Signal Reconstruction from Partial or Modified Linear Time-Frequency Representations

David Manuel Baptista Lopes

Master Copy

A thesis submitted for the degree of

Doctor of Philosophy

University of Southampton

Faculty of Engineering and Applied Science

Institute of Sound and Vibration Research

February 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE

INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

SIGNAL RECONSTRUCTION FROM PARTIAL OR MODIFIED LINEAR TIME-FREQUENCY REPRESENTATIONS

By David Manuel Baptista Lopes

The changing spectral qualities of a non-stationary 1D signal can be analysed using 'Time-Frequency Representations'. TFRs are real or complex 2D functions of time and frequency. The aim of this thesis is to explore both traditional and signal dependant TFRs, with particular emphasis on the potential of a given representation to be inverted to reconstruct the original time series. Invertibility is the ability to reconstruct a 1D signal from a given TFR. Typical methods of inversion assume that the TFR to be inverted is valid, that is the TFR has an exact inverse. Two commonly used signal dependant Time-Frequency methods, Cohen-Posch and the Reassigned approach, modify a TFR to produce a new distribution, which conforms to a set of desirable characteristics. The first section of this thesis reviews such signal dependant techniques and describes a method of inversion for the Reassigned Spectrogram through use of its link to the phase of the Short Time Fourier Transform (ST-FT).

In order to achieve this inversion, signal reconstruction is required using only knowledge of the phase of the ST-FT. Such an inversion is shown to be possible to within an overall amplitude constant given mild conditions upon the degree of overlap used in the computation of the ST-FT. Complimentary to signal reconstruction from knowledge of the phase of the ST-FT, is reconstruction from its magnitude. Two distinct approaches are described. The first previously given in the literature, iteratively applies the Minimum Least Squares (MLS) inversion of the complete (amplitude and phase) ST-FT. The second uses an analogous approach to that used for signal reconstruction from phase. Although the first approach proves more robust to non-valid input TFR, both of these techniques require a similar degree of overlap to the phase case, and both reconstruct the signal to within an overall phase constant. The description of the theory concludes with a discussion of the set of Generalised Wavelet Transforms (GWTs), of which both the ST-FT and WT are members. After defining the set of GWTs, descriptions are given for the MLS inversion for complete GWT information, and for signal reconstruction from either phase or amplitude.

The thesis concludes by using the MLS based technique to create signals from modified or synthetic spectrograms generated using heart sounds. The first application of this is to extend the duration of heart sounds. Temporal extension in this fashion has the ability to extend the signals in time without affecting the spectrum of the sound. In addition, it does not require the use of a model as in matching pursuit based methods described in the literature. The second application is to create a time-series from a synthetic spectrogram constructed by averaging the spectrograms of a patient's heart murmur. Such averaging cannot take place in the time-domain owing to the random nature of the flow noise in a murmur.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to thank my supervisor Dr. P. R. White for his guidance and support throughout my research in the ISVR, and Prof. T. Leighton for the moral support he provided during the research for this thesis. I would also like to thank Prof. J. K. Hammond for introducing me to Time-Frequency analysis and Dr. T. Leung for introducing me to the field of heart sounds.

On a personal note, I would like to thank my family, who have supported me all the way through my academic career. Finally, I would like to thank my wife-to-be Louise, for always being there to listen to me and for her optimism when all I could muster was pessimism.

Abbreviations

Abbreviation	*Full Title*
AF	Ambiguity Function
AP	Ambiguity Plane
ASD	Atrial Septal Defects
CC	Cohen's Class
CW	Choi-Williams TFR
DFT	Discrete Fourier Transform
DC	Direct Current
FFT	Fast Fourier Transform
FM	Frequency Modulated
FS	Sampling Frequency
FT	Fourier Transform
GD	Group Delay
GWT	Generalised Wavelet Transform
Hz	Hertz
IDFT	Inverse Discrete Fourier Transform
IF	Instantaneous Frequency
LHS	Left Hand Side
MLS	Minimum Least Squares
MSE	Mean Squared Error
POCS	Projection onto Convex Sets
PWD	Pseudo Wigner Distribution
RHS	Right Hand Side
RPM	Revolutions per Minute
S1	First Heart Sound
S2	Second Heart Sound
Sec	Seconds
ST-FT	Short Time Fourier Transform
SVD	Singular Value Decomposition
TF	Time-Frequency
TFR / D	Time-Frequency Representation / Distribution
VSD	Ventricular Septal Defects
WD	Wigner Distribution
WT	Wavelet Transform

Chapter 1 Introduction

Signal processing can be considered as the act of taking a signal and transforming it to reveal previously unknown information. In its most general form, no conditions should be placed upon either the type of input signal, or the nature of the information extracted. For example, tasting (transform) an apple (signal) to see if it is ripe (information) would be one example, just as a Fourier transform (transform) of the output from an accelerometer (signal) might reveal if a vibration level is too high (information).

Typically the signals examined throughout this thesis are 1D time series. Examples of such signals include those obtained from a microphone or an accelerometer. Signals can be classified as either stationary or non-stationary. Statistical properties of stationary signals do not vary with time, as opposed to those of non-stationary signals, which do. One example of a stationary signal is the sound generated by a rotating fan having constant RPM. The blade passing frequency of the fan does not alter from moment to moment. However the sound is measured as the fan rotates from rest up to its operating speed is not stationary, since the changing blade speed results in a temporally changing auto-correlation function. A more formal definition of stationarity is given in Chapter 2 Section 2.2.7.

In the process of extracting information from a signal, the stationarity of the signal, needs to be considered. Continuing with the fan example, if the information to be extracted is the blade passing frequency, then if the fan is rotating at constant RPM one number fully describes the required information. However a function of time is required to describe the blade passing frequency as the fan spins up. In general for non-stationary signals, any extracted information requires an additional time variable to describe the temporal changes.

The increasing performance of the modern digital computer has meant that more and more complex signal processing algorithms can now be applied to a wider range of signals. One of the major signal processing techniques to benefit from this

increased computation resource is Time-Frequency (TF) analysis. This technique decomposes a 1D signal into a 2D image showing the changing spectral content present in the signal.

This thesis will investigate a number of different TF approaches each with their own benefits and limitations. Ultimately in a practical sense all these methods are limited by Heisenburg's uncertainty principle, which states that the more precisely a signal is defined in time, the less precisely it is defined in frequency. Beginning with an introduction to spectral and non-stationary signal analysis, this thesis reviews the key ideas behind TF analysis, beginning with the Short-Time Fourier Transform (ST-FT) (for example [Portnoff80, Cohen95, Hammond96]). The ST-FT is one of two currently popular linear TFRs, the other being the Wavelet Transform (WT) [Rioul91]. Although linear, these functions are usually complex valued. Thus in order to display these functions simply the squared modulus (called the Spectrogram for the ST-FT and the Scalogram for WT) is usually plotted. The Spectrogram is part of a wider set of bilinear TFRs referred to as Cohen's Class (CC). Over recent years the popularity of CC has increased with ever more sophisticated kernel functions, enabling the TFR to have a number of user-defined characteristics.

One of the limitations of CC of TFRs is that they do not alter their characteristics to match those present in the signal. As such these distributions are called signal independent TFRs. Chapter 3 investigates current signal dependant methods, which attempt to improve the performance of CC of TFR [Loughlin95a-b, Jones95]. One data adaptive technique that has grown in popularity recently is the 'reassigned' approach [Kodera76-78, Moss94, Auger95]. Although originally developed in the late 1970's using the ST-FT (called the reassigned spectrogram), the method has been extended to all members of CC.

One highly desirable property for any TFR to posses is invertibility, enabling the reconstruction of a signal from the TFR. Mathematically simple inverses exist for all the signal independent approaches [Cohen95]. One goal of this work is to develop such an inverse for the reassigned spectrogram. The computation of the reassigned spectrogram can be achieved in a number of different ways, with the original method [Kodera76] using the difference between successive phase points in the ST-FT.

In Chapter 4 the link between the phase of the ST-FT and the reassigned spectrogram is exploited to allow reconstruction of real signals from their reassigned co-ordinates.

In order to achieve this, an algorithm is developed for reconstructing a signal purely from the phase of its ST-FT. This inversion is achieved to within an overall amplitude constant.

Although inversions exist for members of CC, their usefulness is limited. These approaches assume that the TFR to be inverted is 'valid'. The concept of validity in terms of TFRs is now defined. A TFR is a 2D function created via a TF transform of a 1D function. For a TFR to be valid it must be the result of the application of a TF transform to some 1D signal [Qian96]. In general, an arbitrary 2D function is not a valid TFR since there exists no 1D signal, which could have created it via a TF transform. Hence, inversion of an invalid TFR can only be performed in an approximate manner. This leads to an optimisations problem. Direct application of the CC inversion formulae to invalid TFRs generally results in sub-optimal solutions.

The question is now posed: Why would we wish to recover a signal from a TFR which has been altered? The TF plane offers the possibility of easily defining time-varying filtering operations, simply by modifying a signal's TFR. Any modifications to a valid TFR almost certainly render it invalid. Robust procedures are required which work well upon invalid input TFRs. The alteration of the TF to create a new signal with 'desirable' characteristics has been considered by many authors in the literature for a range of different TFRs [Krattenthaler93, Hlawatsch94, 91, Griffin83, Portnoff80, etc.]. Operations such as the masking out of unwanted signal components, or alteration of the time or frequency support of a signal are examples of such alterations.

Chapter 5 focuses upon the problem of signal reconstruction from the magnitude of the ST-FT, the dual problem of that discussed in Chapter 4. Griffin *et al.* [Griffin83-84] describe a method of signal reconstruction based upon the Minimum Least Squares (MLS) algorithm for the ST-FT. This technique is robust to invalid spectrograms, and can therefore be used as the basis for time-varying filtering. Other previously described techniques are reviewed, and a new approach to the problem (based upon a quadratic form) is presented. This new approach is compared to previously described algorithms.

With a view to generalising the approaches given in Chapter 4 and 5 to a wider set of linear TFRs, Chapter 6 presents the set of Generalised Wavelet Transforms (GWTs) of which both the ST-FT and WT are members. It then proceeds to develop the

techniques for signal reconstruction from either phase or modulus, as well as a MLS algorithm for signal reconstruction from full (complex) representation.

Chapter 7 demonstrates an application for the methods presented in earlier chapters. Two different applications are presented, the temporal extension of heart sounds (without altering its spectral content) and the construction of a typical murmur sound from an averaged spectrogram. Both applications use the MLS algorithm for the spectrogram (reviewed in Chapter 5) as the basis for the reconstruction.

Chapter 2 Classical Time-Frequency Analysis

2.1 Introduction

This chapter begins with a brief description and discussion of the Fourier Transform (FT). This function transforms a 1D signal (typically in the time-domain) to the frequency domain, allowing its spectral content to be computed. The FT uses the whole of the signal to construct the spectrum of the signal. The use of the FT for non-stationary signals is limited since it does not easily provide information regarding the time at which a given frequency component occurred. In order to compute the changing signal characteristics a decomposition of the signal in terms of both time *and* frequency is required. Such a decomposition is called a Time-Frequency Representation (TFR). The most basic linear TFR is a simple extension of the FT, called the Short-Time FT (ST-FT). Basic theory and ideas surrounding the ST-FT and another linear TFR, the Wavelet Transform (WT) are presented. The WT differs from the ST-FT by decomposing the signal in scale rather than frequency. The squared magnitudes of the ST-FT and WT, called the Spectrogram and Scalogram respectively, are both 'bilinear' TFRs. The chapter also describes a wider set of bilinear TFRs called Cohen's Class (CC), of which the spectrogram is a member. The chapter concludes with an examination of CC of TFRs, including the relationship between members of the set and their limitations.

2.2 Fourier Transform

2.2.1 Introduction

The Fourier Transform (FT) is one of the most fundamental of all signal processing techniques. It is the decomposition of a signal in terms of oscillating harmonic waves. Informally, the FT allows the computation of the spectral or frequency content of a signal. Such a decomposition is useful because it can provide information regarding the process used to create the signal, for example the blade passing frequency of a fan, or the resonant frequency of a vibrating plate. Beginning with the definition of the continuous form of the FT, this section then proceeds to

develop the discrete FT by discretising first time, and then both time and frequency variables.

2.2.2 Continuous Fourier Transform

Consider a 1-D signal (real or complex) denoted $x(t)$, the equation for its FT is given in equation (2-1a), where t and ω are the continuous time (in seconds) and frequency (in radians per second) variables respectively. The link between ω and frequency f in Hertz, is given by $\omega = 2\pi f$. The function $X(\omega)$ is called the FT of the input signal $x(t)$. (All integrals are taken over the limits $-\infty, \infty$ unless otherwise stated).

There are a number of alternative definitions of the FT, all simply related to one another. The FT is a one-to-one mapping from the time to frequency domain. The uniqueness of the transform for a given signal means that the transform can be inverted (2-1b).

$$X(\omega) = \frac{1}{\sqrt{2\pi}} \int x(t) e^{-j\omega t} dt = F[x(t)] \tag{2-1a}$$

$$x(t) = \frac{1}{\sqrt{2\pi}} \int X(\omega) e^{j\omega t} dt = F^{-1}[X(\omega)] \tag{2-1b}$$

The measure of the energy in the signal, as defined by (2-2), is the same, regardless of whether energy is measured in the time or frequency domain. This relation is called Parseval's or Rayleigh theorem [James95].

$$E = \int |x(t)|^2 dt = \int |X(\omega)|^2 d\omega \tag{2-2}$$

The FT is a decomposition of an input signal in terms of a set of complex exponential harmonic wave basis functions. Each of these complex exponentials can be thought of as a filter possessing an infinitely narrow bandwidth, centred at the frequency ω, and acting over the whole of the signal. The FT integrates the output of the each of these filters. The squared magnitude of the these filters is the same as the squared magnitude of the FT (2-1a) and is called the spectrum.

2.2.3 Properties of the Continuous Fourier Transform

2.2.3.1 Windowing and Convolution Effects

Consider the signal given in (2-3). This signal is the result of the multiplication of two signals in the time domain. In the frequency domain, the result is that the signal's FT are 'convolved' as defined in (2-4) [James95]. This process in called 'convolution'. A similar relation also holds with convolution in the time domain, resulting in a multiplication in the frequency domain.

$$x(t) = y(t) \cdot z(t) \tag{2-3}$$

$$X(\omega) = \int Y(\gamma) Z(\omega - \gamma) d\gamma = Y(\omega) * Z(\omega) \tag{2-4}$$

To demonstrate the effect of multiplication in the time domain, consider the windowed pure harmonic signal as given in (2-5).

$$x(t) = e^{j\omega_0 t} . h(t) \tag{2-5}$$

This signal is multiplied by two different windowing functions and their effect upon the frequency domain is considered. In this example, a Hanning and rectangular windows are used. The time and frequency domain versions of these windows are given in Table 1. The definition of both of these windowing functions has an additional parameter, T_0. T_0 is used to tailor the support of each of the functions in the time domain. Altering this parameter affects both the time and frequency domain versions of the windowing function. In the following examples T_0 is set to $T_0 = 2$ for both windowing functions.

Figure 1 shows the plot of the real and imaginary parts of the complex tone ($\omega_0 = 15$). The sharp start and stop effect of rectangular window can be seen at –1 and 1 seconds. As described by equation (2-4), multiplication in the time domain results in convolution in the frequency domain. The FT of a complex tone is a Dirac delta function and thus the convolution operation simply shifts the FT of the windowing so that it is centred at the fundamental frequency of the tone, ω_0. This is shown in Figure 2, where the FT of the rectangular window function is shifted to be centred at 15. Figure 3 show the same harmonic windowed using the Hanning window, the result in the frequency being given in Figure 4. Comparing FT versions

of the signals, it can be seen that the main lobe (the part closed to $\omega = 15$) in Figure 2 is narrower than the main lobe seen in Figure 4. Conversely, the amplitude of Figure 4 rolls off much more quickly with increased frequency as compared to Figure 2. Mathematically the -3dB bandwidth of the main lope for a rectangular window is $0.88/T_0$, whereas for a Hanning window is it $1.4/T_0$ [White82]. The asymptotic roll off of a rectangular window is 6dB per octave as compared to the Hanning window which rolls off at 18dB per octave [White82]. It follows therefore that if a signal needs to be windowed prior to analysis, as for example when using the Short-Time FT (see Section 2.3.2.2) careful consideration should be given to the type and size of the windowing function used.

Window Type	Time Domain Version	Frequency Domain Version
Rectangular	$h(t) = \begin{cases} 1 & -\frac{T_0}{2} > t > \frac{T_0}{2} \\ 0 & \text{otherwise} \end{cases}$	$H(\omega) = \frac{\sqrt{\frac{2}{\pi}} \sin\left(\frac{T_0 \omega}{2}\right)}{\omega}$
Hanning	$h(t) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{2\pi t}{T_0}\right)\right] & -\frac{T_0}{2} > t > \frac{T_0}{2} \\ 0 & \text{otherwise} \end{cases}$	$H(\omega) = \frac{T_0}{2}\left(\frac{\sin\left(\frac{\omega T_0}{2}\right)}{\left(\frac{\omega T_0}{2}\right)\left(1 - \left(\frac{T\omega}{2\pi}\right)^2\right)}\right)$

Table 1 Time and Frequency Domain definitions of Hanning and Rectangular Windows

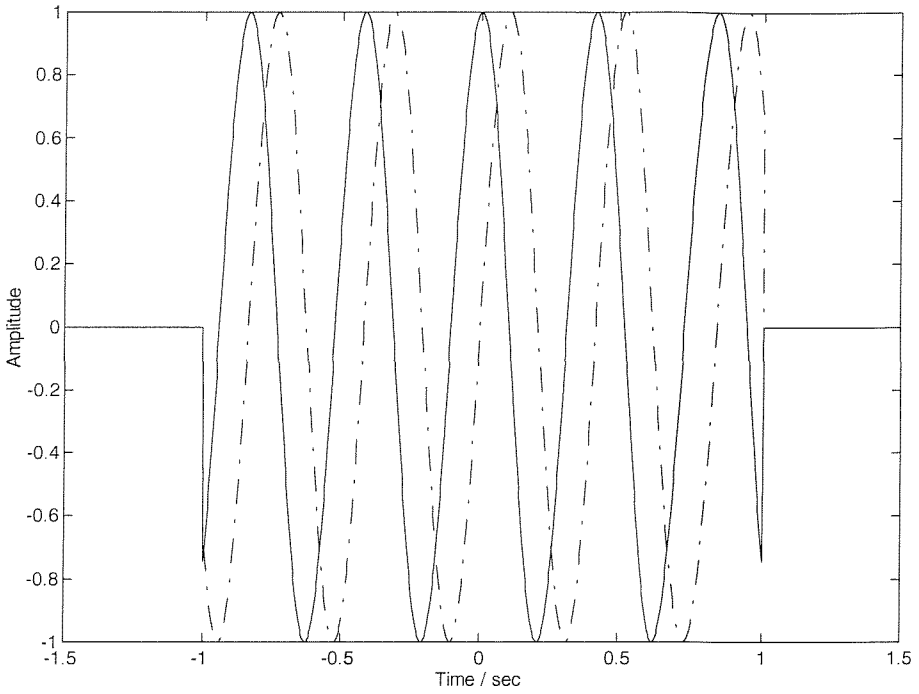

Figure 1 Complex Harmonic Windowed by a Rectangular Window (Real Part Solid Line, Imaginary Part Dashed Line)

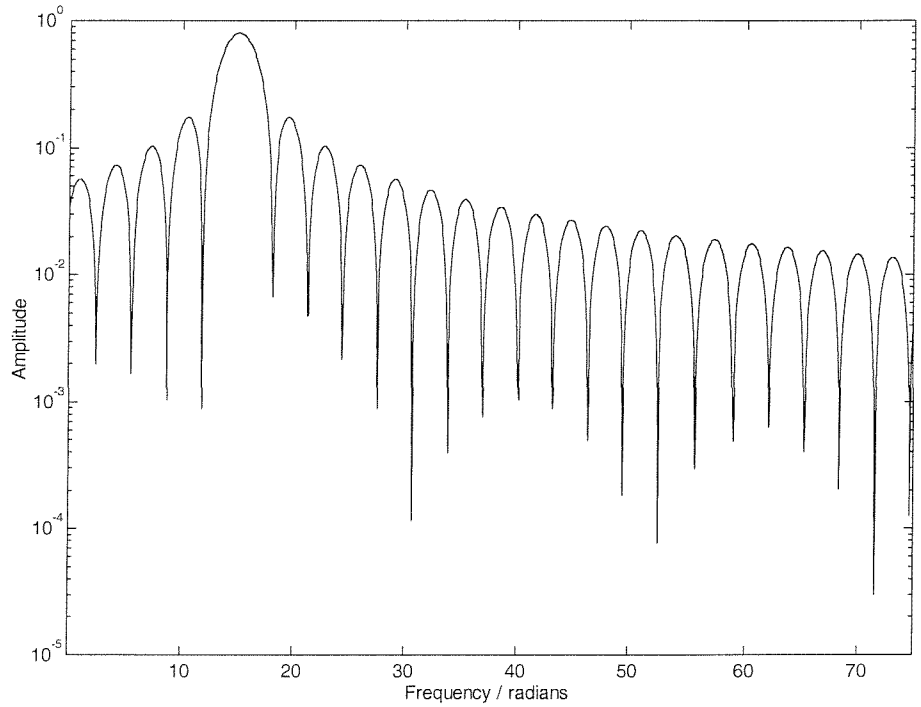

Figure 2 The Amplitude of the FT of a Rectangular Windowed Complex Harmonic

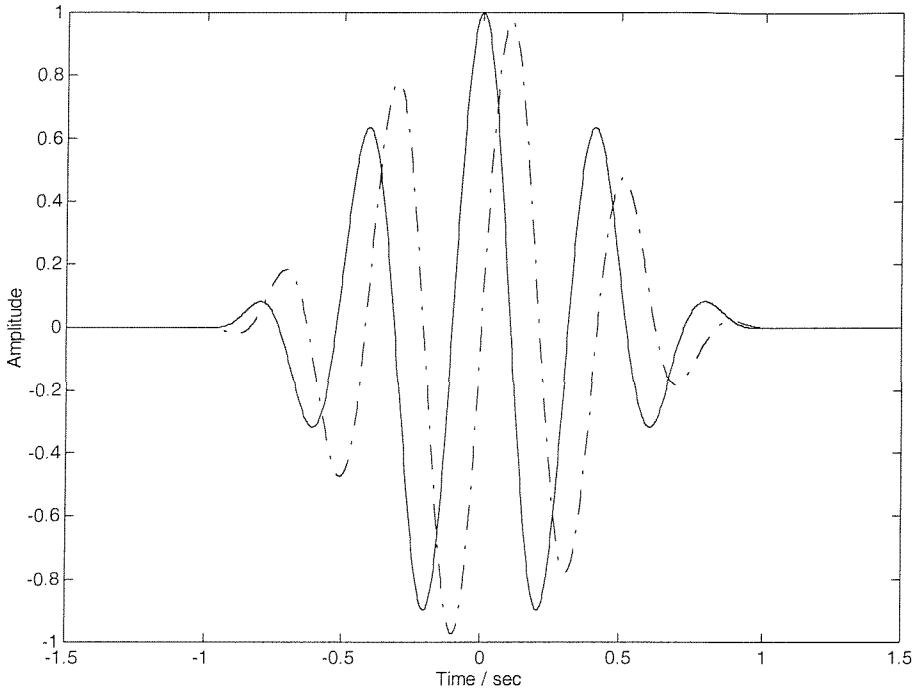

Figure 3 Complex Harmonic Windowed by a Hanning Window (Real Part Solid Line, Imaginary Part Dashed Line)

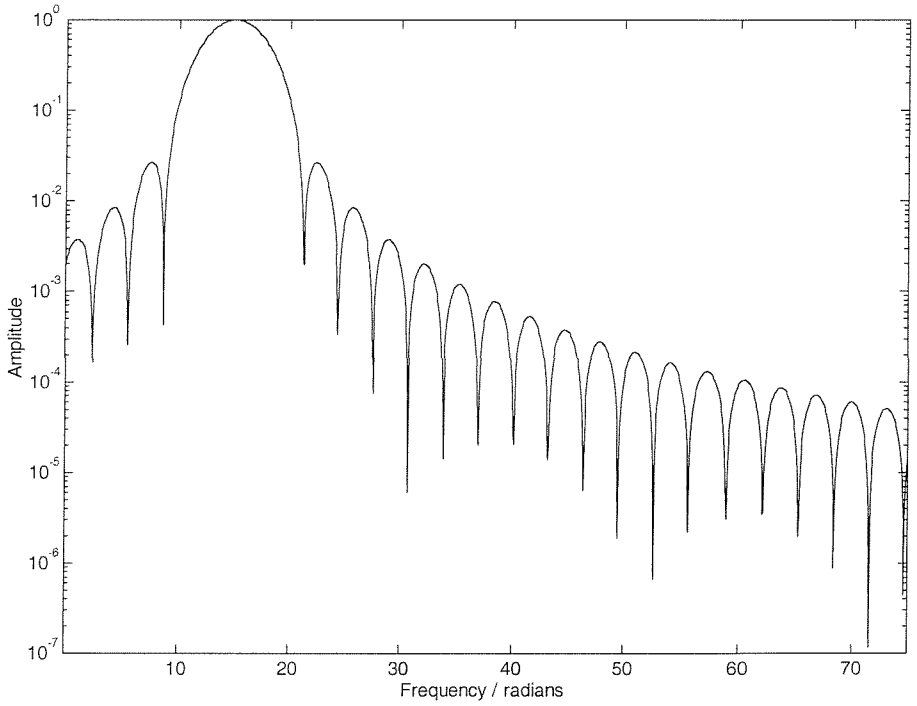

Figure 4 The Amplitude of the FT of a Hanning Windowed Complex Harmonic

2.2.3.2 Contraction and Expansion

This section describes the effect on the FT of rescaling the time axis. The relationship is shown in (2-6) [Qian96, etc.] and proven via a change of variables.

$$F\left[x(at)\right]=\frac{1}{\sqrt{2\pi}}\int x(at)e^{-j\omega t}dt=\frac{1}{\sqrt{2\pi}}\frac{1}{|a|}\int x(\tau)e^{-j\left(\frac{\omega}{a}\right)\tau}d\tau=\frac{1}{|a|}X\left(\frac{\omega}{a}\right) \tag{2-6}$$

By contracting the signal in the time domain, the frequency domain representation is expanded and scaled. This is intuitive and can be illustrated via a simple example of a pure tone in time of known frequency. By expanding and contracting the signal (or by replaying the signal at a slower or faster rate), the period of oscillation of the tone changes and thus so does its spectrum. When the signal is expanded (a is greater than one), the frequency decreases; and as the signal is contracted (a is less than one), the frequency increases.

This property essential to the Wavelet Transform (WT) as defined later in this chapter.

2.2.3.3 The Uncertainty Principle

In relation to the FT, the uncertainty principle can be derived in a number of ways. Cohen [Cohen95] discusses the relationship by saying "The density in time is $|x(t)|^2$ and the density in frequency is $|X(\omega)|^2$; but $x(t)$ and $X(\omega)$ are related and hence one should not be surprised to find that there is a relation between the densities. The relation is such that if one density is narrow the other is broad." There are a number of different ways to prove the relationship between the resolution in time and frequency [Gabor46, Cohen95, Qian96, Rioul91].

For a given signal, the 'bandwidth', $\Delta\omega$, of the FT is defined as given in (2-7), where the dominator is the energy of the signal. The variable $\Delta\omega$ represents the resolution in the frequency domain. In a similar fashion, the spread in time is denoted as Δt and is given by (2-8) where once again the denominator is the energy of the signal. The variable Δt represents the temporal resolution. Using these definitions it is possible to show that (2-9) must always hold [Gabor46, Cohen95, Hammond96]. Equation (2-9) states that the product of the bandwidth in frequency $\Delta\omega$ and the time width Δt is never less than ½. This relation defines the uncertainty principle. The more precisely a signal is specified in time, is less precisely it can be specified in

frequency, and vice versa. This is analogous to the relationship found in quantum physics between displacement and momentum.

$$\Delta\omega^2 = \frac{\int(\omega-\bar{\omega})^2|X(\omega)|^2\,d\omega}{\int|X(\omega)|^2\,d\omega} \tag{2-7}$$

$$\Delta t^2 = \frac{\int(t-\bar{t})^2|x(t)|^2\,dt}{\int|x(t)|^2\,dt} \tag{2-8}$$

$$\Delta\omega\cdot\Delta t \geq \frac{1}{2} \tag{2-9}$$

The product of the bandwidth of a signal in time and frequency is called the Bandwidth-Time (BT) product.

2.2.3.4 The Analytic Signal

The Hilbert transform is used to transform real input signals into 'analytic signals'. Analytic signals have the virtue that they have a defined instantaneous frequency (detailed in the next section) and a lower BT product. The analytic signal derives its name from that fact that complex functions defined in this manner satisfy the Cauchy-Rieman conditions for differentiability and functions which satisfy this have been traditionally called analytic functions [Cohen95].

The analytic signal is defined in (2-10) and is characterised by being zero for negative frequencies.

$$\tilde{x}(t) = F^{-1}\left\{\begin{matrix} 2X(\omega) & \omega>0 \\ X(\omega) & \omega=0 \\ 0 & \omega<0 \end{matrix}\right\} \tag{2-10}$$

The analytic signal is usually computed with the aid of the Hilbert transform [Papoulis91]. By using the analytic form of a signal, the BT product can be reduced. Furthermore, use of the analytic signal can help reduce the number of cross-terms generated when using CC TFRs.

2.2.3.5 Instantaneous Frequency and Group Delay

Consider a signal of the form given in (2-11). If the functions $a(t)$, the amplitude, $\phi(t)$, the phase, are such that the inequality given in (2-12) holds, *i.e.* the amplitude is slowly varying, then the signal is said to be (time) asymptotic [Guillemain96]. The signal given in (2-11) is approximately analytic if it is asymptotic [Tchamitchian91].

$$x(t) = a(t)e^{j\phi(t)} \tag{2-11}$$

$$\left| \frac{1}{a(t)} \frac{da(t)}{dt} \right| << \left| \frac{d\phi(t)}{dt} \right| \tag{2-12}$$

In general, this signal is non-stationary, possessing a time-varying spectrum.

Such a signal possesses only one frequency component at any given instant in time, and one may ask: What is the signal's frequency at a given time? One, intuitive solution, is provided by the instantaneous frequency (IF) as defined in (2-13). Similarly one may ask: When does a frequency component occur? The corresponding solution is the Group Delay (GD) defined in (2-14).

$$x(t) = a(t)e^{j\phi(t)} \Rightarrow \omega_i(t) = \frac{d\phi(t)}{dt} \tag{2-13}$$

$$X(\omega) = A(\omega)e^{j\Phi(\omega)} \Rightarrow t_g(\omega) = -\frac{d\Phi(\omega)}{d\omega} \tag{2-14}$$

In order for the IF and GD to unambiguously describe a signal in time and frequency, each of these functions must be monotonic. If this were not the case, then it would be possible for a signal to have the same IF at two different times. The result of this would be the need for the GD function to be multi-valued at one frequency, in order to fully describe the signal. Since the GD function is single valued, it follows that in order the GD and IF functions to be interpreted reliably then they must be monotonic.

Although the IF and GD essentially answer the same question, whether or not these two parameters describe the same curves depends upon how well (2-12) is satisfied. This will be shown in Chapter 3 Section 3.4.3.1. Further use will be made of the IF

and GD functions in the following chapter in the context of the Reassigned Spectrogram.

2.2.4 Discrete-Time Fourier Transform

In order to be efficiently stored and manipulated on a modern digital computer, continuous signals need to be discretised. This is achieved via sampling the continuous signal at a finite number of discrete regularly spaced points. This sampling must take place at a sufficiently high rate (the sampling interval should be sufficiently small). Wittaker [Wittaker35] showed that the sampling rate should be twice the highest frequency component of the signal. The highest frequency which can be accurately measured is half the sampling rate. This frequency is called the 'Nyquist' or 'folding' frequency. The discrete version $x[n]$ of the continuous time series $x(t)$ is given in (2-15), where n is the discrete time variable, and f_s is the sampling rate (reciprocal of sampling interval). This equation uses a *Dirac Delta* function to sample the continuous time function. Equation (2-16) defines this delta function in terms of the limit of a function using two Heaviside step functions.

$$x[n] = \int_{-\infty}^{\infty} x(t)\delta\left(\frac{n}{f_s} - t\right)dt \tag{2-15}$$

$$\delta(t) = \lim_{\varepsilon \to 0}\left(\frac{1}{2\varepsilon}(H(t+\varepsilon) - H(t-\varepsilon))\right) \tag{2-16}$$

$$H(t) = \begin{cases} t \geq 0 & 1 \\ t < 0 & 0 \end{cases}$$

Using the discrete definition of the input time series, a new form of the FT can be derived. The version given in (2-17) is discrete in time, but continuous in frequency. This is called Discrete-Time Fourier Transform [Lim88] and is a mid-way point between the fully continuous or fully discrete versions of the FT traditionally used.

$$X(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \tag{2-17}$$

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)e^{j\omega n}d\omega$$

2.2.5 Discrete Fourier Transform

For implementation on a digital computer, both the input time series and the output frequency representation must be discrete. In order to achieve this, the input data is assumed to have finite length N, here assumed to start at $n=0$. This reduces the infinite sum in (2-17) to a finite sum. The length of the input data defines the length of the frequency domain version of the signal. By defining the continuous frequency variable, ω in form given in (2-18), equation (2-17) can be shown to reduce to the Discrete Fourier Transform (DFT) as given in (2-19) [Marven96].

$$\omega = \frac{2\pi k}{N} \tag{2-18}$$

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi nk}{N}\right)} \tag{2-19}$$

$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] e^{j\left(\frac{2\pi nk}{N}\right)}$$

The DFT is discrete in both time and frequency domains. It is computed via a finite sum and can therefore be easily implemented, and the results stored upon a digital computer. The length of the sequence generated by the DFT is the same as the length of the signal, *N*. In terms of the frequency, the axis spans the space from DC (0Hz) to the half the sampling frequency in *N/2* discrete steps. An efficient method of computing the DFT has been created, called the Fast Fourier Transform (FFT) [Marven96]. Since the only difference between the DFT and FFT is the method of implementation, further exploration of the FFT is not given here, (see [Marven96] for further details). The FFT implementation of the DFT has allowed Fourier analysis to be applied to a wide range of real world problems. Applications of the FFT include estimation of transfer functions [Herlufsen84], pitch period estimation in speech [Rabiner87], and active noise control [Nelson94].

2.2.5.1 Resolution

The signal and its DFT are of equal length. Therefore the greater the length of the signal, the more closely space the frequency samples. The width of each cell in the frequency domain is given by (2-20). The spacing of the samples is linear in the frequency domain, so the bandwidth of frequencies covered per cell does not change with frequency.

$$\Delta k = \frac{f_s}{N} \tag{2-20}$$

This relationship is sometimes referred to the 'resolution' of the DFT. It is preferred to reserve the term resolution for the ability to able to resolve two closely spaced sinusoids. For example, appending a number of zero samples to its end can extend a discrete signal sequence. By equation (2-20) then spacing between successive points along the DFT has been reduced. However the ability to resolve two closely spaced sinusiods in the signal has not been improved, and as such (2-20) should not be considered as the resolution of the DFT.

2.2.6 Stochastic and Deterministic Signals

The 1D signals used through this thesis can be considered as the result of a measured experiment. If each time the experiment is run, the result is a different output waveform, then the system and signal are called stochastic (or random) [Bendat86]. In contrast, if each time the experiment is conducted the output waveform is the same, then the system and signal are called deterministic.

If only one realisation is available, then without prior system knowledge, it is not possible to determine whether the signal comes from a deterministic or stochastic system. Throughout this thesis, the signal is usually considered to come from a deterministic system. The distinction between stochastic and deterministic systems is important when considering whether a signal is stationary or non-stationary, as detailed in the following section.

2.2.7 Stationary and Non-Stationary Signals

The concepts of stationary and non-stationary are most appropriately presented in context of stochastic signals. However, since throughout this thesis signals are considered to be deterministic, these concepts are introduced from this framework. The time-dependant auto-correlation for a single realisation of a signal, $x(t)$ is given in (2-21) [Bendat86].

$$R(t,\tau) = x\left(t+\frac{\tau}{2}\right)x^*\left(t-\frac{\tau}{2}\right) \tag{2-21}$$

A necessary but not sufficient condition for a signal to be (wise sense) stationary is that the auto-correlation function is a function purely of the lag variable, τ *i.e.*

$R(t,\tau)=f(\tau),\forall t$ (where $f(\tau)$ is an arbitrary function). In order to prove stationarity, then multiple realisations of the stochastic signal should be considered. If the auto-correlation function changes over time as well as lag then the signal is non-stationary. Most of the signals considered in this thesis are non-stationary, in that their time-dependant auto-correlation functions and thus their spectra vary with time. This changing spectral content over time, highlights the need for a decomposition of the signal in terms of both time and frequency.

2.3 Linear Time-Frequency Distributions

2.3.1 Introduction

This section will detail two linear time-frequency representations (TFRs), which achieve a signal decomposition in terms of time and frequency. These are the Short-Time Fourier Transform (ST-FT), an extension of the FT and the Wavelet transform (WT). Prior to defining these, there follows the definition of linearity in the context of TFRs.

For an arbitrary function $f(t)$, to be linear it must satisfy (2-22), where c_1 and c_2 are constants. In order for a unspecified TFR, $P_x(t,\omega)$, to be linear with respect to the input signal, $x(t)$, then it must obey an analogous law as given by (2-23) [Hlawatsch92].

$$f(c_1x(t)+c_2y(t))=c_1f(x(t))+c_2f(x(t)) \tag{2-22}$$

$$x(t)=c_1x_1(t)+c_2x_2(t)\Rightarrow P_x(t,\omega)=c_1P_{x_1}(t,\omega)+c_2P_{x_2}(t,\omega) \tag{2-23}$$

Although linearity in TFRs might be considered as desirable, these TFRs cannot be interpreted as energy decompositions. In other words, the value of a particular point on a linear TFR cannot be considered to be the energy present in the input signal at that particular time and at that particular frequency. This is due to energy being a quadratic rather than a linear quantity.

