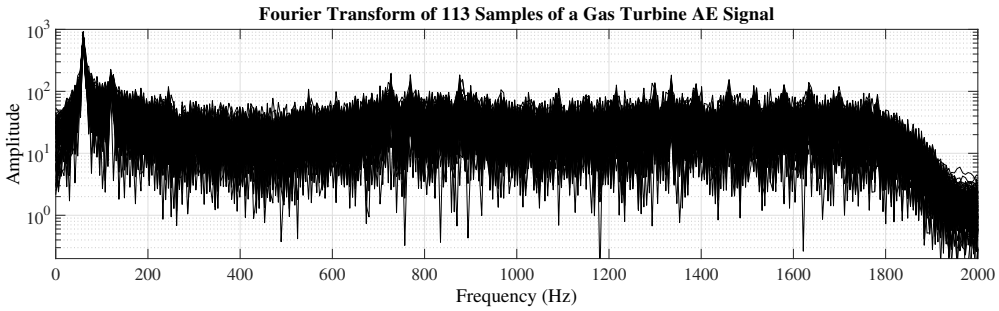

Figure 3.5: Variations of the Fourier coefficients for a gas turbines AE signal.

$$= E\big[x(t)^2\big]$$

since $E\big[x(t)\big] = 0$. Substituting for $x(t)$ from equation (3.10), we have

$$\sigma^2_{x(t)} = E\Big[\sum_{k=-\infty}^{+\infty} X_k e^{j2\pi kt/T} \sum_{k'=-\infty}^{+\infty} X_{k'} e^{j2\pi k't/T}\Big]$$

$$= E\Big[\sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} X_k X_{k'} e^{j2\pi kt/T} e^{j2\pi k't/T}\Big]$$

$$= E\Big[\sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} X_k X_{k'} e^{j2\pi (k+k')t/T}\Big]$$

$$= \sum_{k=-\infty}^{+\infty}\sum_{k'=-\infty}^{+\infty} E\Big[X_k X_{k'} e^{j2\pi (k+k')t/T}\Big]$$

or

$$\sigma^2_{x(t)} = \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] \tag{3.27}$$

which is the same result as derived in equation (3.16). That is, $\sigma^2_{x(t)}$ is equal to the sum of energies of the individual components, or simply the energy of the signal. Also, since X_ks are mutually orthogonal, if $x(t)$ is divided into sub-bands so that $x(t) = \sum_k x_k(t)$,

then the energy of the k^{th} sub-band would be $|X_k|^2$. In a similar manner $R_{[x(t)]_{T_x}}(\tau)$ can be defined as:

$$R_{[x(t)]_{T_x}}(\tau) = \frac{1}{T_x}\int_{-T_x/2}^{T_x/2}\Big[x(t)x^*(t-\tau)\Big]dt$$

$$= \sum_{k=-\infty}^{+\infty} X_k\big[X_k^* e^{j2\pi k\tau/T} * Sinc(\pi k T_x/T)\big]$$

as per equation (3.24). Therefore,

$$\sigma^2_{[x(t)]_{T_x}} = R_{[x(t)]_{T_x}}(0) = \sum_{k=-\infty}^{+\infty} X_k\big[X_k^* * Sinc(\pi k T_x/T)\big] \tag{3.28}$$

$$\approx \sum_{k=-\infty}^{+\infty} |X_k|^2, \quad \text{if} \quad T_x >> T$$

Therefore, if $[x(t)]_{T_x}$ is filtered into N sub-bands so that $[x(t)]_{T_x} = \sum_{k=1}^{N}[x_k(t)]_{T_x}$, then the energy of the k^{th} sub-band would be,

$$\frac{1}{T_x}\int_{-T_x/2}^{T_x/2} x_k^2(t)dt \approx |X_k|^2 \tag{3.29}$$

This is an important result that will be used later in the subsequent chapters, where we use the energy of the sub-bands for the purpose of condition monitoring. Another important result that is derived from equation (3.26) is,

$$\lim_{T_x\to\infty} E\Big[\big\langle R_x(\tau)\big\rangle_{T_x}\Big] = \lim_{T_x\to\infty} E\Big[\big\langle R_{[x(t)]_{T_x}}(\tau)\big\rangle_{T_x}\Big] = R_x(\tau) \tag{3.30}$$

Equation (3.30) can be simply expressed as

$$\bar{R}_{[x(t)]_{T_x}}(\tau) \approx R_x(\tau), \quad \text{for } |\tau| << T_x, \text{ when } T_x >> T \tag{3.31}$$

where $\bar{R}_{[x(t)]_{T_x}}(\tau)$ denotes the mean of $R_{[x(t)]_{T_x}}(\tau)$ taken over many subsamples $[x(t)]_{T_x}$ of $x(t)$. One can summarize the above derivations as following. However, we first need to formally state our assumption.

Assumption 3.2: *Assume that $x(t)$ is a random signal representing a cyclostationary process with a period T, whose conditions (refer to the Definition 1.1 in Chapter 1) are constant throughout a certain time interval $[0, T_p]$, where $E[x(t)] = 0$. Let us also assume that $[x(t)]_{T_x}$ denotes a sub-signal of $x(t)$ taken within $[t_s, t_s + T_x] \subset [0, T_p]$, where t_s is also randomly chosen having a uniform distribution. Additionally, it is assumed that $\mathscr{C}_{x(t)}$ denotes the condition of the cyclostationary process (refer to Definition 1.1), that is fully accessed through the autocorrelation function of $x(t)$.*

We can now state our main result of this chapter.

Theorem 3.1: *If $x(t)$ and $[x(t)]_{T_x}$ are random signals satisfying the conditions given in Assumption 3.2, then $[x(t)]_{T_x}$ can be decomposed into orthogonal monocomponents $[x_k(t)]_{T_x}$, so that $[x(t)]_{T_x} = \sum_k [x_k(t)]_{T_x}$ and $1/T_x \int_{t_s}^{t_s+T_x} [x(t)]_{T_x}^2 dt = 1/T_x \sum_k \int_{t_s}^{t_s+T_x} [x_k(t)]_{T_x}^2 dt$, where,*

$$\frac{1}{T_x}\int_{t_s}^{t_s+T_x} [x_k(t)]_{T_x}^2 dt = |X_k|^2 \quad \textit{and} \quad X_k = \frac{1}{T_x}\int_{t_s}^{t_s+T_x} [x(t)]_{T_x} e^{-j2\pi kt/T} dt$$

Additionally, if $[x(t)]_{T_x}$ is sufficiently long, i.e., $T_x >> T$, then the monotonic energy

components $|X_k|^2$ converge to the deterministic quantities (denoted by $E[|X_k|^2]$) that can optimally represent $\mathscr{C}_{x(t)}$, where,

$$E[|X_k|^2] = \lim_{T_p \to \infty} \frac{1}{T_p} \int_0^{T_p} [x_k(t)]_{T_p}^2 dt$$

Let us investigate the significance of Theorem 3.1. Assumption 3.2 presumes that the machine condition $\mathscr{C}_{x(t)}$ can be fully accessed through $R_x(\tau)$ (as discussed in this chapter). Theorem 3.1 then implies that $\mathscr{C}_{x(t)}$ can be optimally represented by a vector of $[x_k(t)]_{T_x}$ energies, in a sense that no other set of features could estimate $\mathscr{C}_{x(t)}$ better. This is due to the fact that the energy of $[x_k(t)]_{T_x}$s approach $E[|X_k|^2]$ as $T_x \to \infty$, and that $R_x(\tau)$ can be perfectly reconstructed once $E[|X_k|^2]$s are available. We will elaborate this more clearly in the next chapter (Definitions 4.3 and 4.4).

3.6 Summary

In this chapter, we have tried to demonstrate that the gas turbine AE signals (as in most mechanical signals derived from rotary equipment such as pumps, compressors, gears, torque converters, bearings, etc.) are examples of cyclostationary random processes, in which the autocorrelation function of the signal exhibits a periodic waveform. We have seen several examples of gas turbine AE autocorrelations of different gas turbines taken at different times (refer to Fig. 3.1). In these examples the amount of noise added to the

pure sinusoidal waveform have seemed to differentiate the average behaviour of a particular gas turbine among other machines of similar make, type and size, but operating under different conditions.

Hence, we suggested the autocorrelation function as a possible measure discriminating operating conditions of a machine. In Fig. 3.1 we have seen that although the samples were belonging to different gas turbines (of similar make, type and size) operating at different days and under different operating circumstance, all autocorrelation waveforms exhibit equivalent period being the period of the cyclic process, whereas the differentiating factor is the amount of noise added to the waveform. Therefore, it may seem that the energy of the fundamental component of an autocorrelation function can be indicative of the main process if the process is cyclostationary, and the noise level that is a measure of the randomly behaving surrounding conditions, can be an indication of non-cyclic anomalies. However, in order to evoke the fundamental component we have to *wide-sense-stationarize* the AE signal by taking random piecewise sub-signals (refer to Assumption 3.2).

Property 3.1 proves that by giving a random phase to a cyclostationary signal, it can be made *wide-sense-stationary* (WSS). A WSS process is one that its statistical parameters (mean and autocorrelation) remain time-invariant throughout the course of the

signal. Therefore, one can use Fourier analysis to determine the instantaneous frequencies of its *monotonic* ingredients. Since a WSS cyclostationary process is *mean square periodic*, it can be represented in a mean square sense by a sum of monotonic components, which proves the *linearity* of a WSS cyclostationary process. The 'orthogonality' of the coefficients of the monotonic components then render another important derivation stating that *a gas turbine AE signal is composed of independent energy sources* as discussed on page 104.

Next, we have shown *ergodicity* of an AE signal, which allows us to take sub-signals of a limited duration from an AE signal, and make inference about the average behaviour of the process. However, taking several sub-signals of a process that is in a non-varying operating condition yields different autocorrelation functions $R_x(t)$. Therefore, the autocorrelation of a cyclostationary process $x(t)$ although containing useful information do not converge to consistent results. Therefore, Fourier transform of samples of a unique AE signal also do not yield same frequency spectrum, even under constant operating conditions. This is probably a reason for the common belief of inapplicability of Fourier analysis in condition monitoring problems, and the general approach toward nonlinear and nonstationary signal processing techniques, as discussed in Chapter 2. We have shown that this problem is an inherent characteristic for periodogram samples and has

a straightforward remedy, that guarantees convergence of different samples of a cyclo-stationary signal in a constant condition.

The ultimate conclusion of the derivations in this chapter finally yields a theorem under a set of assumptions, as discussed in Assumption 3.2. In the subsequent chapters we will focus on formulating a practical solution for discrimination between operating conditions of a gas turbine by using Theorem 3.1.

Chapter 4

Condition Space and Hypothesis Testing

In the preceding chapter we have seen that a gas turbine AE signal, as an example of a cyclostationary process, can be represented by a linear combination of monotonic energy sources. We have further seen that the ensemble mean of these energy sources can represent the average behaviour of the signal. In this chapter we will try to introduce a concept that is referred to as the *condition space* that can help us discriminate the different operating conditions in a turbomachinery through the AE signals. With the help of the condition space we will also be able to present an official definition for a *condition* as an alternative to the definition that was presented in Chapter 1 (Introduction).

It is important to note again that from equation (3.16), we have

$$VAR\big[x(t)\big] = \sum_{k=-\infty}^{+\infty} E\Big[|X_k|^2\Big] \neq 0$$

whereas from equation (3.19),

$$VAR\Big[\langle R_x(\tau)\rangle_{T_x}\Big] = 0$$

The fact that $\sigma_x^2 \neq 0$ explains the reason why taking samples from an AE signal (even under unchanging conditions) each result into different Fourier components, and is also the reason why it is not meaningful to use the Fourier coefficients of a random signal $x(t)$ as some characteristic quantities, regardless of the length of the signal $x(t)$. However, the case is different for $\langle R_x(\tau)\rangle_{T_x}$, since $\sigma^2_{\langle R_x(\tau)\rangle_{T_x}} = 0$. As discussed in equation (3.31) this implies that by taking sub-signals $[x(t)]_{T_x}$ from an AE signal $x(t)$, the average autocorrelation function of these subsignals denoted by $\bar{R}_{[x(t)]_{T_x}}(\tau)$ converge to $R_x(\tau)$. In other words if one takes m independent subsamples ($[x(t)]^{(i)}_{T_x}$, where $i = 1, 2, ..., m$) from $x(t)$ under some constant condition $\mathscr{C}_{x(t)}$, then for a sufficiently large m we have

$$\bar{R}_{[x(t)]_{T_x}}(\tau) = \frac{1}{m}\sum_{i=1}^{m} R_{[x(t)]^{(i)}_{T_x}}(\tau) \approx R_x(\tau), \quad \text{for } |\tau| << T_x, \text{ when } T_x >> T$$

The practical computation of the feature components in Theorem 3.1 is through a simple procedure, considering that,

$$E\Big[|X_k|^2\Big] = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} \Big(|X_k|^2\Big)^{(i)} \tag{4.1}$$

where as we seen before $\left(|X_k|^2\right)^{(i)}$s were energies of the k^{th} monotonic component that is derived as:

$$\left(|X_k|^2\right)^{(i)} = \mathscr{F}\left(\langle R_x(\tau)\rangle_{T_x}\right)^{(i)} = \mathscr{F}\left(R_{[x(t)]_{T_x}}(\tau)\right)^{(i)} \tag{4.2}$$

that is the Fourier transform of the autocorrelation function of a periodic signal (with period T_x) made by repeating $\left[x(t)\right]_{T_x}^{(i)}$ infinite times. However, as per equation (4.1), we can estimate the features $E\left[|X_k|^2\right]$ as:

$$\widehat{E\left[|X_k|^2\right]} = \frac{1}{m}\sum_{i=1}^{m}\left(|X_k|^2\right)^{(i)} \approx E\left[|X_k|^2\right] \text{ for a sufficiently large } m \tag{4.3}$$

(where $\widehat{E\left[|X_k|^2\right]}$ denotes the estimate of $E\left[|X_k|^2\right]$) simply by taking the average of $|X_k|^2$s. Since $R_x(\tau)$ can be uniquely reconstructed through $E\left[|X_k|^2\right]$s, and since $\widehat{E\left[|X_k|^2\right]} \approx E\left[|X_k|^2\right]$ for a large number of samples, we can assume the *normalized* $\widehat{E\left[|X_k|^2\right]}$ as a feature, as per Definition 1.2, because we have assumed as per Assumption 3.2 that $\mathscr{C}_{x(t)}$ is fully accessible through $R_x(\tau)$. Hence, we can now state the following.

Definition 4.1: *If $x(t)$ and $[x(t)]_{T_x}$ are random signals, and if the conditions in Assumption 3.2 hold,*

$$\mathbb{X}_k = E\left[|X_k|^2\right] / \sum_{i=-\infty}^{\infty} E\left[|X_i|^2\right]$$

will be referred to as a feature of $x(t)$, and a vector composed of the features $\mathbb{X}_k$ will be referred to as the feature vector of $x(t)$ and denoted by $\mathbb{X}$. Additionally, an estimate of the feature vector $\mathbb{X}_k$ under condition $\mathscr{C}_{x(t)}$, that is approximated by N features, is

denoted by,

$$\widehat{\mathbb{X}}\big|\mathscr{C}_{x(t)} = [\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N]^T$$

and referred to as feature vector estimates, where

$$\widehat{\mathbb{X}}_k = \int_{t_s}^{t_s+T_x} [x_k(t)]^2_{T_x} dt \Big/ \int_{t_s}^{t_s+T_x} [x(t)]^2_{T_x} dt$$

and is referred to as the feature estimate.

One may note that Theorem 3.1 implies that $\widehat{\mathbb{X}}_k$ can be (theoretically) made as close as desired to $\mathbb{X}_k$ by increasing T_x. However, this is not practically possible since the machine condition $\mathscr{C}_{x(t)}$ cannot be held constant for too long, which invalidates Assumption 3.2.

Definition 4.2: *Since the features $\mathbb{X}_k$ are real numbers and $\mathbb{X}_k \in [0, 1]$, therefore the feature vector lies in an N dimensional space denoted by $\mathbb{R}^N_{[0,1]}$, and referred to as the feature space.*

In Section 4.2 we will show how a machine condition such as $\mathscr{C}_{x(t)}$ can be uniquely associated to a subset in $\mathbb{R}^N_{[0,1]}$. This is an important step toward machine condition monitoring. This is due to the fact that once a unique subspace in $\mathbb{R}^N_{[0,1]}$ is associated to a unique condition, each time the feature vector of an AE signal falls inside that subspace, one can conclude that the machine is in the condition $\mathscr{C}_{x(t)}$.

In Fig. 3.1 we saw that the autocorrelation function of an AE signal is noisy sinusoidal waveforms of a constant frequency that is exactly equal to the rotor speed, since all gas turbines (actually real-life machines) were connected to the same power grid through synchronous generators, therefore locked to the speed of the network. Since all 'purposeful' or 'useful' processes in a gas turbine correspond to a cyclic nature, they are positively contributing to the fundamental component of the autocorrelation function of the AE signal. Therefore, the fundamental component running at the speed of the rotor can represent the useful process converting fuel into mechanical energy, whereas anomalies or changes in machine condition tend to evolve at some instance, therefore create non-fundamental frequencies that represent signatures of non-cyclic phenomena. Therefore, the feature vector defined in Definition 4.1 can be justified as an indication of the machine health or the machine condition $\mathscr{C}_{x(t)}$.

4.1 Experiments

Before proceeding to the notion of *condition space* and the formal definition of *condition*, it is worth examining a few actual experiments using the feature vector defined in Definition 4.1. In the next chapter, we will put the theories discussed so far, all into practical form to evoke the feature vector of a signal through a simple multi-rate

band-limited filtering system with N channels to classify the operating conditions of a turbo-machine, or identify a new machine condition.

Example 1: We will see how Theorem 3.1 can resolve the scattering of the Fourier coefficients as explained in Section 3.5. We have taken the average of $|X_k|$s to make convergence of scattered Fourier transforms of 113 AE samples taken from a gas turbine under unchanging conditions. The results are shown in Fig. 4.1, where the first (top) subplot (i.e., subplot (a)) shows the variations in $|X_k|^{(i)}$, for $i = 1$ to 113. The second subplot (b) as evident is a much narrower band (compared to subplot (a)) which is the average of 10 $|X_k|^{(i)}$s clearly showing a better convergence. In subplot (c) (i.e., the third graph from the top), the convergence is even more, since it is an average over 30 $|X_k|^{(i)}$s (instead of 10 in subplot (b)). Finally, subplot (d) is the average of all 113 $|X_k|^{(i)}$. This example shows that as we increase the number of samples m in equation (4.3), the average of the samples becomes closer to a mean, with an exception that in this example the average of $|X_k|$s are taken rather than $|X_k|^2$s.

Example 2: In Fig. 4.2, we have computed the feature vector estimates with $N = 4000$, i.e., $\widehat{\mathbb{X}} = [\widehat{\mathbb{X}}_1, ..., \widehat{\mathbb{X}}_{4000}]^T$, for a pair of gas turbines according to equations (4.1) and (4.2). The gas turbine No. 1 (GT-1) is the same machine that is used in Example 1. However, this time we have taken two AE samples, at two different days of different months of

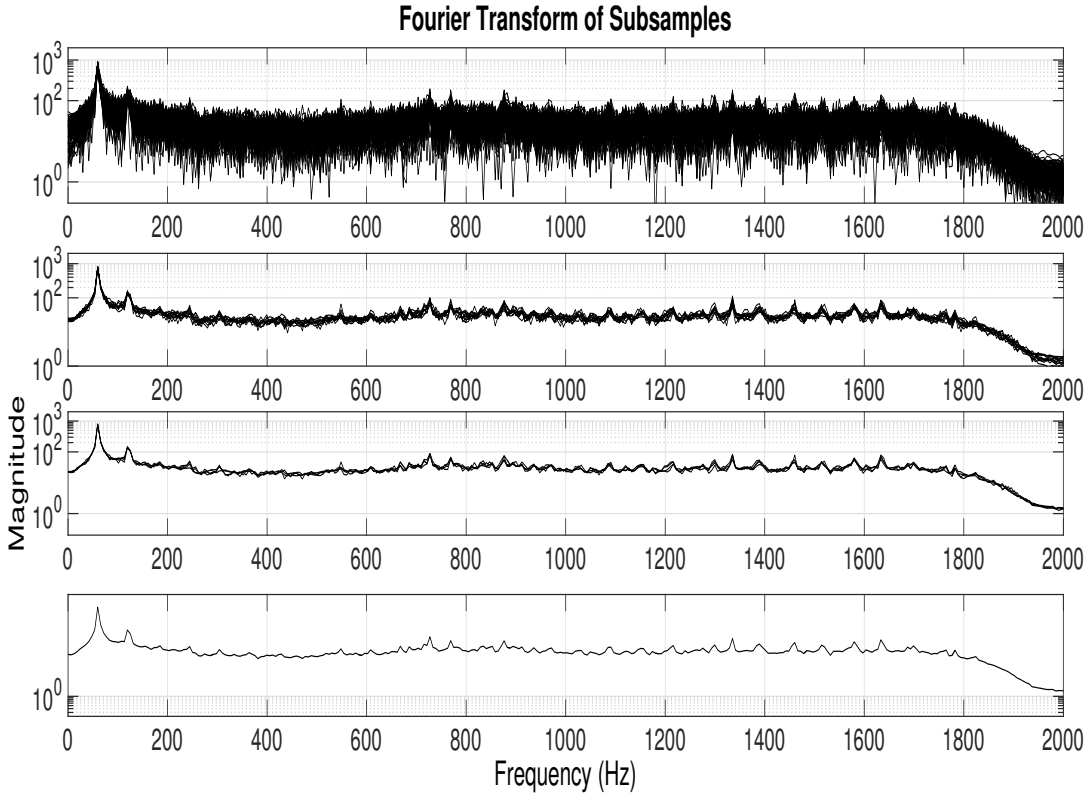

Figure 4.1: Fourier transform of (a) 113 samples of a gas turbine AE signal under unchanging operating conditions (note the scattering), (b) Same 113 samples divided into 11 vectors, each averaged over 10 subsamples, (c) Same 113 samples divided into 4 vectors, each averaged over 28 subsamples, and (d) Average of all the 113 samples. It follows that as m is increased in equation (4.1), the features are converging more.

a year, thus representing different climatic conditions. Still we can see that the feature vector between the two cases (upper and lower plots on left side of Fig. 4.2) resemble one another, showing the same machines (the *autocorrelation coefficient* is used as a similarity test between feature vectors as in Table 4.1 below). The right hand side of Fig. 4.2

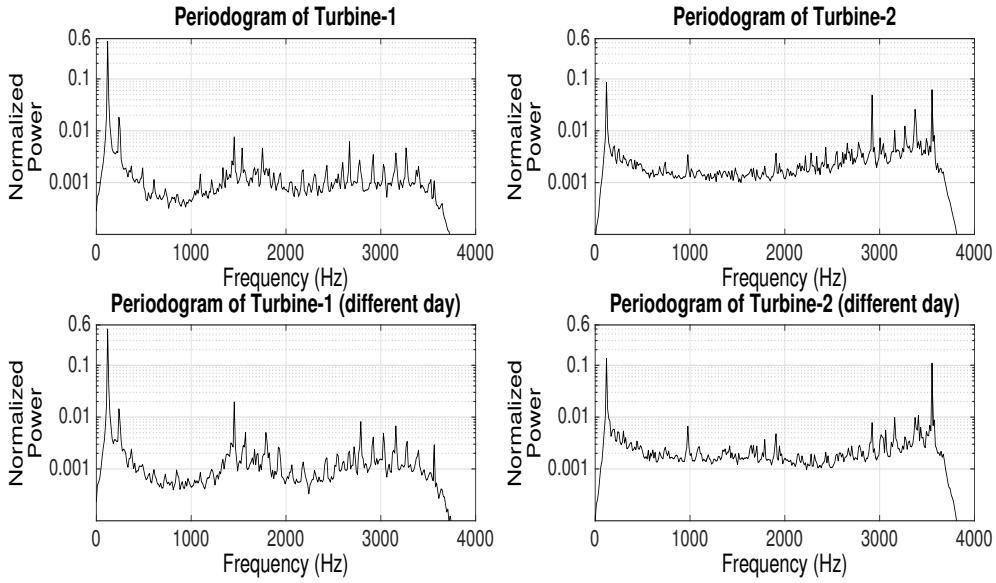

Figure 4.2: Feature vector estimate $\widehat{\mathbb{X}}$ of AE samples belonging to a pair of gas turbines (i.e., GT No. 1 and No. 2). For each gas turbine we have taken two vectors in different days with different signal lengths, under different climatic conditions. Plots belonging to the same machine are similar whereas plots of different machines are not.

however is the feature vector of a different machine (GT-2) also taken at different days and month of a year. Again one can note resemblance between the vectors representing GT-2, except for an energy source (a feature) at frequency 2920Hz which accounts for 4% of the acoustic energy in the upper subplot but less than 1% in the lower right hand subplot. However, it is worth noting that the highest energy source in all four subplots is belonging to an energy source at 119Hz which is the rotor speed, around 7150rpm. The rotor energy is the useful energy delivered to the rotor that provides the torque. The rest of the energy is emitted from other acoustic sources, representing *process noise*.

Therefore, the energy distribution among the vector components are an indication of the machine condition. As evident from the 'Normalized Power' axis in Fig. 4.2, for instance 60% of the acoustic energy in GT-1 is emitted from the rotor which is the useful energy, leaving only 40% to the process noise. In GT-2 however, around 10% of the acoustic energy is emitted from the rotor and 90% from the process noise. It is obvious that GT-1 is in a more stable condition than GT-2, as far as acoustic component energies are concerned.

Taking the microphone inside the gearbox and recording AE samples, reveals that the energy sources at the high end of the frequency spectrum in GT-2 is emitted from the gearbox (Fig. 4.3), because these sources of energy exhibit more power in comparison with the recordings taken outside the compartment. This way it is possible to clue the relation between energy sources in the acoustic spectrum and the physical phenomenon (it is sufficient to find the location where the acoustic emissions are strongest). For instance, lower part of the spectrum (below 1670Hz) is attenuated inside the gearbox compartment, therefore low frequency sources of energy are somewhere outside the gearbox.

Spikes in the gearbox of GT-2 is a very unfavourable phenomenon showing excessive involvement of the gear teeth that reduces the machine efficiency. This thesis is

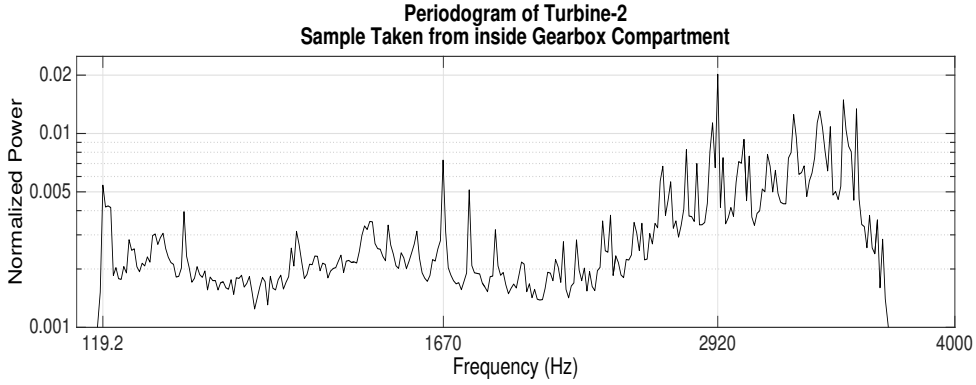

Figure 4.3: Feature estimates (i.e., $\widehat{\mathbb{X}}_k$) of AE Samples taken from inside the Gearbox compartment of GT-2. Comparison of the powerful sources of energy with Fig. 4.2 indicates that the acoustic source at 2920Hz in Fig. 4.2 was belonging to a malfunction in the gearbox.

however focused on the theoretical background of feature extraction and condition monitoring, rather than fault analysis, therefore we do not intend to analyze faults. Hence, the merit of this example was to serve as a comparison between GT-1 and GT-2, as well as identification of a change in the condition of GT-2 between its two samples. One can see that GT-1 had been operating under equivalent machine conditions (as we will subsequently obtain the correlation coefficient $\rho_{GT1(1),GT1(2)} \approx 1$), whereas the condition in GT-2 had changed due to a gearbox problem ($\rho_{GT2(1),GT2(2)} \approx 0.9$). The example also hints that the feature vectors are correctly representing gas turbine conditions, because similarities are very well evident between samples taken from the same machines, whereas there is little similarity between samples corresponding to different machines.

Table 4.1: Correlation coefficient between samples in Fig. 4.2.

ρ	GT-1(1)	GT-1(2)	GT-2(1)	GT-2(2)
GT-1(1)	1	0.9993	0.6828	0.7637
GT-1(2)	0.9993	1	0.6852	0.7647
GT-2(1)	0.6828	0.6852	1	0.9026
GT-2(2)	0.7637	0.7647	0.9026	1

In order to quantify the 'change' in a machine condition, one can use a similarity index that indicates the closeness of the feature vectors between two machine samples. The correlation coefficient, ρ, measures the similarity of the machine condition represented by signals x_1 and x_2, and is defined as:

$$\rho_{x_1 x_2} = \frac{E[X_1 X_2] - E[X_1]E[X_2]}{\sigma_{X_1}\sigma_{X_2}} \tag{4.4}$$

where X_1 and X_2 denote the feature vectors of the samples x_1 and x_2, respectively, and σ_{X_1} and σ_{X_2} are the standard deviations of their respective feature vectors.

Table 4.1 summarizes the correlation coefficients between the samples that are depicted in Fig. 4.2. GT-i(j) denotes the j^{th} sample of GT-i. For instance, GT-2(1) corresponds to the upper right hand plot, whereas GT-1(2) is the sample of the lower left hand subplot. In Table 4.1, the similarity between the AE patterns of a sample

with itself is 100% (i.e., $\rho = 1$). The similarity between the features of the two separate samples from GT-1 is also almost equal to 1 (i.e., 0.9993), implying that the machine condition has not changed on the days the samples were taken. The correlation coefficient between the two samples for GT-2 is however different, showing $\rho = 0.9026$ which indicates a change in the features and therefore the condition. And finally, the correlation coefficients between GT-1 and GT-2 are in the range of 0.68 to 0.76 that indicate very different conditions as expected.

Example 3: This example is meant to demonstrate very mild rotor bowing or misalignment problem that causes radial forces on the shaft. These situations together with unbalance are among the most common malfunctions in turbomachinery. Radial forces cause the rotor to be displaced from its original position to high eccentricity ranges triggering nonlinear effects, causing the response of the rotor to contain synchronous components together with 2^{nd} and 3^{rd} (and higher) harmonics. A gas turbine startup and warm-up cycle often imitates similar behaviour temporarily. As the engine approaches thermal stability, radial forces diminish and harmonic components fade away.

Fig. 4.4 illustrates the start-up cycle of GT-1 (that was analyzed in Examples 1 and 2). The lower graph (i.e., subplot (c)) of Fig. 4.4 displays the low frequency features of GT-1 in a 'cold' condition (that is just 2 hours after the engine was started and loaded

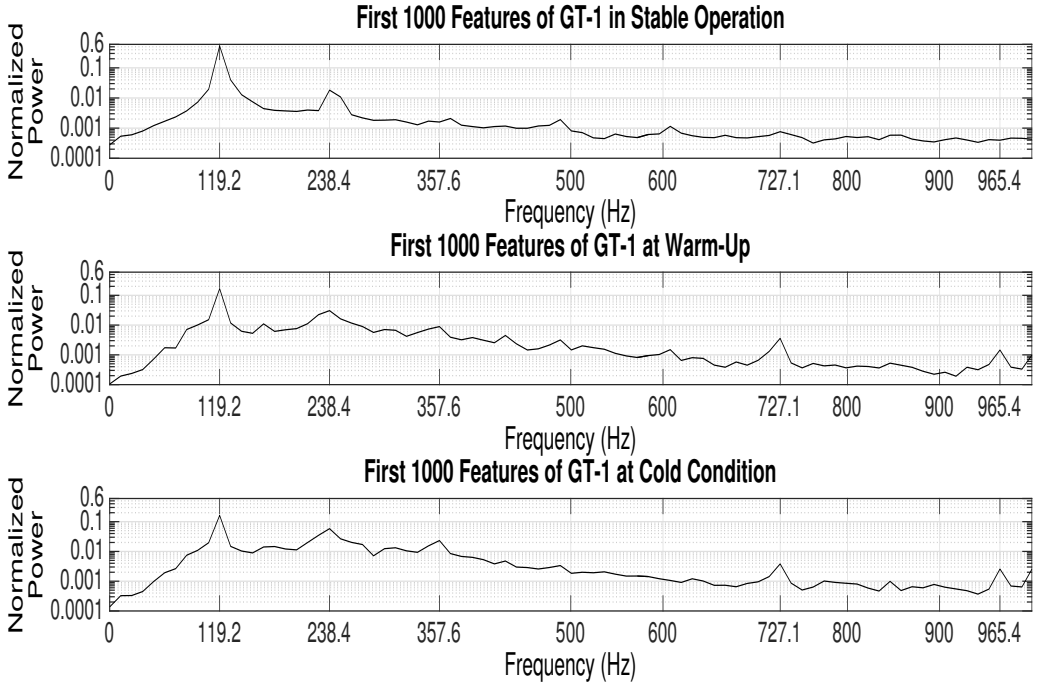

Figure 4.4: The first 1000 features of GT-1 AE samples when (a) GT-1 is operating under thermally stable condition (as in Example 2), (b) When GT-1 was in warm-up condition, i.e., 4 hours after the start-up, and (c) When GT-1 was in a cold state, i.e., 2 hours after the start-up.

from a long shut down). In this graph, the energy of the synchronous component (at 119.2Hz) accounts for 16.27% of the total acoustic energy of the AE sample. Due to high radial forces, the 2^{nd} and 3^{rd} harmonics (at 238.4Hz and 357.6Hz) take up 5.82% and 2.31% of the energy, respectively. As the engine warms up radial forces cease, and the harmonic energy sources fade away. The features of GT-1 after 4 hours from cold start are depicted in the middle graph of Fig. 4.4. Under this condition the synchronous

component takes up 16.75% of the energy, whereas the 2^{nd} and 3^{rd} harmonics have relatively relaxed to 3% and 0.9% in the total energy only. For the upper subplot of Fig. 4.4 (subplot (a)), the engine has fully stabilized thermally, after several hours in operation. Accordingly, the synchronous component now constitutes 53% of the acoustic energy of the signal, where as the 2^{nd} and 3^{rd} harmonics have percentage energies of 1.8% and 0.1%, respectively. As we can see from this example, the energy distribution of the features of the acoustic signal of this gas turbine clearly monitors the radial forces in the rotor.

Complete condition monitoring of a gas turbine is only possible after exact determination of the physical sources that are associated with the features. This of course requires continuous data collection and processing taken from individual microphones located at certain locations in the train. Comparison between relative energies of individual feature components can hint about the physical phenomena producing the features. For instance a microphone located in the gearbox will generate more powerful features associated with the gear mesh, and this way it is possible to identify the features associated with major gas turbine components. The condition monitoring process is then possible through careful monitoring of the energy emitted from different components.

4.2 Hypothesis Testing and Feature Space

In Section 4.1 we have shown through several examples that a *machine condition* with the definition introduced in Chapter 1 can be characterized by the vector of feature estimates (i.e., $\widehat{\mathbb{X}}_k$), derived through equations (4.2) to (4.1) and Assumption 4.3, where the number of features N can be chosen to define the frequency separation between adjacent features. Proper selection of N is based on the number of machine conditions (i.e., classes), and the nature of the conditions. For instance as in Example 3, if one intends to detect the rotor bowing or misalignment of a gas turbine, it would be at least necessary to isolate a subband centred at 238.4Hz from the first harmonic (at 119.2Hz). Then one may decide to choose a subband from 178.8Hz to 298Hz for this purpose. This is the subband where one expects to see an increased acoustic energy when a misalignment takes place. Assuming that the highest acoustic frequency is 4KHz, then one requires at least $N = 34$ features (since $4KHz/119.2Hz = 33.5$).

As another example let us suppose that one only intends to detect gearbox malfunctioning as in Fig. 4.3, through a dedicated filter designed for this purpose. In this case it could be sufficient to monitor the energy of the acoustic emissions above 2KHz, hence $N = 2$. However, for more accurate gearbox monitoring purposes one may decide to choose $N = 4$ so that one may monitor the energy of the subband 1KHz to 2KHz (to verify the spark at 1670Hz), as well as the 3KHz to 4KHz subband simultaneously, to

reduce false alarms (refer to Fig. 4.3).

Generally, let us suppose that the AE signal is sampled at F_sHz. According to the Nyquist theorem the highest observable frequency component will be at $F_s/2$ which should cover the entire frequency spectrum of the potential faults, say 4KHz, i.e., $F_s/2 \approx$ 4KHz. A separation of 2Hz between adjacent features would require:

$$F_s/2 \approx 4000 = 2 \times N \quad or \quad N = 2000 \text{ features}$$

Therefore, a feature vector comprising 2000 components would facilitate detection of mechanical phenomenon with an accuracy of $\pm 2Hz$.