2.3.2 Short-Time Fourier Transform

2.3.2.1 Introduction

In previous sections, the need for time-varying spectral information for non-stationary signals was justified. However, the FT does not readily provide such information, and therefore in order to be able to express the changing characteristics of the signal, a further development is required. One approach is to temporally limit the FT, allowing the time-varying characteristics to be found. This is the essence of the Short-Time Fourier Transform (ST-FT).

2.3.2.2 Continuous ST-FT

In order to focus the FT to study the properties close to time t, a time-localising window can be used which suppresses the influence of the signal away from this point. A new signal is created as a result of multiplying the signal by a windowing function as given in (2-24) [Cohen95]. Examples of this windowing function include Hanning, Hamming and Gaussian [Owens93].

$$x_t(\tau) = x(\tau)h(t-\tau) \tag{2-24}$$

This new signal is a function of two variables, time t, and lag, τ. Application of the FT to the 'windowed' signal yields an estimate of the spectrum about time t (2-25). By moving the centre of the window along the length of the signal, an estimate of the local spectrum as a function of time can be achieved. This is defined in (2-26) and is the ST-FT. The two different definitions of the ST-FT given in (2-26) although equivalent they allow different interpretations of the same equation. The top definition can be viewed as dragging the window past a fixed signal, whereas the bottom definition can be interpreted as dragging the signal past a fixed window.

$$S(t,\omega) = \frac{1}{\sqrt{2\pi}} \int x_t(\tau) e^{-j\omega\tau} d\tau \tag{2-25}$$

$$S(t,\omega) = \frac{1}{\sqrt{2\pi}} \int x(\tau) h(t-\tau) e^{-j\omega\tau} d\tau \tag{2-26}$$

$$S(t,\omega) = \frac{1}{\sqrt{2\pi}} e^{-j\omega t} \int h(\tau) x(t-\tau) e^{j\omega\tau} d\tau$$

2.3.2.3 Proof of Linearity of the Continuous ST-FT

Consider the signal given in (2-27), comprised of a linear combination of two signals, $x_1(t)$ and $x_2(t)$ (c_1 and c_2 are constants). The proof that the ST-FT of this signal is the same linear combination of the ST-FT of each signal component is given in (2-28), where $S_x(t,\omega)$ is the ST-FT of the signal denoted in the subscript. (This subscript notion will be used throughout this thesis when it is necessary to make it explicitly clear upon which signal the TFR is acting)

$$x(t) = c_1 x_1(t) + c_2 x_2(t) \tag{2-27}$$

$$\begin{aligned} S_x(t,\omega) &= \frac{1}{\sqrt{2\pi}} \int \left(c_1 x_1(\tau) + c_2 x_2(\tau)\right) h(t-\tau) e^{-j\omega\tau} d\tau \\ &= \frac{1}{\sqrt{2\pi}} \left(c_1 \int x_1(\tau) h(\tau - t) e^{-j\omega\tau} d\tau + c_2 \int x_2(\tau) h(\tau - t) e^{-j\omega\tau} d\tau\right) \\ &= c_1 S_{x_1}(t,\omega) + c_2 S_{x_2}(t,\omega) \end{aligned} \tag{2-28}$$

2.3.2.4 Discrete ST-FT

Computation of the ST-FT of an arbitrary sampled signal requires a discrete version of the ST-FT. Just as with the FT, generalisation can take place in a two step process. Firstly consider the discretisation in time of the input signal and windowing function. This can be represented as given in (2-29). (The $1/2\pi$ factor has been omitted for notional simplicity).

$$S(n,\omega) = \sum_{m=-\infty}^{\infty} x[m] h[n-m] e^{-j\omega m} \tag{2-29}$$

Using a sampled version of the windowing function, the infinite summation in (2-29) reduces to a finite sum, defined over the range for which the windowing function is not zero, as given in (2-30). The windowing function is N points long and is centred at the origin. By shifting the window in time such that its maximum value is at $N/2$ rather than at the origin, equation (2-31) can be formulated.

$$S(n,\omega) = \sum_{m=-N/2}^{N/2-1} x[m] h[n-m] e^{-j\omega m} \tag{2-30}$$

$$S(n,\omega)=\sum_{m=0}^{N-1}x[m]h[n-m]e^{-j\omega m} \tag{2-31}$$

In the same fashion as with the FT, the continuous frequency variable is now replaced by a discrete value.

$$S[n,k]=\sum_{m=0}^{N-1}x[m]h[n-m]e^{-j\left(\frac{2\pi mk}{N}\right)} \tag{2-32}$$

At each n the windowing function moves one sample along. This is can be generalised so that it is possible to specify the number of samples moved for each successive n, also known as time-slice (2-33). In the case where $L=1$ (2-33) reduces to (2-32) and the ST-FT is said to be 'maximally overlapped'.

$$S[n,k]=\sum_{m=0}^{N-1}x[m]h[nL-m]e^{-j\left(\frac{2\pi mk}{N}\right)} \tag{2-33}$$

This definition is the form of the discrete ST-FT used widely throughout this thesis.

The ST-FT was developed as a means of showing the changing spectral content of a non-stationary deterministic signal. Figure 5 shows the magnitude of the ST-FT for a linear FM chirp as given in (2-34). For this plot, the sample rate f_s was set to 1000Hz, the starting frequency of the chirp, ω_0 was set to $\omega_0 = 160\pi$ and the rate-of-increase of frequency, α was set to 400π. A 64 point Hanning window was used in the computation of this plot.

$$\begin{aligned} x(t) &= \cos\left((\alpha t+\omega_0)t\right) \\ x[n] &= \cos\left(\left(\alpha\frac{n}{f_s}+\omega_0\right)\frac{n}{f_s}\right) \end{aligned} \tag{2-34}$$

The changing frequency content is evident in Figure 5, and the rising frequency component easily visible.

The ST-FT is generally a complex valued function. In order to view this function easily, the modulus or squared modulus is plotted. The squared modulus of the ST-FT is called the 'Spectrogram'. The spectrogram is not a linear TFR.

All the TFRs used throughout this thesis are displayed in the form given in Figure 5. The x-axis is used for time and the y-axis for frequency. The z-axis shown in the form of a colour is the amplitude of the TFR and is displayed on a logarithmic scale. The scale changes from blue for low amplitude to red for high. The scale is dynamic, altering according to the maximum value present in the TFR.

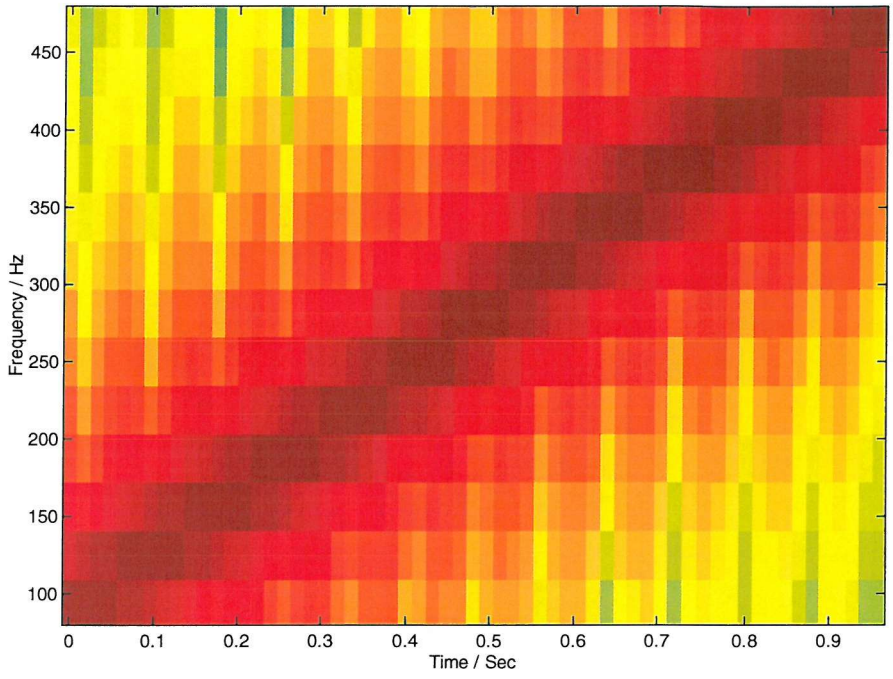

Figure 5 The Spectrogram of a Linear Chirp

2.3.2.5 Limitations of the ST-FT

At the core of the ST-FT is the windowing function. The choice of both shape and size of the window make a large visual difference to the computed spectrogram, the effects of which are briefly described here.

The data window aims to select a desired region, and suppress the signal outside of that region. In order to demonstrate the effect that this window has upon the ST-FT, consider the limiting case where the window is a Kronecker delta function, (2-35).

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2-35}$$

In this case, the window only picks out the time (one sample) of interest, at the expense of all others. However, since the window function in now only 1pt long, the DFT is also only 1point long. A point long DFT offers no insight as to the spectral content of the signal. Therefore, a very fine in time window has come at the expense of frequency resolution.

If good frequency resolution is required, a long time window should be used. If the data is of length l_x, then the longest window which can be used, assuming that the data is not zero-padded, is also of length l_x. However as stated in Section 2.2.5.1, zero-padding the data does not improve the resolution, it merely increases the number of samples used. In the case where the window size is the same as the signal length, the ST-FT has the same frequency resolution as the FT. However, fine frequency resolution has come at the expense of time resolution. This trade-off between time and frequency resolution is another manifestation of Heisenburg's uncertainty principle. An analogous result also holds for the continuous version of the ST-FT.

To demonstrate the differences created by using different window sizes examine the two spectrograms of the same input signal (speech signal) given in Figure 6 and Figure 7. The utterance used is Richard saying 'that human affairs were being watched from the timeless worlds of space', taken from Jeff Wayne's musical version of 'The War of the Worlds'. Figure 6 shows the signal using a short time window (Gaussian window). In this plot, the good time resolution can be seen by observing the abrupt start and stop of components as witnessed by the broad horizontal bands in Figure 6. However, when using the short time window the frequency resolution is poor. In contrast to the short time window, a long time window (Gaussian) was used in the construction of Figure 7. As expected, although this plot displays good frequency resolution, 'smearing' in the time axis indicates poor time resolution. An example of the time-smearing present in long time windows can be seen by considering the distributions close to 0.6sec. When using a short time window (Figure 6) an impulsive component can be seen, ranging from approximately to 0 to approximately 4000 Hz. When the long time window is used (Figure 7), this impulsive component cannot be seen. The poor frequency resolution for a short-time window can be seen when comparing the two figures in the region from 2-2.5sec. The harmonic nature of speech can be clearly seen when using a long time window.

However when using a short time window, these harmonics become blurred into a single component. Without prior signal knowledge, it is not possible to state *a priori* what length the windowing function should take to best match the input signal. Typically the user computes three or four different spectrograms using different window lengths in order to ascertain visually which window length 'best' matches the input signal.

Careful consideration must also be given to the shape of the windowing function used as well as its size. A Gaussian window offers the best time/frequency trade-off, since its time-bandwidth product is a 1/2 [Cohen95]. In order to show the differences that the choice of window can make to the spectrogram, the same speech signal as used previously is analysed using a Gaussian (Figure 8) and a rectangular window (Figure 9). The two windows were selected such that they had the same duration (spread) in time. The two resultant spectrograms look different. The Gaussian window (Figure 8) produces an image devoid of components that start and stop very rapidly. In contrast to this, the spectrogram using the rectangular window (Figure 9) has abrupt starts and stops along the time axis, however it has less clearly defined tonal components in frequency.

The reason different windows produce visually different spectrograms can be traced back to the FT. Previously in this chapter, the effect windowing has been examined in equations (2-3)-(2-4). This stated that windowing a signal in the time domain resulted in a smearing in the frequency domain. The degree of smearing is dependant upon the FT of the windowing function used. Since the Gaussian is the most compact function in time-frequency [Hammond96], it is often selected as the window of choice for the spectrogram.

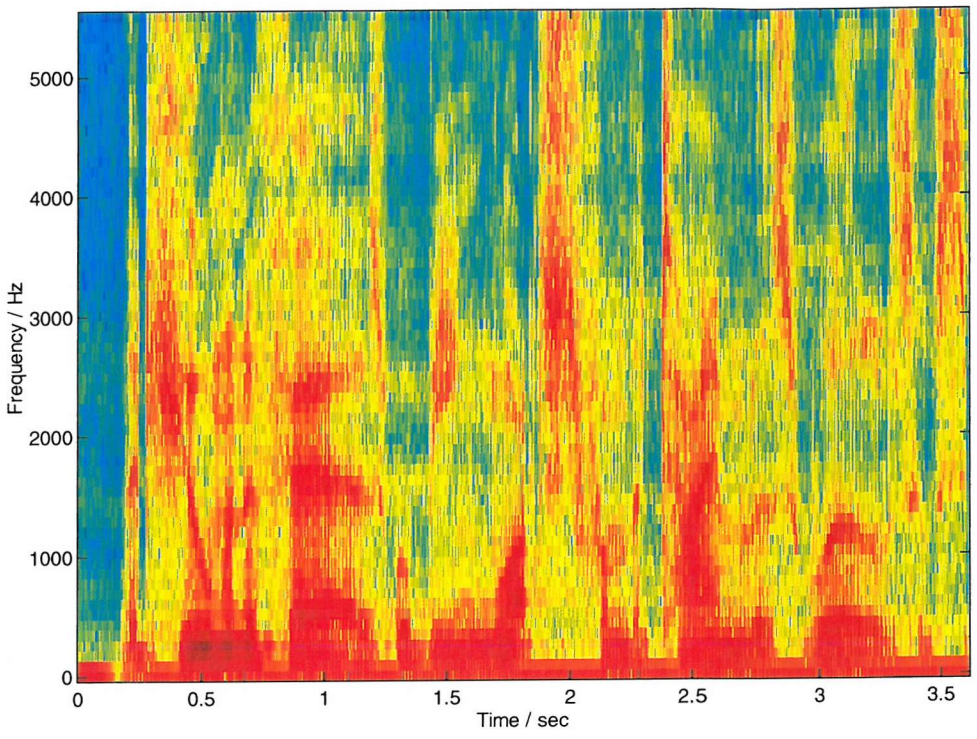

Figure 6 Spectrogram of Speech Signal Short Time Window

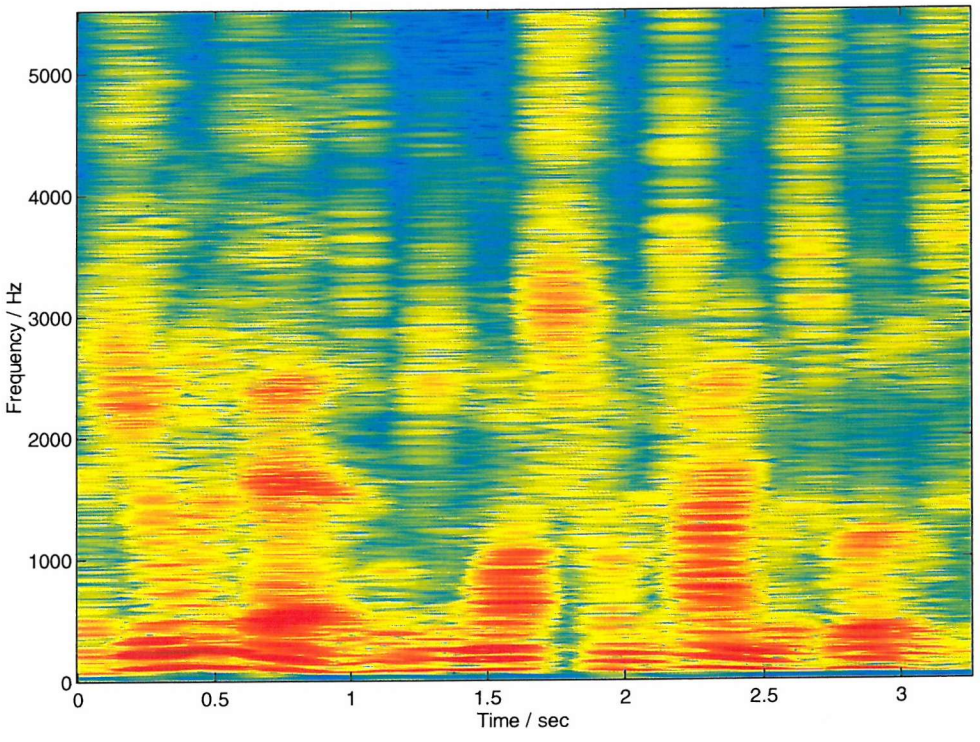

Figure 7 Spectrogram of Speech Signal Long Time Window

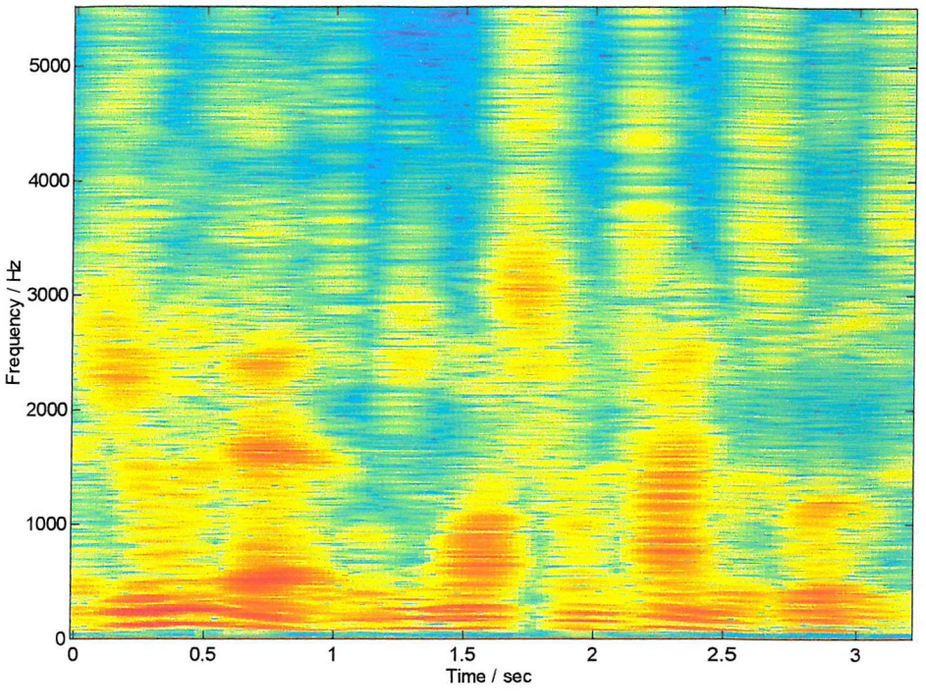

Figure 8 Spectrogram of Speech Signal Gaussian Window

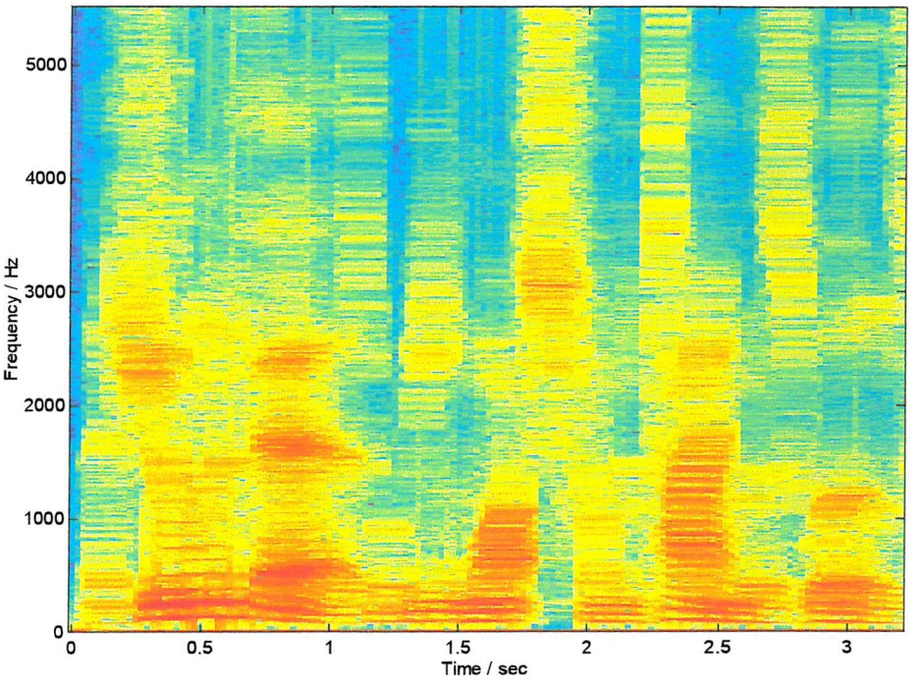

Figure 9 Spectrogram of Speech Signal using Rectangular Window

2.3.3 Continuous Wavelet Transform (WT)

2.3.3.1 Introduction

In the previous section, the basic theory surrounding the ST-FT was outlined. The ST-FT can be viewed simply as a time-localising windowed version of the DFT. One of the inherent properties of the DFT is that the frequency resolution is dependant upon the length of the window, and is constant across the whole spectrum. An alternative approach is to make frequency resolution decrease with centre frequency. The change in frequency resolution with centre frequency results in improved temporal resolution at high frequencies and improved frequency resolution at low frequencies.

If the resolution is proportional to centre frequency then this is called 'Constant Q' analysis [Hlawatsch92] and is at the heart of wavelet analysis. The definition of the WT is as given in (2-36).

$$W(t,a)=\frac{1}{\sqrt{a}}\int x(\tau)h\left(\frac{\tau-t}{a}\right)d\tau \tag{2-36}$$

Typically, the wavelet used is oscillatory in nature. This is in contrast to the ST-FT, where the user specified window used is typically a non-oscillatory, low pass filter of some kind. One example of a wavelet used is the Morlet wavelet as defined by (2-37) where ω_0 is the centre frequency of the wavelet [Guillemain91]. The scale parameter, *a*, provides the contraction (dilation) or expansion of the mother wavelet, which also alters the frequency domain characteristics of the wavelet, as discussed in Section 2.2.3.2. Using a modulating (oscillating) wavelet the WT can be seen as another time-frequency representation due to the link between scale and frequency as given in (2-38) [Rioul91].

$$h(t)=e^{j\omega_0 t}e^{\frac{-t^2}{2}} \tag{2-37}$$

$$a=\frac{\omega_0}{\omega} \tag{2-38}$$

2.3.3.2 Proof of Linearity

As with the ST-FT, the WT is a linear TFR. This is shown in (2-39), where $WT_x(t,a)$ is the WT of the signal $x(t)$ given in (2-36).

$$WT_x(t,a) = \int (c_1 x_1(\tau) + c_2 x_2(\tau)) h\left(\frac{t-\tau}{a}\right) d\tau \tag{2-39}$$
$$= c_1 \int x_1(\tau) h\left(\frac{\tau - t}{a}\right) d\tau + c_2 \int x_2(\tau) h\left(\frac{\tau - t}{a}\right) d\tau$$
$$= c_1 WT_{x_1}(t,\omega) + c_2 WT_{x_2}(t,\omega)$$

2.3.3.3 Digital implementation

In the case of the continuous ST-FT, the conversion from the continuous to discrete domain was clearly defined via the DFT. In the wavelet case the progression is not as clearly defined. The discretisation of the WT as defined in (2-36) can be written as given in (2-40).

$$W[n,a] = \frac{1}{\sqrt{a}} \sum_{m=0}^{l_x} x[m] h\left[\frac{m-n}{a}\right] \tag{2-40}$$

If the wavelet is a sampled version of a known continuous function, then the dilated version of the windowing function required for (2-40) can be easily computed. The wavelet can be pre-computed for a given scale, which reduces the overall computation required for the transform. The variable $h_a[n]$ is used to represent the realisation of the wavelet at scale a as given in (2-41). Substituting this expression into (2-40) results in (2-42).

$$h_a[n] = h\left[\frac{n}{a}\right] \tag{2-41}$$

$$W[n,a] = \frac{1}{\sqrt{a}} \sum_{m=0}^{l_x} x[m] h_a[m-n] \tag{2-42}$$

Unlike the FT case, where the limits of the frequency axis were set implicitly by the FT, in the wavelet case it is the users discretion as to the frequency range used. The lowest frequency which can be analysed is dependant upon the length of the signal and the sampling rate. The highest is the Nyquist frequency.

The squared modulus of the WT is called the 'scalogram'. As with the spectrogram, the scalogram is not a linear TFR. To demonstrate the differences between the spectrogram and scalogram Figure 11 shows the scalogram for the linear chirp shown previously in Figure 5. The previously straight line in TF is now curved

owing to the relationship between frequency and scale as given in (2-38). This plot is also shown on a frequency scale in Figure 12. At high frequency the frequency resolution of the scalogram can be seen to decrease, this phenomenon is described in the following section.

2.3.4 Comparison of the Resolution of the ST-FT and WT

The ST-FT (and thus the spectrogram) is limited by the resolution characteristics of the FT. In order to achieve good frequency resolution, time resolution is sacrificed. The changing window lengths of the WT (scalogram) mean that time-frequency resolution is not fixed. Instead the time-frequency resolution is a function of frequency (or scale), changing as to offer good time / poor frequency at high frequencies, and poor time / good frequency at low frequencies. Although the time/frequency resolution varies with frequency, their product is fixed. This is another manifestation of the Heisenburg's uncertainty principle. An example showing the different resolution tiling of the ST-FT and WT are shown graphically in Figure 10.

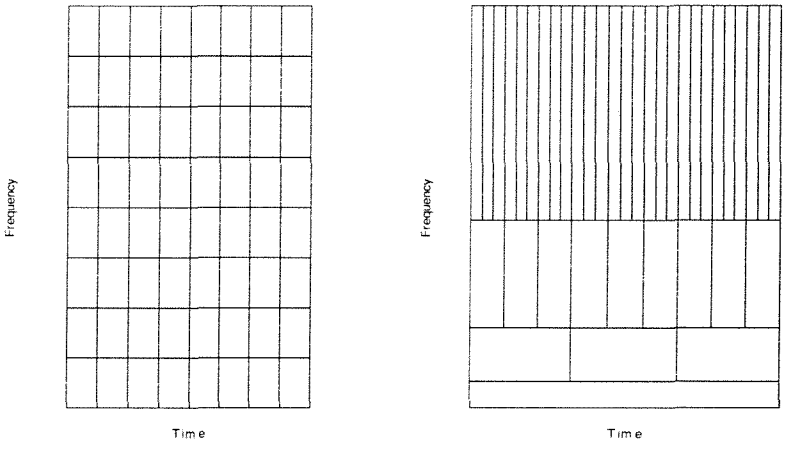

Figure 10 The Resolution Tiling of the ST-FT (Spectrogram) (Left) and WT (Scalogram) (Right)

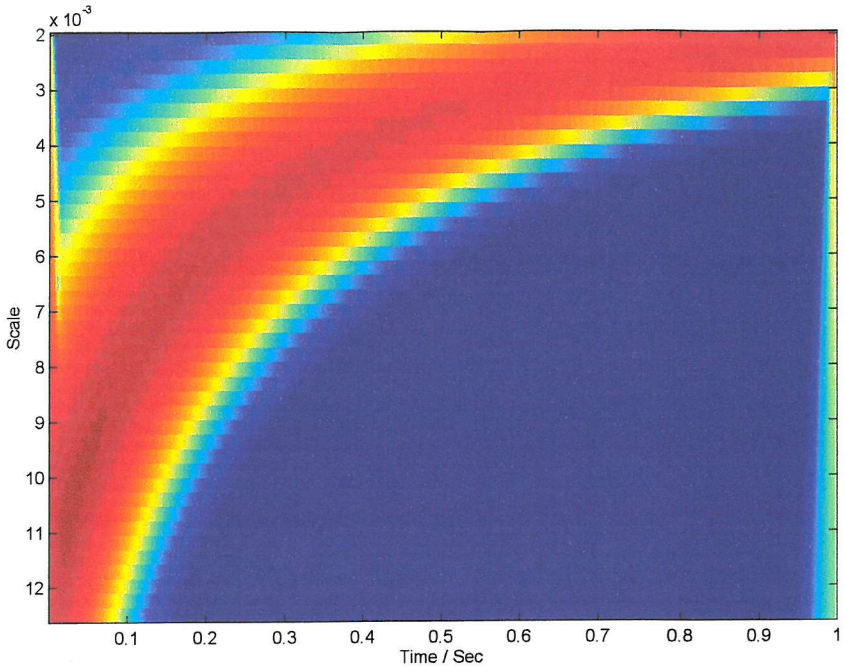

Figure 11 Scalogram of a Linear Chirp using Morlet Wavelet – Linear Scale

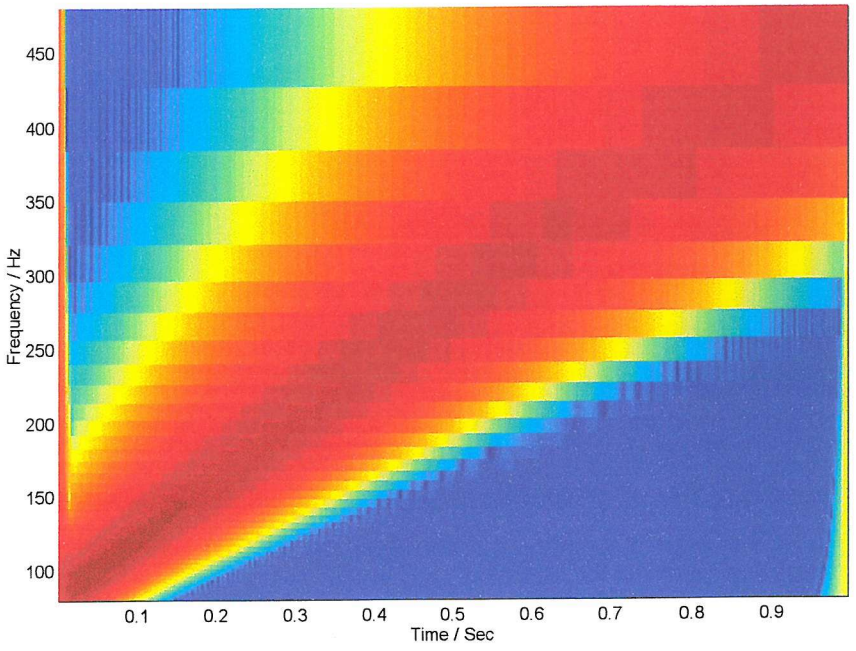

Figure 12 Scalogram of a Linear Chirp using Morlet Wavelet – Linear Frequency

2.4 Cohen Class of Bilinear Time-Frequency Distributions

2.4.1 Introduction

In the previous section, two non-linear TFRs have been discussed: the spectrogram and scalogram. In this section, these are placed in the context of a wider class of TFRs called Cohen's Class. The section begins with the definition of the Wigner TFR, the distribution at the heart of CC and then proceeds to define the rest of CC in terms of a 2D convolution operation upon the Wigner Distribution [Hammond96]. Finally, the limitations of CC of TFRs are considered.

2.4.2 Energetic Time-Frequency Representations

The concept of an energetic TFR is now introduced. A TFR can be considered an energetic TFR if it conforms to the following three conditions, where $P(t,\omega)$ is a real valued TFR [Qian96].

$$P(t,\omega) \geq 0 \quad \forall t,\omega \tag{2-43}$$

$$\int P(t,\omega)\,d\omega = |x(t)|^2 \tag{2-44}$$
$$\int P(t,\omega)\,dt = |X(\omega)|^2$$

The justification for these three requirements is as follows. The TFR must be non-negative for all time and frequency. Since energy is a manifestly non-negative quantity. The second property (2-44) states that integrating the TFR along the frequency axis should yield the energy per unit time at time t. This is called the time-marginal. The third property defines a similar relationship but in frequency. Integrating the TFR along the time axis should yield the energy per unit frequency at frequency ω. This property is called the frequency marginal [Cohen85].

Although guaranteed positive, both the spectrogram and scalogram fail to conform to either the time or frequency marginals. In both cases, this is due to the time-localising window used in the computation. However, these TFRs are still considered by many as energetic TFRs, since they provide an estimate of the energy present in a signal and a particular time and frequency.

2.4.3 Bilinear Time-Frequency Distributions

In Section 2.3.2.3 it was shown that the ST-FT is a linear function, and it was stated (Section 2.3.2.4) that the same property did not hold for the spectrogram. In an analogous manner to that used for the ST-FT and WT cases, consider the spectrogram (2-45) for a two-component signal is given in (2-46). The spectrogram (and the scalogram) can be seen to be the result of adding together the spectrograms (scalograms) of each signal component (so called 'auto-terms') and a series of terms dependant upon the product of ST-FTs of individual components (so called 'cross-terms'). A range of names have been given to the cross-terms, including 'error-terms', and 'interference errors', but they are a necessary result of the interaction between signal components.

$$SP(t,\omega)=\left|S(t,\omega)\right|^2 \tag{2-45}$$

$$\begin{aligned} SP_{x_1+x_2}(t,\omega)&=\left|S_{x_1+x_2}(t,\omega)\right|^2=\left|S_{x_1}(t,\omega)+S_{x_2}(t,\omega)\right|^2 \\ &=\left|S_{x_1}(t,\omega)\right|^2+\left|S_{x_2}(t,\omega)\right|^2+S_{x_1}(t,\omega)S_{x_2}{}^*(t,\omega)+S_{x_1}{}^*(t,\omega)S_{x_2}(t,\omega) \end{aligned} \tag{2-46}$$

The energetic TFRs created in this manner are called 'bilinear' presence of both auto and cross-terms of $x_1(t)$ and $x_2(t)$ [Hlawatsch92]. All members of CC, of which the spectrogram is one, are called 'bilinear'. Their bilinear nature gives rise to the presence of cross-terms. In most bilinear TFRs manifest themselves as visually detracting components in the TFR. In the case of the spectrogram and scalogram, it is the cross-terms which result in the auto-terms becoming spread in TF.

2.4.4 Wigner Distribution

The Wigner Distribution (WD) is given in (2-47) for the continuous input signal $x(t)$. The fact that the WD is a bilinear transform can be easily shown, as given in (2-49) where $WD_x(t,\omega)$ is the WD of the signal $x(t)$ and WD_{x_1,x_2} is called the 'cross-WD', taking in two input signals, x_1 and x_2, as defined in (2-48).

$$WD(t,\omega)=\frac{1}{2\pi}\int x^*(t-\tau/2)x(t+\tau/2)e^{-j\omega\tau}d\tau \tag{2-47}$$

$$WD_{x_1,x_2}(t,\omega)=\frac{1}{2\pi}\int x_1^{*}(t-\tau/2)\,x_2(t+\tau/2)e^{-j\omega\tau}d\tau \tag{2-48}$$

$$\begin{aligned}WD_{x_1+x_2} &= \frac{1}{2\pi}\left(\int\left(\left(x_1^{*}(t-\tau/2)+x_2^{*}(t-\tau/2)\right)\left(x_1(t+\tau/2)+x_2(t+\tau/2)\right)\right)e^{-j\omega\tau}d\tau\right) \\ &= \frac{1}{2\pi}\left(\int\left(x_1^{*}(t-\tau/2)x_1(t+\tau/2)+x_2^{*}(t-\tau/2)x_2(t+\tau/2)\right)e^{-j\omega\tau}d\tau\right)+ \\ &\quad +\frac{1}{2\pi}\left(\int\left(x_1^{*}(t-\tau/2)x_2(t+\tau/2)+x_2^{*}(t-\tau/2)x_1(t+\tau/2)\right)e^{-j\omega\tau}d\tau\right) \\ &= WD_{x_1}+WD_{x_2}+WD_{x_1,x_2}+WD_{x_1,x_2}^{*}=WD_{x_1}+WD_{x_2}+2\,\mathrm{Re}\left[WD_{x_1,x_2}\right]\end{aligned} \tag{2-49}$$

The WD has a number of beneficial characteristics over other TFRs in general, and the spectrogram and scalogram in particular. Consider the linear chirp given in (2-34). Writing the equation in term of modulus and phase, the instantaneous frequency of the signal is found by taking the gradient of the phase, which is given in (2-50). The WD for this signal can easily be shown to be as given in (2-51).

$$x(t)=|x(t)|e^{j\phi(t)} \rightarrow \dot{\phi}(t)=2\alpha t+\omega_0 \tag{2-50}$$

$$WD(t,\omega)=\delta(\omega-(2\alpha t+\omega_0)) \tag{2-51}$$

This is a special case of a fundamental property of the WD, specifically that it is concentrated along the instantaneous frequency for all signals [Cohen95, Qian96]. However, owing to the bilinear nature of the transform, only mono-component linear chirp signals are free from cross-terms.

A useful property of the WD is that it satisfies the 'marginal properties', these as described in (2-44). The WD conforms to three of the four requirements to be considered as an energetic TFR. Unfortunately, the WD it is not guaranteed to be positive. To illustrate this, consider the signal given in (2-52). This is a Gaussian windowed amplitude with linearly increasing frequency. The WD for this signal is given in (2-53) [Cohen95].

$$x(t)=\left(4\alpha^2/\pi\right)^{1/4}te^{-\alpha t^2/2+j\beta t^2+j\omega_0 t} \tag{2-52}$$

$$WD(t,\omega)=\frac{2}{\pi}\left[\alpha t^2+(\omega-\beta t-\omega_0)^2/\alpha-\frac{1}{2}\right]e^{-\alpha t^2-(\omega-\beta t-\omega_0)^2/\alpha} \tag{2-53}$$

This function can be seen to be negative whenever $\alpha t^2 + (\omega - \beta t - \omega_0)^2 / \alpha \leq 1/2$. This makes it difficult to interpret the WD as an energetic TFR.