The main objective in condition monitoring is to identify if any *abnormal* condition is governing the process under question. Abnormality in the context of this thesis is a deviation from the feature vector representing normal operating condition, denoted by $\mathbb{X}\big|\mathscr{C}_{\text{Normal}}$, that is

$$\mathbb{X}\big|\mathscr{C}_{\text{Normal}} = \left[\mathbb{X}_1\big|\mathscr{C}_{\text{Normal}} \quad \mathbb{X}_2\big|\mathscr{C}_{\text{Normal}} \quad ... \quad \mathbb{X}_N\big|\mathscr{C}_{\text{Normal}}\right]^T \tag{4.5}$$

where $\mathbb{X}_k\big|\mathscr{C}_{\text{Normal}}$s are *baseline features* that are derived as per Definition 4.1, when the system (e.g. the gas turbine) is assumed healthy and stable (i.e., immediately after successful commissioning and performance testing, or after a recent major overhaul, when the unit can be considered as new, and for a long enough duration in time, i.e.,

$T_p \to \infty$ as per Assumption 3.2). Later we are interested to see if the estimated feature vectors $\widehat{\mathbb{X}}$ evoked from $x(t)$ indicate deviations from $\mathscr{C}_{\text{Normal}}$ or not. This is generally a problem of *hypothesis testing*, since $\widehat{\mathbb{X}}$ itself is a random vector. We have assumed (just before Section 3.5) that feature estimates are mutually independent, i.e.,

$$E\big[\widehat{\mathbb{X}}_i\widehat{\mathbb{X}}_j\big] = E\big[\widehat{\mathbb{X}}_i\big]E\big[\widehat{\mathbb{X}}_j\big] \quad \text{for all} \quad (i \neq j) \in \{1, 2, ..., N\} \tag{4.6}$$

and that $\widehat{\mathbb{X}}$s are *i.i.d* random variables. Independency assumption for features is not always valid, especially for adjacent features, and/or when two distinct features represent one physical phenomenon. However, we can safely assume that feature estimates are *i.i.d* random variables as long as the samples $\big[x(t)\big]_{T_x}$ are taken from a time duration (i.e., $[t_s, t_s + T_x] \subset [0, T_p]$, refer to Assumption 3.2) in which the operating condition is fairly constant. In order to simplify the derivations let us further assume that the condition $\mathscr{C}_{x(t)}$ is constant for an infinite duration in time, and the AE signal is pre-amplified so that,

$$\lim_{T_p \to \infty} \frac{1}{T_p} \int_{t=0}^{T_p} x^2(t)dt = 1$$

In this case the feature estimates of $x(t)$ reduce to,

$$\widehat{\mathbb{X}}_k = \frac{1}{T_x} \int_{t=t_s}^{t_s+T_x} \big[x_k(t)\big]_{T_x}^2 dt \tag{4.7}$$

and for $T_x = mT$ where $m >> 1$,

$$\widehat{\mathbb{X}}_k = \frac{1}{m}\Big[\frac{1}{T}\int_{t=t_s}^{t_s+T} x_k^2(t)dt + \frac{1}{T}\int_{t=t_s+T}^{t_s+2T} x_k^2(t)dt + ... + \frac{1}{T}\int_{t=t_s+(m-1)T}^{t_s+mT} x_k^2(t)dt\Big] \tag{4.8}$$

where,

$$\frac{1}{T}\int_{t=t_s+kT}^{t_s+(k+1)T} x_k^2(t)dt \approx |X_k|^2$$

or,

$$\widehat{\mathbb{X}}_k \approx \frac{1}{m}\sum_{i=1}^{m}\Big(|X_k|^2\Big)^{(i)} \tag{4.9}$$

Since $\widehat{\mathbb{X}}_k$ as per equation (4.9) is the average of the random $i.i.d$ variables $\big(|X_k|^2\big)^{(i)}$, therefore regardless of the distribution of $|X_k|^2$, due to the Central Limit Theorem, the feature estimates $\widehat{\mathbb{X}}_k$ will have a Gaussian distribution, i.e.,

$$\widehat{\mathbb{X}} \sim \mathcal{N}\Big(\mathbb{X}\big|\mathscr{C}_{\text{Normal}}, \sigma^2_{\widehat{\mathbb{X}}}\big|\mathscr{C}_{\text{Normal}}\Big) \tag{4.10}$$

where $\mathcal{N}$ denotes Guassian distribution. The distribution $\mathcal{N}\big(\mathbb{X}\big|\mathscr{C}_{\text{Normal}}, \sigma^2_{\widehat{\mathbb{X}}}\big|\mathscr{C}_{\text{Normal}}\big)$ provides a probability density function (pdf) for the gas turbine in normal condition, denoted by $p_{\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{Normal}}}\big(\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N\big)$, where $\widehat{\mathbb{X}} = \big[\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N\big]^T \in \mathbb{R}^N_{[0,1]}$. The covariance matrix $\sigma^2_{\widehat{\mathbb{X}}}\big|\mathscr{C}_{\text{Normal}}$ in equation (4.10) is a square diagonal matrix given by

$$\sigma^2_{\widehat{\mathbb{X}}}\big|\mathscr{C}_{\text{Normal}} = \begin{bmatrix} \sigma^2_{\widehat{\mathbb{X}}_1} & & \\ & \ddots & \\ & & \sigma^2_{\widehat{\mathbb{X}}_N} \end{bmatrix}_{N\times N} \tag{4.11}$$

where the diagonal elements are given as,

$$\sigma^2_{\widehat{\mathbb{X}}_k} = E\big[\widehat{\mathbb{X}}_k^2\big] - \Big(E\big[\widehat{\mathbb{X}}_k\big]\Big)^2 \tag{4.12}$$

However, as per equation (4.9)

$$E\big[\widehat{\mathbb{X}}_k\big] = E\Big[\frac{1}{m}\sum_{i=1}^{m}\Big(|X_k|^2\Big)^{(i)}\Big]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\Big(E\big[|X_k|^2\big]\Big)^{(i)} = E\Big[|X_k|^2\Big] = \mathbb{X}_k$$

Therefore,

$$\sigma^2_{\widehat{\mathbb{X}}_k} = E\big[\widehat{\mathbb{X}}_k^2\big] - \mathbb{X}_k^2 \tag{4.13}$$

where,

$$\sigma^2_{\widehat{\mathbb{X}}_k} = \frac{\sigma^2_{|X_k|^2}}{m} \tag{4.14}$$

and where $\sigma^2_{|X_k|^2} = VAR\big[1/T\int_{t=t_s}^{t_s+T} x^2(t)dt\big]$, when t_s is randomly selected in $[0, T_p]$, and where the energy of $x(t)$ is normalized to 1. Assuming that equation (4.6) holds,

$$p_{\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{Normal}}}\big(\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N\big) = \prod_{i=1}^{N} p_{\widehat{\mathbb{X}}_i\big|\mathscr{C}_{\text{Normal}}}\big(\widehat{\mathbb{X}}_i\big) \tag{4.15}$$

where,

$$p_{\widehat{\mathbb{X}}_i\big|\mathscr{C}_{\text{Normal}}}\big(\widehat{\mathbb{X}}_i\big) = \frac{1}{\sqrt{2\pi}\sigma_{\widehat{\mathbb{X}}_i}} e^{-\left[\left(\widehat{\mathbb{X}}_i - \mathbb{X}_i\big|\mathscr{C}_{\text{Normal}}\right)^2 \big/ 2\sigma^2_{\widehat{\mathbb{X}}_i}\right]} \tag{4.16}$$

The objective is now to define a criterion to differentiate 'normal' vs. 'abnormal' operating conditions. The criterion could be based on a desirable *false-alarm* classification rate. Let us assume that the objective is to design a condition monitoring system with a 99% reliability rate. Hence, one can only tolerate $P_{FA} = 1\%$ false-alarms. Therefore, the objective from equation (4.15), would be to find the smallest space $\mathscr{R}_{\text{Normal}} \subset \mathbb{R}^N_{[0,1]}$, so that

$$\int_{\mathscr{R}_{\text{Normal}}} p_{\widehat{\mathbb{X}}|\mathscr{C}_{\text{Normal}}}\big(\widehat{\mathbb{X}}\big)d\widehat{\mathbb{X}} = 1 - P_{FA} \tag{4.17}$$

The space $\mathscr{R}_{\text{Normal}} \subset \mathbb{R}^N_{[0,1]}$ where equation (4.17) holds is defined as *the Condition Space* for $\mathscr{C}_{\text{Normal}}$, and generally we have the following definition.

Definition 4.3: *If* $\widehat{\mathbb{X}} = \left[\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N\right]^T \in \mathbb{R}^N_{[0,1]}$ *denotes the estimated feature vector for the signal* $[x(t)]_{T_x}$ *with the conditions in Assumption 3.2 holding, then the smallest space* $\mathscr{R}_{x(t)} \subset \mathbb{R}^N_{[0,1]}$ *where,*

$$\int_{\mathscr{R}_{x(t)}} p_{\widehat{\mathbb{X}}|\mathscr{C}_{x(t)}}\big(\widehat{\mathbb{X}}\big) d\widehat{\mathbb{X}} = 1 - P_{FA} \tag{4.18}$$

is defined as the Condition Space for $\mathscr{C}_{x(t)}$.

To find the condition space $\mathscr{R}_{x(t)}$ associated with $\mathscr{C}_{x(t)}$, one has to find a feature vector (hereinafter referred to as the *border vector*, formally defined shortly later) $\mu = [\mu_1, \mu_2, ..., \mu_N]^T$ so that,

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{\widehat{\mathbb{X}}_i}} \int_{\mu_i}^{\infty} e^{-\left[\left(\widehat{\mathbb{X}}_i - \mathbb{X}_i\middle|\mathscr{C}_{x(t)}\right)^2 \middle/ 2\sigma^2_{\widehat{\mathbb{X}}_i}\right]} d\widehat{\mathbb{X}}_i = P_{FA}/2 \tag{4.19}$$

By changing variables $\mathscr{X}_i = \frac{\widehat{\mathbb{X}}_i - \mathbb{X}_i|\mathscr{C}_{x(t)}}{\sigma_{\widehat{\mathbb{X}}_i}}$, we obtain

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \int_{\frac{\mu_i - \mathbb{X}_i|\mathscr{C}_{x(t)}}{\sigma_{\widehat{\mathbb{X}}_i}}}^{\infty} e^{-\mathscr{X}_i^2/2} d\mathscr{X}_i = P_{FA}/2 \tag{4.20}$$

or

$$\prod_{i=1}^{N} \Phi\Big(\frac{\mu_i - \mathbb{X}_i|\mathscr{C}_{x(t)}}{\sigma_{\widehat{\mathbb{X}}_i}}\Big) = P_{FA}/2 \tag{4.21}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) for the Gaussian distribution. In order to find the solution for equation (4.21) one has to find the false-alarm

rates $P_{FA}(i) = \Phi\big((\mu_i - \mathbb{X}_i|\mathscr{C}_{x(t)})/\sigma_{\widehat{\mathbb{X}}_i}\big)$ for $i \in \{1, 2, ..., N\}$, so that $\prod_{i=1}^{N} P_{FA}(i) = P_{FA}/2$. The simplest solution for μ_i is to assume that $P_{FA}(i)$s are equal for all i. This results

$$\mu_i = \mathbb{X}_i|\mathscr{C}_{x(t)} + \sigma_{\widehat{\mathbb{X}}_i}\Phi^{-1}\Big((P_{FA}/2)^{1/N}\Big) \tag{4.22}$$

and substituting for $\sigma_{\widehat{\mathbb{X}}_i}$ from equation (4.14), yields

$$\mu_i = \mathbb{X}_i|\mathscr{C}_{x(t)} + \frac{\sigma_{|X_k|^2}}{\sqrt{m}}\Phi^{-1}\Big((P_{FA}/2)^{1/N}\Big) \tag{4.23}$$

where μ_is are the upper border points associated with $\mathscr{C}_{x(t)}$. The vector representing the N dimensional upper point associated with $\mathscr{C}_{x(t)}$, denoted by $\mu^+|\mathscr{C}_{x(t)}$, is therefore given by,

$$\mu^+\big|\mathscr{C}_{x(t)} = \mathbb{X}|\mathscr{C}_{x(t)} + \frac{\Phi^{-1}\big((P_{FA}/2)^{1/N}\big)}{\sqrt{m}}\big(\sigma_{|X|^2}\big) \times I \tag{4.24}$$

where $\sigma_{|X|^2}$ is an $N \times N$ diagonal matrix whose diagonal elements are $\sigma_{|X_k|^2}$, and I is an $N \times 1$ vector whose elements are all 1. The lower border point μ^- in a similar manner can be given as,

$$\mu^-\big|\mathscr{C}_{x(t)} = \mathbb{X}|\mathscr{C}_{x(t)} - \frac{\Phi^{-1}\big((P_{FA}/2)^{1/N}\big)}{\sqrt{m}}\big(\sigma_{|X|^2}\big) \times I \tag{4.25}$$

The vectors $\mu^-|\mathscr{C}_{x(t)}$ and $\mu^+|\mathscr{C}_{x(t)}$ derived from equations (4.24) and (4.25) define the extreme border points in $\mathbb{R}^N_{[0,1]}$ that encloses the condition space $\mathscr{R}_{x(t)}$, i.e.,

$$\mathscr{R}_{x(t)} : \{ \quad \widehat{\mathbb{X}} \in \mathbb{R}^N_{[0,1]} \quad \big| \quad \mu^-|\mathscr{C}_{x(t)} \leq \widehat{\mathbb{X}} \leq \mu^+|\mathscr{C}_{x(t)} \quad \} \tag{4.26}$$

One may notice from equation (4.26) that for a given false alarm rate, P_{FA}, the condition space is a function of the number of the samples m, and indirectly a function of the sub-signal length T_x as per equations (4.7) to (4.9). If one is interested to

monitor the condition of a gas turbine for N_c number of categories, it would be possible to generate equal number of condition spaces, such as $\mathscr{R}_1, \mathscr{R}_2, ..., \mathscr{R}_{N_c}$. The i^{th} space $\mathscr{R}_i$ can be formed when the machine is in condition $\mathscr{C}_i$. If this condition could be retained for a sufficiently long duration in time, the feature vector estimate $\widehat{\mathbb{X}}\big|\mathscr{C}_i$ could be taken from $\big[x(t)\big]_{T_x}$, in accordance with the procedure defined in Definition 4.1. In case $T_x \to T_p \to \infty$, then $\widehat{\mathbb{X}}\big|\mathscr{C}_i \to \mathbb{X}\big|\mathscr{C}_i$, and the condition space $\mathscr{R}_i$ will virtually shrink down in size to the point $\mathbb{X}\big|\mathscr{C}_i$.

Practically, it is sufficient to increase the sample's length so that $\mathscr{R}_i \cap \mathscr{R}_j = \emptyset$, for all $i \neq j \in \{1, 2, ..., N_c\}$. Marginating the condition space is also possible through analytic derivation, by taking m sample sub-signals $\big[x(t)\big]_{T_x}$ with random starting points T_s. The estimated feature vector corresponding to each sample has to be obtained and the mean and standard deviation of the feature estimates calculated, and plugged into equations (4.24) and (4.25), for a given false-alarm rate. Then the extreme border points μ^- and μ^+ would define the borders for $\mathscr{R}_i$. In applications where taking long samples is not feasible, one can alternatively try to increase the dimension of the feature vectors N, which is an alternative way instead of increasing m.

It is important to note that increasing m lowers the variance of $\widehat{\mathbb{X}}_k$. However, there are practical limitations for increasing m, since a gas turbine condition can not be

deemed constant for too long. Hence, one can consider Assumption 3.2 to hold only for a limited duration of time. This practically entails a range $[\mu_i^-, \mu_i^+]$ for every $\widehat{\mathbb{X}}_i \in \mathbb{R}_{[0,1]}$, so that $\widehat{\mathbb{X}}_i \in [\mu_i^-, \mu_i^+] \subseteq \mathbb{R}_{[0,1]}$, where $\mathbb{X}_i$ is the center point in $[\mu_i^-, \mu_i^+]$ as per equation (4.26). Effectively there would be only $[1/(\mu_i^+ - \mu_i^-)]$ practically possible choices for $\widehat{\mathbb{X}}_i$, where $[\cdot]$ denotes the integer value. This gives the maximum number for N_C as,

$$N_C(max) = \prod_{i=1}^{N} \Big[\frac{1}{\mu_i^+ - \mu_i^-}\Big] \tag{4.27}$$

where μ_i^+, μ_i^- can be derived as per equations (4.24) and (4.25), respectively. Equation (4.27) explains the relation between N_C, N, and m. For a given N, increasing m will reduce $\mu_i^+ - \mu_i^-$ and results bigger number of classes N_C. On the other hand, increasing N holding m constant will have the same effect. As per equation (4.27), the parameters N and m can be varied independently.

Theorem 3.1 together with the derivations in this chapter imply that corresponding to every signal $x(t)$ a unique condition space $\mathscr{R}_{x(t)}$ can be formed, with a desirable size, and for a desirable false-alarm rate. Since the condition $\mathscr{C}_{x(t)}$ is also uniquely related to $R_x(\tau)$, there will be a unique association between the condition $\mathscr{C}_{x(t)}$ and the space $\mathscr{R}_{x(t)}$ if the condition spaces are sized so that there is no intersection between any two adjacent condition spaces. We have shown in this chapter that this is possible, by adjusting the parameters m for any given certainty P_{FA}. Therefore, theoretically it would be possible to detect 'any' given condition, with any desirable level of certainty, regardless of the

number of conditions and/or how close they would be in the feature space $\mathbb{R}^N_{[0,1]}$. This is justifiable according to equation (4.27). Theoretically, as $m \to \infty$, $\mu_i^+ - \mu_i^- \to 0$ (equivalently $\mathscr{R}_{x(t)} \to 0$) and $N_C \to \infty$. Moreover, equations (4.24) and (4.25) imply that there is no limitation in selection of a desirable false alarm rate P_{FA}, since the product $\Phi^{-1}\big((P_{FA}/2)^{1/N}\big)\sigma_{|X|^2}$ can be compensated for any P_{FA} simply by changing N. However, as discussed before, it is not always practically possible to increase m beyond a certain limit, since it is essential to assure that all the m samples of $\widehat{\mathbb{X}}$ correspond to a unique machine condition for equations (4.24) and (4.25) to be valid, whilst in gas turbine applications with fast system dynamics (such as military aircrafts) engine conditions remain constant only for a very short duration in time, and do not allow taking too many samples.

Hence, instead of increasing m one has to increase N which excessively complicates the implementation of the CM system (as will be shown in the next chapter). Therefore, in applications with fast dynamics the methodology derived in this thesis delivers poor results (i.e., high alarm rates and low certainty levels). Association of $\mathscr{C}_{x(t)}$ and $\mathscr{R}_{x(t)}$ is symbolically shown below:

$$\mathscr{C}_{x(t)} \longleftrightarrow \mathscr{R}_{x(t)}\big(\mathbb{X}\big|\mathscr{C}_{x(t)}, P_{FA}, m, N\big) \tag{4.28}$$

where $\mathbb{X}\big|\mathscr{C}_{x(t)}$ is the centre point in $\mathbb{R}^N_{[0,1]}$, and $\mathscr{R}_{x(t)}$ is an N dimensional ball in the

feature space, with a radius defined by $\sigma_{|X|^2}$ and m. In equation (4.28), $\mathscr{R}_{x(t)}$ is not shown to be a function of $\sigma_{|X|^2}$, because $\sigma_{|X|^2}$ is itself depending on the length of the samples. We will be providing several numerical results in Section 5.2 to support the derivations introduced in this chapter.

4.3 Numerical Validations

In this chapter we have introduced a few important notions, namely '*feature vector*' and '*feature vector estimate*' (i.e., $\mathbb{X}$ and $\widehat{\mathbb{X}}$ respectively, as per Definition 4.1), '*feature space*' ($\mathbb{R}^N_{[0,1]}$ as defined in Definition 4.2), the '*space border vectors*' ($[\mu^-_{\mathscr{R}}, \mu^+_{\mathscr{R}}]$ derived in equations (4.22) through (4.25)), and the '*condition space*' ($\mathscr{R}$ defined in Definition 4.3). We have further stated that every condition such as $\mathscr{C}_{x(t)}$ can be uniquely associated to a condition space (equation (4.28)) for a given false-alarm rate P_{FA} (refer to equation (4.17) for the definition), and the feature space dimension N, and m (equation (4.8)), as long as the conditions in Assumption 3.2 are holding. The main objective of this chapter was to interrelate these notions (as in equation (4.26)) and to deduce the condition of a gas turbine.

As discussed in Chapter 1 the problem of gas turbine condition monitoring using AE technology is mainly twofold, that is (1) formulating a feature extraction method immune of all random effects, and (2) clustering the different conditions in a perfectly

distinct manner in order to be able to identify a condition without error. The problem is clearly evident in Figure 4.5. In order to illustrate the problem (and the solution proposed in this thesis), we have considered a two-dimensional feature space (i.e., $N = 2$).

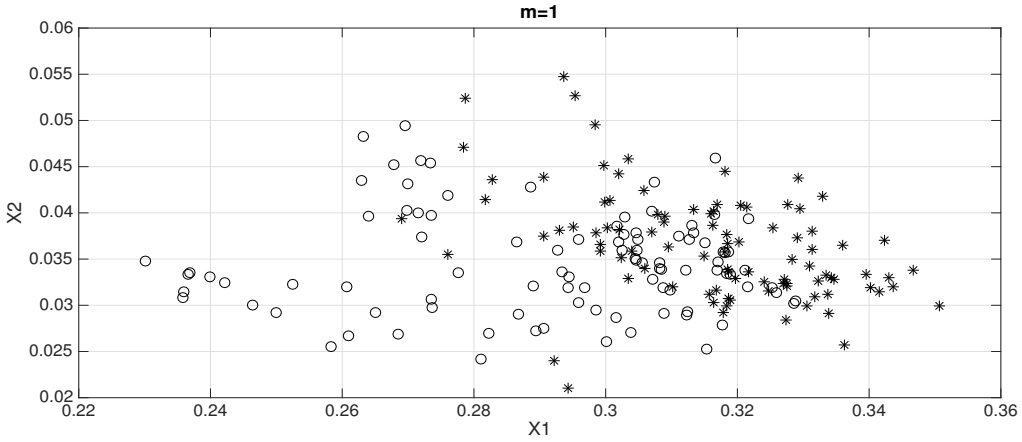

Figure 4.5: Condition spaces of the gas turbines in the 2^{nd} and 3^{rd} subplots of Fig. 3.1, i.e., $\mathscr{R}_{\text{Perfectly Commissioned}}$ (depicted by '$\star$') and $\mathscr{R}_{\text{Poorly Commissioned}}$ (depicted by 'o') respectively, for a two-dimensional feature space $\mathbb{R}^2_{[0,1]}$ with $m = 1$. Each point represents a feature vector estimate $\widehat{\mathbb{X}}_i = [\widehat{\mathbb{X}}_i(1), \widehat{\mathbb{X}}_i(2)]^T$ for $i = \{1, 2, ..., 100\}$.

Marginating the condition spaces $\mathscr{R}_{\text{Perfectly Commissioned}}$ (depicted by '$\star$'s) and $\mathscr{R}_{\text{Poorly Commissioned}}$ (depicted by 'o's) in Fig. 4.5 is not possible since $\mathscr{R}_{\text{Perfect Comm.}} \cap \mathscr{R}_{\text{Poor Comm.}} \neq \emptyset$. However, we have shown in equation (4.14) that $\sigma^2_{\widehat{\mathbb{X}}} \to 0$ as m increases. Hence, we have proposed a solution for marginating any two arbitrary condition spaces regardless of how close the conditions might be, and regardless of the random effect associated with the

AE signals. Figure 4.6 shows how the condition spaces separate as m increases. With $m > 50$ the intersection between the spaces become null enabling one to identify the condition based on $x(t)$.

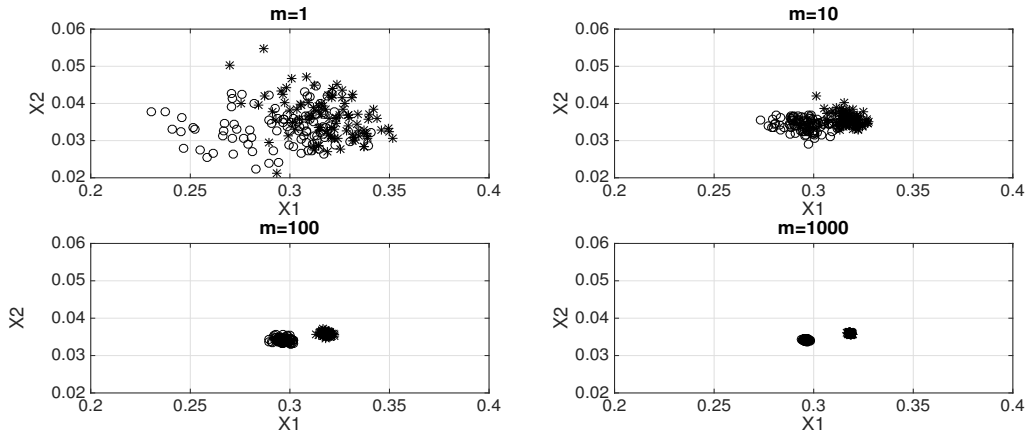

Figure 4.6: Condition spaces for the example illustrated in Figure 4.5 with a varying m.

Now let us examine the condition spaces of the experiments described in Section 4.1 in a two dimensional feature space as illustrated in Fig. 4.7. In order to show a complete demarcation we have chosen $m = 1000$. Let us assume that we have four conditions to monitor, namely a perfectly stable condition denoted by $\mathscr{C}_{\text{stable}}$ (represented by stars), an ageing effect denoted by $\mathscr{C}_{\text{ageing}}$ (represented by circles), a gearbox problem denoted by $\mathscr{C}_{\text{gearbox}}$ (represented by squares), and a rotor imbalance denoted by $\mathscr{C}_{\text{start-up}}$ (represented by diamonds). We have plotted 100 feature vector estimates for each condition. As can be seen the condition spaces have no intersection, implying that certain judgment would

be possible about either of these conditions. This philosophy can be generalized for an N dimensional feature space, as presented in Chapter 5.

It is worth mentioning that even for $N = 2$ we were able to demarcate the conditions

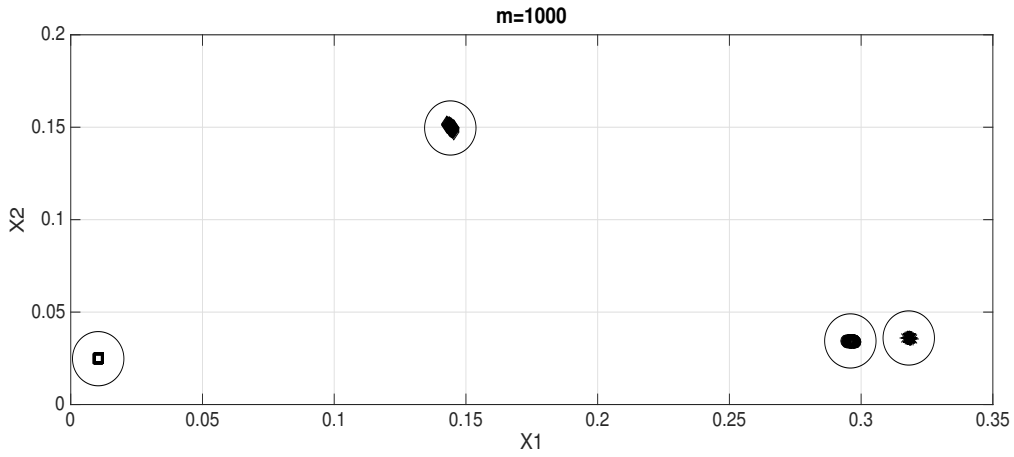

Figure 4.7: Condition spaces for the experiments of Section 4.1. Assume that we have four conditions to monitor, namely a perfectly stable condition denoted by $\mathscr{C}_{\text{stable}}$ (represented by stars), an ageing effect denoted by $\mathscr{C}_{\text{ageing}}$ (represented by circles), a gearbox problem denoted by $\mathscr{C}_{\text{gearbox}}$ (represented by squares), and a rotor imbalance denoted by $\mathscr{C}_{\text{start-up}}$ (represented by diamonds). For each condition the border is illustrated with a circle. In case a feature vector estimate falls inside a circle it can be inferred that the machine is in the corresponding condition.

without intersection. Figure 4.7 also demonstrates the similarity of the conditions. For instance $\mathscr{C}_{\text{stable}}$ and $\mathscr{C}_{\text{ageing}}$ are fairly close both indicating healthy conditions, whereas $\mathscr{C}_{\text{gearbox}}$ and $\mathscr{C}_{\text{start-up}}$ are clearly indicating anomalies being too distant from the 'healthy subspace'.

4.4 Summary

This chapter presents a formal definition for certain quantities referred to as 'features' and 'feature vectors'. It further connects the results of Theorem 3.1 to these quantities, and recommends an estimation method for deriving features from an acoustic signal $x(t)$ through an N dimensional vector. In conclusion one can consider the feature vector derived in this manner to be the optimal representation of the machine condition $\mathscr{C}_{x(t)}$ (based on Definition 1.1) corresponding to $x(t)$, as long as Assumption 3.2 holds.

Additionally, we have defined the set of all possible N dimensional feature vectors as the 'feature space', and have introduced the notion of 'condition space' $\mathscr{R}_{x(t)}$ as a subset of the feature space uniquely associated to $\mathscr{C}_{x(t)}$ with a false alarm rate P_{FA}. Accordingly every acoustic signal $x(t)$ complying with the conditions set forth in Assumption 3.2 can be uniquely transformed into an N dimensional vector such as $\widehat{\mathbb{X}}$ that optimally represents $\mathscr{C}_{x(t)}$ and is denoted by $\widehat{\mathbb{X}}\big|\mathscr{C}_{x(t)}$. We have further shown that the average of a sufficient number of $\widehat{\mathbb{X}}\big|\mathscr{C}_{x(t)}$ samples converge to a point in the feature space that is the center of the condition space $\mathscr{R}_{x(t)}$, and by increasing the number of such samples one can infinitesimally reduce the size of $\mathscr{R}_{x(t)}$.

Hence, it would theoretically be possible to uniquely associate a given condition $\mathscr{C}_{x(t)}$ to a condition space $\mathscr{R}_{x(t)}$ whose size can be freely adjusted by changing N and

the number of samples, for any given P_{FA}. The aim is that one can firstly classify the feature space into N_C condition spaces $\mathscr{R}_i$, where $i \in \{1, 2, ..., N_C\}$, so that $\mathscr{R}_i \cap \mathscr{R}_j = \emptyset$, for all $i \neq j \in \{1, 2, ..., N_c\}$, in a process referred to as *condition spacing* (more precisely addressed in Section 5.2).

Once the condition spacing process is complete every acoustic signal $x(t)$ can be tested against the existing $\mathscr{R}_i$s to verify to which condition it belongs, and it would then be possible to infer about the machine condition that has emitted $x(t)$ based on some arbitrary false alarm rate. Specifically, if $x(t)$ is transformed to $\widehat{\mathbb{X}}$, and if $\widehat{\mathbb{X}} \in \mathscr{R}_i$ then the machine condition corresponds to $\mathscr{C}_i \longleftrightarrow \mathscr{R}_i$ with a probability of P_{FA}. The main advantage of the classification technique developed in this chapter is that one can virtually identify the condition associated with an AE signal among (theoretically) infinite number of conditions, and based on any arbitrary and desirable level of certainty. In the next chapter it will be shown that the entire CM process explained above can implemented through non-sophisticated computational resources and fully online.

In order to provide numerical validations to justify the derivations made in this chapter, it would be necessary to outline an online filter for constructing feature vectors as per Definition 4.1. Therefore, the former part of Chapter 5 is devoted to practical considerations and filtering techniques for evoking feature vectors. Thereafter, we

will demonstrate numerical simulations for condition spacing based on the examples of Section 4.1 (pages 167 to 174), and eventually the results of hypothetical testing of equation (4.28) (summarized in Table 5.1).

Chapter 5

Machine Learning and Autonomous Condition Spacing

In Chapter 4 we have showed that a machine condition $\mathscr{C}_{x(t)}$ corresponding to $x(t)$ can be uniquely associated to a confined space $\mathscr{R}_{x(t)}$ (referred to as the condition space) bidirectionally in the feature space $\mathbb{R}^N_{[0,1]}$, where N is the dimension of the feature space (as per Definition 4.2). That is, given $\mathscr{C}_{x(t)}$ one can uniquely identify $\mathscr{R}_{x(t)}$ (for a given false-alarm rate and sub-signal duration T_x) and vice-versa, where the condition space is represented by a feature vector such as $\mathbb{X}\big|\mathscr{C}_{x(t)}$ in the center of the space. We have seen that $\mathbb{X}\big|\mathscr{C}_{x(t)}$ is in fact the average of the feature vector estimates $\widehat{\mathbb{X}} \in \mathbb{R}^N_{[0,1]}$, when they are taken while $\mathscr{C}_{x(t)}$ is governing and fairly non-changing. We have also showed that m (or T_x) is a dial for controlling the size of the condition space $\mathscr{R}_{x(t)}$.

In this chapter we are mainly concerned with practical considerations evoking $\widehat{\mathbb{X}}$ and $\mathbb{X}\big|\mathscr{C}_{x(t)}$ from $\big[x(t)\big]_{T_x}$. Additionally, we want to realize how perfectly we are able to detect a condition once $\mathscr{R}_{x(t)}$ is defined. Finally, in the latter part of this chapter we will be introducing a typical machine learning concepts for autonomous marginating of $\mathscr{R}_{x(t)}$ adaptively that is defined as the *condition spacing*. However, detail investigation of an optimal machine learning algorithm for condition spacing is left for a future study.

The main objective of introducing machine learning solutions goes even beyond condition spacing. A condition $\mathscr{C}_{x(t)}$ is not a static state in practice and itself is time varying. For instance, a 'healthy' gas turbine is subject to climatic deficiency due to the airborne debris that builds up on the compressor blades and reduces the effective air path (referred to as fouling). This process is easily recoverable (if not fully) by compressor water washing. This is all part of a natural process in a gas turbine during which $\mathscr{R}_{x(t)}$ essentially *travels* in the feature space. Therefore, it is necessary to consider $\mathscr{R}_{x(t)}$ as a time-varying space, hence $\mathscr{R}_{x(t)}(t)$. The concept of the time-varying condition space is no longer explicitly representable and requires machine learning solutions for adaptively relocating the condition space.

In more advanced scenarios, as the condition spaces travel in time, the spacing

between the conditions need to be controlled autonomously so that no intersection takes place among the condition spaces, and multi-agent control theories can be involved, in a sense that each condition space $\mathscr{R}_i(t)$ (where $i \in \{1, 2, ..., N_C\}$) can be regarded as an agent whose space is controlled by $P_{FA}(i)$ and $m(t)$. Then the objective would be to coordinate among the agents so that $P_{FA}(i)$ (i.e., the probability of a false alarm) is minimized for all i, whilst $\mathscr{R}_i(t) \cap \mathscr{R}_j(t) = \emptyset$ for all t.

However, due to time and data insufficiency the author has left these topics for a future study. In this thesis we have provided a simple solution (in contrast to the optimal solution for equation (4.21)) to time-invariant condition spacing problem. Moreover, we have introduced a typical and undeveloped machine learning solution for the condition spacing problem.

5.1 Practical Feature Extraction Methods

As discussed in the preceding chapter, it is possible to represent a condition $\mathscr{C}_{x(t)}$ by decomposing the sub-signal samples $\left[x(t)\right]_{T_x}$ into N frequency sub-bands, and computing the normalized energy of each sub-band (so that the sum of the energy of the sub-bands become 1). The number of sub-bands N (which is the dimension of the feature space) depends on the demarcation we wish to provide between the conditions, and the number of conditions one wishes to isolate. The block diagram of the feature extraction system

can be illustrated as shown in Fig. 5.1.

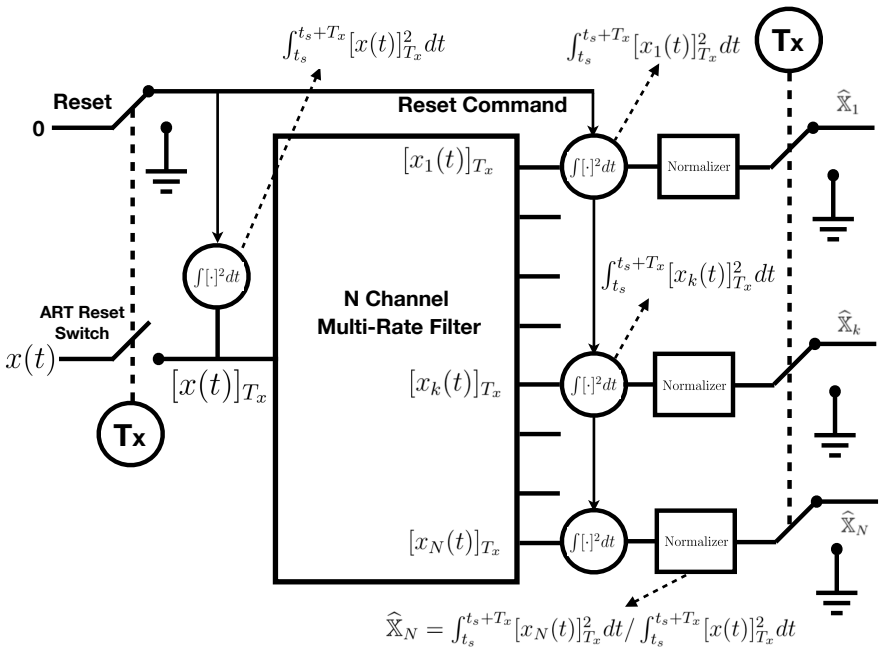

Figure 5.1: Feature extraction block diagram. Commands illustrated by solid arrows to the integrators are used to reset the integrating units to zero, every time the switch is opened.

The AE signal $x(t)$ is injected into the system on the left side of Fig. 5.1. A pulse timer is used to close a switch (referred to as the ART Reset Switch, because this switch will also be used to control the ART-1 neural network to be discussed later) for T_x seconds and then opens the switch for a random but short duration time T_x, before re-closing the ART Reset Switch again. The time T_x is a dial to adjust the variance of the features as discussed earlier. The longer the time T_x, the more the feature vectors

converge (same function as m in equation (4.28)). An energy counter is computing the energy of the AE signal during the closure of the switch. A multi-rate digital symmetric filter then samples $x(t)$ to make $x(n)$ as per the sampling process discussed on page 34, and decomposes the AE signal into N sub-bands with equal bandwidths and after a D/A conversion. Then the energy of each one of the N sub-bands is computed separately in the output of the filter simultaneously during the closure of the switch. The normalized energy of the sub-band k is finally used as an approximation to the k^{th} element of the N dimensional feature vector. As the switch opens at the end of the cycle, all energy meters are reset to zero until the switch re-closes again and another feature vector is derived. The filtering process taking place in Fig. 5.1 is summarized as follows:

$$x(t) \xrightarrow{subsampling} [x(t)]_{T_x} \xrightarrow{A/D} [x(n)]_{T_x} \xrightarrow{discrete filter} [x_k(n)]_{T_x} \xrightarrow{D/A} [x_k(t)]_{T_x}$$

The A/D process (i.e., conversion of $[x(t)]_{T_x}$ to $[x(n)]_{T_x}$) has been discussed on page 34. The D/A conversion is also a simple S/H block followed by a low-pass filter. Details are beyond the scope of this thesis and can be found in discrete signal processing textbooks.

Considering the number of available gas turbine AE samples (and therefore the number of different conditions) collected for this thesis, we have found that a $N = 8$ level filter delivers appropriate performance as it will be shown in the next section. The architecture of an 8-Channel multi-rate filter is shown in Fig. 5.2. In general, multi-rate filters are filters where their different parts operate at different rates. The most clear

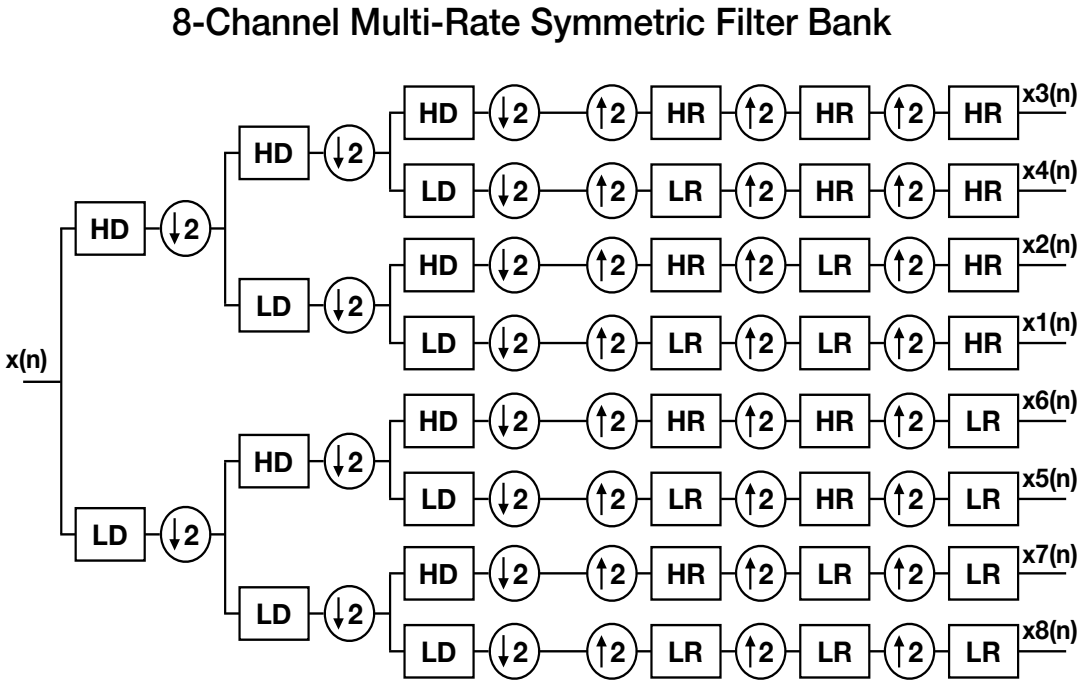

Figure 5.2: Filtering a signal $x(n)$ into 8 symmetric sub-bands $x_1(n)$ to $x_8(n)$ using the wavelet decomposition/reconstruction blocks (as illustrated in Fig. 5.3).

application of such a filter is when the input and output sample rates need to differ (decimation or interpolation) however, multi-rate filters are also often used in design where this is not the case. For example, one may have a system where the input and output sample rates are the same, but internally there is decimation and interpolation occurring in a series of filters such that the final output of the system has the same sample rate as the input. Such a design may exhibit lower cost than could be achieved with a

single-rate filter for various reasons [79]. The reason becomes more clear at the end of this section. The building block of the multi-rate filter used in this thesis is a Wavelet