The only signal for which the WD is positive is a Gaussian modulated chirp as given in (2-54). The WD for this signal is given in (2-55) and pictorially in Figure 13 [Cohen95].

$$x(t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 + j\beta t^2 + j\omega_0 t} \tag{2-54}$$

$$WD(t,\omega) = \frac{1}{\pi} e^{-\alpha t^2 - (\omega - \beta t - \omega_0)^2/\alpha} \tag{2-55}$$

A list detailing various properties of the spectrogram, scalogram and Wigner distribution is given in Table 2 (A more complete list in given in [Hammond96, Hlawatsch92]).

The WD for the speech signal used previously in Section 2.3.2 is shown in Figure 14. Although conforming to the marginal properties and having the correct instantaneous frequency etc., properties which the spectrogram does not possess, for a complicated signal such as speech, the WD can be visually difficult to interpret, as evidenced by Figure 14.

2.4.4.1 The Requirement for Cross-Terms

Although producing visually distracting artefacts, cross-terms are required in order for the WT (and other TFRs) to conform to the marginal properties. This is shown in the following example. Consider a signal comprised of two complex sinusoids as given in (2-56); the energy per unit time of this signal is given in (2-57).

$$x(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \tag{2-56}$$

$$|x(t)|^2 = A_1^2 + A_2^2 + 2A_1 A_2 \cos((\omega_2 - \omega_1)t) \tag{2-57}$$

Through the application of (2-47) the WD for this signal can be derived as given in (2-58). Integrating out frequency yields the signal energy per unit time, given in (2-59). Since the WT has the correct time-marginal, this is equal to that derived directly from the signal.

$$WT(t,\omega) = A_1^2\delta(\omega-\omega_1) + A_2^2\delta(\omega-\omega_2) + 2A_1A_2\delta\left(\omega-\frac{1}{2}(\omega_1+\omega_2)\right)\cos\left((\omega_2-\omega_1)\right) \tag{2-58}$$

$$\int WT(t,\omega)d\omega = A_1^2 + A_2^2 + 2A_1A_2\cos\left((\omega_2-\omega_1)t\right) = |x(t)|^2 \tag{2-59}$$

In this example, the auto and cross-terms of the signal can be seen in the formulation of the WD. A new TFR can be written using only the auto-terms as given in (2-60). Although intuitively correct, this new distribution no longer has the correct time-marginal as shown in (2-61). This illustrates that the cross-terms are necessary for the WD to conform to the marginals and further that the 'intuitive distribution' does not satisfy the marginals.

$$D(t,\omega) = A_1^2\delta(\omega-\omega_1) + A_2^2\delta(\omega-\omega_2) \tag{2-60}$$

$$\int D(t,\omega)d\omega = A_1^2 + A_2^2 \neq |x(t)|^2 \tag{2-61}$$

Properties	Mathematical Description	Wigner Distribution	Spectrogram	Scalogram	Pseudo Wigner Distribution
Reality	$P(t,\omega)=P^*(t,\omega)$	√	√	√	√
Time Marginal	$\int P(t,\omega)d\omega=\lvert x(t)\rvert^2$	√	×	×	×
Frequency Marginal	$\int P(t,\omega)dt=\lvert X(\omega)\rvert^2$	√	×	×	√
Positivity	$P(t,\omega)\geq 0 \quad \forall t,\omega$	×	√	√	×
Time Shift	$x(t-t_0)\Leftrightarrow P(t-t_0,\omega)$	√	√	√	√
Frequency Shift	$x(t)e^{j\omega_0 t}\Leftrightarrow P(t,\omega-\omega_0)$	√	√	√	√
Instantaneous Frequency	$\frac{\int \omega P(t,\omega)d\omega}{\int P(t,\omega)d\omega}=\omega_i(t)$ $\omega_i(t)=\frac{d}{dt}\arg\{x(t)\}$	√	×	×	√
Group Delay	$\frac{\int tP(t,\omega)dt}{\int P(t,\omega)dt}=\tau_g(\omega)$ $\tau_g(\omega)=\frac{d}{d\omega}\arg\{X(\omega)\}$	√	×	×	×

Table 2 Comparison between spectrogram, scalogram and the Wigner distribution

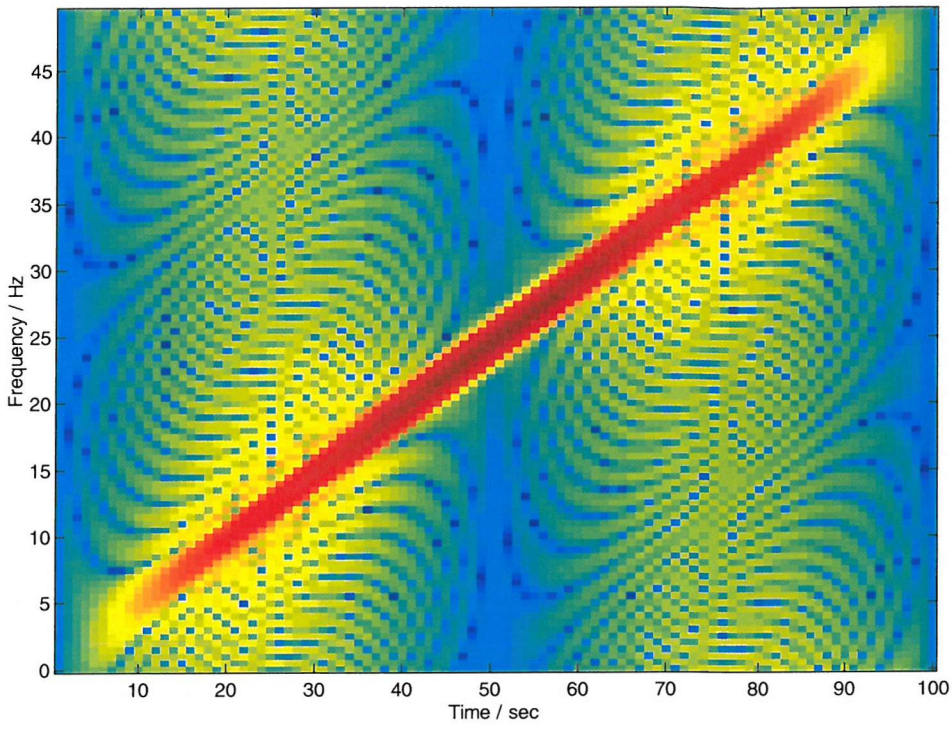

Figure 13 Wigner Distribution for a Gaussian Enveloped Linear Chirp

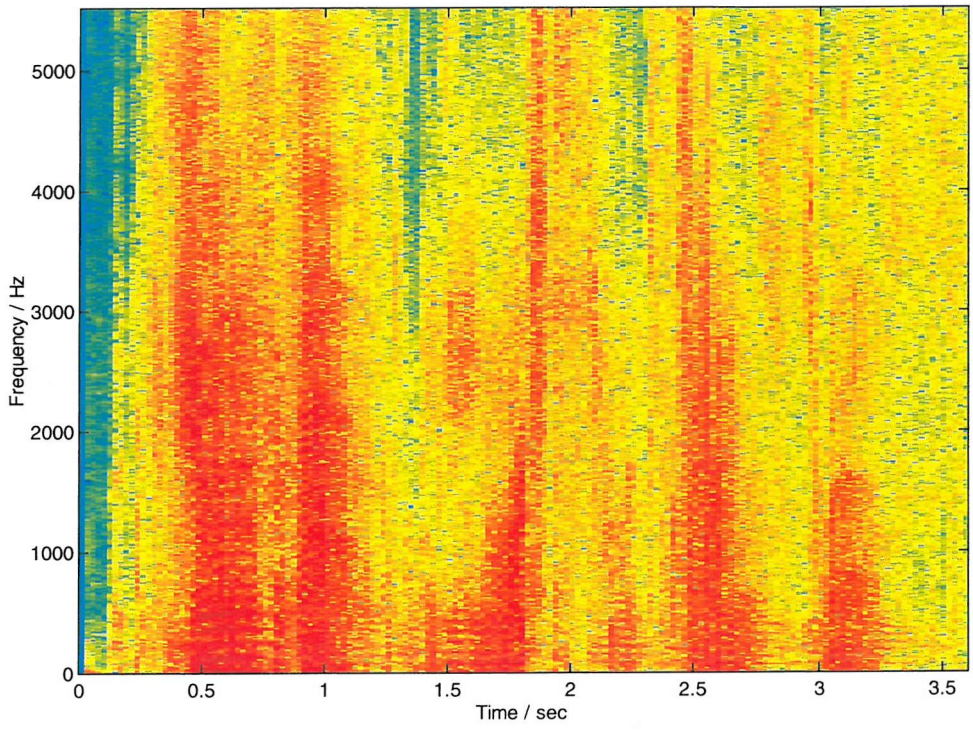

Figure 14 Wigner Distribution of Speech Signal

2.4.5 Relationship between Cohen's Class of Time-Frequency Representations

Previously in this chapter, the WD, spectrogram and scalogram have all been introduced. These TFRs are all bilinear and are members of Cohen's Class (CC) of TFRs. Two equivalent definition's of CC are given in (2-62) and (2-63).

$$P(t,\omega)=\iint WD_x(\tau,\nu)\Pi(\tau-t,\nu-\omega)d\tau\,d\nu \tag{2-62}$$

$$=\frac{1}{2\pi}\iint AF(\tau,\theta)\phi(\tau,\theta)e^{j(\theta t-\omega\tau)}d\tau\,d\omega \tag{2-63}$$

$$AF_x(\tau,\theta)=\frac{1}{2\pi}\iint WD_x(t,\omega)e^{-j(\omega\tau-\theta t)}dt\,d\omega \tag{2-64}$$

The two definitions allow for different interpretations on the relationship between CC TFRs. According to (2-62) a member of CC can be viewed as a 2D convolution operation upon the WD with a kernel function. In equation (2-63) the TFR is defined via the 2D Inverse FT (IFT) of the result of multiplying the Ambiguity Function (AF) (2-64) with another kernel function. This 2D IFT of the TF plane is called the Ambiguity Plane (AP). The relationship between convolution and multiplication was discussed previously for the FT (Section 2.2.3.1). The two different forms of CC follow as a consequence of this. The kernel functions (as defined by (2-63)) for a number of different TFRs are given in Table 3.

The reason for the number of different TFRs is to endeavour to create a TFR with 'desirable' properties, such as marginal properties and finite time/frequency support without the presence of cross-terms.

The cross-terms in the WD tend to be highly oscillatory in nature. This can be seen in the example given in Section 2.4.4.1 where the cross-terms oscillated depending upon the difference in the frequencies of the input sinusoids. It follows that in order to remove them a low-pass kernel function can be used, $\Pi(t,\omega)$ in (2-62). The faster the fall-off rate of the low pass filter used, the greater the degree of smoothing applied to the TF plane and the cross-terms more are suppressed relative to the auto-terms. The smoothing of the TF in this manner has the disadvantage of blurring the auto-terms. As the spectrogram can be shown to be a smoothed version of the WD (Table 3), it too fits into this definition. In the spectrogram case, the cross-terms, while not visually detracting as in the WD, are present in the highly spread auto-

terms. Furthermore, the kernel used to transform the WD to the spectrogram, is the WD of the windowing function itself. Altering the size or shape of the windowing function, changes the size and support of the kernel function and thus its effect in TF.

In order to be able to set the degree of cross-term suppression and auto-term spreading, most CC TFRs (such as the Choi-Williams TFR [Boashash92]) have user selectable parameters.

Table 4 gives the requirements on the kernel function in order for certain time-frequency characteristics to hold. For an arbitrary signal, no TFR can hold all of these properties simultaneously [Cohen95].

The scalogram is not a member of CC as defined by (2-62) [Hlawatsch92], but is related to the WD as given below. As can be seen, the scalogram is related to the WD via affine convolution rather than linear convolution. The kernel function used for the scalogram, WD_h is the same as that used for the spectrogram.

$$|W(t,a)|^2 = \iint WD(\tau,\nu) WD_h\left(\frac{\tau - t}{a}, a\nu\right) d\tau d\nu \qquad (2\text{-}65)$$

$$WD_h(t,\omega) = \int h(t+\tau/2) h^*(t-\tau/2) e^{-j\omega\tau} d\tau$$

Distribution	$\phi(\tau,\theta)$
Wigner Distribution	1
Spectrogram	$\int h^*\left(u-\frac{1}{2}\tau\right)h\left(u+\frac{1}{2}\tau\right)e^{-j\theta u}du$
Choi-Williams	$e^{-\frac{(\tau\theta)^2}{\sigma}}$
Zhao-Atlas-Marks	$g(\tau)\lvert\tau\rvert\frac{\sin a\theta\tau}{a\theta\tau}$
Page	$e^{j\theta\lvert\tau\rvert}$
Born-Jordon	$\frac{2\sin\left(\frac{1}{2}\theta\tau\right)}{\theta\tau}$

Table 3 Definitions of Kernel for Various Time-Frequency Representations

Properties	Kernel
Time-Shift Invariant	Independent of time variable t
Frequency-Shift Invariant	Independent of frequency variable ω
Realness	$\phi(\tau,\theta)=\phi(-\tau,-\theta)$
Time Marginal	$\phi(0,\theta)=1$
Frequency Marginal	$\phi(\tau,0)=1$
Instantaneous Frequency Property	$\phi(0,\theta)=1$ and $\frac{\partial}{\partial\tau}\phi(\tau,\theta)\Big\|_{\tau=0}=0$
Group Delay Property	$\phi(\tau,0)=1$ and $\frac{\partial}{\partial\theta}\phi(\tau,\theta)\Big\|_{\theta=0}=0$

Table 4 Restrictions on the kernel function in order to create a TFR with 'desired' TF properties.

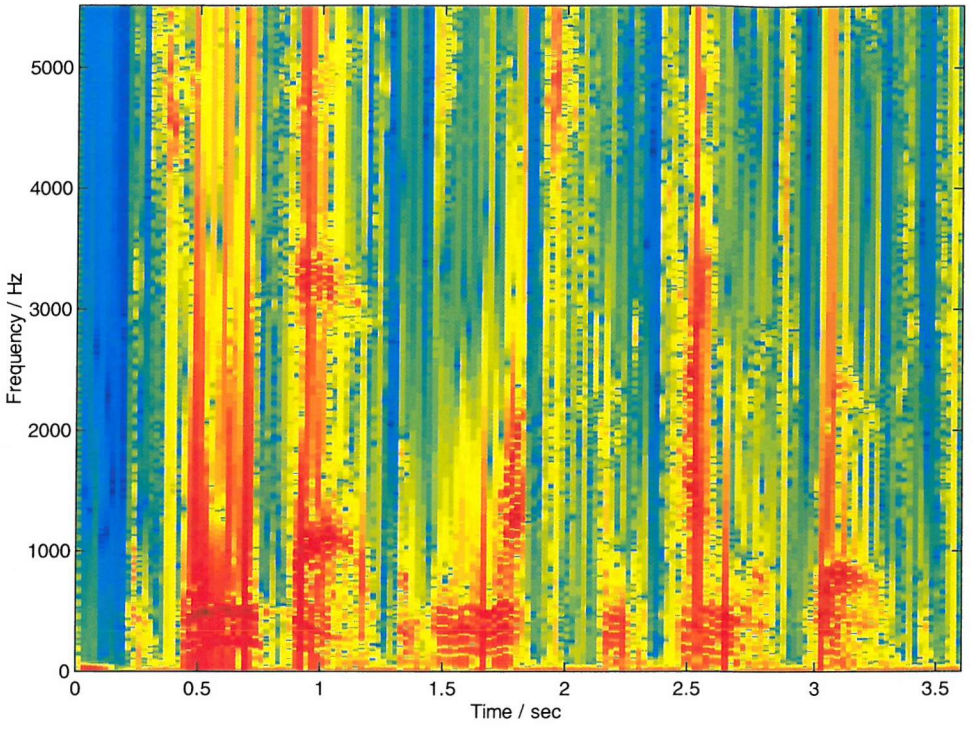

Figure 15 Choi-Williams Distribution for Speech Signal

2.4.6 Inversion

An inversion formula does exist to recover the signal from a CC TFR [Cohen95]. Equation (2-66) shows how a signal can be recovered from a CC TFR, through knowledge of the kernel function. Signal reconstruction can only take place to within an overall phase constant. If the first sample value $x(0)$ is known then (2-66) can be applied directly. Otherwise the phase constant can be selected arbitrarily.

$$x(t) = \frac{1}{2\pi x^*(0)} \iiint \frac{P(\tau,\omega)}{\phi(\tau,\omega)} e^{jt\omega + j\theta\left(\tau - \frac{t}{2}\right)} d\tau \, d\omega \, d\theta \tag{2-66}$$

Implementation of the reconstruction formula is non-trivial if the kernel produced tends to zero. Most AP kernels tend to zero at some locations away from the origin in order to suppress cross-terms, as discussed previously in Section 2.4.5. The result in TF is typically a low pass filter, typically tending to zero away from the origin. The inversion requires the division by the kernel function in the TF domain, and thus numerical errors can result if the kernel has small values.

2.4.7 Limitations of Cohen's Class of Time-Frequency Representations

All practical TFRs must conform to the uncertainty principle, and yet the WD appears to defy this by having well localised harmonics and linear chirps. The WD does however conform to the marginal properties as expressed in Table 3. These place a limit upon the global characteristics of the WD via the FT. As stated previously in this chapter, the FT is itself limited by the uncertainty principle. This restricts frequency resolution according to the length of data. It follows that the frequency resolution of the WD is also limited by the resolution of the FT, and thus the WD does conform to the uncertainty principle.

By designing the kernel function in an appropriate manner, TFR properties which are considered important to the application can be preserved while the cross-terms are suppressed as much as possible [Cohen95]. One example of such a kernel function is the 'Choi-Williams' function. This offers high resolution compared to the spectrogram, and suppresses the cross-terms present in the WD and maintains marginal properties. Figure 15 shows that Choi-Williams (CW) TFR for the speech signal used previously in Section 2.3.2. The improved visual quality can be easily seen when comparing the CW TFR in Figure 15 to the WD in Figure 14 and the spectrogram in Figure 8. The Choi-Williams distribution requires user interaction in the selection of the parameter controlling the degree of cross-term rejection and auto-terms smoothing.

Previously in Section 2.4.2 the properties required for a TFR to be considered as energetic were introduced. These stated that the TFR must be manifestly positive and conform to the marginal properties. The only member of CC which is manifestly positive for all signals is the spectrogram [Qian96]. In order for a TFR to conform to the marginal properties, its kernel function (in the AP) must conform to both $\phi(\tau,0)=1$ (time marginal) and $\phi(0,\theta)=1$ (see Table 4). Since the spectrogram is the only positive member of CC and cannot simultaneously conform to both marginals [Cohen95], it follows that no member of CC can truly be considered as an energetic TFR.

2.5 Conclusions

This chapter has introduced signal invariant TFRs, including the spectrogram, scalogram and WD. The spectrogram and WD can be seen to be members to CC, and as such are bilinear TFRs and exhibit cross-terms. In the extreme case of the WD, these produce distributions that are hard to interpret visually. Although the spectrogram is also bilinear, the cross-terms significantly blur the auto-term components rather than creating visually distracting components. For some members of CC, it is possible to alter a parameter to change the amount of cross-terms present in the resultant distribution. One example of such as TFR is the 'Choi-Williams' distribution. A conceptually simple inverse exists for all CC TFRs.

One key element of CC is that none of the distributions alter their characteristics for a given signal. This characteristic is both a virtue and a limitation. Consider for example a signal which is comprised of tonal components at the start of the signal and is impulsive at the end. The design of the kernel can be targeted to offer good performance for either the tonal or the impulsive component, but not both simultaneously. The inability to be able to alter characteristics to match that of the signal, and the fact that data independent TFRs cannot be considered as an energetic TFRs as defined by (2-43) and (2-44), are the two fundamental limitations of traditional TF analysis.

Chapter 3 Data Adaptive Time-Frequency Analysis

3.1 Introduction

In Chapter 2 the theory of signal independent TF signal analysis was introduced. Although allowing the time-varying characteristics of a signal to be displayed in a manner previously not possible with the FT, all the linear and bilinear introduced TFRs posses some inherent limitations. The limitation for linear distributions is that the decomposition they provide cannot be considered as energetic. For the bilinear CC of TFRs performance is either limited by the presence of cross-terms (as in the case of WD), or by the resolution degrading effect of any kernel function used to suppress them.

A common link between all the TFRs described in Chapter 2 was that they didn't adapt their characteristics to suit the input signal. User intervention is required to optimise the performance of the TFR to suit those of the signal, for example in the case of the spectrogram by selecting the length and shape of the window to match the characteristics of the signal. A method of optimising the TFR to the signal without the need for user intervention is desirable. In this chapter a number of different signal dependant TF techniques are described.

3.2 Data Adaptive Kernel Design

Although the TFRs developed in the previous chapter are not explicitly data adaptive, in order to optimise their performance user intervention in the form of altering some windowing parameter is often required. The altering of the windowing function or kernel parameter implicitly sets the amount of auto and cross terms that are passed in the TF plane and thus alters the visual characteristics of the TFR. An alternative approach is to design a kernel that is algorithmically adapted to the characteristics of the signal. By automatically constructing the kernel in the AP, the kernel could optimise the amount of auto-terms whilst minimising the amount of cross-terms passed to the TF domain. As stated in the previous chapter, the oscillatory nature of cross-terms in TF means that they tend to be located away from

the origin in the AP. This property aids automatic kernel construction in the AP, since auto-terms tend to be less oscillatory in TF being located closer to the origin of the AP as compared to the cross-terms. Baraniuk and Jones [Baraniuk93] first formulated one method of achieving this goal by solving the optimisation problem given in (3-1), where the optimal kernel function $\phi(\tau,\theta)$ is radially non-increasing and centred at the origin.

$$\max_{\Phi} \iint \left| AF(\tau,\theta)\phi(\tau,\theta) \right|^2 d\theta \; d\tau \text{ subject to } \phi(0,0)=1 \tag{3-1}$$

$$\frac{1}{2\pi} \iint \left| \phi(\tau,\theta) \right|^2 d\theta \; d\tau \le \alpha, \; \alpha \ge 0$$

The constraints placed upon the kernel force it to be a low-pass filter in TF of maximum volume α. In order for this approach to be effective the auto and cross terms must be separated from one another, *i.e.* the cross-terms and auto-terms should be located in separate regions of the AP. If the dimensions of the kernel are changed such that it encroaches upon the region between the auto and cross-terms, then the result is an increase in kernel volume with little increase in the signal energy covered by the kernel. Hence, the optimisation procedure favours kernels concentrated at the origin, precisely the region of the AP where auto components are present. Additional constraints can be added to the kernel [Baranuik93] in order that the optimised kernel function conforms to desirable TF characteristics, such as the time and frequency marginals. A fast algorithm for implementing the optimal kernel approach was developed to reduce computation time [Baraniuk94]. Unlike the rest of the data adaptive TFRs covered in the remainder of this chapter, since the technique is based upon kernel design a simple inverse exists. An unmodified version of the inversion formula given in Section 2.4.6 (Equation (2-66)) allows signal reconstruction. A locally adaptive version of the optimal kernel was designed to adapt to signals whose character changes with time was developed by [Jones95]. All these methods have the limitation that they require user specification of the volume for the kernel. In this respect, although allowing the CC TFRs kernel to be signal adaptive, they still require some degree of user intervention.

3.3 Cohen Posch Distributions

Prior to the introduction of data adaptive kernels, Cohen and Zaparovanny proposed a simple data adaptive TFR formulation given in the context of quantum physics (the field in which the WD originated) [Cohen79]. This stated that a TFR function could be formulated as given in (3-2), where Ω is a signal dependant function.

$$P(t,\omega)=|x(t)|^2\,|X(\omega)|^2\,\Omega(u(t),v(\omega)) \tag{3-2}$$

where

$$u(t)=\int_{-\infty}^{t}|x(\tau)|^2\,d\tau \qquad v(\omega)=\int_{-\infty}^{\omega}|X(\zeta)|^2\,d\zeta \tag{3-3}$$

In order for this formulation to offer an improved TFR compared to the WD, three desirable properties are required: the time and frequency marginals (see Table 1) must be conformed to *and* the distribution must be manifestly positive. Cohen and Posch [Cohen85] showed that a simple kernel conforming to these was any positive function Ω satisfying (3-4). Cohen [Cohen95] commented that it was sufficient to define $\Omega(u,v)$ only for $0\le u,v\ge 1$ and that u and v are the cumulative marginals, the sum of the densities up to a given time and frequency value.

$$\int_0^1\Omega(u,v)dv=1 \qquad \int_0^1\Omega(u,v)du=1 \tag{3-4}$$

The TFRs constructed using (3-2) and (3-4) were the first to conform to the marginals and be manifestly positive. TFRs which posses both these properties are called 'Cohen-Posch' (CP) TFRs. Loughlin *et al.* [Loughlin94a] extended this definition to include other desirable properties such as the moments of a TFR defined via the FT.

3.3.1 Numerical Computation

The constraints placed upon the kernel function for CP TFRs are such there are an infinite number of CP TFRs for any given signal. Loughlin *et al.* [Loughlin94a] accomplished the task of finding the most appropriate CP TFR for a given signal by choosing an initial positive estimate TFR (typically a spectrogram) and then minimising the cross entropy subject to the TFR conforming to the marginals. The

algorithm to create a Minimum Cross Entropy (MCE) CP TFR is given in (3-5). Additional constraints other than the simple time and frequency marginals can also be imposed if required, increasing the number of steps in each iteration. Fonollosa [Fonollosa96] extended Loughlin's original approach to include rotated TF (or fractional) marginals.

$$P^{(2n)}(t,\omega) = P^{(2n-1)}(t,\omega)\frac{|X(\omega)|^2}{\int P^{(2n-1)}(t,\omega)dt} \tag{3-5}$$

$$P^{(2n+1)}(t,\omega) = P^{(2n)}(t,\omega)\frac{|x(t)|^2}{\int P^{(2n)}(t,\omega)dt} \tag{3-6}$$

The algorithm works by, at each iteration, enforcing a marginal upon the TFR, the frequency marginal in (3-5) and the time marginal in (3-6). This approach can be shown to be a 'Projection onto Convex Sets' (POCS) algorithm [Clarkson95], and as such is guaranteed to converge to a solution if the system is convex. Appendix 1 proves that the sets, defined by positivity and conforming to the time and frequency marginals, are convex. Using a POCS method a CP TFR for the speech signal used previously is given in Figure 16. Compare Figure 16 to the spectrogram of the speech signal (Figure 8). A marked improvement in terms of the ability to resolve the start and stop of components can be seen. The impulsive components within single words can now be seen, whereas in the spectrogram they were completely lost due to the windowing function. Figure 16 is also free from the visually distracting cross-terms present in the WD and CW TFRs shown previously in Figure 14 and Figure 15. However, the CP TFRs although manifestly positive, can be seen not to track FM components of the signal as well the CW given in Figure 15. This is because the CP TFRs favours harmonic and impulse components over others. This is a limitation of CP TFRs, and discussed in more detail in Section 3.3.3.

An alterative approach to the formulation of positive TFRs was given by [Pitton93]. Using the Evolutionary Spectrum (ES) (defined by Priestley [Priestley81]) Pitton *et al.* [Pitton93] formulated the construction of a CP TFR by suggesting that a spectrogram can be considered as a smoothed CP TFR, see (3-7). In this equation, $H(\omega)$ is the FT of the time-limiting windowing function used and $w(t)$ is a smoothing function

acting only upon time in the TF plane. Pitton *et al.* [Pitton93] proceeded to describe an estimation procedure for $P(t,\omega)$ based on deconvolving $|S(t,\omega)|^2$.

$$|S(t,\omega)|^2 = \iint r(t-\tau,\omega-\theta)P(\tau,\theta)d\tau\, d\theta \tag{3-7}$$

$$\text{where } r(t,\omega) = \frac{1}{2\omega} w(t)|H(\omega)|^2$$

This view of the spectrogram as a blurred positive TFR is returned to in Section 3.4. CP TFRs have been used in a wide range of different applications including speech and condition monitoring [Pitton94].

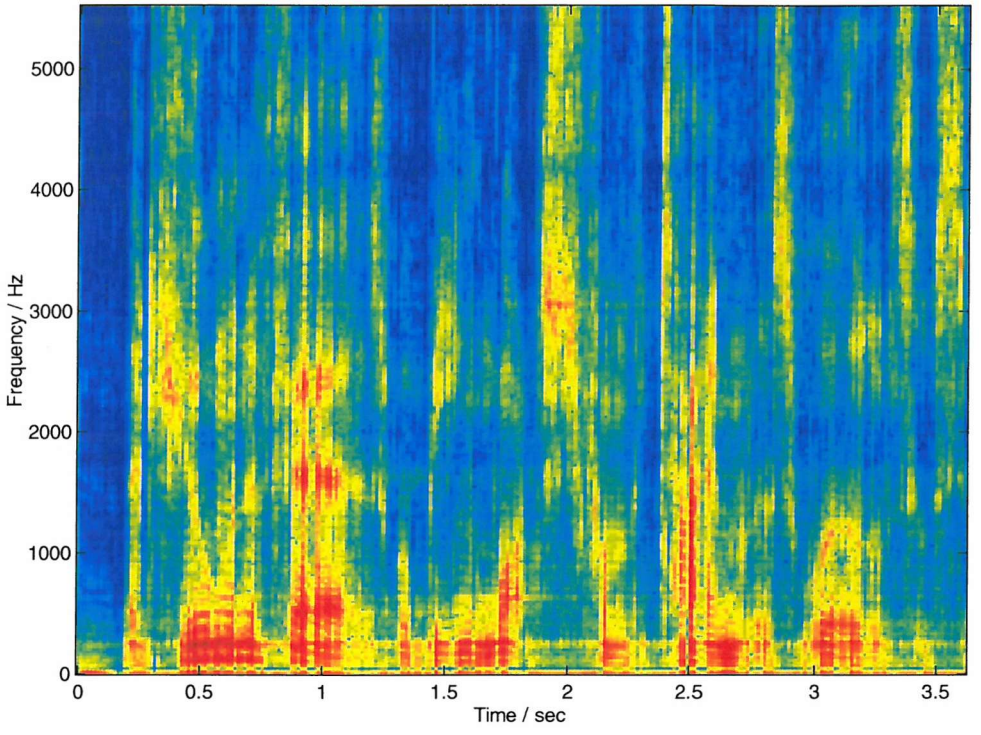

Figure 16 Cohen Posch TFR of Speech Signal

3.3.2 Inversion

Two different approaches for signal reconstruction from CP TFRs have been proposed in the literature. The first [Francos93] is a parameter based approach, which assumes the signal is comprised of a number of amplitude and frequency

modulated harmonic signals. This method extracts the amplitude and frequency modulation laws for each component via a CP TFR. The method is ultimately limited to signals which are well matched to the model. The second approach is a non-parameter based method [Shah98]. This approach is based upon finding a TFR whose squared magnitude conforms to the desired marginals, is invertible and the closest (in a minimum least squared sense) to a given prior estimate. An iterative process is described, imposing at each iteration the time and frequency marginals, as well as the invertibility of the resultant TFR.

3.3.3 Limitations of Cohen Posch Time-Frequency Representations

Although CP TFRs seem to offer good performance characteristics, (in that they are positive and satisfy the marginals in time and frequency), these properties may not be sufficient to guarantee good visual performance. This was demonstrated by Fonollosa [Fonollosa96] where the CP TFR for a linear FM signal was shown to be significantly spread. In fact the CP TFR is ideal for showing the frequency of tones and the time location of impulses, but in general produces poor distributions for FM signals. With a view to improving the visual qualities of CP TFRs for linear FM signals, the requirement for fractional marginals to be conformed can also be introduced [Fonollosa96]. However for non-linear FM signals, no simple way exists to improve the visual quality of the CP TFRs.

3.4 Reassigned Spectrogram

3.4.1 Introduction

Equation (3-7) enabled the spectrogram to be interpreted as a blurred version of a CP TFR. Rihaczek (given in [Ackroyd71]) defined the spectrogram in a similar fashion but using different blurring functions, given in (3-8). In this equation $d(t,\omega)$ is dependant upon the signal, and $g(t.\omega)$ is dependant upon the windowing function used in the computation of the ST-FT (2-36).

$$|S(t,\omega)|^2 = \iint g(t-\tau,\omega-\theta)d(\tau,\theta)d\tau\, d\theta \tag{3-8}$$

where $g(t,\omega)=h(t)H(\omega)e^{-j\omega t}$ and $d(t,\omega)=x(t)X(\omega)e^{-j\omega t}$

Kodera [Kodera76] noted that although the integral (3-8) is taken over the whole range of the TF plane, the main contribution to the spectrogram comes from a limited region, defined by the bandwidth (in time and frequency) of the windowing function $g(t,\omega)$. Kodera argued that since the bandwidth of the windowing function is spreading the underlying distribution, then the energy located the point (t,ω) of the spectrogram would be better placed at the local moment, $\left(t_g(t,\omega),\omega_i(t,\omega)\right)$. By reallocating the energy to this new point, the spreading due to the bandwidth of the windowing function is reduced. There follows a brief justification for this.

Although both $d(t,\omega)$ and $h(t,\omega)$ are complex functions, the imaginary part of the integration (3-8) must be zero since the spectrogram is real. With regards to (3-8) it is useful to define the functions $\eta(\theta,t,\omega)$ and $\mu(\tau,t,\omega)$ given in (3-9) and (3-10), so that (3-11) holds.

$$\eta(\theta,t,\omega)=\int \mathrm{Re}\left[d(\tau,\theta)g(t-\tau,\omega-\theta)\right]d\tau \tag{3-9}$$

$$\mu(\tau,t,\omega)=\int \mathrm{Re}\left[d(\tau,\theta)g(t-\tau,\omega-\theta)\right]d\theta \tag{3-10}$$

$$\left|S(t,\omega)\right|^2=\int \eta(\theta,t,\omega)d\theta=\int \mu(\tau,t,\omega)d\tau \tag{3-11}$$

If $\eta(\theta,t,\omega)$ is sharply peaked about a specific value of frequency, then Kodera [Kodera76] stated that this value will give '[a] better representation of the frequency at which the maximum power is situated'. The peak of the function of $\eta(\theta,t,\omega)$ is assumed to correspond to its first moment, (3-12). A similar procedure can be applied to $\mu(\tau,t,\omega)$ to yield (3-13). (Kodera *et al* later showed [Kodera78] that by using the moment as opposed to the maximum value, that the instantaneous frequency and group delay parameters computed using these functions gave intuitively correct results under a range of signal bandwidths.)

It can be shown that this local moment operation is analogous to computing a value of local instantaneous frequency ω_i and group delay t_g at each point of time-frequency [Kodera76], since (3-12) and (3-13) can be expressed in terms of the derivative of the phase of the ST-FT (3-14). It should be noted that Kodera used a

slightly different definition of the ST-FT, (different phase constant) than given previously in (2-26), this is given in (3-15). This is analogous to the way in which IF and GD were defined in terms of the derivative of the phase of the analytic signal and FT (see Chapter 2 Section 2.2.3.5).

$$\omega_i(t,\omega)=\frac{\int\theta\eta(\theta,t,\omega)d\theta}{\int\eta(\theta,t,f)d\theta} \tag{3-12}$$

$$t_g(t,\omega)=\frac{\int\tau\mu(\tau,t,\omega)d\tau}{\int\mu(\xi,t,\omega)d\tau} \tag{3-13}$$

$$S(t,\omega)=|S(t,\omega)|e^{j\phi(t,\omega)} \rightarrow \begin{matrix}\omega_t(t,\omega)=\dfrac{\partial\phi(t,\omega)}{\partial t}\\ t_g(t,\omega)=t-\dfrac{\partial\phi(t,\omega)}{\partial\omega}\end{matrix} \tag{3-14}$$

$$S(t,\omega)=e^{j\omega\tau}\int x(\tau)h(t-\tau)e^{j2\pi\omega\tau}d\tau \tag{3-15}$$

Equations (3-12) and (3-13) define the reassigned spectrogram co-ordinates (the collective name given to the local instantaneous frequency and group delay parameters). There are a number of alternative ways to compute these functions and they are detailed in Section 3.4.3.

To demonstrate the improved visual qualities that plotting these co-ordinates can produce, Figure 17 shows the reassigned co-ordinates (also called the reassigned vector field [Auger95, Chassande97]) for the linear chirp. The spread of the TFR due to the windowing function used has been reduced, with the co-ordinates being concentrated along the line of the chirp. The chirp has a high BT product and thus the IF and GD curves coincide [Kodera78].

Although the reassigned co-ordinates provide an estimate of the value of the IF and GD at a point in TF, they do not provide amplitude information. The original methodology defined by Kodera [Kodera76] was to reallocate the energy of the signal in TF to a new location which was relatively unaffected by the windowing function. As such the reassigned co-ordinates, although providing the location of these new points, do not provide a value for the energy that should be placed there. Kodera [Kodera76] stated that once the new co-ordinates were computed, that the

energy should be reallocated to this location. A more mathematical definition was provided by Auger *et al.* via a double convolution with a set of delta functions given in (3-16), where $RS(t,\omega)$ is the reassigned spectrogram. The double convolution with a set of a 2D Dirac delta functions centred upon the locations provided by the reassigned co-ordinates, simply picks out these locations in the TF plane. By multiplying the result by the original spectrogram, the result is to create a TFR which only has non-zero values at locations dictated by the reassigned co-ordinates, with the energy at this location provided by the original spectrogram.

The reassigned spectrogram for the linear chirp signal is shown in Figure 18. Figure 19 shows the reassigned spectrogram for the speech signal. As compared to the spectrogram of the same signal given in Figure 6, the reassigned spectrogram appears to offer improved temporal support with tighter, better resolved impulses. The reassigned spectrogram also appears to offer improved frequency resolution, with the frequency modulation law and harmonic components of each word more easily seen than with the spectrogram.

$$RS(t,\omega)=\iint\left|S(\tau,\theta)\right|^2\delta\left(t-t_g(\tau,\theta)\right)\delta(\omega-\omega_i(\tau,\theta))d\tau\,d\theta \tag{3-16}$$

The reassigned spectrogram has been used to improve the visual characteristics of TFRs of speech [Plante98] and of structural response data [Hammond96].