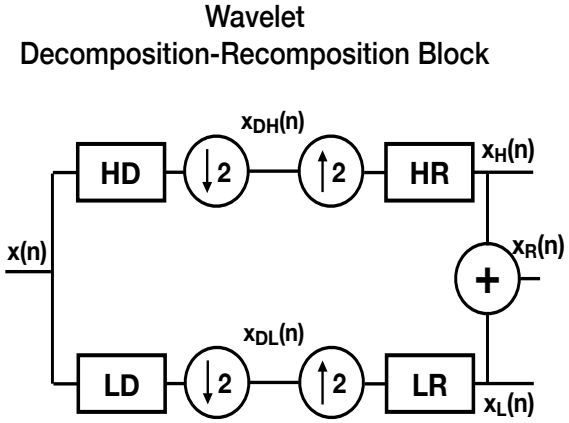

Figure 5.3: Multi-rate sub-band filtering block.

based lowpass/high-pass decomposition and reconstruction unit. The block diagram of such a filter is shown in Fig. 5.3, as discussed in Chapter 2, Fig. 2.1. The highpass decomposition/reconstruction FIR filters are referred to as HD and HR, respectively, and the lowpass FIR filters are referred to as LD and LR, respectively, as depicted in Fig. 5.4. The impulse response of the HD filter is given as:

$$H_{HD}(z) = -0.1115z^{-1} + 0.4946z^{-2} - 0.7511z^{-3} + 0.3153z^{-4} + 0.2263z^{-5}$$

$$-0.1298z^{-6} - 0.0975z^{-7} + 0.0275z^{-8} + 0.0316z^{-9} + 5.538\times10^{-4}z^{-10} - 0.0048z^{-11} - 0.0011z^{-12} \tag{5.1}$$

The $H_{HD}(z)$ filter is a standard 12 point Daubechies wavelet filter [59]. As discussed

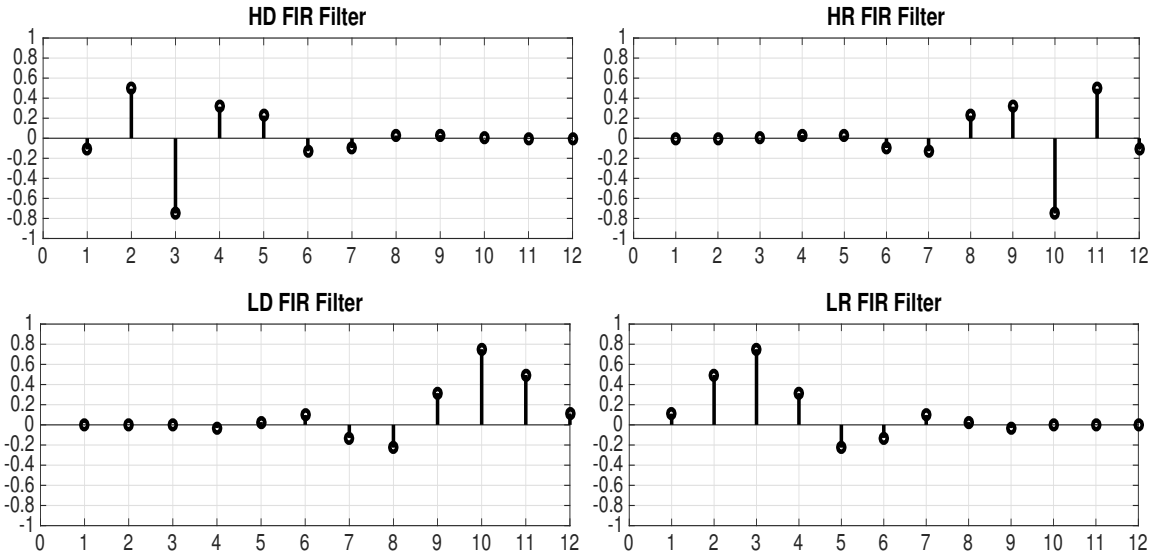

Figure 5.4: Wavelet FIR filters

in Chapter 2, $H_{LR}(z) = (-1)^n H_{HD}(z)$, $H_{HR}(z) = (-1)^n H_{LD}(z)$, and the HR and LR filters are Quadrature mirrors of HD and LD, respectively. The impulse response of the block (in Fig. 5.3) is plotted in Fig. 5.5. In can be seen that the wavelet decomposition and reconstruction block (WDR) can perfectly recompose a delayed version of the input signal with 11 points time delay (refer to Fig. 5.5), by adding its decomposed sub-bands.

It is evident from Fig. 5.6 that $x_H(n)$ and $x_L(n)$ (shown in Fig. 5.3) are the high and low frequency components of the input $x(n)$ (where a signal with n in brackets denote the discrete version of the signal).

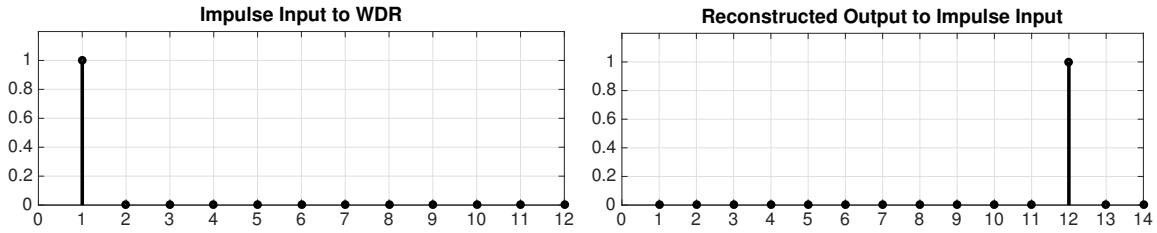

Figure 5.5: The left hand side plot is the input to the WDR block ($x(n) = \delta(n)$) as shown in Fig. 5.3. The right hand side plot is the output of the block $h(n) = x_R(n)$.

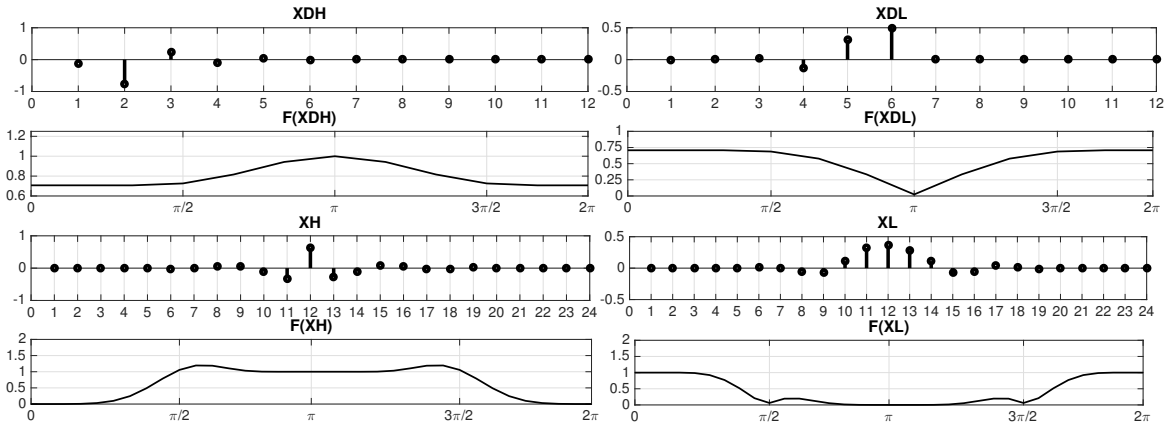

Figure 5.6: The time-domain output of FIR wavelet filters (X_{DH}, X_{DL}, X_H, X_L shown in Fig. 5.3) to an impulse input $x(n) = \delta(n)$, and corresponding frequency domain responses (i.e., Fourier Transforms $\mathscr{F}(X_{DH}), \mathscr{F}(X_{DL}), \mathscr{F}(X_H), \mathscr{F}(X_L)$). The x-axis of the frequency response plots represent normalized frequency and the y-axis represents the gain.

Fig. 5.7 illustrates the gain response of the individual channels. Unfortunately, the gain is not equal in all sub-bands, so not all frequency components are amplified equally, which is a cause for suppression of some feature components (especially H_7). However, since the AE signal is treated equally under all machine condition it is expected that

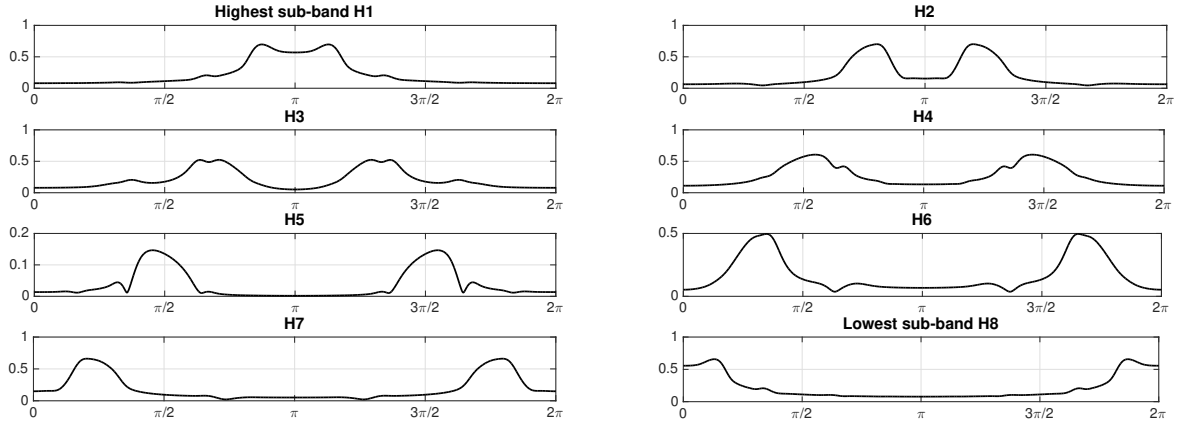

Figure 5.7: Frequency response of the 8-Channel filter of Fig. 5.2. In this figure $H_i = \mathscr{F}(x_i(n))$ (i.e., the Fourier transform of x_i), when the filter input $x(n) = \delta(n)$. The x-axis in all plots represents normalized frequency and the y-axis represents the gain.

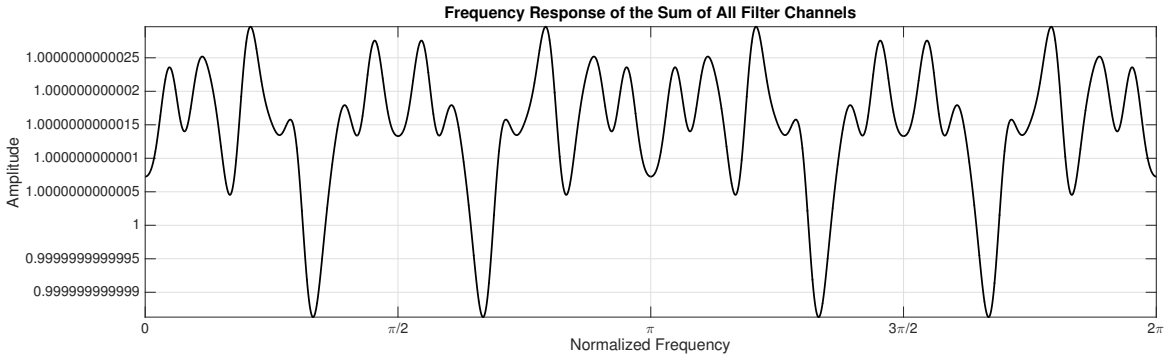

Figure 5.8: Frequency response of the sum of all channels (i.e., $\sum_{j=1}^{8} H_j$) of the multi-rate filter of Fig. 5.2.

the distortion in some of the channels will be tolerable. Fig. 5.8 shows ignorable ripples in the output gain, which is the sum of all the channels. It is also evident from Fig. 5.8 that the sum of energies of individual channels is equal to the energy of the input signal.

In other words, a delayed version of the input signal can be perfectly recomposed by summing the output of the 8 channels. This is a very desirable characteristic for the filter that allows normalization of the energy of the sub-bands and ultimately derivation of the feature vectors of input streams. Finally, Fig. 5.9 shows the wavelets $x_1(n)$ (upper left hand side plot) to $x_8(n)$ (lower right hand side plot), which are simply the impulse responses at nodes x_1 to x_8 in Fig. 5.2. It would be also possible to design the same

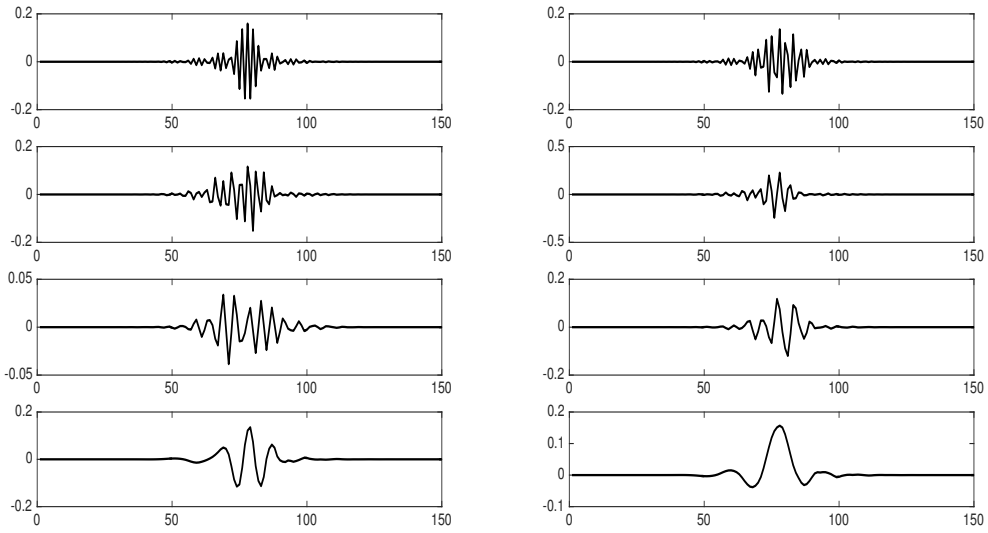

Figure 5.9: Wavelet FIR Filters of the order 150, as an alternative to the multi-rate filtering architecture proposed in Fig. 5.2, from high frequency to low frequency. The plots are x_1 to x_8 in Fig. 5.2 (upper left side plot to the lower right side plot) when the input in $x(n) = \delta(n)$

.

filter having 8 sub-band filters with impulse responses as shown in Fig. 5.9 (if one prefers not to use the multi-rate configuration of Fig. 5.2). As can be seen, the sub-band filters

would then have the order of 150 instead of 12 (as is the case for the WDR blocks). Therefore, the multi-rate filter has helped in simplifying the filter implementation substantially.

Overall, in this section we have shown that it is possible to implement a system for feature extraction of an AE signal with a set of very simple FIR filters. Therefore, the machine condition monitoring proposed in this thesis is technically very feasible and can be implemented online with very little computational resources, which can be considered an advantage of the theories developed in Chapters 3 and 4. In the next section we will see the actual performance of the 8-Channel multi-rate filter designed in this section for feature extraction of the experiments shown in Section 4.1.

5.2 Condition Spacing

In order to evoke features from the AE samples of Section 4.1, we will use the system shown in Fig. 5.1 to obtain a 4-dimensional feature vector (refer to page 137 for selection of N). We will see that with a 4-dimensional vector we will be able to isolate the different machine conditions examined in this thesis perfectly. We will be using the 8-channel multi-rate filter discussed in the previous section to create 4 sub-bands, by recombining higher frequency components together as shown below. Let us again use the notation $\mathbb{X}$ to denote a feature vector. We will define the feature vector $\mathbb{X}$ by combining the outputs

of Fig. 5.1, as given below (note that Xs in the entries of the vector below are the same as shown in the output of the system given in Fig. 5.1), that is

$$\mathbb{X} = \begin{bmatrix} X_1 + X_2 + X_3 + X_4 \\ X_5 + X_6 \\ X_7 \\ X_8 \end{bmatrix}_{4\times 1} \tag{5.2}$$

where,

$$X_i = \sum_{n=0}^{f_s \times T_x} x_i^2(n) \Big/ \sum_{n=0}^{f_s \times T_x} x^2(n) \tag{5.3}$$

and f_s is the sampling rate of the A/D converter used to digitize the AEs (in samples per second), T_x is the duration of ART reset switch closure shown in Fig. 5.1 (in seconds), and $x_i(n)$ denotes the time series response of the i^{th} channel of the filter (shown in Fig. 5.2), and finally $x(n)$ denotes the discrete-time input of the AE signal $x(t)$ to the feature extraction system.

Let us use the notation $\widehat{\mathbb{X}}^{(i)} \big| \mathscr{C}_c$ to denote the i^{th} sample for computing the feature vector $\mathbb{X} \big| \mathscr{C}_c$, when the machine is in condition c. We have computed feature vectors by processing the available AE recordings (gathered for this thesis) through the online feature extraction system as discussed in the previous section.

Considering that the main frequency of all gas turbines (in our experiments) are

119.2Hz (as illustrated in Fig. 4.4) and a sampling rate of 44100 samples/seconds (as explained in 34), one cycle would approximately correspond to 370 time samples or $T = 8.3$ ms. If we also consider $m \geq 500$ appropriate for sizing of the feature spaces (so that $\mathscr{R}_i \cap \mathscr{R}_j = \emptyset$), one should take $T_x \geq 500 \times T \approx 4.1$ seconds. In all the experiments we have assumed $T_x \approx 4.5$ seconds which approximately corresponds to $N_x = 2 \times 10^5$ points (44,100 samples per sec.), for all the sub-signals $\big[x(n)\big]_{N_x}$.

The objective is now to demonstrate how condition spacing can be practically performed, based on all the methodologies developed in this thesis. We will reuse the gas turbine AE samples used in Section 4.1, so that half the samples will be used for marginating the condition spaces, and the remainder used to verify correct classification of the AE samples.

Highly Stable Machine Conditions:

We have obtained 6 AE samples $x(t)$ of 6 gas turbines of the same make, size and type, newly commissioned and put into service. The feature vector $\mathbb{X}\big|\mathscr{C}_{\text{stab}}(i)$ represents the center point of $\mathscr{R}_{\text{stab}}$ for the i^{th} gas turbine ($i \in \{1, 2, ..., 6\}$), that is computed by taking an average of 100 $\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)$s, i.e.,

$$\mathbb{X}\big|\mathscr{C}_{\text{stab}}(i) = \frac{1}{100}\Big([\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)]_1 + [\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)]_2 + ... + [\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)]_{100}\Big)$$

where $[\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)]_{i'}$ is the feature vector estimate resulting from the $(i')^{th}$ closure of the

ART switch in Fig. 5.1 (that is $\widehat{\mathbb{X}}\big|\mathscr{C}_{\text{stab}}(i)$ corresponding to $\left[[x(t)]_{T_x}\right]_{i'}$). However,

$$\mathbb{X}\big|\mathscr{C}_{\text{stab}}(1) = \begin{bmatrix} 2.048 \times 10^{-5} \\ 12.05 \times 10^{-4} \\ 0.0960 \\ 0.9027 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{stab}}(2) = \begin{bmatrix} 1.864 \times 10^{-5} \\ 10.91 \times 10^{-4} \\ 0.085 \\ 0.913 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{stab}}(3) = \begin{bmatrix} 2.086 \times 10^{-5} \\ 9.489 \times 10^{-4} \\ 0.0864 \\ 0.912 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{stab}}(4) = \begin{bmatrix} 1.185 \times 10^{-5} \\ 6.78 \times 10^{-4} \\ 0.0758 \\ 0.923 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{stab}}(5) = \begin{bmatrix} 1.835 \times 10^{-5} \\ 7.83 \times 10^{-4} \\ 0.0819 \\ 0.917 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{stab}}(6) = \begin{bmatrix} 1.969 \times 10^{-5} \\ 6.75 \times 10^{-4} \\ 0.076 \\ 0.922 \end{bmatrix}$$

The maximum variance as computed for the above samples can be given as,

$$\sigma_{\mathbb{X}\big|\mathscr{C}_{\text{stab}}} = \begin{bmatrix} 1.61 \times 10^{-5} \\ 2.35 \times 10^{-4} \\ 0.0186 \\ 0.0188 \end{bmatrix}$$

A probability of less than 1% for false alarm would require $\Phi^{-1}((P_{FA}/2)^{1/4}) = \Phi^{-1}(0.266) \approx 0.65$ (refer to equations (4.24) and (4.25)). Selecting $\Phi^{-1}((P_{FA}/2)^{1/4}) = 1$, the condition

space for $\mathscr{R}_{stab}$ can be given as,

$$\mathscr{R}_{stab} : \{\mathbb{X} \in \mathbb{R}^4_{[0,1]} \Big| \begin{bmatrix} 0 \\ 4.2 \times 10^{-4} \\ 0.057 \\ 0.88 \end{bmatrix} < \widehat{\mathbb{X}} < \begin{bmatrix} 3.7 \times 10^{-5} \\ 14.42 \times 10^{-4} \\ 0.115 \\ 0.943 \end{bmatrix} \} \tag{5.4}$$

The condition space defined in equation (5.4) confines a space, inside of which any estimated feature vector obtained from one of the gas turbines (No. 1 to 6) would infer that the gas turbine is in a stable operating condition, with a probability of more than 99%.