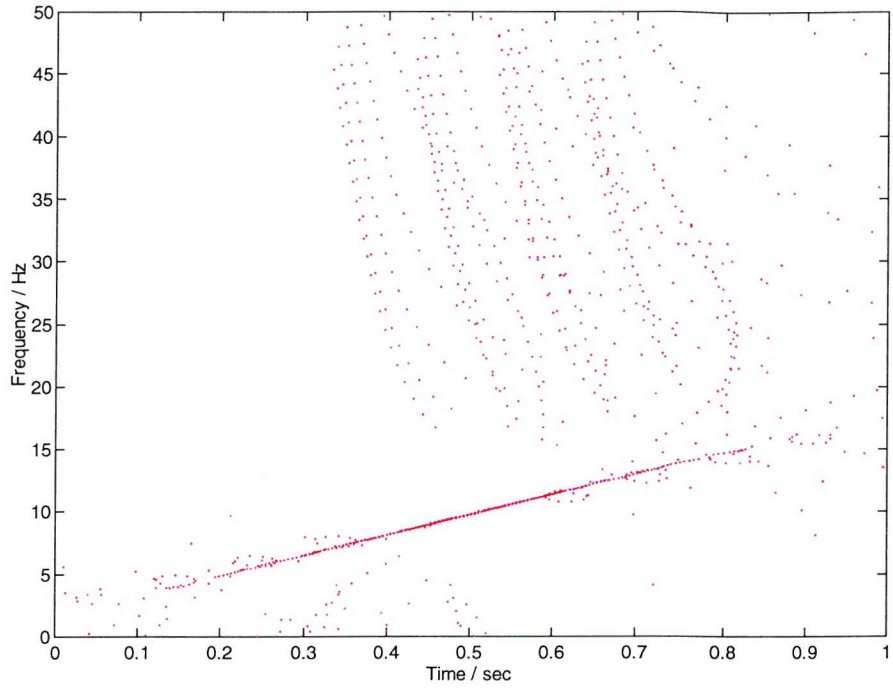

Figure 17 Reassigned Spectrogram Co-ordinates of a Linear FM Signal

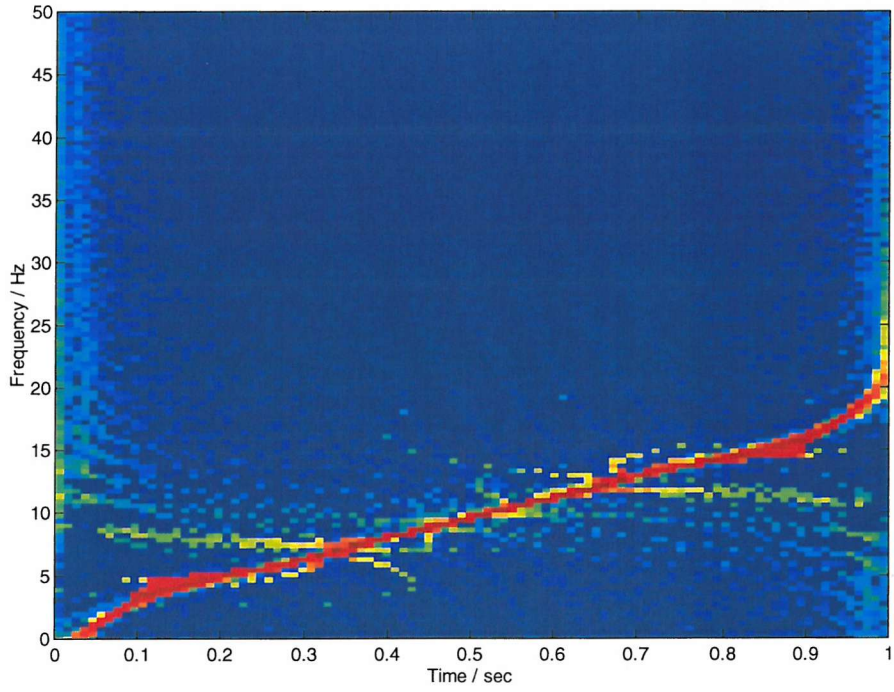

Figure 18 Reassigned Spectrogram of a Linear FM Signal

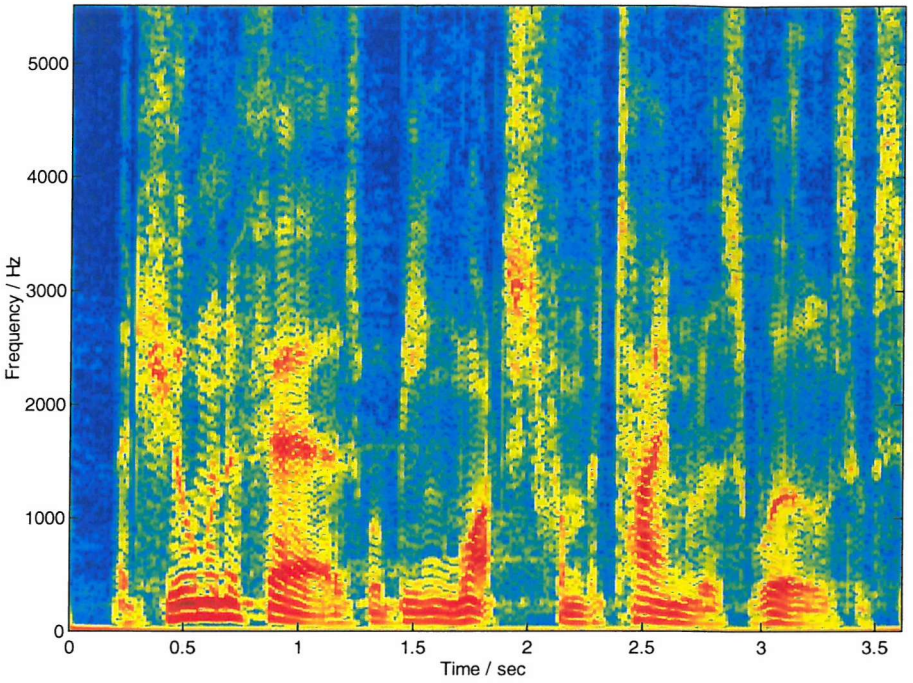

Figure 19 Reassigned Spectrogram of the Speech Signal

3.4.2 Extension to Other Members of Cohen's Class

Given the improved temporal and spectral performance which the reassigned spectrogram appears to offer over the spectrogram, it is logical to ask whether a similar procedure can be applied to other TFRs. The extension to other TFRs was conducted by Moss and Hammond [Moss94] for the pseudo-WD and later by Auger *et al.* [Auger95] for a range of bilinear TFRs. The reassigned co-ordinates for an arbitrary member of CC of TFRs (as defined by (2-66)) are given in (3-17). The reassigned CC TFR is defined in (3-18) and is analogous to the method used for the spectrogram.

$$t_g(t,\omega) = t - \frac{\iint \tau\, \Pi(\tau,\theta)\, WD(t-\tau,\omega-\theta)\, d\tau\, d\theta}{\iint \Pi(\tau,\theta)\, WD(t-\tau,\omega-\theta)\, d\tau\, d\theta} \qquad (3\text{-}17)$$

$$\omega_i(t,\omega) = \omega - \frac{\iint \theta\, \Pi(\tau,\theta)\, WD(t-\tau,\omega-\theta)\, d\tau\, d\theta}{\iint \Pi(\tau,\theta)\, WD(t-\tau,\omega-\theta)\, d\tau\, d\theta}$$

$$RP(t,\omega)=\iint P(\tau,\theta)\delta\left(t-t_g(\tau,\omega)\right)\delta\left(\omega-\omega_i(\tau,\omega)\right)d\tau\ d\theta \tag{3-18}$$

The objective of the reassigned procedure is to mitigate the effect of the windowing function used. It follows that if no windowing function is used in constructing the TFR then reassignment is unnecessary. This leads to the observation that the reassignment operation has no effect upon the WD. In the case of the Pseudo-WD, where only time domain smoothing is used, then no change is made to the frequency co-ordinate [Moss94].

This thesis shall concentrate on the reassigned spectrogram in preference to other reassigned distributions. One reason for this is that the resultant reassigned TFRs all appear similar for different distributions [Moss94, Auger96].

3.4.3 Theoretical Performance of the Reassigned Spectrogram

3.4.3.1 Example of the Instantaneous Frequency and Group Delay Computed via the Fourier Transform

In Chapter 2 Sections 2.2.3.3 and 2.2.3.5 the concepts of the Bandwidth-Time (BT) product and Instantaneous Frequency (IF) and Group Delay (GD) were introduced. It was noted that, although IF and GD attempt measure the same thing, they do not always agree. This is illustrated by in the following example. The continuous signal given in (3-19) is a Gaussian windowed linear chirp, the FT of which is given in (3-20) [Cohen95].

$$x(t)=e^{-\frac{\gamma t^2}{2}}e^{j\left(\omega_0 t+\frac{\beta t^2}{2}\right)} \tag{3-19}$$

$$X(\omega)=\sqrt{\frac{j\beta+\gamma}{\beta^2+\gamma^2}}\,e^{\frac{j(\beta-j\gamma)(\omega-\omega_0)^2}{2(\beta^2+\gamma^2)}} \tag{3-20}$$

The IF and GD of this signal (as computed via equations (2-13), (2-14)) are given in (3-21) and (3-22). For both (3-21) and (3-22) to describe the same line in TF, it can be seen that $\beta^2 >> \gamma^2$. This condition states that for the IF and GD to be the same, the rate of change of the phase of the signal must vary much faster than the envelope. This is the definition of an asymptotic signal as given in Section 2.2.3.5. These signals

typically have a high BT product and thus both the IF and GD laws describe the same line in TF [Kodera78].

$$\omega_i(t) = \omega_0 + \beta t \tag{3-21}$$

$$t_g(\omega) = \frac{\beta(\omega - \omega_0)}{\gamma^2 + \beta^2} \Rightarrow \omega = \beta\left(1 + \frac{\gamma^2}{\beta^2}\right) t_g + \omega_0 \tag{3-22}$$

3.4.3.2 Performance of the Reassigned Co-ordinates under Changing Windowing Bandwidth

In this section the effect that different windowing bandwidths have upon the reassigned co-ordinates is considered.

Consider the definition of the ST-FT, of a signal with respect to a windowing function, $h(t)$ which is zero outside of the region, $[-\Delta t/2, \Delta t/2]$ as given by (3-23).

$$S(t,\omega) = \frac{1}{\sqrt{2\pi}} \int_{t-(\Delta t/2)}^{t+(\Delta t/2)} x(\tau) h(t-\tau) e^{-j\omega\tau} d\tau \tag{3-23}$$

For sufficiently small Δt, the signal can be assumed to be constant within the region $[t - \Delta t/2, t + \Delta t/2]$. This assumption leads to the expression given in (3-24), *i.e.* the ST-FT is equal to a scaled version of the signal. The reassigned co-ordinates for this signal, given in (3-26).

$$S(t,\omega) = \frac{1}{\sqrt{2\pi}} x(t) h(0) e^{-j\omega t} \tag{3-24}$$

$$x(t) = |x(t)| e^{j\varphi(t)} \tag{3-25}$$

$$\omega_i(t,\omega) = \frac{d\varphi(t)}{dt} \qquad t_g(t,\omega) = t \tag{3-26}$$

where $\varphi(t)$ is the signal's phase. So the phase of the ST-FT is the phase of the signal, therefore the reassigned co-ordinates are assigned along the instantaneous frequency as defined in (2-13). If very fine frequency resolution is required, then $\Delta t \rightarrow \infty$. In this case, (3-23) reduces to the form given in (3-27). Consider the ST-FT written in

terms of frequency domain functions, (3-28). As the windowing function becomes broader in time, then its FT becomes more impulse like. In which case the ST-FT becomes a scaled version of the signals FT. Hence the phase of the ST-FT approaches the phase of the signals FT.

$$S(t,\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x(\tau)h(t-\tau)e^{-j\omega\tau}d\tau \tag{3-27}$$

$$S(t,\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}X(\hat{\omega})H(\hat{\omega}-\omega)e^{j(\hat{\omega}-\omega)t}d\hat{\omega}\approx H(0)X(\omega) \tag{3-28}$$

$$\begin{aligned}\omega_i(t,\omega)&=\omega\\ t_g(t,\omega)&=-\frac{d\vartheta(\omega)}{d\omega}\end{aligned} \tag{3-29}$$

where $\vartheta(\omega)$ is the phase of the FT of the signal. Then the reassigned co-ordinates (3-29) are assigned along the group delay of the signal, as defined by the FT (2-14).

Therefore, as the temporal resolution of the ST-FT varies from very fine to very coarse, the reassigned co-ordinates vary from the IF law to the GD law. Kodera [Kodera78] compared this to the performance of a number of other algorithms, showing that the reassigned spectrogram was the only one to offer intuitively correct performance regardless of the length of the window used.

3.4.3.3 Analytical Example of the Reassigned Spectrogram

This section briefly describes the effect upon the reassigned co-ordinates of the BT product of the signal. The reassigned co-ordinates for the linear chirp, (3-19), using the ST-FT with a Gaussian time-localising window, (3-30), are given in (3-31) and (3-32) [Kodera78].

$$h(t)=\left(\frac{2}{\alpha}\right)^{\frac{1}{4}}e^{-\frac{t^2}{2\alpha}} \tag{3-30}$$

$$t_g(t,\omega)=-\frac{1}{C}\left[(1+\alpha\gamma)t+(\omega-\omega_0)\alpha^2\beta\right] \tag{3-31}$$

$$\omega_i(t,\omega)=\frac{1}{C}\left[\left(\beta t+(1+\alpha\gamma)\omega_0+\alpha^2\beta^2\omega+\alpha\gamma(1+\alpha\gamma)\omega\right)\right] \tag{3-32}$$

$$C=1+2\alpha\gamma+\alpha^2\left(\beta^2+\gamma^2\right) \tag{3-33}$$

If the temporal resolution of the ST-FT (as defined by the spread of the windowing function), Δt, is much smaller than the duration bandwidth of the signal, T, (*i.e.* $\Delta t << T$) then it follows that $\alpha\gamma \to 0$. If this is the case, then (3-32) can be shown to satisfy $\omega_i(t,\omega) \approx \beta t_g(t,\omega)$. Therefore, the reassigned co-ordinates follow the instantaneous frequency law [Koder78].

Similarly, if the reciprocal temporal resolution of the ST-FT, is much larger than the spectral bandwidth of the signal, B, (*i.e.* $\Delta t >> 1/B$) then it can be shown that $\gamma/\alpha\left(\beta^2+\gamma^2\right) \to 0$. If this is the case then $\omega_i(t,\omega) \approx \frac{\beta^2+\gamma^2}{\beta} t_g(t,\omega)$, which is the group delay law [Kodera78].

Kodera [Kodera78] commented than since $BT > 1/2$ that, $\Delta t << T$ or $\Delta t >> 1/B$ is almost always satisfied. Thus the reassigned co-ordinate tend to either follow the instantaneous frequency or group delay laws. It therefore follows that the reassigned spectrogram fails to totally remove the effect of the windowing function, since the reassigned co-ordinates alter with different window bandwidths. However if the signal is asymptotic, so the IF and GD laws coincide, then the reassignment will be invariant to the window used.

3.4.4 Computation of the Reassigned Spectrogram Co-ordinates

This section examines the different computational techniques used to compute the reassigned co-ordinates.

3.4.4.1 Phase Difference Methods

The reassigned co-ordinates can be formulated either in terms of a local moment operation, (3-12) and (3-13), or in terms of the partial derivatives of the phase of the ST-FT (3-14). The latter definition allows for a simpler digital implementation. The most basic of which is to approximate the phase derivative using a first order numerical difference [Kodera76]. Considering the discrete ST-FT, as computed via

(2-36), the reassigned co-ordinates can be estimated as given in (3-34), where f_s is the sampling rate, N is the length of the windowing function used and L is the overlap.

$$\omega_i[n,k]=\left(\frac{f_s}{N-L}\right)(\phi[n+1,k]-\phi[n,k]) \qquad (3\text{-}34)$$
$$t_g[n,k]=\frac{nL-(\phi[n,k+1]-\phi[n,k])}{f_s}$$

Although computationally simple these first order differences offer poor estimates of the derivatives of the phase. This method is also prone to errors due to phase unwrapping problems, which becomes worse as the overlap in the ST-FT is reduced. Alternative methods of computing the reassigned co-ordinates are generally preferred [Auger95].

3.4.4.2 Analytical Derivative Methods

An alternative to the difference method is to use the reassignment operators as defined by Auger and Flandrin [Auger95] in (3-17). By inserting the kernel function that creates the spectrogram from the WD, the expressions for the reassigned co-ordinates can be simplified to (3-35), where $S_{h(t)}(t,\omega)$ is the ST-FT computed with the time window $h(t)$ [Auger95]. Using this formulation a direct estimate of the reassigned co-ordinates can be achieved, requiring the computation of three ST-FTs, rather than one.

$$\omega_i(t,\omega)=\omega-\mathrm{Im}\left[\frac{S_{\frac{dh(t)}{dt}}(t,\omega)S^*_{h(t)}(t,\omega)}{\left|S_{h(t)}(t,\omega)\right|^2}\right] \qquad (3\text{-}35)$$
$$t_g(t,\omega)=t+\mathrm{Re}\left[\frac{S_{t\,h(t)}(t,\omega)S^*_{h(t)}(t,\omega)}{\left|S_{h(t)}(t,\omega)\right|^2}\right]$$

Reductions in the amount of computation required can be achieved if a Gaussian window is used in the computation of the ST-FT. Then the computation of the reassignment operators requires computing only one additional ST-FT, since a Gaussian is simply related to its derivative, as given in (3-36). It follows therefore

that the ST-FT using the windowing function $\frac{dh(t)}{dt}$ can be calculated from the ST-FT using the windowing function $t \cdot h(t)$ via (3-37).

$$h(t) = \sqrt{\alpha} e^{-\frac{t^2}{2\alpha}} \Rightarrow \frac{dh(t)}{dt} = -\frac{1}{\sqrt{\alpha}} te^{-\frac{t^2}{2\alpha}} = \frac{t}{\alpha} h(t) \tag{3-36}$$

$$S_{\frac{dh(t)}{dt}}(t,\omega) = \frac{1}{\alpha} S_{th(t)}(t,\omega) \tag{3-37}$$

This method of computing the reassigned co-ordinates is preferred over the numerical difference method given in Section 3.4.4.1 for a number of reasons. Firstly there is no need to unwrap the phase in order to accurately estimate the gradient. Secondly the derivative is estimated directly, so avoiding the use of finite differences. Thirdly the performance of the algorithm is robust with respect to the overlap used. This is method that was used to generate Figure 17 and Figure 18.

3.4.4.3 Recursive Implementation

Implementing the ST-FT in a recursive manner can further reduce the computational load. Since the ST-FT is computed using an overlapping time window, there is a degree of shared signal information between successive time slices. Richard *et al.* [Richard97] describe how this can be used to reduce the computation required. This method requires that the window has a suitable Fourier series decomposition.

3.4.5 Links with Ridge and Skeleton approach

The reassigned spectrogram reallocates signal energy along a line described by either the IF, GD, or somewhere between the two. An alternative is to attempt to extract the IF laws present in the signal directly, without computing the reassigned co-ordinates. One approach is to use the phase of the ST-FT (or Wavelet) as defined in Chapter 2 [Guillemain91, Delprat92, Lopes1997]. A simple form of this algorithm is outlined here. Consider the tonal signal given in (3-38) the ST-FT can be written as (3-39). By taking the derivative of the phase of the ST-FT an estimate of the frequency of the tone can be obtained. Furthermore, knowledge of the FT of the windowing function used allows estimation of the amplitude of the signal.

$$x(t) = Ae^{j\omega_o t} \tag{3-38}$$

$$S(t,\omega)=\frac{1}{\sqrt{2\pi}}AH(\omega-\omega_0)e^{-j(\omega-\omega_0)t} \tag{3-39}$$

Most signals are multi-component and not perfect tones. Therefore a more algorithmic approach is required in order to extract all IF laws [Guillemain91]. The task reduces to finding the points of the derivative of the phase of the ST-FT such that (3-40) holds. The iterative algorithm given in (3-41) converges towards a solution of the above equation given an initial estimate of ω.

$$\omega=\frac{\partial\phi(t,\omega)}{\partial t} \tag{3-40}$$

$$\omega_{i+1}=\frac{\partial\phi(t,\omega)}{\partial t} \tag{3-41}$$

In order to extract the appropriate number of IF laws, multiple starting values for the iteration (3-41) must be taken, resulting in a number IF laws for each signal. More advanced methods of finding the IF laws of signals are also available [Guillemain96, Delprat92], these are based upon a similar method but using non-harmonic signals. This method of IF extraction has been called the 'Ridge and Skeleton' approach due to the form of the images created [Guillemain96].

The similarity between the ridge and skeleton and the reassigned spectrogram approach has been commented on independently by [Lopes97] and [Chassande97]. It can be seen that the reassigned IF co-ordinate, as calculated through the derivative of the phase of the ST-FT (3-14), is the first stage of the ridge and skeleton approach of (3-41) using regularly spaced initial estimates as defined by the spacing of the DFT.

3.5 Conclusions

In this chapter signal adaptive TFRs have been outlined. By virtue of their ability to match the characteristics of the signal they offer improved visual performance over the signal independent TFRs discussed in Chapter 2. In the case of the optimal kernel function as given by Baraniuk [Baraniuk93], although the kernel function is data adaptive, user intervention is still require to provide a suitable volume for the kernel function. Cohen-Posch TFRs, as originally defined, seem to provide an ideal TFR, by requiring that the distribution be both positive and conform to the TF

marginals [Cohen85]. However, poor performance can result if only the time and frequency marginals are used for chirp or non-linear FM signals [Fonollosa96]. The reassigned spectrogram does not suffer from this limitation, offering good performance for a wide range of signals.

Inversion formulas exist for both the data adaptive kernel functions and CP TFRs, but to date no inversion of the reassigned spectrogram has been published. In the following chapter, one method of recovering the signal from the reassigned co-ordinates is introduced.

Chapter 4 Signal Reconstruction from Phase of the ST-FT

4.1 Introduction

In Chapter 3 the concept of reassigned TFRs was introduced. These TFRs reduce the effect of the windowing function by reallocating the energy of a point in TF to a new location dependant upon a local moment. This process is achieved in two stages. Firstly the new location for a given point in TF is computed (called the reassigned co-ordinates), and secondly the energy at this point is computed. In the case of the spectrogram Kodera [Kodera76] showed that the reassigned co-ordinates for the spectrogram could be found by computing the partial derivatives of the phase of the ST-FT with respect to time and frequency. This link, between the reassigned co-ordinates of the spectrogram and the phase of the ST-FT, is exploited in this chapter to allow signal reconstruction from the reassigned co-ordinates. In order to achieve this inversion, the ability to recover a signal from the phase of the ST-FT is required. This chapter works towards this result by starting with inversion from the phase of the FT and generalising to a piece-wise, and then global, approach for signal reconstruction from the phase of the ST-FT.

4.2 Recovery of the Phase of the ST-FT from the Reassigned Spectrogram Co-ordinates

In order to reconstruct the signal from the reassigned spectrogram co-ordinates, the operation of creating the reassigned co-ordinates must be invertible. Various definitions of the reassigned co-ordinates are given in Chapter 3. These include Kodera's original definition, in terms of a local moment in TF and a definition in terms of the partial derivatives of the phase of the ST-FT. One further definition (if a Gaussian window is used) is given in (4-1) [Chassande97]. This simple formulation allows the recovery of the spectrogram from the reassigned co-ordinates via an

integration, (4-2), where $c_1(t)$ and $c_2(\omega)$ are unknown functions of time and frequency. These constants are required since only the partial derivatives are evaluated in (4-1).

$$\omega_i(t,\omega) = \frac{\partial \log|S(t,\omega)|}{\partial t}$$
$$t_g(t,\omega) = \frac{\partial \log|S(t,\omega)|}{\partial \omega} \tag{4-1}$$

$$\begin{aligned} |S(t,\omega)|^2 &= e^{2\left(c_1(t)+\int \omega_i(t,\omega)dt\right)} \\ &= e^{2\left(c_2(\omega)+\int t_g(t,\omega)d\omega\right)} \end{aligned} \tag{4-2}$$

The two equations given in (4-2), cannot be readily solved for the three unknown functions, $c_1(t)$, $c_2(\omega)$ and $|S(t,\omega)|^2$. A similar set of equations for the phase of the ST-FT, $\phi(t,\omega)$ can be derived from Kodera's definition of the reassigned co-ordinates (4-3). Once more, due to the partial derivatives being used, the system of two equations defines three unknown functions, $\phi(t,\omega)$, $d_1(\omega)$ and $d_2(t)$.

$$\begin{aligned} \phi(t,\omega) &= \int \omega_i(t,\omega)dt + d_1(\omega) \\ &= t - \int t_g(t,\omega)d\omega + d_2(t) \end{aligned} \tag{4-3}$$

It is not expected that the transform that maps the squared magnitude or phase of the ST-FT to the reassigned co-ordinates is invertible. Invertible in this context means the ability to recover the location in TF from which the energy at the reassigned coordinate originated. In the previous section, the reassigned co-ordinates for a Gaussian windowed linear chirp were given in equations (3-31) and (3-32). A pure tone is a special case obtained by setting the chirp rate to zero, *i.e.* $\beta = 0$. The expressions for the reassigned co-ordinates for a pure tone are given in (4-4). Previously it was stated that if the temporal resolution of the ST-FT is much smaller than the bandwidth of the signal, then $\alpha\gamma \to 0$. If this is the case, then the expressions for the GD and IF parameters simplify to those given in (4-5).

$$t_g(t,\omega) = \frac{t}{1+\alpha\gamma} \qquad (4\text{-}4)$$
$$\omega_i(t,\omega) = \frac{\alpha\gamma\omega + \omega_0}{1+\alpha\gamma}$$

$$t_g(t,\omega) = t \qquad (4\text{-}5)$$
$$\omega_i(t,\omega) = \omega_0$$

In this case, the reassigned IF co-ordinate is not a function of frequency. This is an example of a many-to-one function and thus is not invertible.

Discrete methods of computing the reassigned co-ordinates were introduced in Chapter 3 Section 3.4.4. The most basic of these methods is to compute the phase of the discrete ST-FT, and estimate its partial derivatives with respect to time and frequency using difference methods. The reassigned time co-ordinate (3-33) is given (for convenience assume $f_s = 1$) in (4-6). Rearranging this equation so that $\phi[n,k+1]$ is on the RHS, results in (4-7). This equation has two phase components present, $\phi[n,k+1]$ and $\phi[n,k]$. As such this equation is recursive. Making a change of variables in (4-7), equation (4-8) can be shown. Substituting equation (4-8) back into (4-8) yields the first expansion of the recursive formula, (4-9). Continuing the expansion leads to a simplified into the form given in (4-10).

$$t_g[n,k] = nL - (\phi[n,k+1] - \phi[n,k]) \qquad (4\text{-}6)$$

$$\phi[n,k+1] = (nL - t_g[n,k]) + \phi[n,k] \qquad (4\text{-}7)$$

$$\phi[n,k] = (nL - t_g[n,k-1]) + \phi[n,k-1] \qquad (4\text{-}8)$$

$$\phi[n,k+1] = (nL - t_g[n,k]) + (nL - t_g[n,k-1]) + \phi[n,k-1] \qquad (4\text{-}9)$$

$$\phi[n,k+1] = \left(\sum_{\lambda=1}^{k} nL - t_g[n,\lambda]\right) \qquad (4\text{-}10)$$

Equation (4-10) allows the recovery of the phase of the ST-FT via knowledge of the reassigned time co-ordinate, as defined via a numerical difference and the knowledge of the phase point $\phi[n,0]$.

A similar formula to (4-10) could be constructed using just the reassigned IF co-ordinate (3-33) (once more for convenience assuming $f_s = 1$) (4-11).

$$\phi[n+1,k] = (N-L)\left(-\sum_{\lambda=1}^{n}\omega_i[\lambda,k]\right) + \phi[0,k] \tag{4-11}$$

Rather than requiring $\phi[0,n]$, equation (4-11) requires knowledge of the phase of the first time slice, $\phi[0,k]$, in order to be able to reconstruct the phase of the ST-FT completely.

For real signals, the value of the phase at $\phi[n,0]$ can be shown (see Section 4.5.1) to be either 0 or π. For real signal the phase for each time-slice can therefore be recovered to within a known constant (0 or π), whereas for complex signals the phase can only be recor.structed to within an arbitrary unknown constant. However, in order to recover $\phi[0,k]$, the signal used in the construction of the first time-slice must be known. Therefore since more signal knowledge is required to reconstruct the phase from the reassigned IF co-ordinate than the GD co-ordinate, (4-10) is used as the phase reconstruction formula of choice.

4.3 Signal Reconstruction from the Phase of the DFT

Once the phase of the ST-FT has been recovered from the reassigned co-ordinates, then the signal still has to be reconstructed from this phase. The next two sections discuss signal reconstruction from the ST-FT phase. This begins with a review of previously published techniques for signal reconstruction from the DFT phase. These methods are then extended as to enable signal reconstruction from the phase of the ST-FT.

In the literature, two methods of signal reconstruction from the phase of DFT have been described [Quatieri81, Hayes80a-b]. The first technique assumes that the input signal is either maximum or minimum phase [Quatieri81]. Alternatively, Hayes *et al.* [Hayes80a-b] describe a method for arbitrary signal reconstruction by rearranging the definition of the phase of the DFT, and by making the assumption that the signal has at least $N/2$ zero samples. Both iterative and closed forms of this method were

proposed. This section begins with relaxing the assumption of $N/2$ zero samples to $N/2$ known samples.

4.3.1 Iterative Solution

The modulus and argument form of an N point DFT of the complex signal sequence, can be denoted in modulus and phase form as given in (4-12).

$$X[k] = |X[k]| e^{j\phi[k]} \tag{4-12}$$

An iterative algorithm to reconstruct the sequence $x[n]$ from N samples of the phase of the DFT, $\phi[k]$, and knowledge of the last M samples ($M \geq N/2$) samples of $x[n]$ is now outlined. The known samples of the input sequence $x[n]$ will be denoted as $d[n]$. This three step process is now detailed.

Step 1: Beginning with an initial (non-zero) estimate of the amplitude of the DFT, $|\hat{X}_0[k]|$, construct an estimate of the complex spectrum of the signal. Compute the IDFT, this is an estimate of the input sequence, denoted $x_p[n]$, where p is the iteration number.

$$X_p[k] = |\hat{X}_{p-1}[k]| e^{j\phi[k]} \tag{4-13}$$

$$x_p[n] = F^{-1}[X_p[k]] \tag{4-14}$$

Step 2: From $x_p[n]$ construct a new sequence, $\hat{x}_p[n]$ as defined by (4-15). This ensures that the last M samples are the same as those of the desired known sequence, $d[n]$.

$$\hat{x}_p[n] = \begin{cases} x_p[n] & 0 \leq n \leq N-M-1 \\ d[n] & N-M \leq n \leq N-1 \end{cases} \tag{4-15}$$

Step 3: Perform an N point DFT upon $\hat{x}_p[n]$, and use the modulus of this as the new estimate of the signals modulus.

Repetition of Steps 1-3 defines the iteration.

The error between the desired signal, $x[n]$, and the current estimate of the signal, $x_p[n]$, is given in (4-16). In this iterative procedure, the squared error, E_p, is non-increasing at every iteration. To prove this the method used by Hayes *et al.* [Hayes80b] is followed.

$$E_p = \sum_{n=0}^{N-1} \left| x[n] - x_p[n] \right|^2 \tag{4-16}$$

Previously defined in the continuous domain in Section 2.2.2, Parseval's theorem can be applied in the discrete domain to the DFT. Application of Parseval's theorem to (4-16) results in equation (4-17). Since both $X[k]$ and $X_p[k]$ have the same phase, (4-17) can be rewritten as (4-18). Since $\left|X_p[k]\right| = \left|\hat{X}_{p-1}[k]\right|$ by definition (see (4-13)), equation (4-18) can be written as (4-19). From the definition of the magnitude of $X_p[k]$ and application of the triangle inequality for vector differences, equation (4-20) can be derived.

$$E_p = \sum_{k=0}^{N-1} \left| X[k] - X_p[k] \right|^2 \tag{4-17}$$

$$E_p = \sum_{k=0}^{N-1} \left(\left|X[k]\right| - \left|X_p[k]\right| \right)^2 \tag{4-18}$$

$$E_p = \sum_{k=0}^{N-1} \left(\left|X[k]\right| - \left|\hat{X}_{p-1}[k]\right| \right)^2 \tag{4-19}$$

$$E_p \le \sum_{k=0}^{N-1} \left| X[k] - \hat{X}_{p-1}[k] \right|^2 \tag{4-20}$$

Equality holds in equation (4-20) if $\phi[k] = \phi_{p-1}[k]$. Application of Parseval's theorem to (4-20) results in (4-21).

$$E_p \le \sum_{k=0}^{N-1} \left| x[n] - \hat{x}_{p-1}[n] \right|^2 \tag{4-21}$$

From (4-15) it follows that the error between $\hat{x}_{p-1}[n]$ and $x[n]$ must be less than the error between $x_{p-1}[n]$ and $x[n]$ (4-22) and thus

$$\sum_{n=0}^{N-1}\left|x[n]-\hat{x}_{p-1}[n]\right|^2 \leq \sum_{n=0}^{N-1}\left|x[n]-x_{p-1}[n]\right|^2 \tag{4-22}$$

$$\sum_{n=0}^{N-1}\left|x[n]-x_{p-1}[n]\right|^2 = E_{p-1} \tag{4-23}$$

and hence

$$E_p \leq E_{p-1} \tag{4-24}$$

Equality holds in equation (4-24) if and only if $\hat{x}_{p-1}[n]=x_{p-1}[n]$ *i.e.* if the system has converged. Although this is not sufficient to guarantee convergence of the algorithm, numerical trials have verified that the system tends to the correct solution.

4.3.2 Closed Form Solution

Although the iterative algorithm described in Section 4.3.1 appears to converge towards the desired sequence, a closed form solution is preferable, such an algorithm is now presented. This is an extension to the method proposed by Hayes *et al.* [Hayes80a-b] to relax the condition that the known samples are zero. From the definition of the DFT, (2-19), the DFT phase can be written as given in (4-25). In this equation, the numerator is the imaginary part of the DFT and the denominator is the real part.

$$\tan\left(\phi[k]\right) = \frac{\mathrm{Im}\left[\sum_{n=0}^{N-1} x[n]e^{-j\left(\frac{2\pi nk}{N}\right)}\right]}{\mathrm{Re}\left[\sum_{n=0}^{N-1} x[n]e^{-j\left(\frac{2\pi nk}{N}\right)}\right]} \tag{4-25}$$

Since $x[n]$ is an arbitrary complex sequence, it can be separated into its real and imaginary parts as given in (4-26). Substitution of (4-26) into (4-25) leads to (4-27). By re-writing $\tan\left(\phi[k]\right)$ in terms of sine and cosine, (4-28) can be derived. Further manipulation using standard sine and cosine identities leads to (4-29).

$$x[n]=x_r[n]+jx_i[n] \tag{4-26}$$

$$\tan\left(\phi[k]\right)=\frac{\sum_{n=0}^{N-1}x_i[n]\cos\left(\frac{2\pi nk}{N}\right)-x_r[n]\sin\left(\frac{2\pi nk}{N}\right)}{\sum_{n=0}^{N-1}x_r[n]\cos\left(\frac{2\pi nk}{N}\right)+x_i[n]\sin\left(\frac{2\pi nk}{N}\right)} \tag{4-27}$$

$$\begin{aligned}&\sin\left(\phi[k]\right)\left(\sum_{n=0}^{N-1}x_r[n]\cos\left(\frac{2\pi nk}{N}\right)+x_i[n]\sin\left(\frac{2\pi nk}{N}\right)\right)\\&-\cos\left(\phi[k]\right)\left(\sum_{n=0}^{N-1}x_i[n]\cos\left(\frac{2\pi nk}{N}\right)-x_r[n]\sin\left(\frac{2\pi nk}{N}\right)\right)=0\end{aligned} \tag{4-28}$$

$$\left(\sum_{n=0}^{N-1}x_r[n]\sin\left(\phi[k]+\frac{2\pi nk}{N}\right)\right)-\left(\sum_{n=0}^{N-1}x_i[n]\cos\left(\phi[k]+\frac{2\pi nk}{N}\right)\right)=0 \tag{4-29}$$

It is assumed that (the last) M samples of the signal are known. In which case (4-29) can be re-written as (4-30), where d_r and d_i are the real and imaginary parts of the known signal.

$$\begin{aligned}&\sum_{n=0}^{N-M-1}x_r[n]\sin\left(\phi[k]+\frac{2\pi nk}{N}\right)-x_i[n]\cos\left(\phi[k]+\frac{2\pi nk}{N}\right)\\&=\sum_{n=N-M}^{N-M-1}d_r[n]\sin\left(\phi[k]+\frac{2\pi nk}{N}\right)-d_i[n]\cos\left(\phi[k]+\frac{2\pi nk}{N}\right)\end{aligned} \tag{4-30}$$

Using (4-30), all the unknown parameters of the signal are on the left hand side, whilst the right hand side is a function of the known parameters. This system of equations is linear in the unknown parameters, $x_r[n]$ and $x_i[n]$, and can be written in a matrix form as given by (4-31).

$$\mathbf{Ax}=\mathbf{b} \tag{4-31}$$

where

$$\mathbf{A}=\begin{bmatrix}\sin(\mathbf{Q}) & -\cos(\mathbf{Q})\end{bmatrix} \tag{4-32}$$

$$\mathbf{Q}=\begin{bmatrix} \phi[0] & \phi[0] & \cdots & \phi[0] \\ \phi[1] & \phi[1]+\frac{2\pi}{N} & \cdots & \phi[1]+\frac{2\pi(N-M-1)}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi[N-1] & \phi[N-1]+\frac{2\pi}{N} & \cdots & \phi[N-1]+\frac{2\pi(N-1)(N-M-1)}{N} \end{bmatrix}$$

$$\mathbf{x}=\begin{bmatrix} x_r[0] & x_r[1] & \cdots & x_r[N-M-1] & x_i[0] & x_i[1] & \cdots & x_i[N-M-1] \end{bmatrix}^T$$

$$\mathbf{b}=\begin{bmatrix} \sin(\mathbf{P}) & \cos(\mathbf{P}) \end{bmatrix}\mathbf{d}$$

$$\mathbf{P}=\begin{bmatrix} \phi[0] & \phi[0] & \cdots & \phi[0] \\ \phi[1]+\frac{2\pi(N-M)}{N} & \phi[1]+\frac{2\pi(N-M+1)}{N} & \cdots & \phi[1]+\frac{2\pi(N-1)}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi[N-1]+\frac{2\pi(N-1)(N-M)}{N} & \phi[N-1]+\frac{2\pi(N-1)(N-M+1)}{N} & \cdots & \phi[N-1]+\frac{2\pi(N-1)(N-1)}{N} \end{bmatrix}$$

$$\mathbf{d}=\begin{bmatrix} d_r[0] & d_r[1] & \cdots & d_r[M-1] & d_r[0] & d_i[1] & \cdots & d_i[M-1] \end{bmatrix}$$

The sine (or cosine) of a matrix is assumed to refer to the matrix composed of the sine (or cosine) of the elements of the matrix.