Moderately Stable Machine Conditions:

Same as the case for stable conditions four AE recordings were taken of gas turbines that were operating normally, but either not commissioned properly, or not in a clean or new state. Again 100 estimated vectors were evoked from each recording in a random manner and the averages denoted by $\mathbb{X}\big|\mathscr{C}_{\text{Mod.}}(i)$ were taken:

$$\mathbb{X}\big|\mathscr{C}_{\text{Mod.}}(1) = \begin{bmatrix} 1.116 \times 10^{-5} \\ 38.23 \times 10^{-4} \\ 0.3023 \\ 0.6938 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{Mod.}}(2) = \begin{bmatrix} 0.943 \times 10^{-5} \\ 35.87 \times 10^{-4} \\ 0.28 \\ 0.716 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{Mod.}}(3) = \begin{bmatrix} 3.108 \times 10^{-5} \\ 32.95 \times 10^{-4} \\ 0.274 \\ 0.722 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{Mod.}}(4) = \begin{bmatrix} 1.682 \times 10^{-5} \\ 23.08 \times 10^{-4} \\ 0.1954 \\ 0.8022 \end{bmatrix}$$

The variance is computed as,

$$\sigma_{\mathbb{X}\big|\mathscr{C}_{\text{mod.}}} = \begin{bmatrix} 1.28 \times 10^{-5} \\ 2.52 \times 10^{-4} \\ 0.0143 \\ 0.0145 \end{bmatrix}$$

so that the condition space for a moderately stable state with the same false-alarm rate of less that 1% was obtained as,

$$\mathscr{R}_{mod.stable} : \{\mathbb{X} \in \mathbb{R}^4_{[0,1]}\Big| \begin{bmatrix} 0 \\ 20.5 \times 10^{-4} \\ 0.181 \\ 0.679 \end{bmatrix} < \widehat{\mathbb{X}} < \begin{bmatrix} 4.4 \times 10^{-5} \\ 40 \times 10^{-4} \\ 0.32 \\ 0.82 \end{bmatrix}\} \tag{5.5}$$

By comparison between equations (5.4) and (5.5), it becomes evident that the condition spaces for $\mathscr{C}_{\text{mod. stable}}$ and $\mathscr{C}_{\text{stable}}$ have no intersection. Therefore, the two conditions can be detected and perfectly isolated with a probability of more than 99%.

Problem in the Gearbox:

We have analyzed a problem that was detected in the gearbox of a gas turbine in Example 2 in Chapter 4. Now let us see the feature vectors that are obtained from processing the same data through the filtering method as discussed in this chapter. Let us denote the average features of the machine condition with gearbox problem as $\mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(i)$. The features are given below:

$$\mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(1) = \begin{bmatrix} 1.054 \times 10^{-5} \\ 69.2 \times 10^{-4} \\ 0.416 \\ 0.577 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(2) = \begin{bmatrix} 1.12 \times 10^{-5} \\ 68.7 \times 10^{-4} \\ 0.346 \\ 0.647 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(3) = \begin{bmatrix} 1.9 \times 10^{-5} \\ 63.5 \times 10^{-4} \\ 0.437 \\ 0.556 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(4) = \begin{bmatrix} 1.76 \times 10^{-5} \\ 65 \times 10^{-4} \\ 0.34 \\ 0.59 \end{bmatrix}, \mathbb{X}\big|\mathscr{C}_{\text{Gearbox}}(5) = \begin{bmatrix} 2.12 \times 10^{-5} \\ 70 \times 10^{-4} \\ 0.461 \\ 0.531 \end{bmatrix}$$

and

$$\sigma_{\mathbb{X}\left|\mathscr{C}_{\text{gearbox}}\right.} = \begin{bmatrix} 1.93 \times 10^{-5} \\ 13.12 \times 10^{-4} \\ 0.04 \\ 0.041 \end{bmatrix}$$

Therefore, the condition space $\mathscr{R}_{gearbox}$ can be partitioned with a false alarm rate of less than 1%,

$$\mathscr{R}_{gearbox} : \{\mathbb{X} \in \mathbb{R}^4_{[0,1]} \Big| \begin{bmatrix} 0 \\ 50.4 \times 10^{-4} \\ 0.305 \\ 0.49 \end{bmatrix} < \widehat{\mathbb{X}} < \begin{bmatrix} 4.05 \times 10^{-5} \\ 83.1 \times 10^{-4} \\ 0.502 \\ 0.69 \end{bmatrix} \} \tag{5.6}$$

Due to the fact that $\mathbb{X}\left|\mathscr{C}_{\text{gearbox}} > \mathbb{X}\right|\mathscr{C}_{\text{mod. stable}}$ (refer to equation (5.5) for comparison), the condition $\mathscr{C}_{\text{gearbox}}$ would have no intersection with $\mathscr{C}_{\text{mod. stable}}$, therefore it is possible to demarcate with a very low false alarm rate of 1%.

Start-up Conditions:

Finally, let us examine the spacing of the feature vectors representing start-up conditions

as discussed in Example 3 in Chapter 4, and denoted by $\mathbb{X}\big|\mathscr{C}_{\text{Startup}}(i)$.

$$\mathbb{X}\big|\mathscr{C}_{\text{Startup}}(1) = \begin{bmatrix} 120 \times 10^{-5} \\ 100 \times 10^{-4} \\ 0.204 \\ 0.785 \end{bmatrix}, \quad \mathbb{X}\big|\mathscr{C}_{\text{Startup}}(2) = \begin{bmatrix} 44.9 \times 10^{-4} \\ 57.3 \times 10^{-4} \\ 0.103 \\ 0.89 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{Startup}}(3) = \begin{bmatrix} 72.6 \times 10^{-4} \\ 114 \times 10^{-4} \\ 0.138 \\ 0.842 \end{bmatrix}, \quad \mathbb{X}\big|\mathscr{C}_{\text{Startup}}(4) = \begin{bmatrix} 54.2 \times 10^{-5} \\ 32.3 \times 10^{-4} \\ 0.082 \\ 0.914 \end{bmatrix}$$

$$\mathbb{X}\big|\mathscr{C}_{\text{Startup}}(5) = \begin{bmatrix} 52 \times 10^{-5} \\ 88 \times 10^{-4} \\ 0.141 \\ 0.849 \end{bmatrix}, \quad \mathbb{X}\big|\mathscr{C}_{\text{Startup}}(6) = \begin{bmatrix} 12 \times 10^{-5} \\ 32 \times 10^{-4} \\ 0.11 \\ 0.87 \end{bmatrix}$$

and

$$\sigma_{\mathbb{X}\big|\mathscr{C}_{\text{startup}}} = \begin{bmatrix} 6 \times 10^{-5} \\ 14.2 \times 10^{-4} \\ 0.021 \\ 0.022 \end{bmatrix}$$

And finally,

$$\mathscr{R}_{startup} : \{\mathbb{X} \in \mathbb{R}^4_{[0,1]} \Big| \begin{bmatrix} 6 \times 10^{-5} \\ 17.8 \times 10^{-4} \\ 0.061 \\ 0.762 \end{bmatrix} < \widehat{\mathbb{X}} < \begin{bmatrix} 732 \times 10^{-5} \\ 128 \times 10^{-4} \\ 0.22 \\ 0.93 \end{bmatrix} \} \tag{5.7}$$

Once again because of the higher relative energy of the high spectrum band in $\mathscr{C}_{\text{startup}}$, $\mathbb{X}\big|\mathscr{C}_{\text{Startup}} > \mathbb{X}\big|\mathscr{C}_{\sim\text{Startup}}$, where $\mathscr{C}_{\sim\text{startup}}$ implies any condition other than a startup condition. The ultimate conclusion is that all four conditions are capable of being perfectly isolated and detected with a 1% or lower false alarm rate.

The condition spaces given by equations (5.4) to (5.7) are definitions for the machine conditions $\mathscr{C}_{\text{stable}}$ to $\mathscr{C}_{\text{startup}}$, respectively. In these definitions we have attempted to mitigate the probability of false alarm and a *miss* (that is the probability of a sample belonging to a condition and not being detectable) to less than 1%, by separating the condition spaces so that the intersection between any two given subsets would be null. We were able to do that by selecting the time dial T_x in Fig. 5.1 appropriately (i.e., 4.5 seconds as explained earlier). Let us denote $P[\mathscr{C}_i\big|\mathscr{C}_j]$ as the probability of condition $\mathscr{C}_i$ being detected when the machine was actually operating under condition $\mathscr{C}_j$ for a time dial equal to T_x, where $T_x = m/f_s$ and m is the number of samples, as given by equation (4.9) and $f_s = 44,100$ (samples/second) is the sampling rate in the A/D converter

Table 5.1: Confusion matrix for detecting the machine conditions.

	Stable	Moderately Stable	Gearbox Fault	Start-up	Undefined
Stable	$P[\mathscr{C}_{stable}\|\mathscr{C}_{stable}]$ $= 95\%$	$P[\mathscr{C}_{stable}\|\mathscr{C}_{mod.}]$ $= 0\%$	$P[\mathscr{C}_{stable}\|\mathscr{C}_{gear}]$ $= 0\%$	$P[\mathscr{C}_{stable}\|\mathscr{C}_{startup}]$ $= 0\%$	$P[\mathscr{C}_{stable}\|\mathscr{C}_{Undef.}]$ $= 5\%$
Moderately Stable	$P[\mathscr{C}_{mod.}\|\mathscr{C}_{stable}]$ $= 0\%$	$P[\mathscr{C}_{mod.}\|\mathscr{C}_{mod.}]$ $= 100\%$	$P[\mathscr{C}_{mod.}\|\mathscr{C}_{gear}]$ $= 0\%$	$P[\mathscr{C}_{mod.}\|\mathscr{C}_{startup}]$ $= 0\%$	$P[\mathscr{C}_{mod.}\|\mathscr{C}_{Undef.}]$ $= 0\%$
Gearbox Fault	$P[\mathscr{C}_{gear}\|\mathscr{C}_{stable}]$ $= 0\%$	$P[\mathscr{C}_{gear}\|\mathscr{C}_{mod.}]$ $= 0\%$	$P[\mathscr{C}_{gear}\|\mathscr{C}_{gear}]$ $= 100\%$	$P[\mathscr{C}_{gear}\|\mathscr{C}_{startup}]$ $= 0\%$	$P[\mathscr{C}_{gear}\|\mathscr{C}_{Undef.}]$ $= 0\%$
Start-up	$P[\mathscr{C}_{startup}\|\mathscr{C}_{stable}]$ $= 0\%$	$P[\mathscr{C}_{startup}\|\mathscr{C}_{mod.}]$ $= 0\%$	$P[\mathscr{C}_{startup}\|\mathscr{C}_{gear}]$ $= 0\%$	$P[\mathscr{C}_{startup}\|\mathscr{C}_{startup}]$ $= 100\%$	$P[\mathscr{C}_{startup}\|\mathscr{C}_{Undef.}]$ $= 0\%$
Undefined	$P[\mathscr{C}_{Undef}\|\mathscr{C}_{stable}]$ $= 5\%$	$P[\mathscr{C}_{Undef}\|\mathscr{C}_{mod.}]$ $= 0\%$	$P[\mathscr{C}_{Undef}\|\mathscr{C}_{gear}]$ $= 0\%$	$P[\mathscr{C}_{Undef}\|\mathscr{C}_{startup}]$ $= 0\%$	$P[\mathscr{C}_{Undef}\|\mathscr{C}_{Undef.}]$ $= 95\%$

(digitizing the AE recordings).

We have validated the condition spaces given by equations (5.4) to (5.7) by confirming correct classification of a new set of AE recordings consisting 1000 samples, and for $T_x = 4.5$ seconds. The result is given in Table 5.1. Except for $\mathscr{R}_{stable}$ we have attained no misses and no false alarms. However, we have observed an error of 5% for $P[\mathscr{C}_{startup}|\mathscr{C}_{startup}]$. Since there are no errors elsewhere in the confusion matrix we may conclude that the error in $P[\mathscr{C}_{startup}|\mathscr{C}_{startup}]$ is due to a momentary change in the machine condition to an unknown condition during the AE recordings.

Now let us examine the effect of m on the condition spacing and condition classification. Recalling (from page 139) that $T_x = mT$ that is to examine the effect of T_x (considering that T is always assumed constant as per Assumption 3.2). Therefore, we have tested $P[\mathscr{C}_{startup}|\mathscr{C}_{startup}]$ and $P[\mathscr{C}_{\sim startup}|\mathscr{C}_{startup}]$ varying T_x in 1000 trials. The results are given below:

$$T_x = 0.22sec: \quad P[\mathscr{C}_{startup}|\mathscr{C}_{startup}, T] = 89.6\%, \quad P[\mathscr{C}_{\sim startup}|\mathscr{C}_{startup}, T] = 0\%$$

$$T_x = 0.45sec: \quad P[\mathscr{C}_{startup}|\mathscr{C}_{startup}, T] = 94.2\%, \quad P[\mathscr{C}_{\sim startup}|\mathscr{C}_{startup}, T] = 0\%$$

$$T_x > 3.4sec: \quad P[\mathscr{C}_{startup}|\mathscr{C}_{startup}, T] = 100\%, \quad P[\mathscr{C}_{\sim startup}|\mathscr{C}_{startup}, T] = 0\%$$

Since there are several conditions to be isolated (as demonstrated in Table 5.1), we have now redefined *False-Alarm* and *Miss*. In the new context, $P[\mathscr{C}_{\sim startup}|\mathscr{C}_{startup}]$ is defined as the probability of a false-alarm, and $1 - P[\mathscr{C}_{startup}|\mathscr{C}_{startup}]$ is the probability of a miss. The false-alarm is characterized by the spacing between the condition spaces, and the miss is in fact replacing the P_{FA} as defined in Chapter 4. Experiments based on all available AE samples have shown that for $T_x = 4.5sec$, the probability of a false-alarm is zero due to the fact that the condition spaces have no intersection, whereas the probability of a miss is almost 0 (less than 5 faults in 1000 feature vector estimates).

The excellent results obtained in our hypothesis testings is all due to the fact that we were able to extend the length of the sub-signals to $T_x = 4.5$ sec. for complete

convergence of the feature vector estimates. In applications where the rate of change in the condition is too fast in comparison with T_x, such as in aerospace applications, we are not able to take such long sub-signals, therefore as a last resort would be to have no choice, but increasing the dimension of the feature space N beyond 4, which dramatically increases the computational resources.

In real practice, the number of machine conditions are much more that what we have experimented (given our limited available AE samples). In order to classify large number of conditions it is not possible to manually perform the analysis we have shown in Section 4.2. Therefore, we will try to develop a machine learning algorithm in the next section, to be able to monitor and alarm abnormal machine conditions, and/or even anomalies and faults.

5.3 Machine Learning and Challenges (Future Work)

In this section we are introducing a typical machine learning algorithm for autonomous condition spacing and classification. The purpose is not to suggest a particular neural network for gas turbine CM, since no attempt has been made to compare the performance of different machine learning solutions. However, we are aiming to propose new areas of research for extending the scope of this thesis to address the concerns stipulated

on page 156. Specifically, in this section we wish to highlight some challenges that interested researchers may confront applying the techniques introduced in this thesis, and further propose certain remedies for converting these techniques into a machine learning algorithm.

Let us begin by explaining why machine learning solutions are practically required for CM. The number of conditions in a practical condition monitoring application can be enormously high as there will be a new condition associated with every speed the rotor is experiencing from the time the engine is started until its shut-down. Probably every loading level of the engine will also be represented by a condition different with all the other stable and healthy conditions. All these conditions can be considered to be equivalent (in a sense that they are all 'healthy and stable' conditions), although occupying different regions in the feature space. An example has been demonstrated in the previous section where start-up conditions of a given gas turbine engine have exhibited quite different feature vectors. This implies that in an actual industrial environment it would be only possible to identify the conditions of an engine, if a machine learning system is deployed to learn new conditions as the gas turbine evolves in a non-supervised manner. This means that the machine learning system should have the capability to autonomously group similar 'patterns' into same 'classes'. Accordingly, the neurons in

an artificial neural network (ANN) are each gradually assigned with feature vectors, associating a point in the feature space to a condition. As the gas turbine system evolves in time, more and more neurons are assigned until the condition space is almost occupied with neurons, each one located in a small proximity to the other. After a while, new feature vectors will begin to fall onto locations already occupied with a previous neuron, which is a state in which the learning process is over and the ANN is fully trained with a particular condition. At this point, if the gas turbine experiments similar operating conditions, the neurons already trained in the process explained above will be *resonating* and the system can recognize the particular machine condition.

In case the machine condition is changed, a new region in the feature space begins to form with a new set of neurons that are labelled differently (corresponding to the new condition), through a separate process (not covered in this thesis). That process is essentially correlating a physical phenomenon to a particular region in the feature space, based on instruments mounted on the gas turbine and an engineering know how. For instance, we have illustrated an example on correlating problems associated with the gearbox to a particular region in the feature space before. In order to interpret the entire feature space, a detailed knowledge about gas turbine mechanics is required which is beyond the scope of this thesis.

Another characteristic of the machine conditions is that the condition space travels in time. For instance, let us consider a gas turbine going through an aging process. The performance of the machine is deteriorated as the parts wear and tear. Still the condition might be considered a stable and healthy one, as long as natural aging effects are governing. In these circumstances we require the machine learning to be *adaptive* to gradually adapt to the changing conditions. A well suited family of ANNs, for these type of applications might be *Adaptive Resonance Theory* (ART) ANNs [80].