Examination of the $\mathbf{A}$ matrix in equation (4-31) shows that this matrix is square if $M = N/2$. If $M < N/2$ is smaller than this value, then the system of the equations is under-determined and thus no unique solution can be found. If $M = N/2$ then numerical trials have shown that the $\mathbf{A}$ matrix has full rank. In this case, signal reconstruction can take place by computing the inverse of $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, and computing the vector product $\mathbf{A}^{-1}\mathbf{b}$. If $M > N/2$ then the $\mathbf{A}$ matrix is over-determined and thus computation of the pseudo-inverse of $\mathbf{A}$ is required. Denoting the pseudo-inverse of $\mathbf{A}$ as $\mathbf{A}^{\#}$, then computing the vector product $\mathbf{A}^{\#}\mathbf{b}$ reconstructs the signal.

Prior to describing signal reconstruction from the phase of the ST-FT, it should be noted that (4-29) can for written as an alternative matrix formulation as given in (4-33), where * denotes a complex conjugate.

$$\mathrm{Im}[\mathbf{C}\mathbf{x}^*]=0 \tag{4-33}$$

Where

$$\mathbf{C}=\{c_{r,s}\}$$

$$c_{r,s}=e^{j\left(\phi[r]+\frac{2\pi rs}{N}\right)}$$

$$\mathbf{x}=\begin{bmatrix}x[0] & x[1] & \cdots & x[N-1]\end{bmatrix}$$

Although a more compact form of the equation, this equation is not used throughout the remainder of the thesis since it does not add significantly to the development of the ideas and mathematics.

4.4 Signal Reconstruction from the Phase of the ST-FT

4.4.1 Plecewlse Approach to Signal Reconstruction from the Phase of the ST-FT

The results given in the previous section are extended to allow signal reconstruction from the phase of the discrete ST-FT.

Considering just one time-slice (or frame) of the ST-FT leads to (4-34).

$$S[n,k]=\sum_{m=0}^{N-1}x_n[m]h[m]e^{-j\left(\frac{2\pi mk}{N}\right)} \tag{4-34}$$

where

$$x_n[m]=x[nL+m] \tag{4-35}$$

For each time-slice of the ST-FT the product of the windowing function and the input time sequence is transformed into the frequency domain using the DFT. By considering each time-slice of the ST-FT to be the result of a windowed DFT, the results derived in the previous section can be applied directly to the ST-FT. If the windowing function and sufficient signal samples are known then the phase of the ST-FT is sufficient to reconstruct the signal by analogy with the methods presented in Section 4.3.

4.4.1.1 Piecewise Signal Reconstruction from the Phase of the ST-FT

If the overlap used in the computation of the ST-FT is sufficiently small, then there will be at least $N/2$ samples common between successive time slices. If the overlap is at least $N/2$ then reconstruction of samples of the one time-slice enables the reconstruction of the following time-slice, once the effect of the windowing function has been factored out. It is assumed that at least $N/2$ samples are known at the start of the signal to enable reconstruction of the first time slice. In a synthetic example this condition can be met by adding $N/2$ known samples to the start of the signal and computing the ST-FT on this padded signal.

This piecewise approach to signal reconstruction from the phase of the ST-FT can be summarised. An iterative algorithm is now presented in five steps assuming that the overlap is set to $N/2$.

1. From knowledge of the first $N/2$ samples and of the phase of the ST-FT, recover the unknown samples from the first time slice, via the solution of (4-31).

2. The recovered signal is a windowed version of the input signal. Thus to recover the input sequence divide by the windowing function.

3. Multiply the first $N/2$ samples of the windowing function by the samples recovered in Step 2. Thus these are the same samples used in the construction of the next time slice.

4. Use this new sequence and knowledge of the phase of the ST-FT of the next time slice to recover the next segment of the sequence.

5. Repeat Steps 2-4 until all phase slices have been used.

This description assumes that an overlap of $N/2$ is used, although only a simple modification is required to allow for a greater overlap. The greater the degree of overlap the greater then robustness of the algorithm since the matrix $\mathbf{A}$ in equation (4-31) becomes more over-determined.

4.4.1.2 Limitations to Piecewise Signal Reconstruction

The piecewise algorithm suffers from two limitations. Firstly, in order to recover the signal defined by the phase of the first time-slice, additional signal knowledge is

required. This is because the formulation is based upon signal reconstruction from the phase of the DFT. This requirement for some *a priori* signal knowledge means that the phase of the ST-FT alone is not sufficient to reconstruct the signal using this method. Secondly the open loop form of the iteration results in poor numerical characteristics. Specifically small errors in each time-slice tend to accumulate and cause divergence from the correct solution. This effect can be seen in later in this chapter in Section 4.5.2.3.

Both these limitations stem from the fact that the solution is constructed in a piece-wise fashion. In the next section a new global approach to the problem is presented, attempting to solve the problem in single step.

4.4.2 Global Approach to Signal Reconstruction from the Phase of the ST-FT

The closed form global solution to signal reconstruction from the phase of the ST-FT can be developed in an analogous manner to that employed in the development of the matrix form for solving the DFT problem (Section 4.3.2). Assuming a, real, time limiting window, equation (4-34) can be written in terms of the phase of the ST-FT given in (4-36).

$$\tan\left(\phi[n,k]\right)=\frac{\sum_{m=0}^{N-1}x_{n_i}[m]h[m]\cos\left(\frac{2\pi mk}{N}\right)-x_{n_r}[m]h[m]\sin\left(\frac{2\pi mk}{N}\right)}{\sum_{m=0}^{N-1}x_{n_r}[m]h[m]\cos\left(\frac{2\pi mk}{N}\right)+x_{n_i}[m]h[m]\sin\left(\frac{2\pi mk}{N}\right)} \tag{4-36}$$

$$x_n[m]=x_{n_r}[m]+jx_{n_i}[m]$$

Manipulation of this equation via trigonometric identities leads to simplified form given in (4-37). Unlike (4-30), since it is assumed that no signal information is available (4-37) cannot be split into known, and unknown parts.

$$\left(\sum_{m=0}^{N-1}x_{n_r}[m]h[m]\sin\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right)-\left(\sum_{m=0}^{N-1}x_{n_i}[m]h[m]\cos\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right)=0 \tag{4-37}$$

Where the signal component $x_n[m]$ fits into the whole signal is dependant upon the overlap used. Substituting (4-35) back into (4-37) yields

$$\left(\sum_{m=0}^{N-1} x_r[nL+m]h[m]\sin\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right) \\ -\left(\sum_{m=0}^{N-1} x_i[nL+m]h[m]\cos\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right)=0 \tag{4-38}$$

The equation is now linear in terms of the signal components and can be defined in terms of the matrix formulation given in (4-39).

$$\begin{bmatrix}\mathbf{A}_1 & \vdots & \mathbf{A}_2\end{bmatrix}\begin{bmatrix}\mathbf{x}_r \\ \mathbf{x}_i\end{bmatrix}=\mathbf{0} \tag{4-39}$$

where

$$\mathbf{A}_1=\begin{bmatrix}\boldsymbol{\psi}_1 & \mathbf{Z} & \cdots & \mathbf{Z} \\ \mathbf{Z} & \boldsymbol{\psi}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z} & \mathbf{Z} & \cdots & \boldsymbol{\psi}_P\end{bmatrix} \qquad \mathbf{A}_2=\begin{bmatrix}\boldsymbol{\Lambda}_1 & \mathbf{Z} & \cdots & \mathbf{Z} \\ \mathbf{Z} & \boldsymbol{\Lambda}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z} & \mathbf{Z} & \cdots & \boldsymbol{\Lambda}_P\end{bmatrix} \tag{4-40}$$

$$\boldsymbol{\Psi}_s=\begin{bmatrix} h[0]\sin(\phi[s,0]) & h[1]\sin(\phi[s,0]) & \cdots & h[N-1]\sin(\phi[s,0]) \\ h[0]\sin(\phi[s,1]) & h[1]\sin\left(\phi[s,1]+\frac{2\pi}{N}\right) & \cdots & h[N-1]\sin\left(\phi[s,1]+\frac{2\pi(N-1)}{N}\right) \\ \vdots & \vdots & \ddots & \vdots \\ h[0]\sin(\phi[s,N-1]) & h[1]\sin\left(\phi[s,N-1]+\frac{2\pi(N-1)}{N}\right) & \cdots & h[N-1]\sin\left(\phi[s,N-1]+\frac{2\pi(N-1)(N-1)}{N}\right)\end{bmatrix}$$

$$\boldsymbol{\Lambda}_s=\begin{bmatrix} h[0]\cos(\phi[s,0]) & h[1]\cos(\phi[s,0]) & \cdots & h[N-1]\cos(\phi[s,0]) \\ h[0]\cos(\phi[s,1]) & h[1]\cos\left(\phi[s,1]+\frac{2\pi}{N}\right) & \cdots & h[N-1]\cos\left(\phi[s,1]+\frac{2\pi(N-1)}{N}\right) \\ \vdots & \vdots & \ddots & \vdots \\ h[0]\cos(\phi[s,N-1]) & h[1]\cos\left(\phi[s,N-1]+\frac{2\pi(N-1)}{N}\right) & \cdots & h[N-1]\cos\left(\phi[s,N-1]+\frac{2\pi(N-1)(N-1)}{N}\right)\end{bmatrix}$$

$$\mathbf{x}_r=\begin{bmatrix}x_r[0] & x_r[1] & \cdots & x_r[l_x-1]\end{bmatrix}^T \qquad \mathbf{x}_i=\begin{bmatrix}x_i[0] & x_i[1] & \cdots & x_i[l_x-1]\end{bmatrix}^T$$

In (4-40) l_x is the length of the input time sequence, and P is the number of time slices in the ST-FT. The matrix $\mathbf{Z}$ is a (typically non-square) matrix of zeros which has N rows and L columns. The matrix alters size according to the overlap used, providing the correct spacing between the $\boldsymbol{\Psi}_s$ and $\boldsymbol{\Lambda}_s$ for a given overlap.

The system of equations (4-39)is homogeneous. One method of finding a non-trivial solution is to force the first value of the vector $\mathbf{x}$ to unity, as shown in (4-41). This extension has the result that the first sample is always reconstructed to be unity. It thus follows that the signal is only reconstructed to within an overall scale constant, which is dependant upon the value of the first sample of the real part of the signal. It follows that for inversion to take place, this sample is assumed to be non-zero.

$$\hat{\mathbf{A}}\mathbf{x} = \mathbf{b} \tag{4-41}$$

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ [\mathbf{A}_1 & \vdots & & \mathbf{A}_2] \end{bmatrix}$$

$$\mathbf{b} = [1 \quad 0 \quad \cdots \quad 0]^T$$

That signal reconstruction from phase can only take place to within an overall scale constant is to be expected. Since the ST-FT is a linear transform, scaling the amplitude of the signal has no effect upon the phase. It follows that for two signals, one which is an amplitude scaled version of the other, the phase of their ST-FT's will be identical. Since the phase does not encode this overall amplitude information, any signal reconstructed purely from the phase can only be recovered to within an overall constant.

If a signal is comprised of two components, separated by at least N zero samples, then reconstruction of *each* signal component will take place to within an multiplicative amplitude constant. This is because there is no continuum of information between the two components. The signal can be considered as two separate signals, and thus the linearity of the ST-FT means that the phase of the ST-FT only allows for reconstruction to within an amplitude constant for each separate signal. If any noise is present (for example, recording noise) above the level of the numerical precision of the machine, then the signal will be reconstructed to within an overall, rather than a local amplitude constant. Therefore, since this phenomenon is only present on synthetic examples it is not considered as a severe limitation.

The solution to (4-41) can be achieved by matrix inversion, assuming $\hat{\mathbf{A}}$ is non-singular. Empirical studies have shown that an overlap $L < 0.3N$ is sufficient for $\hat{\mathbf{A}}$ to be non-singular.

An alternative (related) method for solving this system of equations is based on Singular Value Decomposition (SVD) of the $\mathbf{A}$ matrix defined in (4-40) [Golub89]. The singular vector associated with the zero singular value is the solution to the equation. The use of this method removes the requirement for the first sample to be non-zero, as required in the method described by (4-41).

4.4.3 Examples of Signal Reconstruction from Phase of ST-FT

As an illustration of signal reconstruction from the phase of the ST-FT, two examples are presented. The first is the familiar linear chirp example. In order to show that no structure is required in the signal, the second example is a pseudo-random Gaussian sequence. Other examples showing this technique in use are given in Lopes *et al.* [Lopes98].

4.4.3.1 Example 1 Linear Chirp

The ST-FT for the linear FM signal shown in Figure 20 was computed using a maximally overlapped 32pt Gaussian window. The phase of the ST-FT is shown in Figure 21. The ridge (line of large magnitude in TF) of the linear chirp appears in this phase figure as the chequered region. It is the zero-crossing of the phase in this region that the 'ridge and skeleton' algorithms and the reassigned spectrogram pick out. Using only the phase information, and through the application of equation (4-41), the signal was reconstructed. Both the recovered and the original signals were then normalised with respect to their first samples. The amplitude difference between the two signals is plotted in Figure 22. The error between the amplitudes of the two signals can be seen to be of the order of 10^{-14}, which represents a -280dB error. A similar performance measure can be seen when comparing the phase of the original signal to that of the reconstructed signal as shown in Figure 23.

4.4.3.2 Example 2 Pseudo Random Sequence

As stated in the introduction to this section, in order to show that there is no specific requirement upon the type of signal used, a pseudo random Gaussian white noise signal is considered. The input sequence is shown in Figure 24, the phase of the ST-FT generated using 32pt Gaussian window is shown in Figure 25. Unlike the case of the linear chirp, no clear ridge can be seen in this plot of the phase, since no FM signal is present in the signal. The error between the original and reconstructed

signals for the pseudo-random noise case is of the same order as the linear chirp example; The amplitude and phases errors are displayed in Figure 26 and 27.

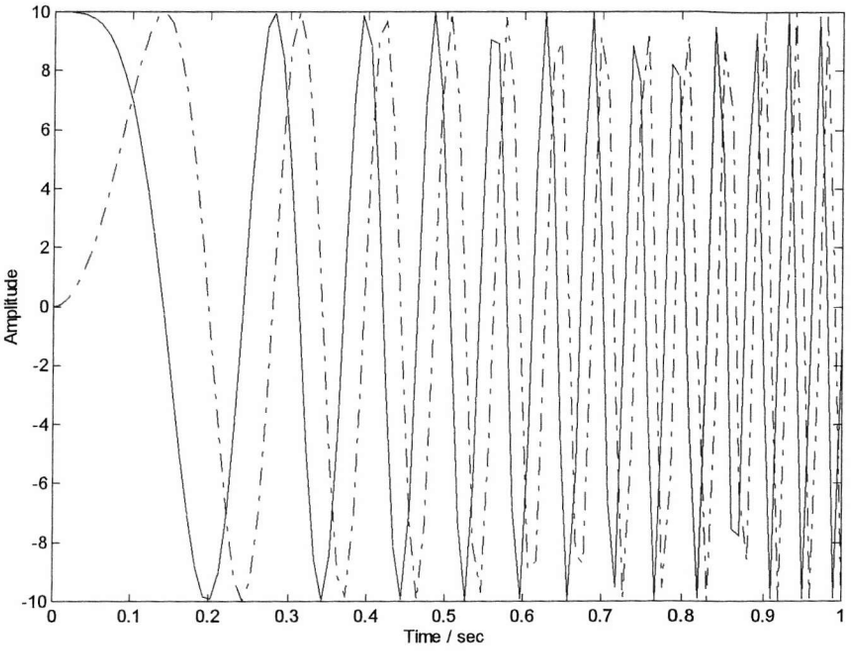

Figure 20 Linear FM Harmonic Wave (Real Part Solid, Imaginary Part Dashed Line)

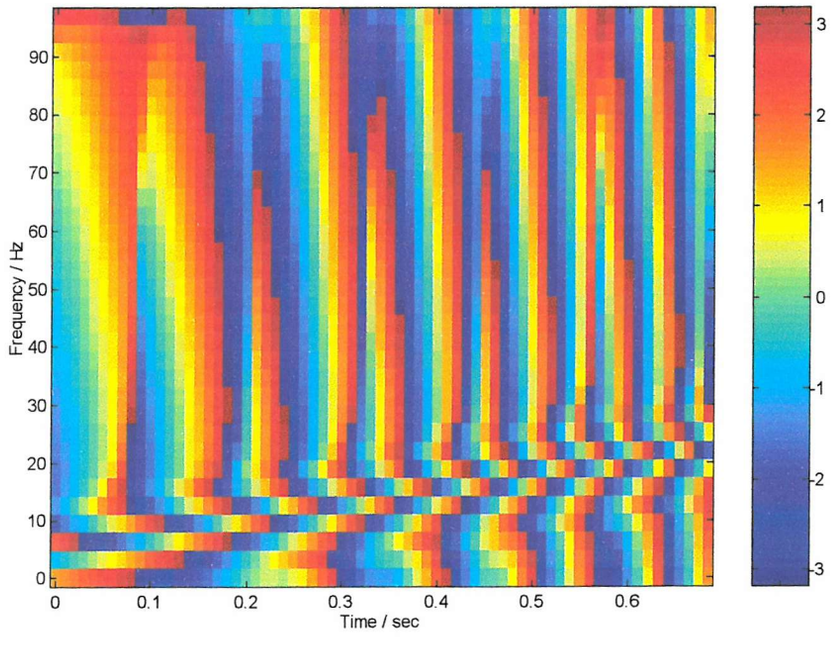

Figure 21 Phase of the ST-FT of Linear Chirp Computed using 32pt Gaussian Window (Maximally Overlapped)

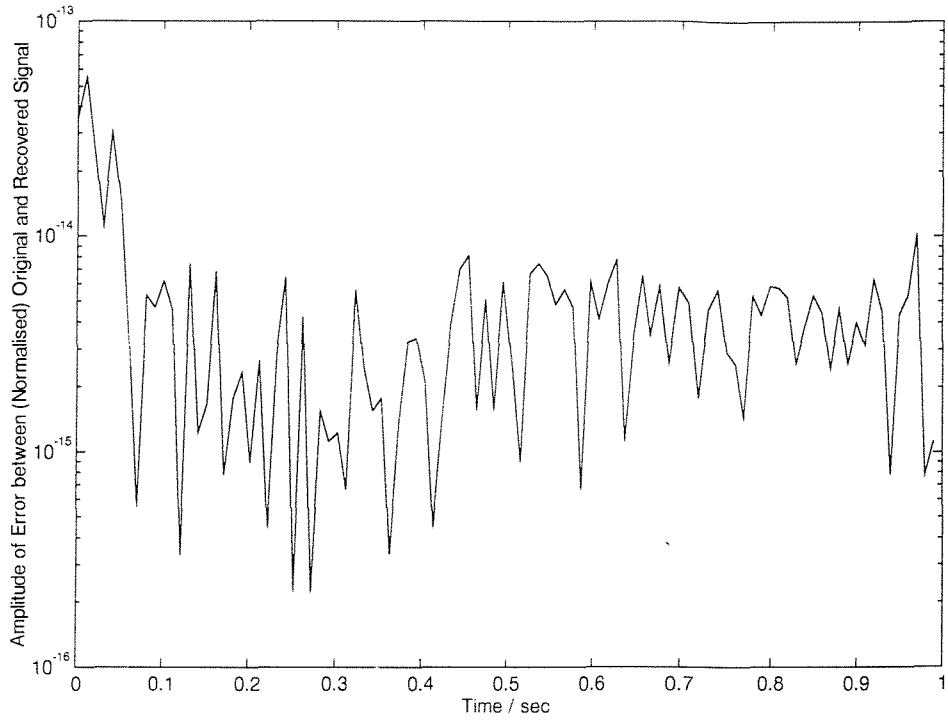

Figure 22 Error Between Original and Reconstructed Amplitude - Linear Chirp

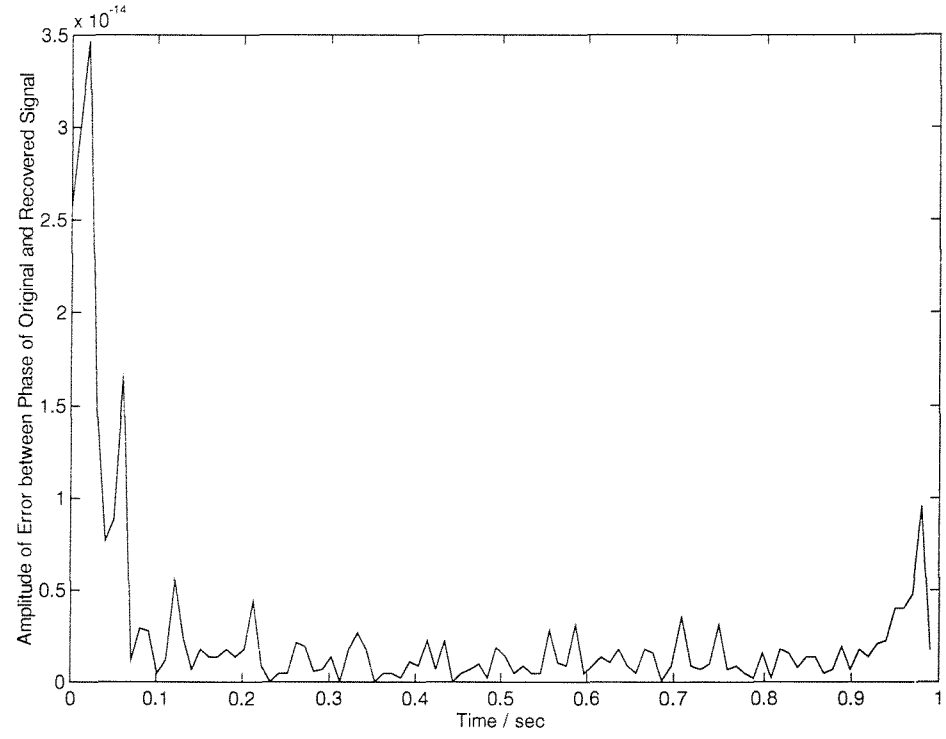

Figure 23 Error Between Original and Reconstructed Phase - Linear Chirp

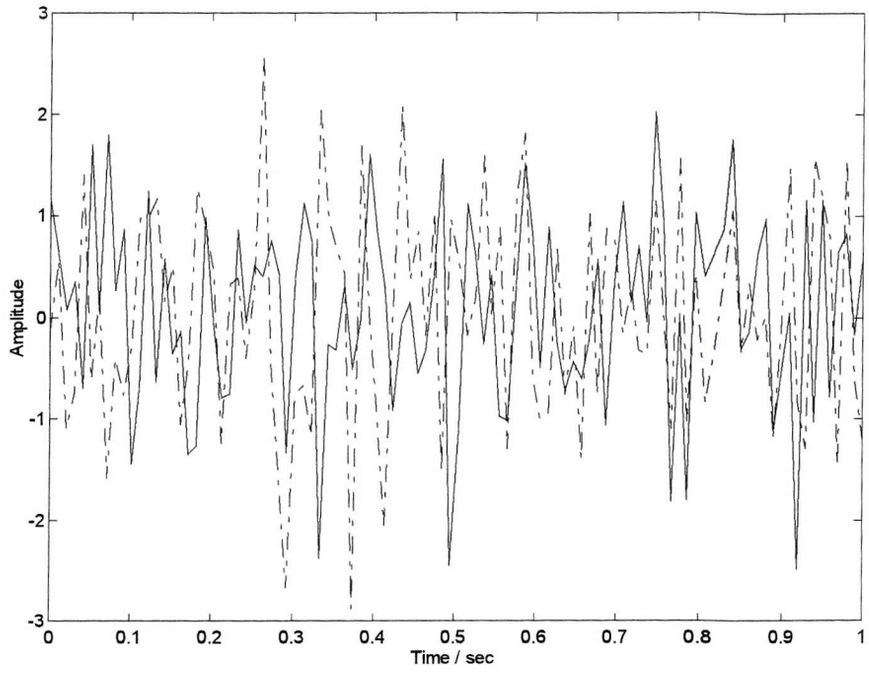

Figure 24 Pseudo Random Gaussian White Noise (Real Part Solid Line, Imaginary Part Dashed Line)

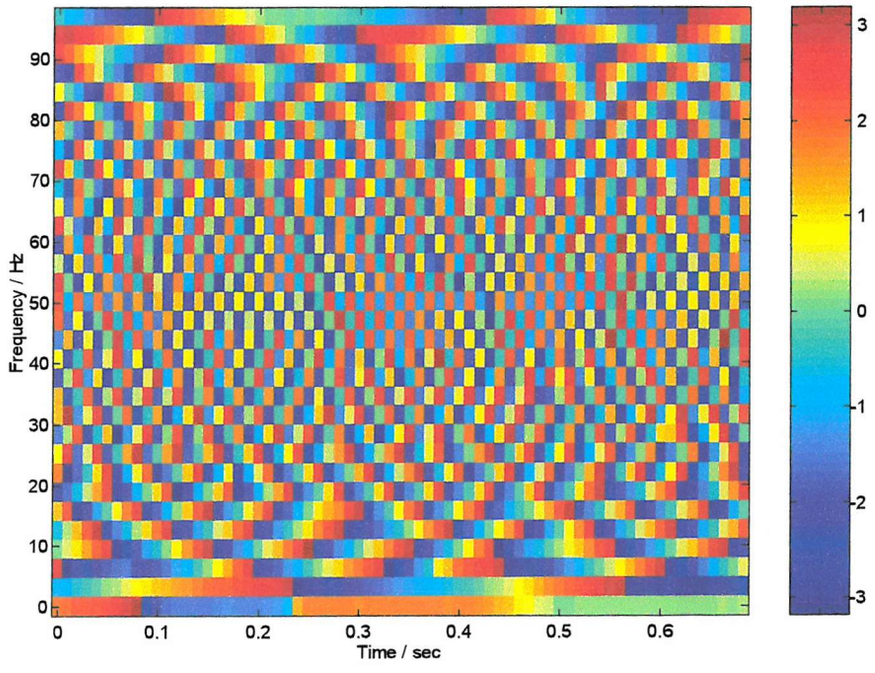

Figure 25 Phase of the ST-FT of Pseudo Random Gaussian Noise Computed using 32pt Gaussian Window (Maximally Overlapped)

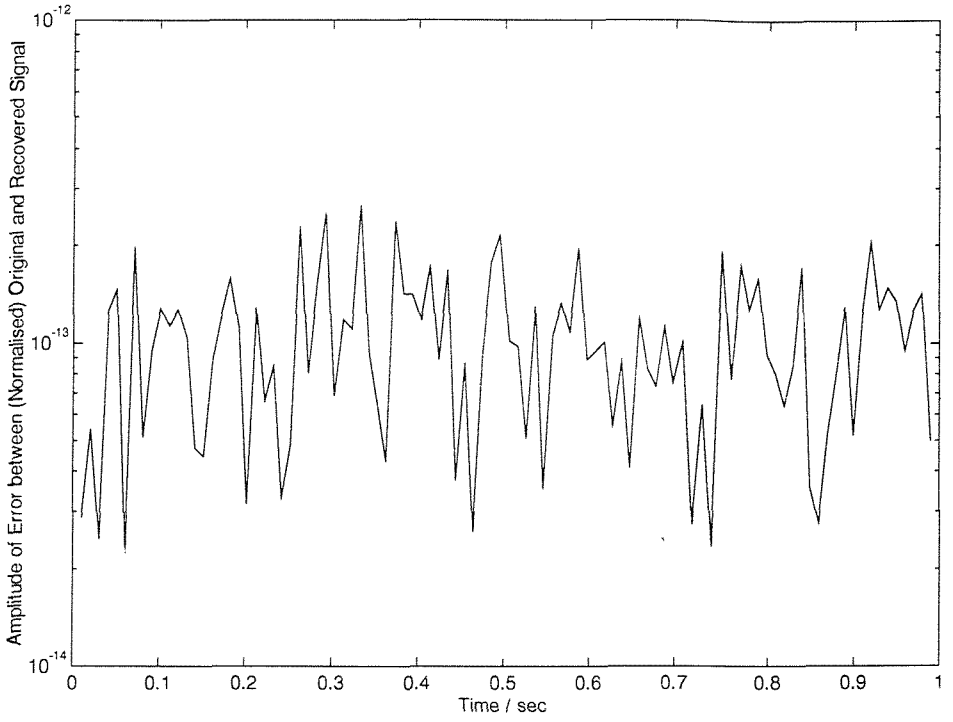

Figure 26 Error Between Original and Reconstructed Signal Pseudo Random Gaussian White Noise

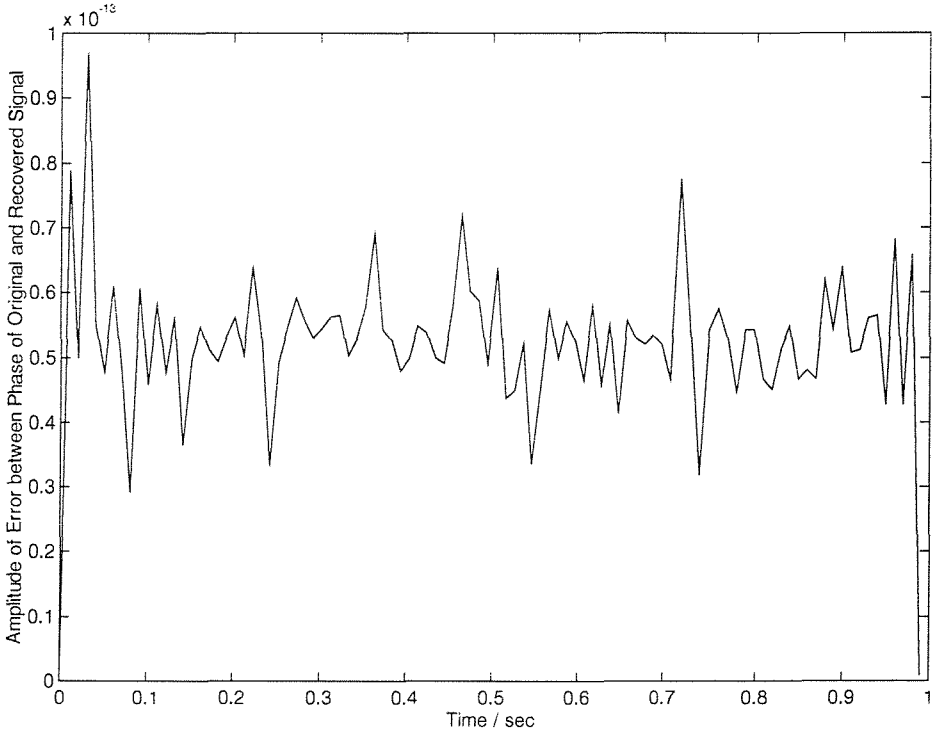

Figure 27 Error Between Original and Reconstructed Phase – Gaussian Noise

4.5 Signal Reconstruction from the Reassigned Spectrogram Co-ordinates

4.5.1 Real Signal Reconstruction from the Reassigned Spectrogram Co-ordinates

Section 4.2 outlined a method of reconstructing the signal from the reassigned spectrogram co-ordinates. The recursive equation given in (4-10) enables the reconstruction of the phase of the ST-FT from the reassigned GD co-ordinates if computed via a numerical difference method.

From the definition of the DFT given in Chapter 2 equation (2-19), the first sample point of the DFT can be seen to be equal the sum of the signal samples (4-42). If the signal is real, then the result of the summation must also be real. The modulus is defined as being real and positive, so the value of the phase at this point can only be 0 or π. It follows for real signals that (4-10) reduces to (4-43), where $c[n]$ is function whose value is zero or one depending upon whether the sum of the signal (multiplied by the windowing function) in the n'th time slice is positive or negative.

$$|X[0]|e^{j\phi[0]} = \sum_{n=0}^{N-1} x[n] \tag{4-42}$$

$$\phi[n,k] = \left(\sum_{\lambda=1}^{k} nL - t_g[n,\lambda]\right) + c[n]\pi \tag{4-43}$$

Equation (4-43) establishes a one-to-two mapping between the reassigned Group Delay (time) co-ordinate and the phase of the ST-FT. The effect that this additional π offset has upon the signal reconstruction from phase, as defined in Section 4.3.2, is now explored.

By allowing the additional variable $c[n]\pi$ into (4-38), (4-44) can be derived. Since the function $c[n]$ can only take the value of zero or one, (4-45) follows. By expanding (4-44) using trigonometric identities and simplifying the result, (4-46) can be derived. The result of variable $c[n]$ can now be seen to change the overall sign of

the result of the summation. This variable can thus be removed without affecting the solution of the systems of equations since the RHS is zero.

$$\left(\sum_{m=0}^{N-1} x_r[nL+m]h[m]\sin\left(\phi[n,k]-c[n]\pi+\frac{2\pi mk}{N}\right)\right) - \left(\sum_{m=0}^{N-1} x_i[nL+m]h[m]\cos\left(\phi[n,k]-c[n]\pi+\frac{2\pi mk}{N}\right)\right)=0 \tag{4-44}$$

$$\sin\left(c[n]\pi\right)=0$$
$$\cos\left(c[n]\pi\right)=-1^{c[n]} \tag{4-45}$$

$$\left(-1^{c[n]}\right)\left(\sum_{m=0}^{N-1} x_r[nL+m]h[m]\sin\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right) - \left(-1^{c[n]}\right)\left(\sum_{m=0}^{N-1} x_i[nL+m]h[m]\cos\left(\phi[n,k]+\frac{2\pi mk}{N}\right)\right)=0 \tag{4-46}$$

Since the variable $c[n]$ does not affect the solution, it can, without loss of generality, be assumed to be zero. This leads to the simplification of (4-43) as given in (4-47). Equation (4-47) expresses a one-to-one mapping between the reassigned group delay co-ordinate and the phase of the ST-FT. In this manner knowledge of the reassigned group delay co-ordinates, computed via a numerical difference method, allows the phase of the ST-FT to be recovered. Thus if the original ST-FT is computed using sufficient overlap, the signal can now be reconstructed using the previously described method.

$$\phi[n,k]=\left(\sum_{\lambda=1}^{k} nL-t_g[n,\lambda]\right) \tag{4-47}$$

To illustrate this method, the real part of the linear chirp given in Figure 20, is reconstructed from its reassigned co-ordinates, shown in Figure 28. Using the reassigned GD parameter, and assuming that the input is real, the phase of the ST-FT, and then the signal, was reconstructed. The error in amplitude between the reconstructed signal and the normalised input signal is given in Figure 29. The error is of the order of 10^{-14} smaller than the maximum amplitude.

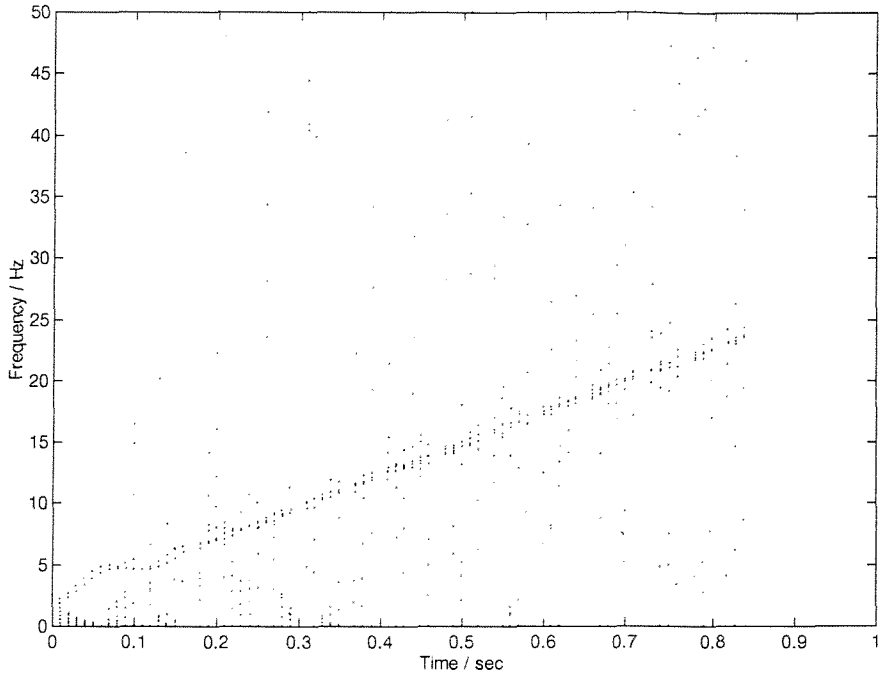

Figure 28 Reassigned Co-ordinates of Real Linear Chirp Computed using Numerical Difference Method, based a 16pt Maximally Overlapped Gaussian Windowed ST-FT.