An ART-1 network consists of two fields: a Comparison layer denoted as F-1 and a Recognition layer denoted as F-2, as shown in Fig. 5.10 [80]. Initially the input layer consists of no feature vector that deactivates the recognition layer through the ART reset Switch. As a pattern (feature vector) is presented in the input layer (simultaneously with every opening edge of the ART switch as shown in Fig. 5.1), Gain-1 activates the comparison layer which fans out the same pattern through the long-term memory traces. A competition then takes place in the Recognition layer among the neurons each representing a self-organized class by previously trained patterns, and the neuron being closest to the applied input fires, deactivating all other neurons in the F-2 field. This enforces Gain-1 to zero which puts the already stored pattern of the particular fired exposed to F-1 layer for comparison. This pattern is then compared with the input pattern and if a sufficient match is recognized by means of a vigilance test taking part

in the orienting subsystem, the input is then stored in the memory, overwriting the old pattern with recent features. If however the match is found insufficiently close to the

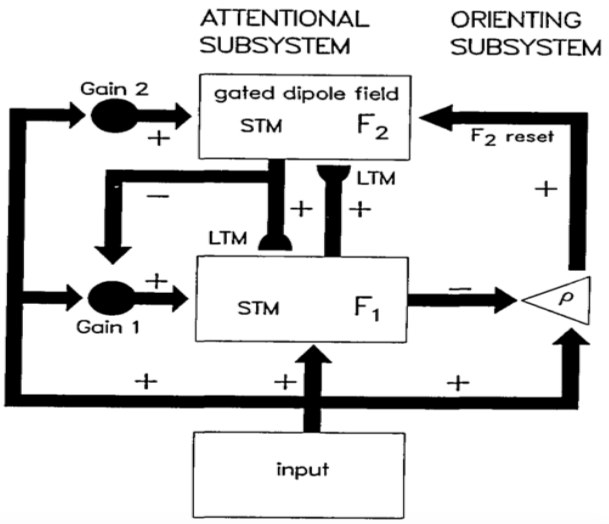

Figure 5.10: ART-1 network architecture. Rectangles represent fields where Short Term Memory (STM) patterns are stored. Semi-circles represent adaptive filter pathways, and arrows indicate paths which are not adaptive. Filled circles represent gain control nuclei, which sum input signals. Output paths are not specified in the sense that at any given time a uniform signal is sent to all nodes in a receptor field. Gain Control F_1 and F_2 coordinates STM processing with input presentation rate [78].

former pattern, the particular neuron will be deactivated through the F2 Reset signal, and another competition begins among the remaining neurons still active, until another close match is found. This again fires a new neuron which reflects its own pattern to the Comparison layer for a cross check with the input data through a second vigilance test, which if unsuccessful will deactivate the neuron and another competition again takes

place until all neurons are exhausted. If there is eventually no pattern close enough with the level of vigilance associated for the comparison, the input pattern will be stored as a new class and long-term traces of both fields will be modified accordingly.

5.3.1 Vigilance Vector

As explained in the preceding subsection, in order for the ART-1 ANN to classify feature vectors the closest neuron to the input vector is found first. Then a vigilance test takes place to decide if the input vector can be classified into the class represented by the closest neuron, or else if a new neuron is to be assigned to the new vector. If the contents of the short-term and long-term memories (STM, and LTM) corresponding to the k^{th} neuron, are denoted by $\mathscr{C}_i$ and $\mathbb{X}_{k|\mathscr{C}_i}$ respectively, then the vigilance test is a comparison between $|\widehat{\mathbb{X}} - \mathbb{X}_{k|\mathscr{C}_i}|$ versus $\mathbb{X}_v$ (hereinafter referred to as *the vigilance vector*), where $\widehat{\mathbb{X}}$ and $\mathbb{X}_v$ denote the input feature vector estimate and the vigilance vector, respectively. If,

$$\left|\widehat{\mathbb{X}} - \mathbb{X}_{k|\mathscr{C}_i}\right| \leq \mathbb{X}_v \implies \mathbb{X}_{k|\mathscr{C}_i} \leftarrow \widehat{\mathbb{X}}$$

the new vector estimate $\widehat{\mathbb{X}}$ will be classified into the condition $\mathscr{C}_i$, and will replace $\mathbb{X}_{k|\mathscr{C}_i}$. If the vigilance test is not satisfied for all the available neurons, a new neuron will be assigned and the human expert would have to decide if a new condition has emerged or if $\widehat{\mathbb{X}}$ still belongs to an existing condition. Hence, it can be considered as if the feature space is effectively divided into a number of grids, where the grid is sized based on the vigilance vector $\mathbb{X}_v$. If $\mathbb{X}_v$ is selected too small, the number of neurons to cover the entire

feature space will be too large, requiring excessive memory capacity. If $\mathbb{X}_v$ is selected too large, conditions may overlap. Therefore, choosing an appropriate $\mathbb{X}_v$ requires *a priori* knowledge about the separation of the conditions and the time dial T_x. For instance, if the machine conditions under investigation were only those defined by equations (5.4) to (5.7), then the feature space would have been confined by the limits,

$$\text{Feature-Space} = \begin{bmatrix} 0 \\ 4 \times 10^{-4} \\ 0.057 \\ 0.49 \end{bmatrix} \text{ to } \begin{bmatrix} 732 \times 10^{-5} \\ 128 \times 10^{-4} \\ 0.502 \\ 0.943 \end{bmatrix} \tag{5.8}$$

By investigating through equations (5.4) to (5.7), one can also identify a suitable choice for the vigilance vector,

$$\mathbb{X}_v = \begin{bmatrix} 10^{-5} \\ 10^{-4} \\ 10^{-2} \\ 10^{-2} \end{bmatrix} \tag{5.9}$$

which is also sizes the grids of the feature space. Accordingly, the feature space can be considered to be divided into a number of grids, to be estimated as Feature-Space/$\mathbb{X}_v$,

that is,

$$\text{Segments on each dimension} = \begin{bmatrix} 732 \\ 124 \\ 45 \\ 90 \end{bmatrix}$$

and the total number of grids can be estimated as,

$$\text{No. of Grids} = 732 \times 124 \times 45 \times 90 = 367,610,400 \tag{5.10}$$

Equation (5.10) shows that for implementation of the ART-1 ANN, 367,610,400 neurons are required.

5.3.2 Non-Homogenous Grids and Training Time

As demonstrated by equation (5.10), the number of neurons required to train an ART-1 ANN with all available AE samples collected for this thesis is simply unaffordable. This is clear if we consider $T = 4.5$ seconds for every neuron to be trained, which requires over 50 years. An alternative technique that can be used to diminish the number of neurons is to deploy *non-homogenous grids*, that is a feature space divided into grids of non-uniform size.

For instance, let us note by investigating equations (5.4) to (5.7) that the first entry of the feature vectors for all the conditions except $\mathscr{C}_{\text{Startup}}$ vary between $0 \sim 4.4 \times 10^{-5}$,

whereas for $\mathscr{C}_{\text{Startup}}$ they vary between $6\times10^{-5} \sim 732\times10^{-5}$. Therefore, it seems suitable to divide the feature-space into only $7\times124\times45\times90$ (instead of equation (5.10)), except that this time not all the grids within the feature space are of the same size. Particularly the grids in $\mathscr{R}_{\text{startup}}$ are sized differently, i.e.,

$$\mathbb{X}_{v|\mathscr{C}_{\text{startup}}} = \begin{bmatrix} (732-6)\times10^{-5} \\ 10^{-4} \\ 10^{-2} \\ 10^{-2} \end{bmatrix} \tag{5.11}$$

and

$$\mathbb{X}_{v|\sim\mathscr{C}_{\text{startup}}} = \begin{bmatrix} 10^{-5} \\ 10^{-4} \\ 10^{-2} \\ 10^{-2} \end{bmatrix} \tag{5.12}$$

In this way we will still be able to demarcate the conditions as good as before, but with a far less number of grids (hence the number of neurons) and the time for training the feature space would be around 183 days (if the ANN is exposed to no repeated patterns in this duration). As we can see we have been able to reduce the training period dramatically by economizing the number of neurons. One drawback is that every class (or condition) may necessitate a unique vigilance vector, that is to be derived according to the variance of the input patterns for that condition. This scheme is an extension to the ART-1 ANN introduced in this thesis.

5.4 Summary

This chapter aims to integrate the mathematical derivations and results obtained in chapters 3 and 4 into a practical online N-dimensional feature extraction system with software controlled flexibility for m (or T_x). Additionally, this chapter aims to demonstrate how machine conditions could be partitioned into non-intersecting condition spaces through an online condition spacing algorithm. Meanwhile, we have introduced shortnesses associated with our results, and have tried to propose certain remedies to be examined in the future (as an extension of this thesis).

Specifically, we have presented an 8-channel filter for decomposing an AE signal $x(t)$ into orthogonal subbands $x_k(t)$ where $k \in \{1, 2, ..., 8\}$. For a simple implementation we have chosen a digital multi-rate symmetric filter bank to process the discrete time signal $x(n)$ (made by sampling $x(t)$ as discussed on page 34) and then converting $x_k(n)$s to $x_k(t)$s. We have explained that the 8-channel multi-rate filter can be extended to an N-channel filter bank with the same method explained in Section 5.1. We have stated that the experiments undertaken in Chapter 4 could be successfully categorized into non-intersecting condition spaces through a 4 dimensional feature space, and finally partitioned all the conditions (discussed in Section 4.2) with clear boundaries.

In order to verify the results obtained in this thesis we have then taken random AE

samples from the conditions used earlier (for the purpose of condition spacing), and manually classified the new samples into one of the condition spaces. Results summarized in Table 5.1 showed no missed-alarms nor false-alarms, proving that our methodology developed for the subject of the thesis performs satisfactory as long as the available AE samples were concerned.

Finally, we have introduced a new area for further extension of the results obtained in this thesis through machine learning solutions. We have tried to underline the probable hassles that the interested researcher may confront applying these techniques into machine learning algorithms, and have tried to address possible remedies.

Chapter 6

Summary and Conclusion

6.1 Summary of the Thesis

The objectives of our work in this thesis was to develop a theoretical framework for evoking optimal features (as defined in Definition 6.2), that uniquely represent the conditions governing the operation of a gas turbine. We have presented an on-line filtering procedure that can estimate features (and therefore assess the machine condition) with any given level of accuracy, involving minimal computational resources. We have also formulated a criteria for inferring about the condition of the machine based on these features. As a result one would be theoretically able to detect a machine condition among a set of available conditions (virtually) without error. Of course, there are mathematical proofs presented to support our idea and the methodology. Finally, a machine learning

system is presented based on mathematical derivations, for learning new conditions and also performing judgement about the condition of the gas turbine, simply by online processing of the AE signal into the system, and based on accumulated experience collected by the neurons.

We have firstly proposed definitions for 'condition' and 'feature'. A Condition (as well as a feature) is a general notion and may be defined differently, but our emphasis is to help our own objective. Therefore, we have customized these notions to suit our purpose as given below:

Definition 6.1: *The average behaviour of an acoustic signal emitted from a cyclic mechanical equipment, under a fairly constant set of inputs, is hereinafter regarded as a 'machine condition', or simply a condition.*

Definition 6.2: *Quantitative characteristics of an AE signal that uniquely represent machine conditions are referred to as features. A feature vector is a vector comprising a set of features.*

In Chapter 2 we have summarized recent activities in the field of machine condition monitoring using acoustic emissions. We have mainly focused on some of the prominent

nonlinear and nonstationary signal processing methods that have been claimed of being effective in extracting features of acoustic signals. These methods were discussed in detail first, and then we have applied the methods to actual AE data collected from real life gas turbine applications in service. We have shown that these methods are generally incapable of representing the different conditions that were governing the gas turbines. Additionally, we have shown that applying these methods are only possible (generally) through extensive computational resources in off-line circumstances.

The reason why Fourier analysis cannot help analyzing AE signals has been explained in length in Chapters 2 and 3. We have seen that the highly random nature of the AE signals are causing non-consistent results when applying Fourier methods. In Chapter 3 we have firstly introduced a general characteristic of random but 'cyclic' mechanical processes. These processes emit *cyclostationary* AE signals and entail certain attributes that enabled us to construct a theory to sample stationary sub-signals from an AE signal, and use these sub-signals to evoke deterministic features that *optimally* represent the average behaviour of the AE signal, in a sense that the features can be used to perfectly reconstruct the autocorrelation function of the signal (which uniquely associates to the machine condition as per Assumption 6.2). Therefore, our first assumption in this thesis is as follows:

Assumption 6.1: *A gas turbine AE signal operating under a set of unchanging operating condition, is assumed to be a cyclostationary random process.*

The theory established in Chapter 3 based on Assumption 6.1 proves a set of very basic characteristics for gas turbine AE signals, such as *ergodicity*. In an ergodic process we are able to use time average and autocorrelation functions, interchangeably for ensemble means and autocorrelations. Hence, we are able to use a sufficiently long AE sample to deduce important derivations such as periodogram estimates, and to deduce deterministic quantities by taking the average of some random variables over a large number of samples. In particular, we have shown that the energy of a certain sub-band of an AE signal does not change unless there is a source of anomaly in that sub-band.

Identifying a fault in a signal with a narrow band-width is a straightforward practice in condition monitoring excelled by many maintenance engineers. Moreover, we have proven that a change in the machine condition is detectable by measuring the energy of a set of sub-bands without exception, if we have access to a reasonably long AE signal. These derivations have been explained through a series of assumptions and a theorem summarized below:

Assumption 6.2: *Assume that $x(t)$ is a random signal representing a cyclostationary*

process with period T, whose conditions (refer to Definition 6.1) are constant through a certain time interval $[0, T_p]$, where $E[x(t)] = 0$. Let us also assume that $[x(t)]_{T_x}$ is a sub-signal of $x(t)$, taken within $[t_s, t_s + T_x] \subset [0, T_p]$, where t_s is also randomly chosen, with a uniform distribution. Additionally assume that $\mathscr{C}_{x(t)}$ denotes the condition of the cyclostationary process, that is fully accessed through the autocorrelation function of $x(t)$. Then:

Theorem 6.1: *If $x(t)$ and $[x(t)]_{T_x}$ are random signals satisfying the conditions in Assumption 3.2, then $[x(t)]_{T_x}$ can be decomposed into orthogonal mono-components $[x_k(t)]_{T_x}$, so that $[x(t)]_{T_x} = \sum_k [x_k(t)]_{T_x}$, and $1/T_x \int_{t=t_s}^{t_s+T_x} [x(t)]_{T_x}^2 dt = 1/T_x \sum_k \int_{t=t_s}^{t_s+T_x} [x_k(t)]_{T_x}^2 dt$, where,*

$$\frac{1}{T_x} \int_{t_s}^{t_s+T_x} [x_k(t)]_{T_x}^2 dt = |X_k|^2, \quad \textit{and} \quad X_k = \frac{1}{T_x} \int_{t=t_s}^{t_s+T_x} [x(t)]_{T_x} e^{-j2\pi kt/T}$$

Additionally, if $[x(t)]_{T_x}$ is sufficiently long, i.e., $T_x >> T$, then the monotonic energy components $|X_k|^2$, converge into deterministic quantities (denoted by $E[|X_k|^2]$) that can optimally represent $\mathscr{C}_{x(t)}$, where,

$$E[|X_k|^2] = \lim_{T_p \to \infty} \frac{1}{T_p} \int_{t=0}^{T_p} [x_k(t)]_{T_p}^2 dt$$

We have used Theorem 6.1 to formally define a feature and a feature vector as

given below:

Definition 6.3: *If $x(t)$ and $[x(t)]_{T_x}$ are random signals, and if the conditions in Assumption 6.2 hold, then*

$$\mathbb{X}_k = E\big[|X_k|^2\big] / \sum_{i=-\infty}^{\infty} E\big[|X_i|^2\big]$$

will be referred to as a feature of $x(t)$, and a vector composed of the features $\mathbb{X}_k$ will be referred to as the feature vector of $x(t)$ and denoted by $\mathbb{X}$. Additionally, an estimate of the feature vector $\mathbb{X}_k$ under condition $\mathscr{C}_{x(t)}$, that is approximated by N features, is denoted by,

$$\widehat{\mathbb{X}}\big|\mathscr{C}_{x(t)} = [\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N]^T$$

Definition 6.4: *Since the features $\mathbb{X}_k$ are real numbers, and $\mathbb{X}_k \in [0, 1]$, therefore the feature vector lies in an N dimensional space denoted by $\mathbb{R}^N_{[0,1]}$, and referred to as the feature space.*

In Chapter 4 we have shown that the feature vector estimates (defined in Definition 6.3) are random variables with a Gaussian distribution. This provides a confined space in the feature space (defined in Definition 6.4), inside which the probability of the feature estimates to be placed, can be determined. This space is defined as the condition space as explained in Definition 6.4:

Definition 6.5: *If* $\widehat{\mathbb{X}} = \left[\widehat{\mathbb{X}}_1, \widehat{\mathbb{X}}_2, ..., \widehat{\mathbb{X}}_N\right]^T \in \mathbb{R}^N_{[0,1]}$ *denote the estimated feature vector for the signal* $[x(t)]_{T_x}$ *with the conditions in Assumption 3.2, the smallest space* $\mathscr{R}_{x(t)} \subset \mathbb{R}^N_{[0,1]}$ *where,*

$$\int_{\mathscr{R}_{x(t)}} p_{\widehat{\mathbb{X}}|\mathscr{C}_{x(t)}}\left(\widehat{\mathbb{X}}\right) d\widehat{\mathbb{X}} = 1 - P_{FA} \tag{6.1}$$

is defined as the Condition Space for $\mathscr{C}_{x(t)}$.

The very useful attribute associated with the condition space is that it can be sized as desirable. By taking long sub-signals $\left[x(t)\right]_{T_x}$ one can reduce the size of the condition space, and by taking shorter lengths of sub-signals, the space inflates. When there are numerous conditions in the machine to be isolated, it is suitable to make the condition spaces sized so that there are no intersection among the condition spaces. Then it will be possible to demarcate the conditions from one another without a false-alarm.

The dimension of the feature space is a key parameter for controlling the probability of 'missing' a condition. In fact the certainty associated with the hypothesis testing can be controlled by selecting the dimension of the feature space. A high dimensional feature space dramatically increases the computational resources required for extraction of the feature vectors. Practically, feature extraction is accomplished through an FIR filter. A low dimensional feature space reduces the order of the FIR filter, but it sacrifices the level of the certainty associated with hypothesis testing.

Chapter 4 shows that any machine condition can be uniquely mapped into a confined space in the feature space. This reduces the condition monitoring process to a simple location finding of the feature vector estimates evoked from a signal. If $x(t)$ corresponds to the feature vectors that fall in $\mathscr{R}_{x(t)}$, then the system condition would be the condition associated with $\mathscr{R}_{x(t)}$, with a probability of $1 - P_{FA}$.

Marginating condition spaces is a simple task. It is sufficient to input a sufficiently long AE signal, to the feature extraction unit. The resulting vector would be a point in $\mathbb{R}^N_{[0,1]}$ that represents the center of the condition space $\mathscr{R}_{x(t)}$, and the variance of the vectors determine the radius of the ball representing the condition space. By choosing longer AE sub-signals it is possible to form smaller balls so that there are no intersection with the neighbouring balls associating other conditions.

In Chapter 5 we have introduced an N channel FIR filter for decomposing a signal into N orthogonal band-limited signals. We are effectively decomposing a signal $x(t)$ into N signals $x_1(t)$ to $x_N(t)$, so that each sub-band contains a distinct frequency band, in a way that $x(t) = \sum_{i=1}^{N} x_i(t)$, and $x^2(t) = \sum_{i=1}^{N} x_i^2(t)$. We have shown in Chapter 3 that this would be possible as long as $x(t)$ fulfills the requirements set forth in Assumption 6.2. The next step is computing the feature vector estimates with the procedure

explained in Definition 6.3.

In Chapter 4, we have analyzed several machine conditions using actual and real life AE signals obtained from different gas turbines in a qualitative manner. In Chapter 5, we have used the same data and extracted feature vectors for a four dimensional feature space, i.e., $N = 4$ using an 8 channel FIR filter system that was fully developed in this thesis. Then we have formulated the condition spaces corresponding to each set of data and we have shown that the condition spaces were having no intersection with one another, when the AE samples were taken for a length of 4.5 seconds.

Finally, the latter part of Chapter 5 is devoted to verification of the methodologies developed in this thesis. We have firstly used a condition spacing algorithm for marginating the condition spaces in the feature space. Secondly, we have demonstrated perfect classification of random AE samples into their corresponding conditions without any missed-alarms or false-alarms.

6.2 Future Direction of Research

In conclusion, the objectives set forth in this thesis were fully accomplished to the extent of the available data. However, in reality and in a long run, condition spaces are

not static and travel in time, therefore the condition monitoring problem evolves into a more advanced level of monitoring 'dynamic conditions'. In this context $\mathscr{R}_{x(t)}$ will be $\mathscr{R}_{x(t)}(t)$. Then the machine learning algorithm will become more complex, as the condition spacings already formed would need to be controlled in a continuous manner with new conditions emerging.

In the meantime, the feature extraction system proposed in Chapter 5 consists of a hardware digital filter with predefined number of channels N. Although we have foreseen a flexible reset switch for adaptive adjustment of T_x (or m), the system lacks flexibility for changing N. This means that the CM system design is based on *a priori* knowledge on all future conditions under which the gas turbine is expected to operate.