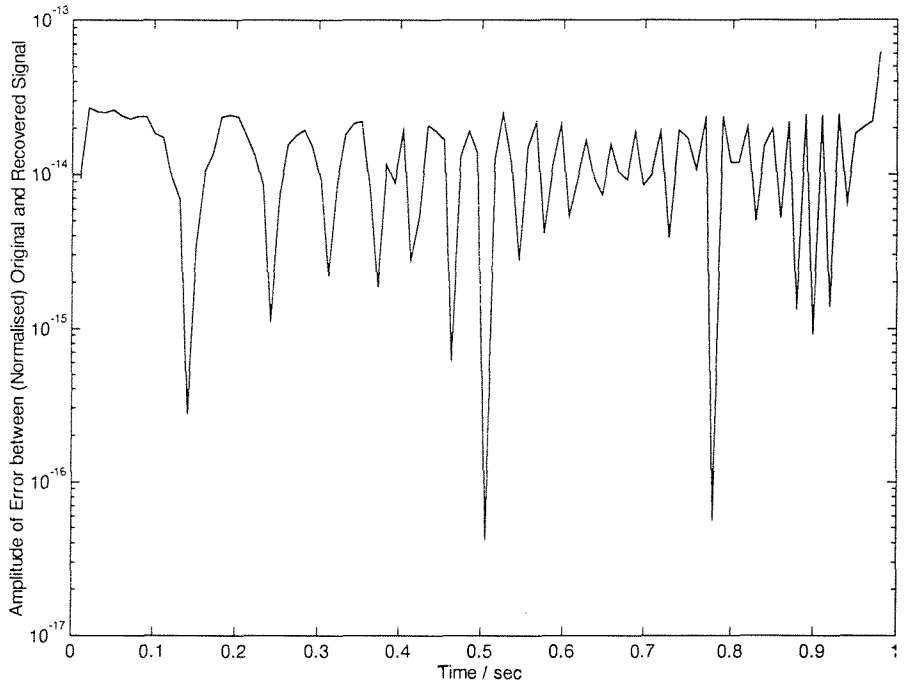

Figure 29 Error in Amplitude between Real Normalised Linear Input Signal, and Signal Reconstructed via the Group Delay Reassigned Co-ordinate.

4.5.2 Complex Signal Reconstruction from the Reassigned Spectrogram Co-ordinates

This section will investigate the reconstruction of complex valued signals from their reassigned GD co-ordinate.

4.5.2.1 Signal Reconstruction from the Phase of the DFT known to within a Constant

If the input signal is real, then a known (or inferred) point of the phase is sufficient to allow reconstruction of the signal from its reassigned GD co-ordinate (Section 4.5.1). If the signal is complex, then the reassigned GD co-ordinate can only reconstruct the phase of the ST-FT to within an additive constant per time-slice. The method of reconstruction from phase given previously needs to be modified to take this additional constant into account. As before, signal reconstruction is considered using the phase of the DFT and then generalised to the ST-FT case.

Application of equation (4-10) for complex signals results in the reconstruction of each phase to within an additive constant per time slice. This can be represented as given in (4-48), where $\phi'[k]$ is the reconstructed phase for one time-slice through application of (4-10), $\phi[k]$ is the correct phase and ϕ_c is a constant. Substituting (4-48) into the signal reconstruction from the phase of the DFT (4-29) produces (4-49). Expanding this expression using standard trigonometric identities, equation (4-50) can be deduced.

$$\phi'[k] = \phi[k] - \phi_c \tag{4-48}$$

$$\left(\sum_{n=0}^{N-1} x_r[n]\sin\left(\phi'[k]+\phi_c+\frac{2\pi nk}{N}\right)\right)-\left(\sum_{n=0}^{N-1} x_i[n]\cos\left(\phi'[k]+\phi_c+\frac{2\pi nk}{N}\right)\right)=0 \tag{4-49}$$

$$\begin{aligned}&\cos(\phi_c)\left(\sum_{n=0}^{N-1} x_r[n]\sin\left(\phi'[k]+\frac{2\pi nk}{N}\right)-x_i[n]\cos\left(\phi'[k]+\frac{2\pi nk}{N}\right)\right)\\&-\sin(\phi_c)\left(\sum_{n=0}^{N-1} x_r[n]\sin\left(\phi'[k]+\frac{2\pi nk}{N}\right)+x_i[n]\cos\left(\phi'[k]+\frac{2\pi nk}{N}\right)\right)=0\end{aligned} \tag{4-50}$$

As previously, this equation can now be split into known and unknown parts. It is assumed than the last M samples of the real and imaginary parts of the signal are

known. Using the definitions provided by (4-51) and (4-52), (4-50) can be separated into two summations as given in (4-53). This equation can be written in terms of a matrix equation that can be solved for the unknown parameters, $\tilde{x}_r, \tilde{x}_i, \hat{x}_r, \hat{x}_i$ and $\cos\phi_c, \sin\phi_c$.

$$\begin{aligned} \tilde{x}_r[n] &= x_r[n]\cos\phi_c \quad \tilde{x}_i[n] = x_i[n]\cos\phi_c \\ \hat{x}_r[n] &= x_r[n]\sin\phi_c \quad \hat{x}_i[n] = x_i[n]\sin\phi_c \end{aligned} \tag{4-51}$$

$$\begin{aligned} a_r[n] &= x_r[n]\sin\left(\phi'[k] + \frac{2\pi nk}{N}\right) \quad b_r[n] = x_r[n]\cos\left(\phi'[k] + \frac{2\pi nk}{N}\right) \\ a_i[n] &= x_i[n]\sin\left(\phi'[k] + \frac{2\pi nk}{N}\right) \quad b_i[n] = x_i[n]\cos\left(\phi'[k] + \frac{2\pi nk}{N}\right) \end{aligned} \tag{4-52}$$

$$\begin{aligned} \sum_{n=0}^{N-M-1} (\tilde{x}_r[n] + \hat{x}_i[n])\sin\left(\phi'[k] + \frac{2\pi nk}{N}\right) + (\hat{x}_r[n] - \tilde{x}_i[n])\cos\left(\phi'[k] + \frac{2\pi nk}{N}\right) \\ - \sum_{n=N-M}^{N-1} (a_r[n] - b_i[n])\cos\phi_c + (a_i[n] + b_r[n])\sin\phi_c = 0 \end{aligned} \tag{4-53}$$

The matrix formulation for (4-53) is given in (4-54).

$$\mathbf{Sx} = \mathbf{0} \tag{4-54}$$

where $$\mathbf{S} = [\sin(\mathbf{\Phi}) \vdots -\cos(\mathbf{\Phi}) \vdots (\mathbf{A}_r - \mathbf{B}_i) \vdots (\mathbf{A}_i + \mathbf{B}_r)]^T$$

$$\mathbf{\Phi} = \begin{bmatrix} \phi'[0] & \phi'[0] & \cdots & \phi'[0] \\ \phi'[1] & \phi'[1] + \frac{2\pi}{N} & \cdots & \phi'\left[\frac{2\pi(N-1)}{N}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \phi'[N-M-1] & \phi'[N-M-1] + \frac{2\pi(N-M-1)}{N} & \cdots & \phi'[N-M-1] + \frac{2\pi(N-M-1)(N-1)}{N} \end{bmatrix}$$

$$\gamma = \begin{bmatrix} \phi'[N-M] & \phi'[N-M] + \frac{2\pi(N-M)}{N} & \cdots & \phi'[N-M] + \frac{2\pi(N-M)(N-1)}{N} \\ \phi'[N-M+1] & \phi'[N-M+1] + \frac{2\pi(N-M+1)}{N} & \cdots & \phi'[N-M+1] + \frac{2\pi(N-M+1)(N-1)}{N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi'[N-1+1] & \phi'[N-1] + \frac{2\pi(N-1)}{N} & \cdots & \phi'[N-1] + \frac{2\pi(N-1)(N-1)}{N} \end{bmatrix}$$

$$A_r = \left(\sin(\gamma)\right)\begin{bmatrix} x_r(N-M) \\ x_r(N-M+1) \\ \vdots \\ x_r(N-1) \end{bmatrix} \quad A_i = \left(\sin(\gamma)\right)\begin{bmatrix} x_i(N-M) \\ x_i(N-M+1) \\ \vdots \\ x_i(N-1) \end{bmatrix}$$

$$B_r = \left(\cos(\gamma)\right)\begin{bmatrix} x_r(N-M) \\ x_r(N-M+1) \\ \vdots \\ x_r(N-1) \end{bmatrix} \quad B_i = \left(\cos(\gamma)\right)\begin{bmatrix} x_i(N-M) \\ x_i(N-M+1) \\ \vdots \\ x_i(N-1) \end{bmatrix}$$

$$\mathbf{x} = \left[\tilde{\mathbf{x}}_\mathbf{r} + \hat{\mathbf{x}}_\mathbf{i} \quad \hat{\mathbf{x}}_\mathbf{i} - \tilde{\mathbf{x}}_\mathbf{r} \quad \cos(\phi_c) \quad \sin(\phi_c)\right]^T$$

As previously, in order to avoid the trivial solution, an additional row is appended, producing the modified form as given below.

$$\mathbf{x} = \hat{\mathbf{S}}^{-1}\mathbf{b} \tag{4-55}$$

where

$$\hat{\mathbf{S}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ & \mathbf{S} & & \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

Since the system of equations is homogeneous, then the four unknowns, $\tilde{x}_r + \hat{x}_i$, $\hat{x}_r - \tilde{x}_i$, $\cos\phi_c$ and $\sin\phi_c$ are only solved to within a multiplicative constant. The value of this constant, denoted α, means that the two equations (4-56) need to be solved to recover ϕ_c. Equation (4-57) provides the solution to this problem, allowing the amplitude to be recovered to within a sign constant.

$$\begin{aligned} \gamma_1 &= \alpha \sin\phi_c \\ \gamma_2 &= \alpha \cos\phi_c \end{aligned} \tag{4-56}$$

$$\phi_c = \pm\cos^{-1}\left(\frac{\gamma_1}{\sqrt{\gamma_1^2 + \gamma_2^2}}\right) \tag{4-57}$$

$$\alpha = \pm\sqrt{\gamma_1^{\,2} + \gamma_2^{\,2}} \tag{4-58}$$

Through knowledge of the recovered signal components, $\tilde{x}_r + \hat{x}_i$ and $\hat{x}_r - \tilde{x}_i$, the signal can be recovered as given in (4-59).

$$\begin{aligned} x_r[n] &= (\tilde{x}_r[n] + \hat{x}_i[n])\cos(\phi_c) + (\hat{x}_r[n] - \tilde{x}_i[n])\sin(\phi_c) \\ x_i[n] &= (\tilde{x}_r[n] + \hat{x}_i[n])\sin(\phi_c) - (\hat{x}_r[n] - \tilde{x}_i[n])\cos(\phi_c) \end{aligned} \tag{4-59}$$

Thus the unknown sequence can be recovered, along with the unknown additive phase constant. Numerical inversion is used to implement (4-55). Trials have shown that typically $N/2+2$ signal samples are required to be known to ensure that the matrix inversion is well defined.

4.5.2.2 Signal Reconstruction from the Phase of the ST-FT known to within a constant per time slice

In Section 4.5.2.1 signal reconstruction from the phase of the DFT known to within a constant was described. This approach can be applied to the complex signal reconstruction from the reassigned group delay problem. Since the phase of each time slice can be reconstructed to within an additive constant, (4-53) can be extended to included phase constants from multiple time slices. In (4-60) this extension is defined, where $\phi[m,k]$ is the phase from the m^{th} time slice, and $\phi_c[m]$ is the unknown phase constant for the m^{th} slice.

$$\begin{aligned} &\sum_{n=0}^{N-M-1} (\tilde{x}_r[m,n] + \hat{x}_i[m,n])\sin\left(\phi'[m,k] + \frac{2\pi nk}{N}\right) + (\hat{x}_r[m,n] - \tilde{x}_i[m,n])\cos\left(\phi'[m,k] + \frac{2\pi nk}{N}\right) \\ &+\cos(\phi_c[m]) \sum_{n=N-M}^{N-1} \sin\left(\phi'[m,k] + \frac{2\pi nk}{N}\right)(\tilde{x}_r[m-1,n] + \hat{x}_i[m-1,n]) \\ &+\sin(\phi_c[m]) \sum_{n=N-M}^{N-1} \cos\left(\phi'[m,k] + \frac{2\pi nk}{N}\right)(\hat{x}_r[m-1,n] + \tilde{x}_i[m-1,n]) = 0 \end{aligned} \tag{4-60}$$

$$\begin{aligned} \tilde{x}_r[m,n] &= \cos(\phi_c[m])x_r[n] \quad \tilde{x}_i[m,n] = \cos(\phi_c[m])x_i[n] \\ \hat{x}_r[m,n] &= \sin(\phi_c[m])x_r[n] \quad \hat{x}_i[m,n] = \sin(\phi_c[m])x_i[n] \end{aligned}$$

This formulation allows the computation of the unknown signal component, purely in terms of components of the previous time slice. Unlike the real signal case, this expression cannot be generalised into a global solution. This is due to the fact that the reconstruction formulation is non-linear in terms of the reconstructed

components. In order to extract $\phi_c[m]$ the products $(\tilde{x}_r[m-1,n]+\hat{x}_i[m-1,n])\cos(\phi_c[m])$ and $(\hat{x}_r[m-1,n]+\tilde{x}_i[m-1,n])\sin(\phi_c[m])$ must be computed. This expression is non-linear in terms of ϕ_c, and therefore post-processing is required to compute its value. It thus follows that a simple linear matrix form of equation (4-60) cannot be written, since it is not possible to extract the signal without previously calculating the phase shift, ϕ_c.

Although it is not possible to find a closed form solution to reconstruct the whole of the signal in a single step, a 'piece-wise' algorithm can still be developed. As with signal reconstruction from phase of ST-FT, this has the advantage of dividing the reconstruction procedure into smaller, less computationally demanding elements. However is also has the disadvantage that, in order to begin the reconstruction procedure some knowledge of the signal is required. Furthermore, such a procedure is numerical unstable, with numerical errors propagating through the solution.

4.5.2.3 Complex Signal Reconstruction from the Reassigned Group Delay Co-ordinate

One iterative approach to complex signal reconstruction follows. In order for the matrix formulation of (4-60) to have a solution, the overlap, M, between successive time slices must be greater than $N/2+2$ (*i.e.* $M > (N/2+2)$). In order to recover the first time slice, it is assumed that M samples are known over the length of window. This is sufficient signal data for equation (4-60) to be able to reconstruction the signal corresponding to the first time-slice. If the overlap used is much less than $N/2$, more than one time slice is used to recover a given signal sample. To reduce numerical errors, an average of the computed sample values reconstructed from each of time slices can be constructed. An iterative scheme incorporating this approach is given below.

1. Through application of (4-60), using the recovered phase known to within a constant and known sample, recover the unknown samples of the first segment.

2. Application of (4-57) and (4-59) will lead to the recovery of the real and imaginary parts of the signal, along with the phase off-set for that time slice.

3. Using this recovered signal information, compute the unknown part of next time slice, and the phase offset.

4. Repeat Part 3) averaging if overlap is sufficient until all the signal samples have been reconstructed.

To illustrate this algorithm, the complex chirp shown in Figure 20 is used as an example. The phase of the ST-FT is reconstructed to within an additive constant per slice, via inversion of the reassigned group delay parameter (4-10). The reassigned co-ordinates are shown in

Figure 30. The error between the reconstructed and normalised original signal is shown in

Figure 31. As can be seen, the more time-slices reconstructed, the greater the error between the recovered and original signal. As commented previously (Section 4.4.1.2), this is due to the build up of numerical errors. Empirical studies have shown increasing the size of the window function used reduces this error, by increasing the overlap between successive time slices and thus reducing the total number of iterations the algorithm has to perform.

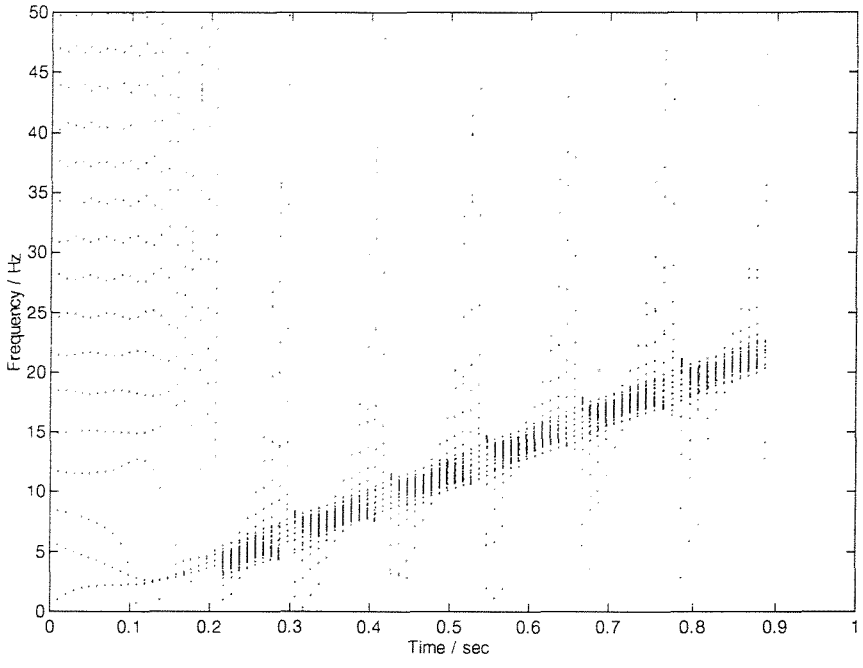

Figure 30 Reassigned Co-ordinates of Complex Linear Chirp Computed using Numerical Difference Method Based upon a 32pt Maximally Overlapped Gaussian Window ST-FT.

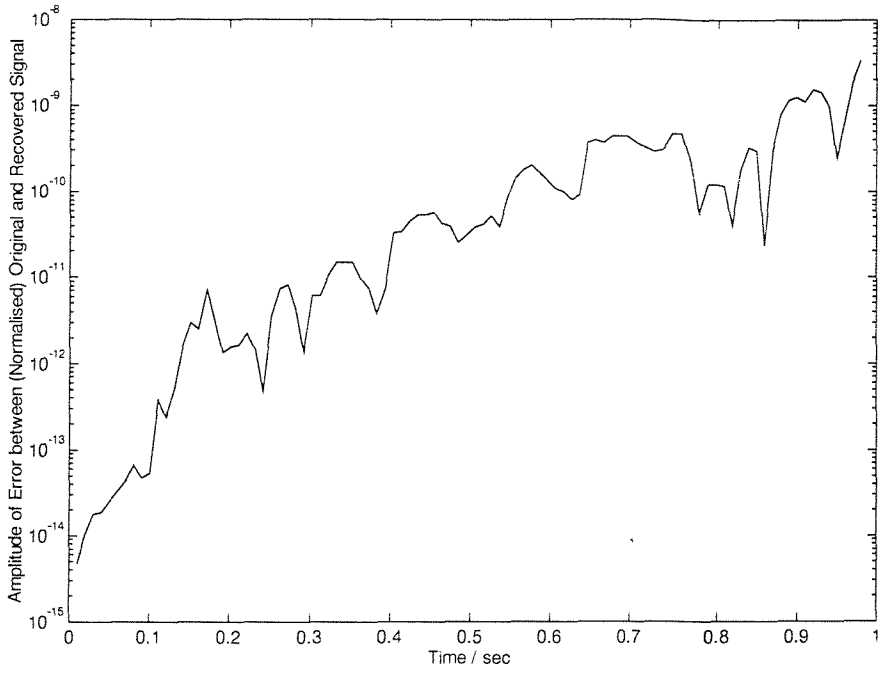

Figure 31 Error in Amplitude between Normalised Complex Linear Input Signal, and Signal Reconstructed via the Group Delay Reassigned Co-ordinate.

4.6 Conclusions

This chapter has reviewed signal reconstruction techniques from the phase of the DFT and presented new algorithms for reconstructing a signal from the phase of the ST-FT. Two approaches were derived; firstly signal reconstruction took place in an iterative piece-wise manner and secondly in a global closed linear form. Although computationally simpler, the piece-wise approach suffers from two limitations. These are that in order to reconstruct the first slice of the signal some prior knowledge is required, and that numerical errors propagate through the solution, increasing over the length of the signal.

The global approach does not suffer from either of these limitations. Signal reconstruction purely from the phase of the ST-FT can be achieved if the number of common samples between successive time-slices is sufficient large. Examples of signal reconstruction using a linear chirp and a pseudo random noise signal were given.

The link between the reassigned co-ordinates, as computed via a numerical difference method and the phase of the ST-FT, was established in a recursive equation. In the case where the signal is known to be real, the phase of the ST-FT can be assumed to start at zero across all time-slices. This information permits the recovery of the phase of the ST-FT from knowledge of the reassigned GD parameter, as computed via numerical differences. Once the phase has been recovered, the signal itself can be reconstructed through use of the global reconstruction algorithm. It follows therefore that real signals can be reconstructed from their reassigned GD co-ordinates.

For complex signals, the situation is more complicated. Since the phase reconstruction equation only yields the phase of each time slice to within an additive unknown constant, modifications are required to take this constant into account. If sufficient overlap is used then the system can still be solved. However, a global, closed, linear approach is no longer possible. A piece-wise approach is thus adopted, resulting in complex signal reconstruction from the reassigned group delay parameter. As previously where the use of a piece-wise approach results in two limitations, namely the requirement for some signal knowledge and the poor numerical performance.

The dual of signal reconstruction from the phase of the ST-FT is reconstruction from the magnitude. This is the subject of the following chapter.

Chapter 5 Signal Reconstruction from the Magnitude of the ST-FT

5.1 Introduction

In Chapter 4 algorithms to reconstruct a signal from the phase of its ST-FT were developed. It was found that, if the overlap between successive time slices was sufficiently high, then the phase of the ST-FT uniquely identified the signal to within an overall scale constant. In this chapter, analogous results for the modulus of the ST-FT are presented. Unlike the phase case, signal reconstruction from the magnitude of the ST-FT has described previously in the literature [Anderson92, 93, 94a-b, Nawab82, 83a-b].

Beginning once more with signal reconstruction from the DFT, this chapter reviews previously described techniques to reconstruct the signal in a piece-wise and closed form [Quatieri81]. These approaches for signal reconstruction from the magnitude of the DFT lead directly to a piece-wise approach to signal reconstruction from the magnitude of the ST-FT. Two global methods of reconstructing the signal from the magnitude of the ST-FT are then presented. Both of these techniques are iterative, but approach the problem in different ways.

5.2 Signal Reconstruction from the Magnitude of the FT

5.2.1 Minimum and Maximum Phase Signals

The z-transform of a discrete input sequence, $x[n]$ of length l_x is given in (5-1) [Oppenheim75]. Evaluation of the z-transform about the unit circle, (5-2), leads to the Discrete-Time version of the FT described previously in Chapter 2 Section 2.2.4.

$$X(z)=\sum_{n=0}^{l_1} x[n]z^{-n} \tag{5-1}$$

$$X(\omega)=X\left(e^{j\omega}\right)=X(z)\Big|_{z=e^{j\omega}} \tag{5-2}$$

Restricting the z-transform of a sequence to be rational for notional simplicity, Yegnanarayana *et al.* [Yegnanarayana84] expressed the z-transform as given in (5-3), where A is a real constant, n_0 is an integer, and $N(z)$ and $D(z)$ are polynomials in z^{-1}.

$$X(z)=Az^{-n_0}\frac{N(z)}{D(z)} \tag{5-3}$$

A signal is said to have 'minimum phase' if in (5-3) $n_0=0$ and if all the roots of $N(z)$ and $D(z)$ of $X(z)$ are inside the unit circle. Conversely, a signal is said to be 'maximum phase' if, for a given n_0, the poles and zeros of (5-3) lie outside the unit circle.

5.2.2 Maximum and Minimum Phase Signal Reconstruction from either Magnitude or Phase of the FT

In parallel with the development of algorithms to recover an arbitrary signal from the magnitude of the DFT, a class to algorithms were presented to reconstruct a signal from the magnitude (or phase) of its DFT if the signal was minimum or maximum phase [Oppenheim75, Quatieri81,Yegnanarayana83, 84].

The reconstruction of the time-series from partial (magnitude or phase) information is made possible because, for maximum or minimum phase signals there is a direct relationship between the magnitude and the phase of the DFT. The link between the modulus and phase of a minimum phase signal was given by Oppenheim *et al.* [Oppenheim75]. Quatieri *et al.* [Quatieri81] developed these ideas further, deriving iterative algorithms for signals reconstruction for minimum phase signals from either the magnitude or the phase of the DFT. Non-iterative algorithms were later derived [Yegnanarayana83, 84] for the reconstruction of signals which have minimum or

maximum phase, from either the magnitude or phase of the DFT. These algorithms used the relationship between the magnitude and phase of the DFT and the 'cepstral' coefficients [Oppenheim75].

For a minimum phase signal, then log-magnitude and phase can be written in terms of the cepstral coefficients as given in (5-5) and (5-6), where $\lambda(\omega)$ is a function which unwraps the phase.

$$X(\omega) = |X(\omega)| e^{j\phi(\omega)} \tag{5-4}$$

$$\ln|X(\omega)| = \frac{c(0)}{2} + \sum_{n=1}^{\infty} c(n)\cos(n\omega) \tag{5-5}$$

$$\phi(\omega) = -\sum_{n=1}^{\infty} c(n)\sin(n\omega) - 2\pi\lambda(\omega) \tag{5-6}$$

For a maximum phase signal, the same relationship exists between the log magnitude and the cepstral coefficients; however, the relationship between the phase and the cepstral coefficients is slightly altered (5-7).

$$\phi(\omega) = \sum_{n=1}^{\infty} c(n)\sin(n\omega) - 2\pi\lambda(\omega) \tag{5-7}$$

From knowledge of the modulus or phase, the cepstral coefficients can be solved for directly. If the magnitude of the DFT is known, then application either of (5-6) or (5-7) reconstructs the principle value of the phase of the DFT. If the phase is known, then application of (5-5) recovers the magnitude. Once both the magnitude and phase of the FT are known, the signal can be reconstructed directly through the IDFT.

5.2.3 Arbitrary Signal Reconstruction from the Magnitude of the DFT

Arbitrary signal reconstruction from the magnitude of the DFT has also previously been described in the literature [Hayes80]. In a similar fashion to the phase case, in order to be able to reconstruct a signal from the magnitude of the FT additional signal information is required. An iterative procedure, almost identical to that used to reconstruct the signal from the phase of the DFT (see Chapter 4 Section 4.3), was

implemented by Quatieri *et al.* [Quatieri81]. This procedure imposes known zero-samples and the desired DFT magnitude at every iteration. An alternative approach based upon the auto-correlation function was extended to allow signal reconstruction from the magnitude of the ST-FT [Nawab92, 93a-b]. It is this approach which is detailed here for signal reconstruction from the DFT.

The discrete auto-correlation function of the complex input signal $x[n]$ is given in (5-8). The relationship between the auto-correlation function and the squared magnitude of the signal's DFT is given in (5-9). This relationship is easily shown by expressing (5-8) as a convolution.

$$r[n] = \sum_{m=-N-1}^{N-1} x[m]x[m+n] \tag{5-8}$$

$$\left|X[k]\right|^2 = F\left[r[n]\right] \tag{5-9}$$

Since (5-9) defines the auto-correlation function in terms of the magnitude of the DFT, it follows that reconstructing the signal from the auto-correlation function is analogous to reconstructing the signal from the magnitude of the DFT.

Consider the $(N-1)^{\text{th}}$ lag of the autocorrelation function, $r[N-1]$, the summation for which is given in (5-10). Since $x[n]$ is only defined over the range, $0..N-1$, this summation can be fully expressed as given in (5-11). Assuming that the first signal sample $x[0]$ is known, then (5-12) allows the reconstruction of the last sample $x[N-1]$, from knowledge of the auto-correlation function at $r[N-1]$.

$$r[N-1] = \sum_{m=-N-1}^{N-1} x[m]x\left[(N-1)+m\right] \tag{5-10}$$

$$r[N-1] = x[0]x[N-1] \tag{5-11}$$

$$x[N-1] = r[N-1]/x[0] \tag{5-12}$$

Successive sample points can be reconstructed by solving a system of linear equations, given in (5-13). Nawab [Nawab83] showed that it is possible to reconstruct P unknown samples of $x[n]$, where $P \geq \lceil M/2 \rceil$ (where $M = N+2$

and $\lceil \alpha \rceil$ is the smallest integer greater or equal to α). Since (5-13) uses only the auto-correlation function to recover the unknown samples, (5-13) defines a method of reconstructing the signal of knowledge of the squared magnitude of the DFT. The requirements for signal reconstruction, that at least half the signal must be known, are in direct analogy with the conditions required for signal reconstruction from the phase DFT given in the previous chapter (Chapter 4 Section 4.3.2).

$$\begin{bmatrix} x[0] & & \\ x[1] & x[0] & \\ x[2] & x[1] & \\ \vdots & \vdots & \\ x[(M/2)-1] & x[(M/2)-2] & \cdots \quad x[0] \end{bmatrix} \begin{bmatrix} x[M-1] \\ x[M-2] \\ \vdots \\ x[M/2] \end{bmatrix} = \begin{bmatrix} r[M-1] \\ r[M-2] \\ \vdots \\ r[M/2] \end{bmatrix} \tag{5-13}$$

In the above it was assumed that the first sample and second half the signal are known. If the first sample is not known then signal reconstruction can still take place, to within an overall complex constant. Since the signal must have the same energy in both time and frequency domains (Parseval's Theorem, Equation (2-2)) the amplitude component of the complex constant can be removed. This resulting signal is thus correct to within an overall phase constant.

5.3 Piece-Wise Signal Reconstruction from the Magnitude of the ST-FT

In the case of signal reconstruction from the phase of the ST-FT, two distinct approaches were outlined, one piece-wise and one global. In considering signal reconstruction from the modulus of the ST-FT, the same distinction can be drawn. Two further distinct piece-wise approaches to the problem of recovering a time series from the magnitude of the ST-FT are given by Anderson [Anderson92, 93, 94a,b] and Nawab [Nawab82, 83a-b].

Through a series of papers Anderson [Anderson84, 85, 92, 93, 94a,b] derived the conditions under which a complex causal signal can be reconstructed from the magnitude of its ST-FT [Anderson92] and WT [Anderson93]. Anderson establishes the fact that as few as three points per time-slice of the spectrogram or scalogram may be sufficient to recover a complex causal signal. Details of this technique are not given here. However, in outline it works in a similar manner to all piece-wise

approaches, using signal information (in this case the causality of the signal) to recover information associated with the first time-slice, and then recovers the remainder of the signal based on this. Reconstruction of each sample takes place by identifying the common root of two quadratic equations, with reconstruction taking place to within an overall sign constant if the signal is real, and an overall phase constant if the signal is complex [Anderson94b]. In the following chapter, Anderson's method for signal reconstruction from the magnitude of the WT is described more fully.

The technique described by Nawab [Nawab82, 83a-b] makes use of the relationship between the auto-correlation function of a signal and the inverse DFT of the magnitude of the ST-FT. By assuming the signal to be causal, reconstruction of the signal samples from the first time-slice of the magnitude can be undertaken as described in Section 5.2. Once the effect of the windowing function has been removed, the reconstructed samples of one time-slice are then used in the recovery of the samples of the next time-slice. As with the phase case (Chapter 4 Section 4.4.1), in order for reconstruction to take place, the number of common samples between successive time-slices must be greater than half the window length used. Also in an analogous manner to the phase case, this iterative piece-wise approach allows numerical errors to propagate through the solution. An improved version was later developed [Nawab83] with greater error suppression characteristics arising from the least squares approach being adopted. Once more signal reconstruction takes place to within an overall sign constant if the signal is real, and an overall phase constant if the signal is complex. Quatieri [Quatieri83] showed that if the signal is maximally overlapped then as few as one or two spectrogram samples are required per time-slice to permit reconstruction through use of this auto-correlation method.

Both of these methods suffer from two limitations. Firstly they assume that the signal is causal in order to recover information encoded in the first time-slice. Although this assumption is mild, it does require the user to possess some *a priori* signal knowledge. Hence signal reconstruction does not take place purely from information provided from the magnitude of the ST-FT. The second limitation is of a numerical nature, *i.e.* that numerical errors in the reconstruction build up in an unstable fashion. This phenomena has already been demonstrated for the piece-wise reconstruction of a complex signal from the reassigned group delay parameter (Section 4.5.2)

In the phase case, in order to overcome similar limitations (Chapter 4 Section 4.4.1.2), global methods of signal reconstruction were developed. A new novel global approach for signal reconstruction from the magnitude of the ST-FT is presented in the following section.

5.4 Global Signal Reconstruction from the Magnitude of the ST-FT

As stated in Chapter 2, the spectrogram is a member of Cohen's Class of TFRs. Equation (2-70) defines a simple inversion for members of CC's of TFRs to recover the original signal (to within a multiplicative phase constant). This method of signal reconstruction from the spectrogram is not used in practice, since it requires division by the kernel function. As the kernel function of a spectrogram tends to zero away from the origin, this division operation becomes numerical unstable.

In this section two alternative global approaches are presented to the problem of signal reconstruction from the modulus of the ST-FT. The first approach previously described in the literature, employs the least-squares inversion of the ST-FT and uses an iterative procedure to converge towards the solution using only the modulus of the ST-FT [Griffin83, 84]. The second approach follows a similar strategy to that adopted in Chapter 4 Section 4.4.2 for the development of signal reconstruction from the phase of the ST-FT.

5.4.1 Least Squares Inversion of the ST-FT

The task of inverting the complete ST-FT in order to recover the original signal should be simple. A number of different methods have been proposed in the literature, so called 'overlap and add' and 'weighted overlap and add' techniques [Allen77, Portnoff80, Crochiere80]. Although all these techniques recover the signal from the ST-FT, their performance when altered or synthetic ST-FT are used, is not optimal. Optimal in this sense means minimising the Mean Squared Error (MSE) between the desired ST-FT and that created by the ST-FT of the constructed signal. Griffin [Griffin83, 84] described such a minimum least squares (MLS) procedure to compute a signal from a synthetic or modified ST-FT. Although originally defined in the continuous domain, the work is presented here in discrete form. Beginning with the definition of the discrete ST-FT as described by (2-33), equation (5-14) can be derived, where $F_n\left[x[n]\right]$ represents the DFT of the function $x[n]$ with respect to the

variable n, L is the overlap used between successive time-slices, and N is the length of the real windowing function $h[n]$.

$$X[m,k]=\sum_{l=0}^{N-1} x_{nL}[l]e^{-j\left(\frac{2\pi lk}{N}\right)}=F_l[x_{nL}[l]] \tag{5-14}$$

$$x_{mL}[l]=x[l]h[mL-l] \tag{5-15}$$

Consider a modified (or synthetic) ST-FT called $Y[m,k]$, the inverse FT of one time-slice of which is given in (5-16).

$$y_{mL}[l]=F^{-1}{}_k[Y[m,k]]=\frac{1}{N}\sum_{k=0}^{N-1} Y[m,k]e^{-j\left(\frac{2\pi lk}{N}\right)} \tag{5-16}$$

In general the function $Y[m,k]$ is not a valid ST-FT[1]. As such depending upon which method of reconstruction is chosen, a different time history will result. Therefore rather than simple inversion, the inversion of $Y[m,k]$ should be chosen to minimise the error between it and a valid ST-FT, $X[m,k]$. The distance measure between the desired ST-FT and a valid ST-FT is given in (5-17), where P is the number of time-slices present in the ST-FT. Application of the discrete form of Parseval's theorem (continuous form given previously in (2-2)) results in the form of the distance measure given in (5-18). Note that the equivalent equations in Griffin [Griffin83, 84] assumed real input sequence (thus removing the modulus brackets from the (5-18)). Here the more general result for a complex signal is developed. This equation computes a distance measure between the FT of the desired ST-FT, $y_{mL}[l]$ and the FT of a valid ST-FT, $x_{mL}[l]$.

$$D[x[n],Y[m,k]]=\sum_{m=0}^{P-1}\frac{1}{N}\sum_{k=0}^{N-1}|X[m,k]-Y[m,k]|^2 \tag{5-17}$$

$$D[x[n],Y[m,k]]=\sum_{m=0}^{P-1}\sum_{l=0}^{N-1}|x_{mL}[l]-y_{mL}[l]|^2 \tag{5-18}$$

[1] A valid ST-FT is a ST-FT which could have been created through application of the ST-FT equation upon a signal. An arbitrary 2-D function is not generally a valid ST-FT.

The distance measure is quadratic with respect to the signal, $x[n]$ and thus the minimisation of $D[x[n],Y[m,k]]$, obtained by setting the gradient with respect to $x[n]$, to zero has a closed form. This defines the optimal signal is a MLS sense. The minimisation of the distance measure with respect to the complex sequence $x[n]$ can be regarded as the minimisation of the distance measure with respect to the real and imaginary signal parts as independently described in (5-19). Therefore finding the minimum reduces to solving (5-20) and (5-21).

$$\frac{\partial(D[x[n],Y[m,k]])}{\partial x[n]} = \frac{\partial(D[x[n],Y[m,k]])}{\partial x_r[n]} + j\frac{\partial(D[x[n],Y[m,k]])}{\partial x_i[n]} \tag{5-19}$$

$$\frac{\partial(D[x[n],Y[m,k]])}{\partial x_r[n]} = 0 \tag{5-20}$$

$$\frac{\partial(D[x[n],Y[m,k]])}{\partial x_i[n]} = 0 \tag{5-21}$$

Consider (5-20), this can be expressed as given in (5-22). Expanding the right hand side, and simplifying, results in (5-23), where * denotes complex conjugation. The relationships (5-24) can readily be shown. Application of these relationships to equation (5-23) results in (5-25).

$$\frac{\partial\left(D[x[n],Y[m,k]]\right)}{\partial x_r[n]} = \frac{\partial\left(\sum_{m=0}^{P-1}\sum_{l=0}^{N-1}\left|x_{mL}[l]-y_{mL}[l]\right|^2\right)}{\partial x_r[n]} \tag{5-22}$$

$$\frac{\partial\left(D[x[n],Y[m,k]]\right)}{\partial x_r[n]} = \sum_{m=0}^{P-1}\sum_{l=0}^{N-1}\left(x_{mL}[l]-y_{ml}[l]\right)\frac{\partial x_{mL}^*[l]}{x_r[n]} + \left(x_{mL}[l]-y_{ml}[l]\right)^*\frac{\partial x_{mL}[l]}{x_r[n]} \tag{5-23}$$

$$\begin{aligned}\mathrm{Re}[z_1 z_2] &= \mathrm{Re}[z_1]\mathrm{Re}[z_2]-\mathrm{Im}[z_1]\mathrm{Im}[z_2] \\ 2\,\mathrm{Re}[z_1] &= z_1 + z_1^* \\ 2\,\mathrm{Im}[z_1] &= z_1 - z_1^*\end{aligned} \tag{5-24}$$

$$\frac{\partial\left(D[x[n],Y[m,k]]\right)}{\partial x_r[n]} = 2\sum_{m=0}^{P-1}\mathrm{Re}\left[\left(x_{mL}[n]-y_{ml}[n]\right)h[mL-n]\right] = 0 \tag{5-25}$$

Performing a similar analysis on the imaginary part of the signal results in (5-26). Combining the expressions for the real and imaginary parts back yields (5-27). Rearranging this equation to solve directly for the signal results in the MLS solution as given in (5-28).

$$\frac{\partial\left(D\left[x[n],Y[m,k]\right]\right)}{\partial x_i[n]} = 2\sum_{m=0}^{P-1}\mathrm{Im}\left[\left(x_{mL}[n]-y_{ml}[n]\right)h[mL-n]\right]=0 \tag{5-26}$$

$$\sum_{m=0}^{P-1}\left(x_{mL}[n]-y_{ml}[n]\right)h[mL-n]=0 \tag{5-27}$$

$$\sum_{m=0}^{P-1}\left(x[n]h[mL-n]\right)h[mL-n]=\sum_{m=0}^{P-1}y_{ml}[n]h[mL-n]$$

$$x[n]=\frac{\sum_{m=0}^{P-1}h(mL-n)\,y_{mL}[n]}{\sum_{m=0}^{P-1}\left(h[mL-n]\right)^2} \tag{5-28}$$

Thus from an invalid ST-FT, the optimal reconstructed signal in a MLS sense can be computed. A similar procedure will be employed in Chapter 6 to enable MLS signal reconstruction from a generalised version of the WT.

The implementation of (5-28) requires little additional computation over other previously defined methods [Portnoff80, Crochiere80]. If the windowing function chosen such that (5-29) is satisfied then some reduction in computation is possible [Griffin83, 84].

$$\sum_{m=0}^{P-1}\left(h[mL-n]\right)^2=1 \tag{5-29}$$

5.4.2 Signal Reconstruction based on MLS Inversion of the ST-FT

Beginning with the MLS inversion of the ST-FT, Griffin [Griffin83, 84] then proceeded to develop a method for signal reconstruction from the magnitude of the ST-FT. This will be referred to as the least squares ST-FT method of signal reconstruction. Through repeated application of MLS solution for the discrete ST-FT, an iterative procedure was described. A discrete version of Griffin's iteration is

given in (5-30), where $|Y[m,k]|$ is the desired modulus, and $X^p[m,k]$ is the ST-FT of the current signal estimate, $x^p[n]$. The algorithm is now described.

1. Choose an initial estimate for the signal, denoted $x^0[n]$, set $p=0$.
2. Using $x^p[n]$ construct the ST-FT through the application of equation (2-33) denoted $X^p[m,k]$.
3. Impose the required amplitude $|Y[m,k]|$ upon the phase of $X^p[m,k]$ as given in (5-31). This variable is given the name $\hat{Y}[m,k]$.
4. Use $\hat{Y}[m,k]$ to compute a new estimate of the signal, $x^{p+1}[n]$ through the application of (5-30).

Repeated application of Steps 2-5 defines the iteration. The algorithm is stopped when the error between the amplitude of the ST-FT, $|X^p[m,k]|$, and the desired amplitude, $|Y[m,k]|$, defined in (5-32) is considered to have converged in some fashion.

Since the distance measure, (5-32) is not quadratic in terms of the signal, it may have more than one minimum. Convergence to a solution is guaranteed [Griffin83], however there is no guarantee that this is a global minimum.

$$x^{p+1}[n]=\frac{\sum_{m=0}^{P-1}h[mL-n]\left(\frac{1}{N}\sum_{k=0}^{N-1}\hat{X}^p[m,k]e^{j\left(\frac{2\pi nk}{N}\right)}\right)}{\sum_{m=0}^{P-1}\left(h[mL-n]\right)^2} \tag{5-30}$$

$$\hat{X}^p[m,k]=|Y[m,k]|\frac{X^p[m,k]}{|X^p[m,k]|} \tag{5-31}$$

$$D\left[x[n],|Y[m,k]|\right]=\frac{1}{N}\sum_{k=0}^{N-1}\sum_{m=0}^{P-1}\left(|X[m,k]|-|Y[m,k]|\right)^2 \tag{5-32}$$

This method of signal reconstruction from the modulus does not assume that the signal is causal as in the piece-wise reconstruction ([Quatieri83] amongst others).

Since the distance measure is a function of the modulus of the ST-FT obtained from the current estimate of the reconstructed signal and of the desired modulus, it is invariant to overall phase shifts. The result of this is that the signal can only be reconstructed to within an overall phase constant.

The iteration is run until the distance measure (5-32) falls below a user specified threshold or until a predefined number of iterations is exceeded. The method requires computation of the MLS signal reconstruction algorithm given in (5-28) and a discrete ST-FT every iteration. Even for moderately long signals, this is not computationally prohibitive.

Application of this method of signal reconstruction from the magnitude of the ST-FT has been used in speech modification and synthesis [Griffin84] and is applied in Chapter 7 to the extension in time of heart sounds.

5.4.2.1 Examples of Signal Reconstruction from the Spectrogram (Least Squares ST-FT Method)

Two examples are now given to illustrate signal reconstruction from the modulus of the ST-FT. The two signals are a linear FM chirp sequence, and a Gaussian pseudo random sequence. Both signals are 100 points long, and are assumed to be sampled at 100 Hz. In order to compute the ST-FT a 32-point maximally overlapped Gaussian window function is used.

5.4.2.1.1 Example 1 Linear FM Chirp Sequence

The time series of the linear FM chirp is shown in Figure 32. The spectrogram for this signal is depicted in Figure 33. Through use of the modulus (square root of the spectrogram) and application of (5-30), the signal was reconstructed. The error at each successive iteration is given in Figure 34, and can be seen to be monotonically decreasing, until levels of the order of 10^{-8} error (at which point numerical precision affects the result). The difference in amplitude between the original and reconstructed signal are as given in Figure 35, and the error in the phase between the two signals is given in Figure 36. The magnitude has been reconstructed to within a small error (below 3×10^{-7}), and the phase to within an overall constant.

5.4.2.1.1 *Example 2 Pseudo Random Gaussian White Noise Sequence*

The second example signal is a pseudo-random Gaussian signal, the time series and spectrogram of which is given in Figure 37 and Figure 38 respectively. From Figure 39 the algorithm can be seen to have converged, resulting in a small error between the desired and reconstructed signals. Similar performance to that for the linear chirp can also be seen in the difference between the amplitudes (Figure 40) and the phases (Figure 41).

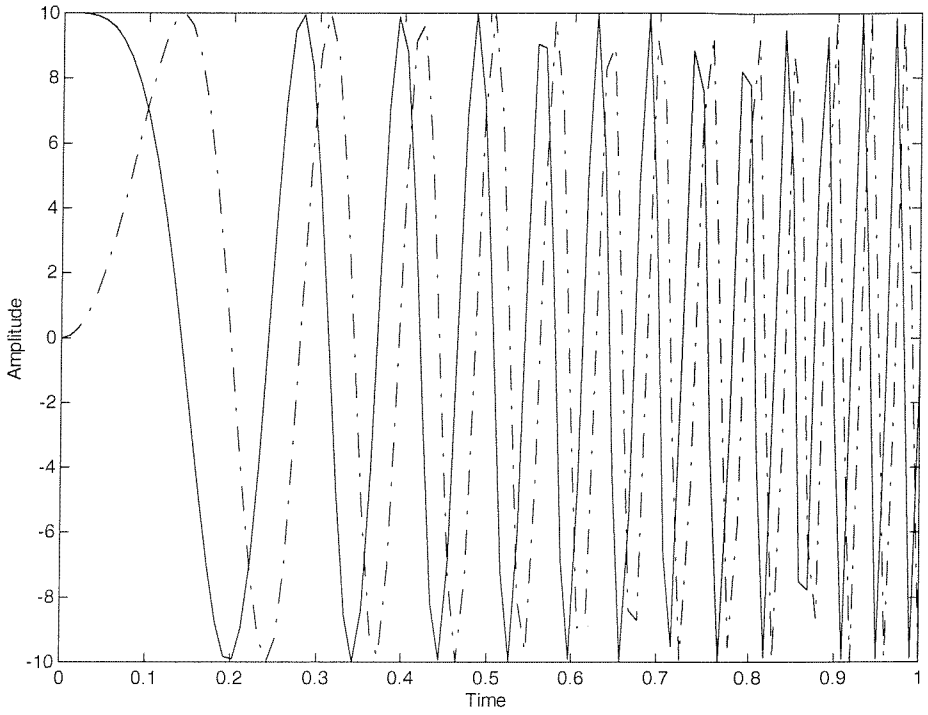

Figure 32 Linear FM Harmonic Signal (Real Part Solid, Imaginary Part Dashed Lines)

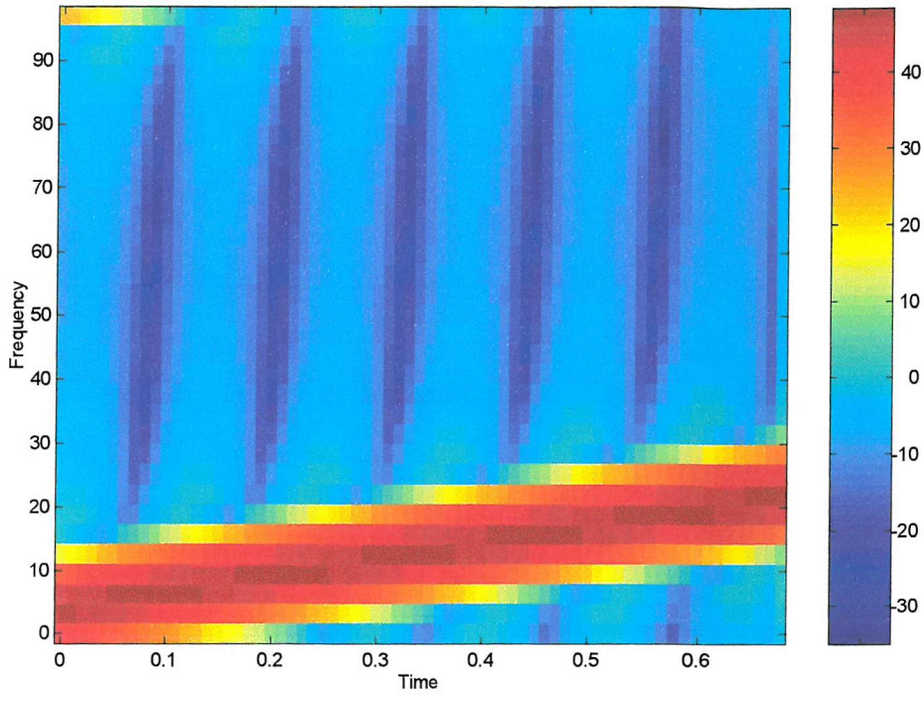

Figure 33 Spectrogram of Linear FM Harmonic Signal

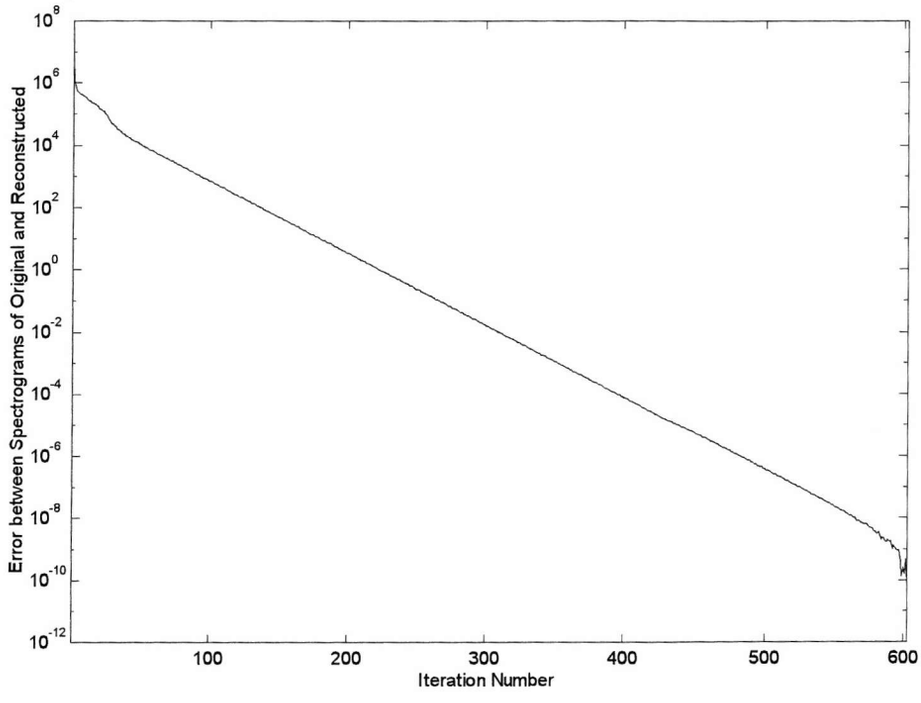

Figure 34 Error between Spectrogram of Original and Current Estimate of the Signal – Linear Chirp

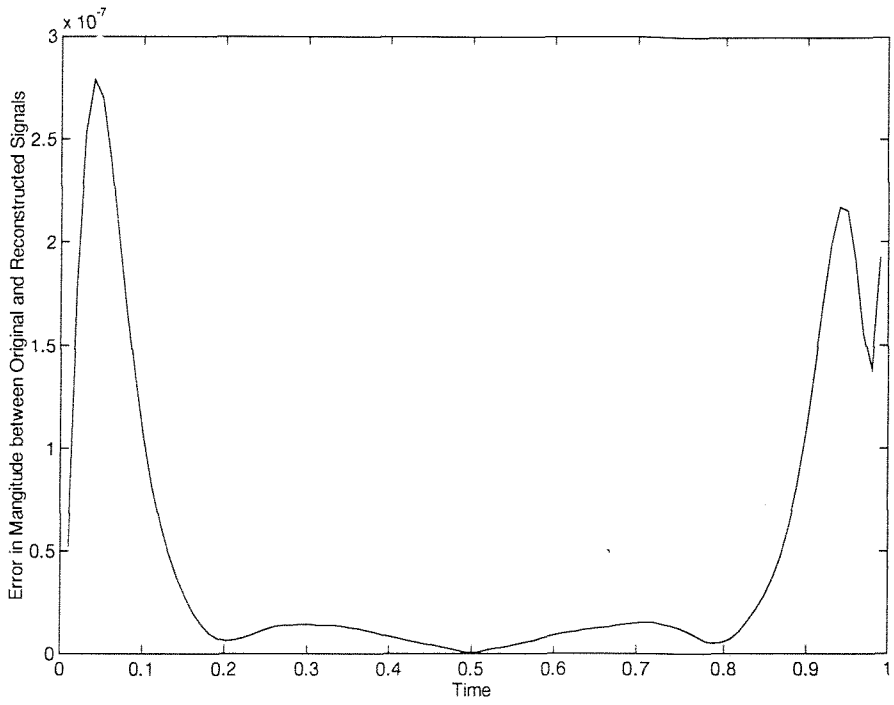

Figure 35 Error Between Amplitude of Original and Reconstructed Signal - Linear Chirp

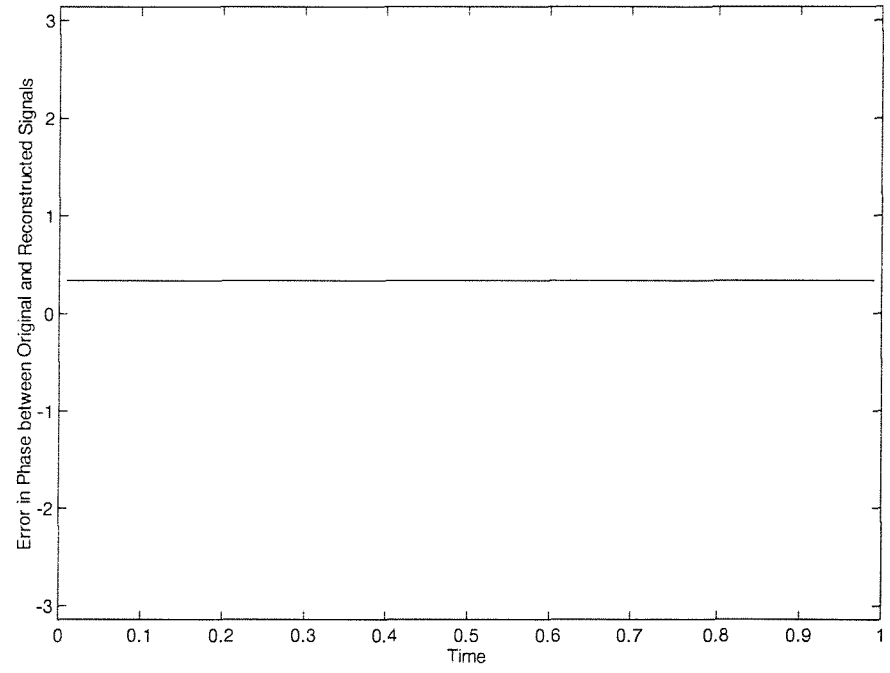

Figure 36 Error Between Phase of Original and Reconstructed Signal - Linear Chirp

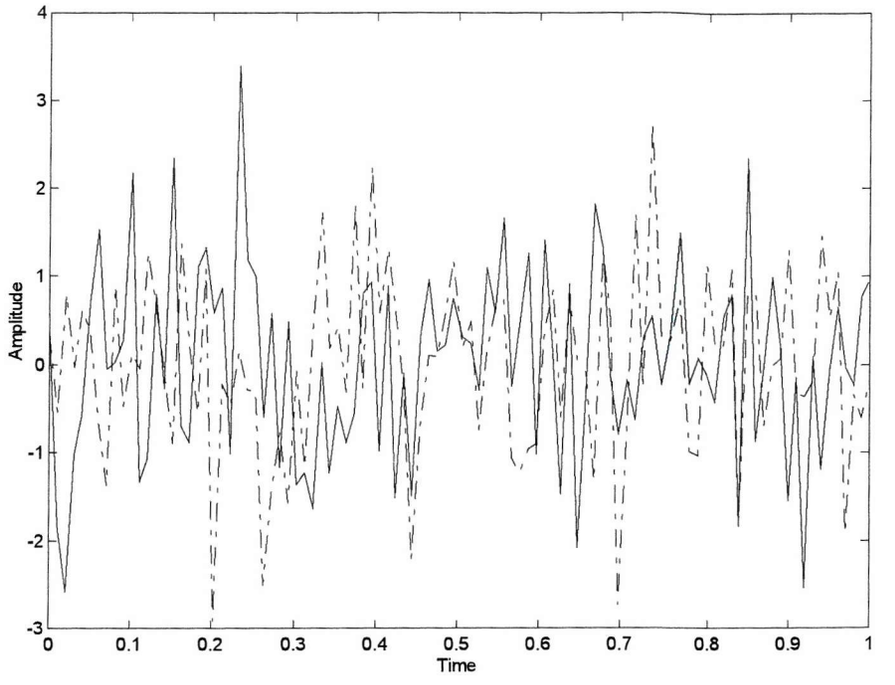

Figure 37 Pseudo-Random Gaussian White Noise Signal (Real Part Solid, Imaginary Part Dashed Lines)

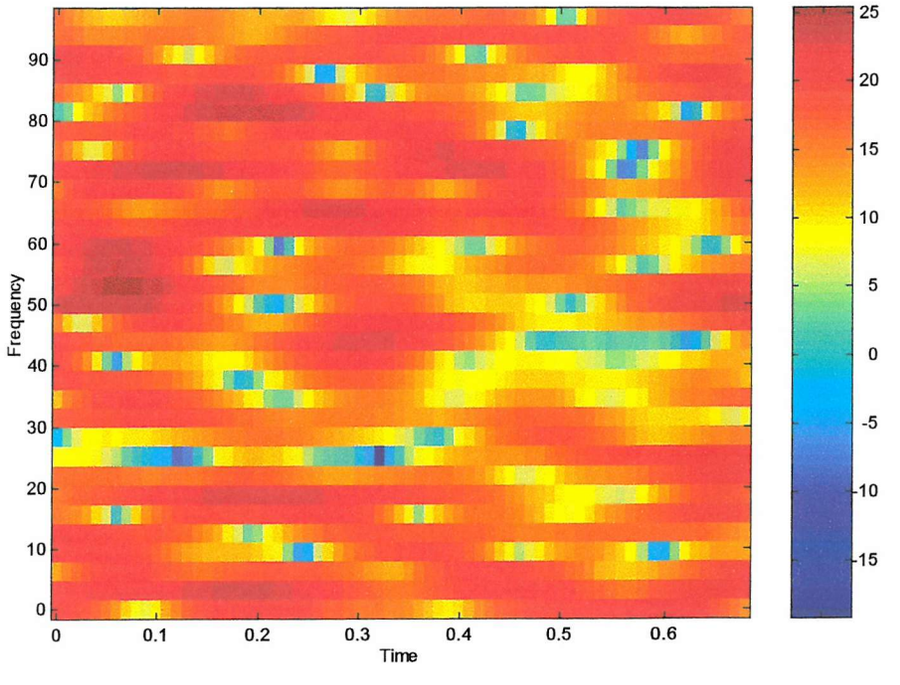

Figure 38 Spectrogram of Gaussian White Noise Signal

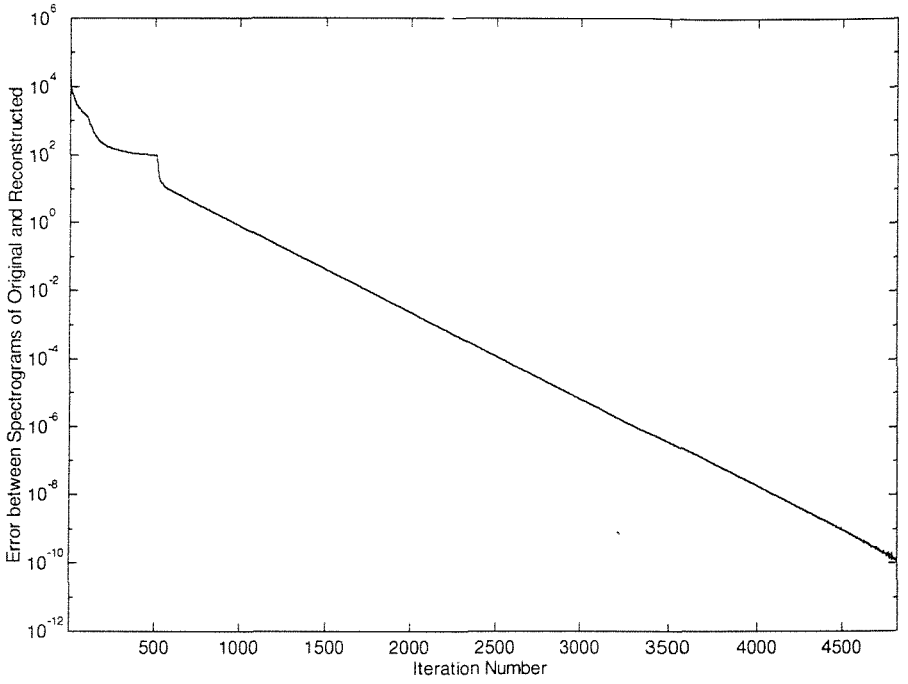

Figure 39 Error between Spectrogram of Original and Current Estimate of the Signal – Gaussian White Noise

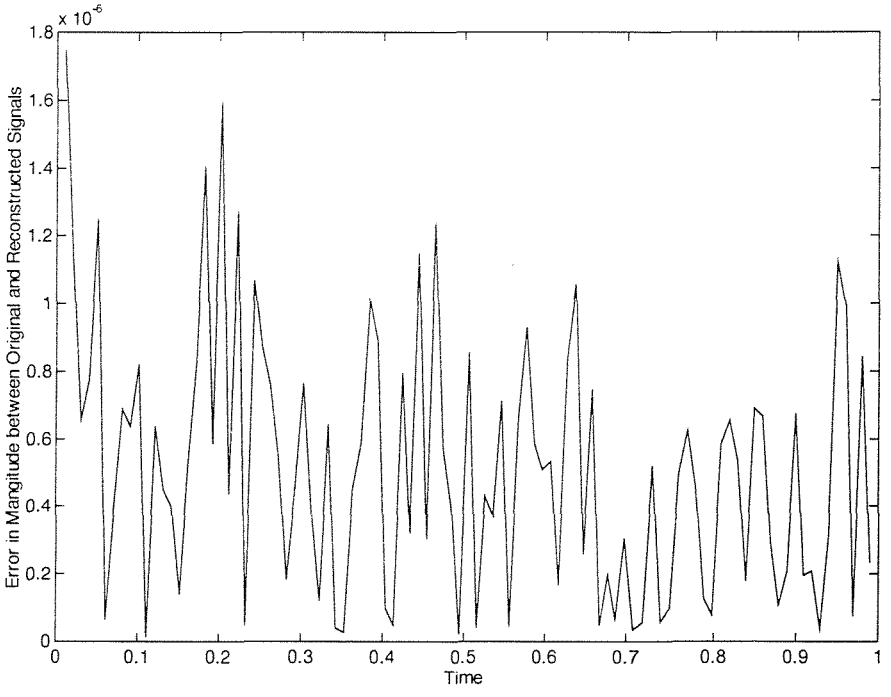

Figure 40 Error Between Amplitude of Original and Reconstructed Signal – Gaussian White Noise

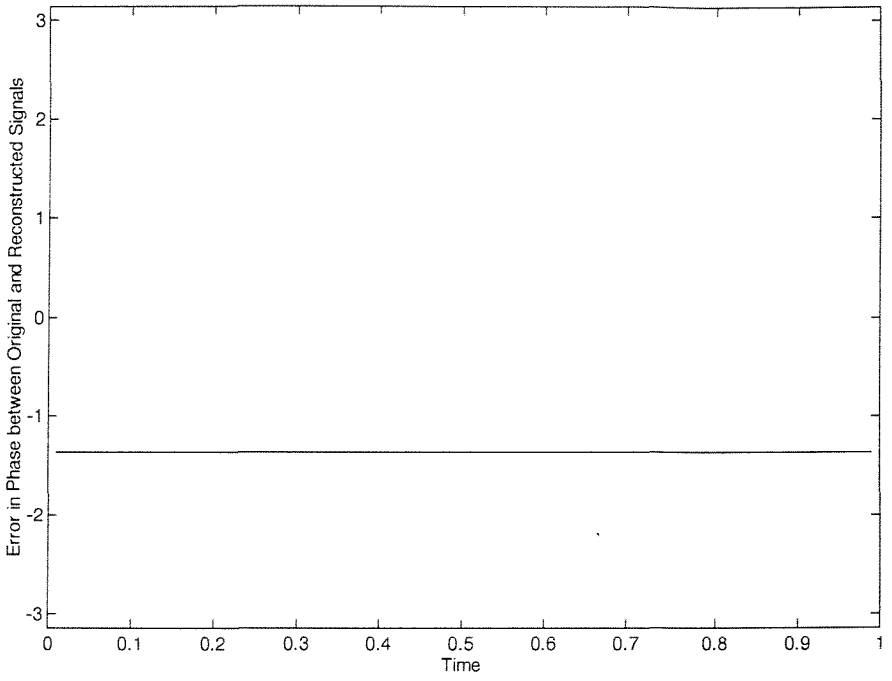

Figure 41 Error Between Phase of Original and Reconstructed Signal – Gaussian White Noise

5.4.3 Quadratic Form of the Magnitude of the DFT

An alternative global approach to the method discussed in Section 5.4.2 is now presented. In the case of signal reconstruction from the phase of the ST-FT an expression was derived for the signal by expanding the phase into terms involving the real and imaginary parts of the transform. The analogous expansion for the squared magnitude of the DFT of a complex input sequence can be created (5-33).

$$|X[k]|^2 = \mathrm{Re}[X[k]]^2 + \mathrm{Im}[X[k]]^2 \tag{5-33}$$

$$|X[k]|^2 = \left(\sum_{n=0}^{N-1} x_r[n]\cos\left(\frac{2\pi nk}{N}\right) + x_i[n]\sin\left(\frac{2\pi nk}{N}\right)\right)^2 + \left(\sum_{n=0}^{N-1} x_i[n]\cos\left(\frac{2\pi nk}{N}\right) - x_r[n]\sin\left(\frac{2\pi nk}{N}\right)\right)^2$$

By considering just one element in frequency, equation (5-33) can be written as a quadratic form [Anton91]. A matrix expression forming this relationship is given in (5-34).

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \left|X\left[k\right]\right|^2 \tag{5-34}$$

where

$$\mathbf{x} = \left[x_r[0] \quad x_r[1] \quad \cdots \quad x_r[N-1] \quad x_i[0] \quad x_i[1] \quad \cdots \quad x_i[N-1]\right]^T$$

$$\mathbf{A}_{n,m}(k) = \begin{cases} n > m & \mathbf{A}_{m,n}(k) \\ n = m & 1 \\ m = n + N & 0 \\ m < N & \cos\left(\dfrac{2\pi k}{N}(m-n)\right) \\ m > N \quad m < n+N & -\sin\left(\dfrac{2\pi k}{N}(m-n)\right) \\ m > N \quad m > n+N & \sin\left(\dfrac{2\pi k}{N}(m-n)\right) \end{cases}$$

Equation (5-34) expresses the relationship between the sample points in the input data, and just one point in the frequency domain. In order to extend (5-34) to cover all provided frequency points, its simple quadratic form is lost. This new formulation, making use of all frequency points, is given in (5-35):

$$\hat{\mathbf{x}}\hat{\mathbf{A}}\mathbf{x} = \mathbf{X} \tag{5-35}$$

where

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^{\mathrm{T}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}^{\mathrm{T}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}^{\mathrm{T}} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} \left|X[0]\right|^2 \\ \left|X[1]\right|^2 \\ \vdots \\ \left|X[N-1]\right|^2 \end{bmatrix} \qquad \hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A(0)} \\ \mathbf{A(1)} \\ \vdots \\ \mathbf{A(N-1)} \end{bmatrix}$$

and $\mathbf{0}$ is a vector of zeros with the same dimensions as vector $\mathbf{x}^{\mathrm{T}}$. Although not longer in a quadratic form as described in (5-34), this formulation is still bilinear in terms of the signal. Such a non-linear system cannot be easily solved, and a linearised form is sought. One linearisation of (5-35) is given in (5-36):

$$\hat{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{X} \tag{5-36}$$

where

$$\tilde{\mathbf{x}} = \left[x_r[0]^2 \quad x_r[0]x_r[1] \quad \cdots \quad x_r[0]x_r[N-1] \quad x_r[0]x_i[0] \quad \cdots \quad x_r[0]x_i[N-1] \right.$$
$$\left. \cdots \quad x_i[N-2]x_i[N-1] \quad x_i[N-1]^2 \right]$$

The vector $\tilde{\mathbf{x}}$ contains all the products between the real and imaginary components of the input signal. It follows therefore that the size of this vector is $4N^2$ where N the length of the complex vector $\mathbf{x}$. In fact there are only $2N^2 + N$ unique variables.

Since the vector containing the magnitude of the DFT, $\mathbf{X}$, is N points long, it follows that this system of equations is under-determined. Thus there are an infinite number of solutions to the problem. Each known signal sample reduces the set of unknowns by N. Equation (5-37) expresses the relation between the number of unknowns ($2N^2 + N$), and the required number of complex samples which must be known ($2\alpha N$) such that there are fewer unknowns that the number of points provided by the DFT (N). Solving for the number of required samples, α, results in the need for at least N real, or $N/2$ complex, samples to be known (5-38). This requirement, for there to be at least $N/2$ known samples is the same as that required by Nawab's algorithm described in previously in this section, and is also the same requirement as was derived in the case were only phase information from the DFT was known (Chapter 4 Section 4.4.1.1).

$$2N^2 + N - 2N\alpha \geq N \tag{5-37}$$

$$\alpha \leq N \tag{5-38}$$

5.4.4 Extension of the Quadratic Form to the Spectrogram

Equation (5-35) is a bilinear equation expressing the relationship between the squared magnitude of the DFT and the signal that generated it. The reformulation of the bilinear form into a linear form (5-36) showed that some signal knowledge was required in order for the squared magnitude of the DFT to uniquely identify the signal. This method is now extended from the DFT to the ST-FT. Although the formulation given in (5-36) is a matrix form, it is bilinear in terms of the signal samples, and not suitable for extension to the spectrogram. Thus it is the

formulation provided by (5-35), that is extended. Using the definition of the ST-FT (2-36) and by extending (5-35) over multiple time-slices yields (5-39):

$$\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x} = \mathbf{S} \tag{5-39}$$

Where $\mathbf{x} = \begin{bmatrix} x_r[0] & x_r[1] & \cdots & x_r[l_x-1] & x_i[0] & x_i[1] & \cdots & x_i[l_x] \end{bmatrix}^T$

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_0 & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_1 & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{P-1} \end{bmatrix}$$

$N-1$ Columns

$$\mathbf{S} = \begin{bmatrix} |S[0,0]|^2 \\ |S[0,1]|^2 \\ \vdots \\ |S[0,N-1]|^2 \\ |S[1,0]|^2 \\ \vdots \\ |S[T,N-1]|^2 \end{bmatrix}$$

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{hx}_r\left[(Ln)\ldots(Ln+(N-1))\right] & \mathbf{Z}\left[l_x-(Ln+(N-1))\right] & \mathbf{hx}_i\left[(Ln)\ldots(Ln+(N-1))\right] \end{bmatrix}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{SA}[0] \\ \mathbf{SA}[1] \\ \vdots \\ \mathbf{SA}[T-1] \end{bmatrix} \qquad \mathbf{SA}[n] = \begin{bmatrix} \mathbf{PA}[n,0] \\ \mathbf{PA}[n,1] \\ \vdots \\ \mathbf{PA}[n,N-1] \end{bmatrix} \qquad \mathbf{A}(k) = \begin{bmatrix} \mathbf{RR}(k) & \mathbf{RI}(k) \\ \mathbf{IR}(k) & \mathbf{II}(k) \end{bmatrix} \tag{5-40}$$

$$\mathbf{PA}_{(Ln)\ldots(L.n+(N-1)),(L.n)\ldots(L.n+(N-1))}[n,k] = \mathbf{RR}(k).\mathbf{H}$$
$$\mathbf{PA}_{(L.n)\ldots(L.n+(N-1)),(L.n+l_x)\ldots(L.n+l_x+(N-1))}[n,k] = \mathbf{RI}(k).\mathbf{H}$$
$$\mathbf{PA}_{(L.n+l_x)\ldots(L.n+l_x+(N-1)),(L.n)\ldots(L.n+(N-1))}[n,k] = \mathbf{IR}(k).\mathbf{H}$$
$$\mathbf{PA}_{(L.n+l_x)\ldots(L.n+l_x+(N-1)),(L.n+l_x)\ldots(L.n+l_x+(N-1))}[n,k] = \mathbf{II}(k).\mathbf{H}$$

$$\mathbf{H} = \begin{bmatrix} h[0] & h[1] & \cdots & h[N-1] \\ h[0] & h[1] & \cdots & h[N-1] \\ \vdots & \vdots & \ddots & \vdots \\ h[0] & h[1] & \cdots & h[N-1] \end{bmatrix}$$

where $\mathbf{Z}[n]$ is a vector containing n zeros, $\mathbf{PA}_{i,j}$ is the i, j element of the matrix $\mathbf{PA}$, l_x is the length of the input signal, P is the number of time-slices, L is the overlap between successive time-slices and the vector $\mathbf{h}$ contains the samples of the

time-limiting window function used. The matrices, $\mathbf{RR},\mathbf{RI},\mathbf{IR},\mathbf{II}$ are defined by splitting up the $\mathbf{A}$ matrix defined in (5-34). Let us dwell on the form of the $\tilde{\mathbf{A}}$ and $\hat{\mathbf{x}}$ matrices. Previously the $\hat{\mathbf{A}}$ matrix in (5-35) was created by concatenating together the smaller $\mathbf{A}$ matrices defined by (5-34). In the ST-FT case, each of these matrices must only operate upon a small sub-set of the whole signal, the length of which depends on the size of the data window used, and the location of which depends on the time-slice number and the overlap being used. In order to correctly position the smaller $\mathbf{A}$ matrices in terms of the $\mathbf{x}$ vector, zero vectors are used. Furthermore the original $\mathbf{A}$ matrices need to be split into components which operate only on products of the real with the real, the real with the imaginary etc. (5-40).

Although the formulation given in (5-39) is simple (albeit if the required matrix definitions are long winded), there are two principal problems with this definition. The first of which relates to the rapid increase in the size of $\hat{\mathbf{A}}$ as the signal length increases. For each additional complex sample, the size of the matrix $\hat{\mathbf{x}}$ will increase. The dimensions of the matrix are $\left(P.N,2T.l_x.N\right)$. If the signal is maximally overlapped, then for each additional sample there is an additional time-slice in the ST-FT. This increases the overall size of the matrix by $\left(N,\left(P+l_x+1\right)N\right)$. This rate of increase in size of matrix is prohibitive.