Finally, the work presented in this thesis merely encompasses cyclostationary processes and will not be applicable in non-cyclic processes. Moreover, accurate classification is only expected under the conditions set forth in Assumption 6.2 and $T_x >> T$ (i.e., slow system dynamics). The author considers it a very substantial advancement if future researchers can develop a theoretical framework for an accurate (i.e. low false-alarms and misses) classification of gas turbine AE signals in dynamic applications such as military jet engines.

Bibliography

[1] R. M. Douglas, S. Beugn, M. D. Jenkins, A. K. Frances, J. A. Steel, R. L. Reuben and P. A. Kew. Monitoring of Gas Turbine Operating Parameters Using Acoustic Emission, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK., Bridge of Don, Aberdeen, UK., 2004

[2] Armor, A. F. Graham, L. J. Frank, Acoustic Emission Monitoring of Steam Turbines. Joint ASME/IEEE Power Generation Conference, St. Louis, USA, Oct. 4-8, 1981.

[3] I. Sato, Rotating Machinery Diagnosis with Acoustic Emission Techniques. Electrical Engineering in Japan, Vol. 100 (2), Pages 115-127, 1990.

[4] D. Mba, and L.D. Hall, The Transmission of Acoustic Emission Across Large-Scale Turbine Rotors. NDT and E International, Vol. 35 (8), Pages 529-539, 2002.

[5] C. Zuluaga-Giraldo, D. Mba, M. Smart, Acoustic Emission During Run-Up and Run-Down of a Power Generation Turbine, Tribology International, Pages 415-422,

Vol. 37, May 2004.

[6] M. S. Nashed, J. A. Steel and R. L. Reuben, The Use of Acoustic Emission for the Condition Assessment of Gas Turbines: Acoustic Emission Generation from Normal Running. Journal of Process Mechanical Engineering, Vol. 228, Issue-4, Pages 286-308, 2014.

[7] A. Purarjomandlangrudi, and G. Nourbakhsh, Acoustic Emission Condition Monitoring: An Application For Wind Turbine Fault Detection, Queensland University of Technology (QUT), George St Brisbane, QLD 4000, May 2013.

[8] He Kuanfang, Wu. Jigang, W. Guangbin, Time-Frequency Entropy Analysis of Alternating Current Square Wave Current Signal in Submerged Arc Welding, Journal of Computers, Vol. 6, No. 10, Oct. 2011

[9] S. A. Farhat , M. K. Al-Taleb, Combustion Oscillations Diagnostics in a Gas Turbine Using an Acoustic Emissions, Jordan Journal of Mechanical and Industrial Engineering, Vol. 4, No. 3, June 2010

[10] A. Nairac, T. Corbett-Clark, R. Ripley, N. Townsend, L. Tarassenko, Choosing an Appropriate Model for Novelty Detection. In IEE 5th International Conference on Artificial Neural Networks, 1997.

[11] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A.S. Moodie, Q. Zhu, Acoustic emission-Based Condition Monitoring Methods: Review and Application for Low

Speed Slew Bearing, Mechanical Systems and Signal Processing, Volumes 72?73, Pages 134-159, May 2016.

[12] D. Mba and R. B. K. N. Rao, Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating Structures, The Shock and Vibration Digest Vol. 38 (1), Pages 3-16, 2006.

[13] P. R. Aguiar, C.H.R. Martins, M. Marchi, E.C. Bianchi, Digital Signal Processing for Acoustic Emission, Data Acquisition Applications, DOI: 10.5772/48557, August 23, 2012.

[14] Y. H. Kim, A.C.C. Tan, B.S. Yang, Parameters Comparison of Acoustic Emission Signals for Condition Monitoring of Low-Speed Bearings, Australian Journal of Mechanical Engineering, Vol. 6, Issue-1, 2008

[15] A.K.S. Jardine, D. Lin, D. Banjevic, A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance, Mechanical Systems and Signal Processing, Vol. 20, Issue-7, Pages 1483-1510, October 2006.

[16] LJ. Yu, DJ Cleary, PE. Cuddihy, A Novel Approach to Aircraft Engine Anomaly Detection and Diagnostics. In 19th Conference of Uncertainty in Artificial Intelligence, 2003.

[17] W. Kaewwaewnoi, A. Prateepasen, P. Kaewtrakulpong, Measurement of Valve Leakage Rate using Acoustic Emission, King Mongkut's University of Technology Thonburi, Thailand, May 12-13 2005.

[18] N. Jamaludin, D. Mba, Monitoring Extremely Slow Rolling Element Bearings: Part I, NDT and E International, Vol. 35, Issue-6, Pages 349-358, September 2002.

[19] N. Jamaludin, D. Mba, Monitoring Extremely Slow Rolling Element Bearings: Part II, NDT and E International, Vol. 35, Issue-6, Pages 359-366, September 2002.

[20] K.R. Al-Balushi, A. Addali, B. Chamley, D. Mba, Energy Index Technique for Detection of Acoustic Emissions Associated with Incipient Bearing Failures, Applied Acoustics, Vol. 71, Issue-9, Pages 812-821, September 2010.

[21] P. He, P. Li, H. Sun, Feature Extraction of Acoustic Signals Based on Complex Morlet Wavelet, Procedia Engineering, Vol. 15, Pages 464-468, 2011.

[22] C. J. Li, S. Y. Li, Acoustic Emission Analysis for Bearing Condition Monitoring, Wear, Vol. 185, Issues 1-2, Pages 67-74, June 1995.

[23] B. S. Kim, D. S. Gu, J. G. Kim, Y. C. Kim, B. K. Choi, Rolling Element Bearing Fault Detection using Acoustic Emission Signal Analyzed by Envelope Analysis with Discrete Wavelet Transform, Proceedings of the 4th World Conference on Engineering Asset Lifecycle Management, Pages 596-602, Greece, Sep. 28-30, 2009.

[24] R. Ganguli, I. Chopra, DJ. Haas, Helicopter Rotor System Health Monitoring Using Numerical Simulations and Neural Networks, AIAA Modeling and Simulation Technologies Conference, New Orleans, USA, April 29, 1997

[25] S.A. Niknam, T. Thomas, J.W. Hines, R. Sawhney, Analysis of Acoustic Emission Data for Bearings Subject to Unbalance, International Journal of Prognostics and Health Management, ISSN 2153-2648, 015, 2013.

[26] H. Oh, M.H. Azarian, M. Pecht, Estimation of Fan Bearing Degradation using Acoustic Emission Analysis and Mahalanobis Distance, Proceedings of the Technical Program for MFPT: The Applied Systems Health Management Conference, 2011.

[27] S.A. Niknam, V. Songmene, Y.H.J. Au, Proposing a New Acoustic Emission Parameter for Bearing Condition Monitoring in Rotating Machines, CSME-87 No. 12, E.I.C. Accession 3407, February 2013.

[28] M. Ong, X. Ren, G. Allan, V. Kadirkamanathan, HA. Thompson, PJ. Fleming, Decision Support System on the Grid, International Conference on Knowledge-Based Intelligent Information and Engineering Systems, New Zealand, 2004.

[29] X. Liu, X. Wu, C. Liu, A Comparison of Acoustic Emission and Vibration on Bearing Fault Detection, Proceedings of the International Conference on Transportation, Mechanical and Electrical Engineering (TMEE), China, Page 922, December 16, 2011.

[30] HR. DePold, F.D. Gass, The Application of Expert Systems and Neural Networks to Gas Turbine Prognostics and Diagnostics, ASME Journal of Engineering for Gas Turbines and Power, Page 116, 1994.

[31] S. Al-Dossary, R.I.R. Hamzah, D. Mba, Observations of Changes in Acoustic Emission Waveform for Varying Seeded Defect Sizes in a Rolling Element Bearing, Applied Acoustics, Vol. 70, Issue-1, Pages 58-81, January 2009.

[32] S. Simani, Idenfitication and Fault Diagnosis of a Simulated Model of an Industrial Gas Turbine. IEEE Transactions on Industrial Informatics, Vol. 1, Issue-3, Aug. 2005.

[33] J. Antoni, Cyclostationarity by Examples, Mechanical Systems and Signal Processing, Vol. 23(4), Pages 987-1036, May 2009.

[34] B. Eftekharnejad, M.R. Carrasco, B. Charnley, D. Mba, The Application of Spectral Kurtosis on Acoustic Emission and Vibrations from a Defective Bearing, Mechanical Systems and Signal Processing, Vol. 25, Issue-1, Pages 266-284, January 2011.

[35] W. B. Ng, E. Clough, K. J. Syed *et al.*, The Combined Investigation of the Flame Dynamics of an Industrial Gas Turbine Combustor using High-Speed Imaging and an Optically Integrated Data Collection Method, Measurement Science and Technology, Vol. 15, No. 11, October 2004.

[36] G.D. Neill, R.L. Reuben, P.M. Sandford, *et al.*. Detection of Incipient Cavitation in Pumps using Acoustic Emission. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, November 1, 1997.

[37] C. Booth, J.R. McDonald, P. Donald, N. Lines, N. Cooke, C. Smith, Model-Based Condition Monitoring of Gas Turbine Engines for Power Generation Duty, IEEE Power Engineering Review, 2001.

[38] T. Rus, M. Dular, S. Irok, M. Hocevar, I. Kern, An Investigation of the Relationship Between Acoustic Emission, Vibration, Noise, and Cavitation Structures on a Kaplan Turbine, ASME, Journal of Fluids Engineering, Vol. 129, Issue-9, April 2, 2007.

[39] A. Addali, S. Al-Lababidi, H. Yeung, D. Mba, F. Khan, Acoustic Emission and Gas-Phase Measurements in Two-Phase Flow. Proc IMechE, Vol. 224, Part E: Journal of Process Mechanical Engineering, 2010.

[40] P. R. Spina, G. Torella, and M. Venturini, The Use of Expert Systems for Gas Turbine Diagnostics and Maintenance, ASME Turbo Expo 2002: Power for Land, Sea, and Air, Vol. 2, Parts A and B, Amsterdam, The Netherlands, June 3-6, 2002.

[41] J. Austin, J. Kennedy, K. Lees, The Advanced Uncertain Reasoning Architecture, AURA. In Proceedings of the Weightless Neural Network Workshop, August 1995.

[42] A. Smulders, and C. Loob, Machine Condition Monitoring using Multi-Parameter Measurements, Proceedings of COMADEM 94 International Congress. Tata McGraw-Hill Publishing Co. India, Pages 147-152, 1994.

[43] TC. Lieuwen, Investigation of Combustion Instability Mechanisms in Premixed Gas Turbines. PhD thesis, Georgia Institute of Technology, 1999.

[44] O.J.I Major, F.L. Greitzer, LJ. Kangas, TJ. Reeve, An Artificial Neural Network System for Diagnosing Gas Turbine Engine Fuel Faults, In 48th Meeting of Mechanical Failures Prevention Group (MFPG 48), in Wakefield, MA, 19-21 April 1994.

[45] R. Isermann, Fault-Diagnosis Systems, An Introduction from Fault Detection to Fault Tolerance, Springer Science and Business Media, Oct.13, 2005.

[46] L.A. Garca-Escudero, O. Duque-Perez, D. Morinigo-Sotelo, M. Perez-Alonso, Robust Condition Monitoring for Early Detection of Broken Rotor Bars in Induction Motors, Expert Systems with Applications, Vol. 38, Issue-3, Pages 2653-2660, March 2011.

[47] R.R. Schoen, T.G. Habetler, Effects of Time-Varying Loads on Rotor Fault Detection in Induction Machines, IEEE Transactions on Industry Applications, Journal Vol. 31, Issue-4, Paper presented at the 1993 IEEE Industry Applications Society Annual Meeting, Toronto, Ontario (CA), October 2-7, 1993.

[48] R. G. T. de Almeida, S. Alexandra da Silva Vicente, L. Rodrigues Padovese, New Technique for Evaluation of Global Vibration Levels in Rolling Bearings, Shock and Vibration, Vol. 9, Issue 4-5, Pages 225-234, 2002.

[49] Z. Liu, X. Yin, Z. Zhang, D. Chen, W. Chen, Online Rotor Mixed Fault Diagnosis Way Based on Spectrum Analysis of Instantaneous Power in Squirrel Cage Induction Motors, IEEE Transactions on Energy Conversion, Vol. 19, Issue-3, Pages 485-490, 24 August 2004.

[50] M.H. El-Ghamrya, L. Reubenj, A. Steel, The Development of Automated Pattern Recognition and Statistical Feature Isolation Techniques for the Diagnosis of Reciprocating Machinery Faults Using Acoustic Emission, Mechanical Systems and Signal Processing, Vol. 17, Issue-4, Pages 805-823, July 2003.

[51] M. El-Ghamrya, J.A. Steel, R.L. Reubena, T.L. Fog, Indirect Measurement of Cylinder Pressure From Diesel Engines using Acoustic Emission, Mechanical Systems and Signal Processing, Vol. 19, Issue-4, Pages 751-765, July 2005.

[52] M.H. Elghamry, E. Brown, I. G. Ferguson, J. D. Gill, R. L. Reuben, J. A. Steel, M. Scaife , S. Middleton, Gaseous Air-Fuel Quality Identification for a Spark Ignition Gas Engine using Acoustic Emission analysis. Proceedings of COMADEM, Pages 235-244, Monash University, Australia, 1998.

[53] D.M.J. Tax, A. Ypma, R. P. W. Duin, Support Vector Data Description Applied to Machine Vibration Analysis, Machine Learning, Vol. 54, Issue-1, Pages 45-66, 2004.

[54] J. D. Gill, E. R. Brown, M. Twite, G. Horner, R. L. Reuben, J. A. Steel, Monitoring of a Large Reciprocating Compressor. Proceedings of COMADEM Pages 317-326, Monash University, Australia, 1998.

[55] T.J. Holroyd, N. Randall, Field Application of Acoustic Emission to Machinery Condition Monitoring, Proceedings of COMADEM, University of West of England, Bristol. 217, 1993.

[56] C. Trammel, G. Vossler, M. Feldmann, UK Ministry of Defence Generic Health and Usage Monitoring System, Aircraft Engineering and Aerospace Technology, Vol. 69 Issue-5, Pages 414-422, 1997.

[57] N. Ho, P. Lozano, R. Mangoubi, M. Martinez-Sanchez, A Model-Based Vehicle Health Monitoring System for the Space Shuttle Main Engine, 34th AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Cleveland, Ohio, USA, 1998.

[58] J. Z. Sikorska and M. Hodkiewicz, Comparison of Acoustic Emission, Vibration and Dynamic Pressure Measurements for Detecting Change in Flow Conditions on a Centrifugal Pump, 18th International Congress and Exhibition on Condition

Monitoring and Diagnostic Engineering Management, Cranfield, UK, August 31 and 2nd September 2005.

[59] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.

[60] Wikipedia, Acoustic Emission.

[61] O. Av, S. Av, Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1999.

[62] L. S. Law, J. H. Kim, Willey Y. H. Liew, S. K. Lee, An Approach Based on Wavelet Packet Decomposition and Hilbert-Huang Transform (WPD HHT) for Spindle Bearings Condition Monitoring, Processing, Vol. 33, Pages 197-211, November 2012.

[63] Z.K. Peng, P. W. Tse, F. L. Chu, An Improved Hilbert-Huang Transform and its Application in Vibration Signal Analysis, Journal of Sound and Vibration, Vol. 286, Issues 1-2, Pages 187-205, August 23, 2005.

[64] P. W. Tse, W. X. Yang, H. Y. Tam, Machine Fault Diagnosis Through an Effective Exact Wavelet Analysis, Journal of Sound and Vibration, Vol. 277, Issues 4-5, Pages 1005-1024, November 5, 2004.

[65] Alexander D. Poularikas, The Transforms and Applications Handbook, 3rd Edition, by CRC Press, January 19, 2010.

[66] N. E. Huang, Z. Shen, S. R. Long, M. Wu. H. Shih, N. Zheng, C. Yen, C.C. Tung, H.H. Liu, The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis, Proceedings of the Royal Society of London Series, Mathematical Physical and Engineering Sciences, March 8, 1998.

[67] J. P. Eckmann and D. Ruelle, Ergodic Theory of Chaos and Strange Attractors, Reviews of Modern Physics, Vol. 57, Issue-3, Pages 617-656, July 1985.

[68] A. Wolf, J. B. Swift, H. L. Swinney, J. A. Vastano, Determining Lyapunov Exponents from a Time Series, Elsevier Science Publishers, Physica 16D, Pages 285-317, 1985.

[69] P.F. Pai, Nonlinear Vibration Characterization by Signal Decomposition, Journal of Sound and Vibration, Vol. 307, Issues 3-5, Pages 527-544, November 6, 2007.

[70] W. Caesarendra, B. Kosasih, A. K. Tieu, C. A. S. Moodie, Application of the Largest Lyapunov Exponent Algorithm for Feature Extraction in Low Speed Slew Bearing Condition Monitoring, Mechanical Systems and Signal Processing, Vol. 50?51, Pages 116-138, January 2015.

[71] Hassan K. Khalil, Stability of Perturbed Systems, Non Linear Systems, Third Edition, Dec. 18, 2001.

[72] Hassan K. Khalil, Input to State Stability, Non Linear Systems, Third Edition, Dec. 18, 2001.

[73] M. Kedadouche, M. Thomas , A. Tahan, R. Guilbault, Monitoring Gears by Vibration Measurements: Lempel-Ziv Complexity and Approximate Entropy as Diagnostic Tools, Proceedings of the Analyse Vibratoire Exprimentale (AVE), Nov. 2014.

[74] A. Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, Third Edition, Pearson Prentice Hall, 2008.

[75] G. M. Jenkins, D. G. Watts, Spectral Analysis and Its Applications, June 1968.

[76] D. Clifton, Condition Monitoring of Gas Turbine Engines, PhD Thesis, Oxford University, January 2006.

[77] R. Willett, Lecture Notes on Statistical Signal Processing, University of Wisconsin.

[78] Gail A. Carpenter, S. Grossberg, ART-1 Network Architecture, Applied Optics, Vol. 26 No. 23, Dec. 1987.

[79] MATLAB Mathworks, https://www.mathworks.com/help/dsp/ug/multirate-filters.html

[80] G. A. Carpenter, S. Grossberg, Applied Optics Vol. 26 No. 23 Dec. 1987.

[81] International Standards Organization (ISO), DIS Standard 22096, Condition Monitoring and Diagnosis of Machines Acoustic Emission, 2007.

[82] J. Z. Sikorska, D. Mba, Challenges and Obstacles in the Application of Acoustic Emission to Process Machinery, Proceedings In Mechanical Engineering, Vol. 222, Part E: Journal of Process Mechanical Engineering, 2008.

[83] J. Z. Sikorska, The Application of Acoustic Emission Monitoring to the Detection of Flow Conditions in Centrifugal Pumps, PhD Thesis, University of WA, 2006.

[84] J.C. Mossing, T. A. Tuthill, Reduced Interference Distributions for the Detection and Classification of Outside Sound Source Acoustic Emissions, IEEE International Conference on Acoustics, Speech and Signal Processing, Pages 2758-2761, Atlanta, Georgia, USA, 1996.

[85] A. Terci, Y.H. Au, Journal of Acoustic Emission Signal Processing, Measurement and Control, Vol. 34, Pages 240-244, 2001.

[86] J.D. Wu, J.C. Chen, Continuous Wavelet Transform Technique for Fault Signal Diagnosis of Internal Combustion Engines, NDT and E International, Vol. 39, Issue-4, Pages 304-311, 2006.

[87] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault Diagnosis of Rotating Machinery Based on a New Hybrid Clustering Algorithm, Int Joint Adv. Manufacturing Technology, Issue-35, Pages 968-977, 2008.

[88] A. Jawadekar, S. Paraskar, S. Jadhav, G. Dhole, Artificial Neural Network Based Induction Motor Fault Classifier Using Continuous Wavelet Transform, Systems Science and Control Engineering, Vol. 2, Pages 684-690, Oct 2014.

[89] A. K.S. Jardine, D. Lin, D. Banjevic, Wavelets for Fault Diagnosis of Rotary Machines, A Review with Applications, Mechanical Systems and Signal Processing, Vol. 20, Pages 1483-1510, 2006.