The second problem with the definition given in (5-39), is that once more the formulation is non-linear in terms of the input signal. One approach to solve this problem is to employ an iterative procedure to compute the solution. A fixed-point iteration is defined in (5-41), where the current estimate of the signal is denoted by $\mathbf{x}_p$. The matrix $\hat{\mathbf{x}}_p\tilde{\mathbf{A}}$ is generally not square, and thus pseudo-inverse procedures are used, the pseudo-inverse of being denoted by $^{\#}$.

$$\mathbf{x}_{p+1}=\left(\hat{\mathbf{x}}_p\tilde{\mathbf{A}}\right)^{\#}\mathbf{S} \tag{5-41}$$

Convergence is obtained when the spectrogram of the signal $\mathbf{x}_{p+1}$ is the same as the desired spectrogram as given by the vector $\mathbf{S}$. Numerical trials have shown that this procedure is convergent to the required solution if sufficient overlap is used, although no theoretical proof has been developed. As with the phase reconstruction case, the algorithm appears to require $L<0.3N$. In the phase case this requirement

was so that a phase matrix had sufficient rank. In the modulus case an overlap smaller than this results in a lack of convergence.

A number of other different methods could have been used to solve (5-39). Re-arranging (5-39) leads directly to (5-42). Solving this equation can now be thought of as finding the roots to a set of non-linear equations. This can be represented in the form given in (5-43), where $\mathbf{F}$ is the system of equations described by $\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x}-\mathbf{S}$; a function of the signal $\mathbf{x}$.

$$\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x}-\mathbf{S}=\mathbf{0} \tag{5-42}$$

$$\mathbf{F}(\mathbf{x})=\mathbf{0} \tag{5-43}$$

Various numerical methods have been proposed for finding these roots, one of the most popular being Newton's method [Press92, Stoer93 and Kincaid91]. Unlike the fixed-point algorithm described by equation (5-41), application of Newton's method to this problem guarantees convergence if the initial estimate of the signal is sufficient close to a root [Press92].

$$\mathbf{x}_{p+1}=\mathbf{x}_p+\delta\mathbf{x} \tag{5-44}$$

where

$$\delta\mathbf{x}=-\mathbf{J}^{-1}\mathbf{F}(\mathbf{x}) \tag{5-45}$$

Here $\mathbf{J}$ is the Jacobian matrix. The general form of this matrix is given in (5-46), where $\mathbf{F}_n$ is the n^{th} equation and x_n is the n^{th} element of $\mathbf{x}$. The Jacobian for the system of equations defined by $\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x}-\mathbf{S}$ is relatively simple to compute numerically, although it has no simple analytic form.

Unless the number of equations in $\mathbf{F}$ is the same as the number of variables $\mathbf{x}$ then this matrix will not be square and thus no simple inverse will exist. Typically there are more equations than unknowns and thus a pseudo inverse matrix, can be used in (5-45).

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{F}_1}{\partial x_1} & \frac{\partial \mathbf{F}_1}{\partial x_2} & \cdots & \frac{\partial \mathbf{F}_1}{\partial x_n} \\ \frac{\partial \mathbf{F}_2}{\partial x_1} & \frac{\partial \mathbf{F}_2}{\partial x_2} & \cdots & \frac{\partial \mathbf{F}_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{F}_N}{\partial x_1} & \frac{\partial \mathbf{F}_N}{\partial x_2} & \cdots & \frac{\partial \mathbf{F}_N}{\partial x_n} \end{bmatrix} \tag{5-46}$$

Newton's method is guaranteed to converge to a root of the equation in the neighbourhood of the starting point of the iteration. However, since we have no knowledge of the roots of the equation, there is a requirement for an algorithm that converges to the solution, regardless of the starting point. A modified version of Newton's method can be constructed which will guarantee convergence to a solution of the equation from an arbitrary starting point, by backtracking if the error doesn't decrease between successive iterations [Press92]. Although guaranteed to converge, the algorithm may converge to a local rather than global solution. The MLS technique reviewed in Section 5.4.2 suffers from the same limitation.

When reconstructing a signal from the squared magnitude of the ST-FT, the signal is recovered to within an overall phase constant. In the case of signal reconstruction from phase, it was commented (Chapter 4 Section 4.4.2) that if a signal has two components separated by a number of zeros, greater than a window length, then each of the two components would be reconstructed to within an individual magnitude constant. An analogous limitation is also present in the reconstruction from magnitude case, since if the signal is zero for greater than N samples, the ST-FT and thus the spectrogram are also zero. In this case, no information is carried between the two segments, and thus each segment is recovered to within its individual phase offset. Computer trials have shown this to be the case. As commented upon previously the presence of measurement noise is likely to render this unimportant in practice.

5.4.4.1 Examples of Signal Reconstruction from Spectrogram (Bilinear Form)

Two examples are now presented to show the bilinear form of the problem working. As before, the two example signals are a complex linear chirp, and a pseudo-random signal. Due to computational considerations the size of the two signals has been

reduced from those used previously in this section to 50 points. Both signals are assumed to have been sampled at 50Hz. The ST-FT was computed using a 16 point maximally overlapped Gaussian window. The signal were reconstructed using the fixed-point iteration algorithm described in equation (5-41) since this technique appeared to converge with fewer iterations than Newton's method.

5.4.4.1.1 Example 1 Linear FM Chirp Sequence

The ST-FT of the linear FM signal given in Figure 42 was computed. The spectrogram (squared magnitude of the ST-FT) is shown in Figure 43, where the increasing frequency of the FM tone can be seen. Through use of the spectrogram and application of (5-39) the signal was reconstructed. The reducing error between the spectrogram of the original signal and that of the reconstructed estimate is shown in Figure 44. Although the error is not monotonically decreasing, after only 20 iterations the system converges to an error below 1×10^{-16}, which approximately corresponds to the numerical precision of the computer used. The amplitude difference between the original and reconstructed signal is given in Figure 45. As with signal reconstruction from the phase of ST-FT of the recovered signal, the amplitude of the recovered signal closely matches the original. However, unlike that case, no signal normalisation of the original signal is necessary. Since no normalisation has taken place, a phase difference between the original and reconstructed signal is anticipated and can be seen in Figure 46.

5.4.4.1.2 Example 2 Pseudo Random Gaussian White Noise Sequence

The same performance shown for the linear chirp in previous example is now demonstrated for a pseudo random sequence. The time series of the sequence is given in Figure 47, and spectrogram is shown in Figure 48. Similar convergence characteristics can also be seen in Figure 49. Once more, the error between the amplitudes of the original and reconstructed signals is small, (Figure 50), and the reconstruction can be seen to take place within an overall phase constant (Figure 51).

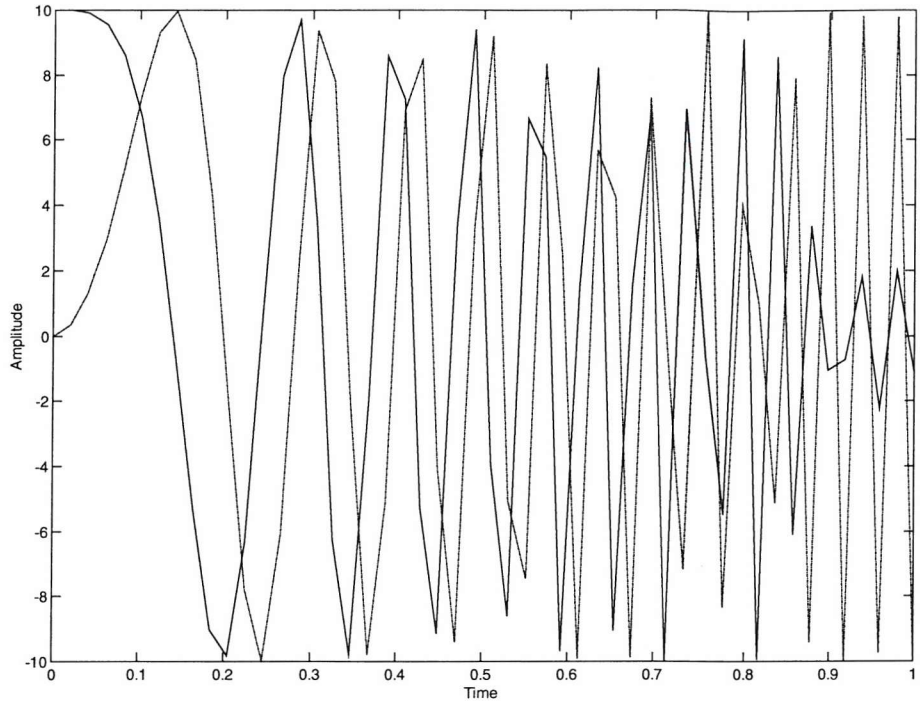

Figure 42 Linear FM Harmonic Signal (Real Part Solid, Imaginary Part Dashed Lines)

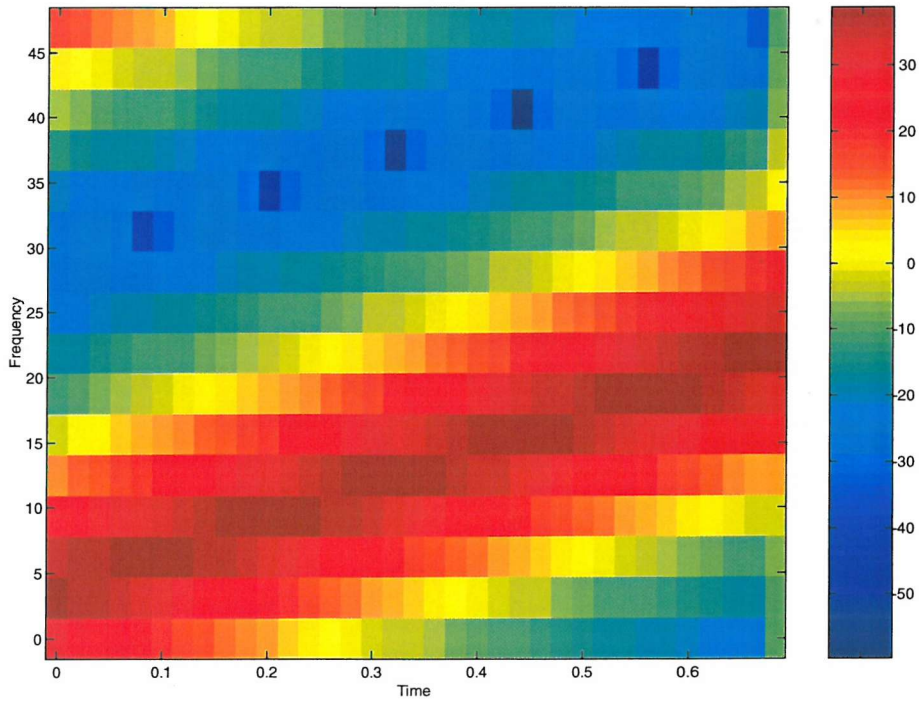

Figure 43 Spectrogram of Linear FM Harmonic Signal

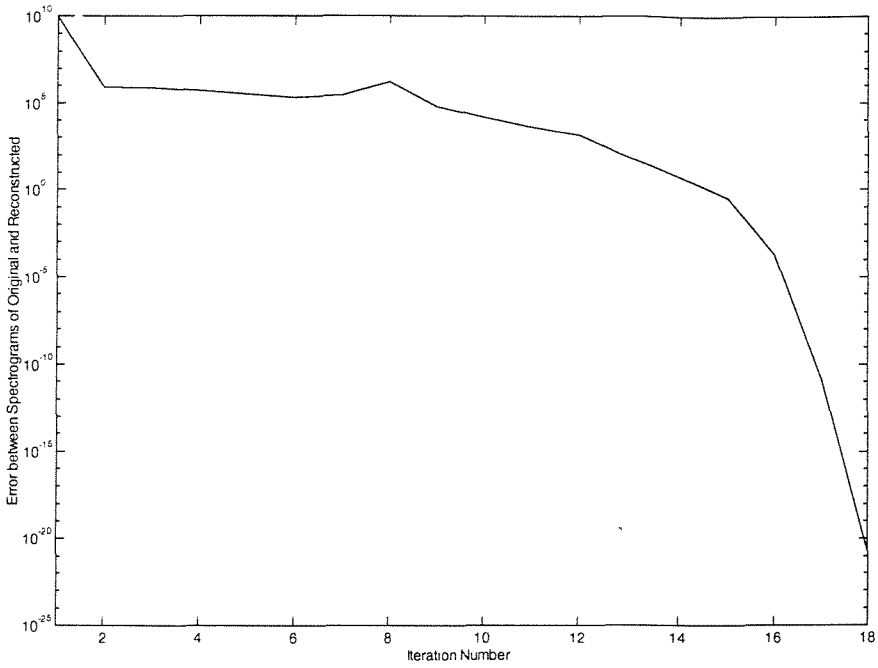

Figure 44 Error between Spectrogram of Original and Current Estimate of the Signal – Linear Chirp

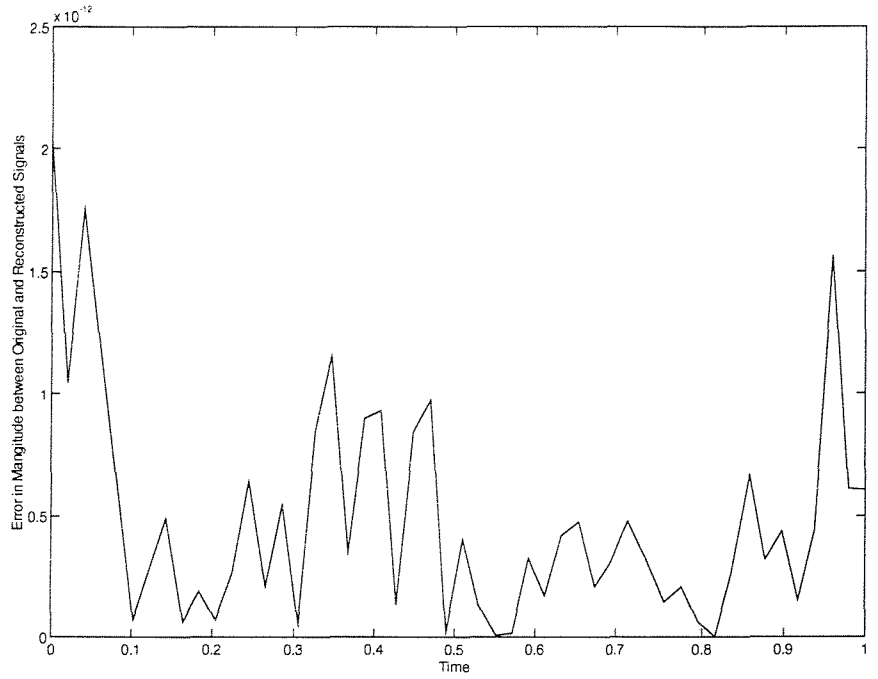

Figure 45 Error Between Amplitude of Original and Reconstructed Signal - Linear Chirp

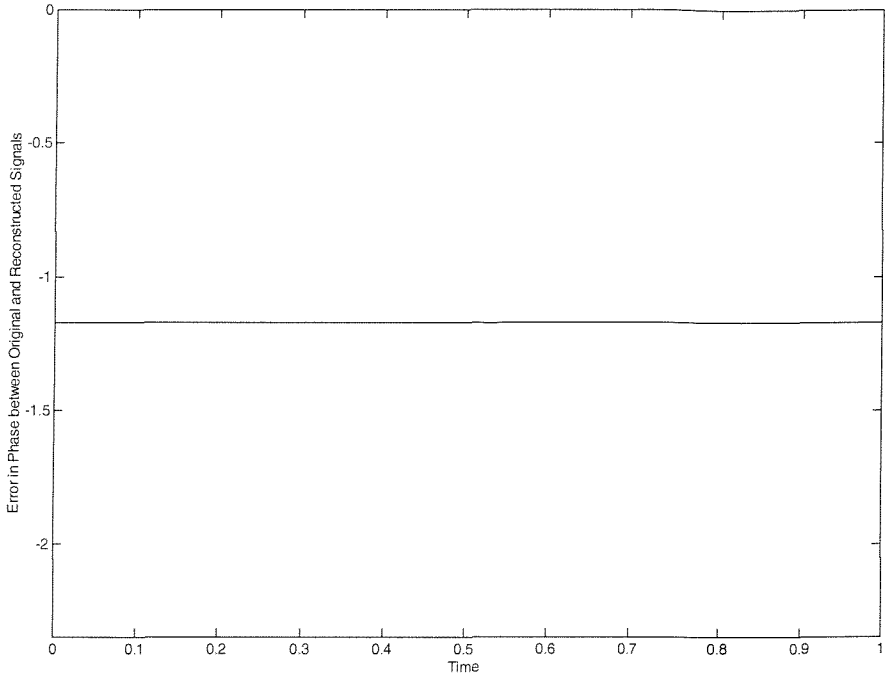

Figure 46 Error Between Phase of Original and Reconstructed Signal - Linear Chirp

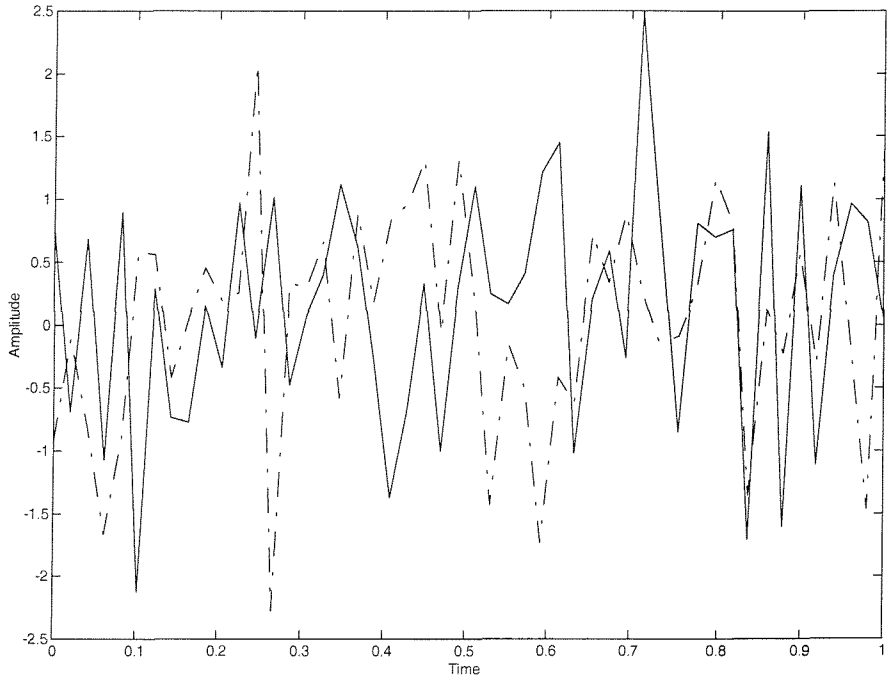

Figure 47 Pseudo-Random Gaussian White Noise Signal (Real Part Solid, Imaginary Part Dashed Lines)

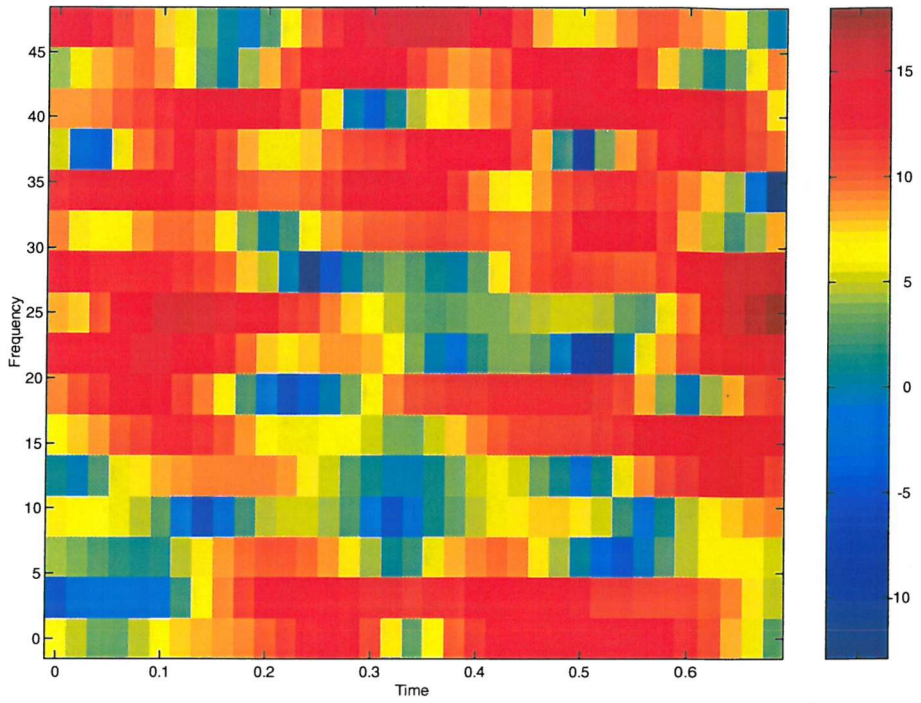

Figure 48 Spectrogram of Gaussian White Noise Signal

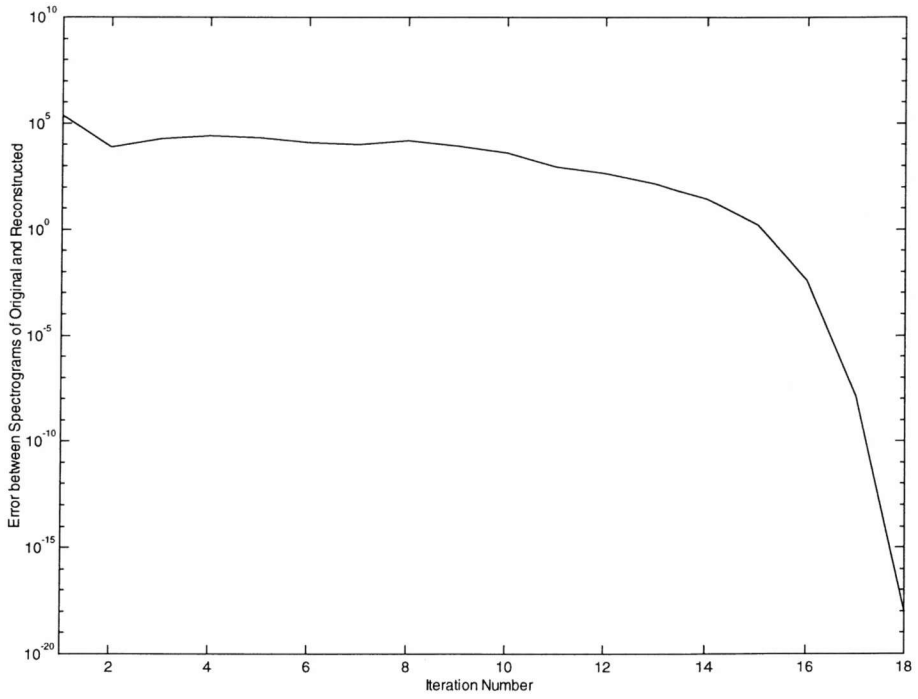

Figure 49 Error between Spectrogram of Original and Current Estimate of the Signal – Gaussian White Noise

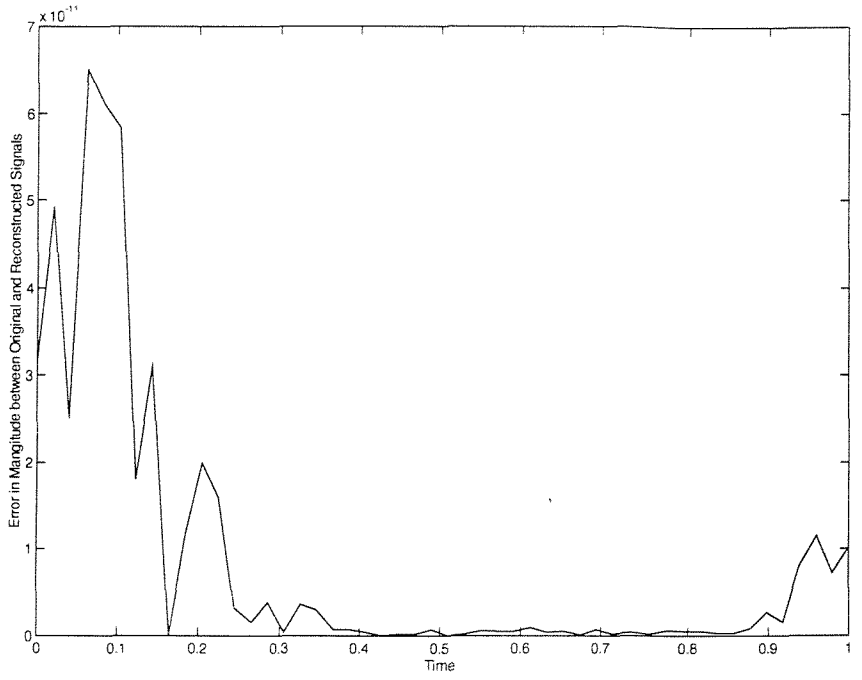

Figure 50 Error Between Amplitude of Original and Reconstructed Signal – Gaussian White Noise

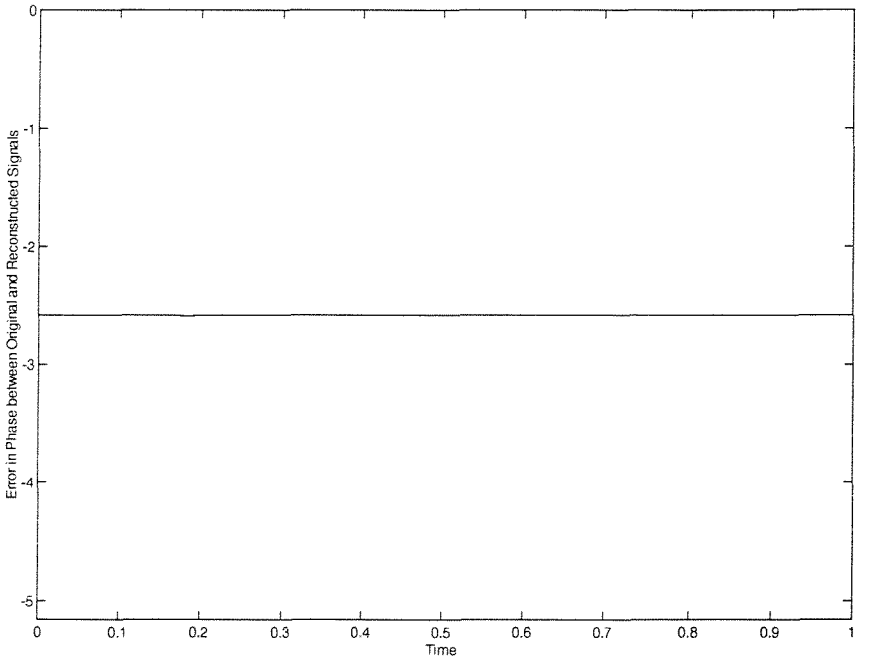

Figure 51 Error between Phase of Original and Reconstructed Signal – Gaussian White Noise

5.4.5 Comparison Between Different Techniques

In Sections 5.4.2 and 5.4.4, two different approaches to signal reconstruction from the modulus have been presented. Both techniques solve the problem in an iterative manner. The first reconstructed the signal by repeatedly applying the MLS algorithm developed for signal reconstruction from ST-FT. The second algorithm was based on a formulation of the squared magnitude of the ST-FT and solved the problem using either a fixed point or Newton Raphson algorithm.

Although fundamentally designed with signal reconstruction from the spectrogram in mind, the method based on the quadratic form described in Section 5.4.4 suffers a number of limitations. The first of which is the computational load required per iteration. For even small signal sizes (100pt), the computational resources required both in terms of memory and processing power are too restrictive on current computers. For a fixed point algorithm the pseudo-inverse matrix, $\left(\hat{\mathbf{x}}_p\tilde{\mathbf{A}}\right)^{\#}$ needs to be computed. For a Newton's based method, the Jacobian, $\mathbf{J}$, its pseudo-inverse and the equation $\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x}-\mathbf{S}$ need to be computed. The calculation of either of these matrices makes this method much more computationally intensive compared to the MLS based algorithm. However, computer trials have shown that, using a fixed-point iteration, the quadratic form based method typically required fewer iterations than the MLS based technique.

The second, more fundamental, difference between the two procedures is the guarantee of convergence of the MLS based algorithm regardless of whether the desired spectrogram is valid or not. A modified version of Newton's algorithm can be guaranteed to reduce the error at each iteration, offering similar performance to the MLS algorithm. However, the computational load required for this method is much greater than both the fixed-point iteration or the MLS based algorithm. Trials of the fixed-point algorithm for the quadratic form based technique have shown that the algorithm always convergences to the correct solution (given sufficient overlap) for valid ST-FT. Furthermore it typically takes fewer iterations for this algorithm to converge than Newton's method. However, the fixed-point algorithm fails to converge for invalid spectrograms. This means that use of this method to reconstruct a signal from a modified spectrogram is not advisable.

Regardless of which method of solving this quadratic form based system of equations, the computational requirements are always greater than if using the MLS based method. As such there are no reasons for opting for the quadratic form based method over the MLS based approach.

5.5 Conclusions

This chapter has followed the form of Chapter 4, presenting signal reconstruction methods for the magnitude, rather than the phase of the ST-FT. In total four different approaches have been presented, two piece-wise and two global solutions. The piece-wise solutions suffer the limitation of requiring that the signal be causal, in order to recover the signal from the first time-slice of the magnitude of the ST-FT.

Two distinct methods of global signal reconstruction from the magnitude were also presented. One method described previously in the literature is based upon a MLS solution to signal reconstruction from the ST-FT. An iterative algorithm using this approach can be used in the reconstruction of the signal from the modulus of the ST-FT. Although computational lean in terms of resources used, this method can converge to non-optimal solutions. This method has the advantage that a solution is always guaranteed regardless of whether the input modulus is a valid ST-FT or not.

An alterative approach was formulated based upon a quadratic form equation. In order to recover the signal from this formulation, it was necessary to find the roots of multi-dimensional non-linear equation. This problem was solved in two ways, either via a fixed-point algorithm or via application of Newton's algorithm. The fixed-point algorithm converged quickly in trials, however offered no guarantee of convergence. Newton's algorithm is guaranteed to converge only when the initial guess of the solution is sufficiently close to the root. A modified version of Newton's algorithm can be constructed which is guaranteed to converge, regardless of the initial guess, however this technique requires more computation per iteration. Both the fixed point and Newton's algorithm require the construction of large matrices for even modest signal sizes, and both require the computation of pseudo-inverses, further adding to the computational load.

Both techniques to solve the quadratic form based definition of the problem require significantly more computer resources, both in terms of time and space, than the MLS based method. The results of which is that in all bar the most academic of

examples that the MLS based method should be used. Two examples showing both of these global methods were presented, a linear FM chirp and a pseudo random Gaussian noise sequence.

The symmetry between signal reconstruction from phase and modulus can now be seen. Fundamental limitations in terms of the ability to reconstruct signals given the amount of overlap, or the signals being separated by a number of zeros greater than the window length are common between both modulus and phase cases. In the phase case, signal reconstruction could only take place to within an overall magnitude constant. The dual of this is that is the modulus case in that signal reconstruction can only take place to within an overall phase constant.

Chapter 6 Signal Reconstruction from the Generalised Wavelet Transform Information

6.1 Introduction

In Chapters 4 and 5 the development of the theory surrounding signal reconstruction from full or partial (magnitude or phase) ST-FT information was presented. The ST-FT is one of the two major linear TFRs used currently, the other being the Wavelet Transform (WT).

The WT was introduced in Chapter 2 (Section 2.3.3) as an alternative to the ST-FT. The principle difference between the ST-FT and the WT is that the WT decomposes the signal using a 'Constant Q' approach, changing the frequency resolution with the centre frequency of each analysis filter. In contrast, the ST-FT uses analysis filters whose resolution remains fixed. The use of such variable bandwidth filters makes the WT more suited to analysis of speech and music sounds, since at high frequencies the ear can be modelled as a set of 'Constant Q' filters [Irino92, 93, Kronland88].

A number of simple inverses exist to recover the signal from its WT, all of which have been developed from the definition of the WT. Implicit to these definitions is the concept that the WT, from which the signal is to be recovered, is valid. This assumption, or rather the lack of provision for signal reconstruction from invalid WTs results in sub-optimal reconstruction of signals when synthetic or invalid WTs are used. This chapter begins by reviewing existing methods of signal reconstruction from the WT.

Although the ST-FT and WT provide different signal decompositions, the basic forms of their definitions are very similar. This similarity is explored in the construction of a generalised form of the standard WT, termed the Generalised Wavelet Transform

(GWT), of which the ST-FT and WT are special cases. Since this transform encompasses a range of linear TFRs, a MLS inversion formula for this set is desired. The construction of such a formula means that for any TFR which can be placed within the GWT framework, a simple MLS inversion formula can be written. MLS inversion for GWT is described in this chapter.

Signal reconstruction from partial (magnitude or phase) GWT information is then considered. In an analogous manner to that used in the derivation of signal reconstruction from the phase of the ST-FT, signal reconstruction from the phase of the GWT is presented. The chapter then completes the theory by reviewing two previously presented methods of signal reconstruction from the magnitude of the WT, and presenting two new methods for signal reconstruction from the magnitude of the GWT. These new methods are extensions of techniques developed for the ST-FT, one based upon a quadratic form type of equation, and the other based upon the MLS approach of signal reconstruction from the GWT.

6.2 Signal Recnstruction from the Complete Wavelet Transform

6.2.1 Direct Inversion Techniques

In the continuous domain, the inversion to the WT can be defined as in two alternative forms, as given in (6-1) and (6-2) [Rioul91, Kronland88], where $h(t)$ is the decomposing wavelet, and c_1 and c_2 are constants depending upon the wavelet used for decomposition.

$$x(t) = c_1 \iint_{a>0} \frac{WT(\tau, a)}{a^2} h\left(\frac{\tau - t}{a}\right) d\tau da \tag{6-1}$$

$$x(t) = c_2 \int WT(t, a) \frac{da}{a^{3/2}} \tag{6-2}$$

Discretised forms of these equations can be used to reconstruct a signal from a discretely defined WT. One form of this discretisation is given in (6-3), where c_3 is a constant [Irino92, 93].

$$x[n] = c_3 \sum_{m=-\infty}^{\infty} \sum_{a=0}^{\infty} \frac{1}{\sqrt{a}} WT[m,a] h\left[\frac{n-m}{a}\right] \tag{6-3}$$

Although (6-3) can be implemented rapidly owing to its simplicity, these equations assume that the WT is valid. With a view to developing a MLS inversion of the WT, a generalised version of the wavelet transform (GWT) is now presented.

6.2.2 Generalised Form of the Wavelet Transform

A generalisation of the discrete WT presented in Chapter 2 (2-42) is given in (6-4) and (6-5). This generalisation allows user selection of the number and scale of each decomposing wavelet. A user-defined function $q(c)$ defines the scale decomposition used, given a discrete scale number, c (6-4), where as previously l_x is the length of the input signal, $x[n]$. In essence, the parameter c describes the amount by which the wavelet is contradicted. The function $g(m)$, defines the amount of overlap used, where m is non-negative variable.

$$a = q(c) \tag{6-4}$$

$$X[m,p] = \frac{1}{\sqrt{q(c)}} \sum_{l=0}^{l_x-1} x[l] h\left[\frac{l-g(m)}{q(c)}\right] \tag{6-5}$$

A further step can be taken to generalise this expression in order to allow different, non-scale decompositions of a signal. In the WT case, the scale variable defines the amount of contraction and expansion undergone by the wavelet. In the formulation given in (6-6), the user has complete control over the decomposition used. The scale variable has been replaced by an arbitrary 'frequency' variable, u.

Both traditional wavelet decompositions (such as the Morlet wavelet) and the ST-FT (illustrated for a Gaussian window) can be written in this form. The appropriate choices of windowing function are given in Table 5 (the sampling rate is assumed to be unity to simplify notation).

$$X[m,u] = \sum_{l=0}^{l_x-1} x[l] h[l-g(m),u] \tag{6-6}$$

This extension is similar to that presented by Rabiner [Rabiner78] and Anderson [Anderson93]. It is evident that all GWTs constructed in this manner are linear TFRs regardless of the decomposing wavelet used[2].

Time-Frequency Decomposition	Windowing Function $h[n,u]$
Morlet Wavelet	$\frac{1}{\sqrt{u}} e^{\frac{1}{2}-\left(\frac{n}{u}\right)^2} e^{\frac{j\omega_0 n}{u}}$
Continuous ST-FT (Gaussian Window)	$\frac{1}{2\pi} e^{\left(\frac{-n^2}{2}\right)} e^{-jnu}$
Discrete ST-FT (Gaussian Window)	$\frac{1}{N} e^{\left(-\frac{n^2}{2}\right)} e^{\frac{-2jn\pi u}{N}}$

Table 5 Windowing Function Definitions for Generalised Wavelet Transform

6.2.3 Minimum Least Squares Inversion

This section details an MLS inversion scheme for the GWT. This method will extend the ST-FT MLS inversion to other GWTs, most importantly it extends to the conventional WT.

Consider a 'desired' GWT given by $Y[m,u]$. In this case the recovered signal should minimise the mean squared error between $Y[m,u]$, and the GWT computed from the reconstructed signal, $X[m,u]$. To obtain an optimum solution, the distance measure given in (6-7) is minimised, where $P(u)$ is the number of time-slices present in the GWT for a given scale u, and U is the total number of 'frequency' points used.

$$D[x[n],Y[m,u]] = \sum_{u=0}^{U-1} \sum_{m=0}^{P(u)-1} |X[m,u] - Y[m,u]|^2 \qquad (6\text{-}7)$$

[2] It is assumed that the windowing function is signal independent.

The input signal $x[n]$ is generally complex, and can be separated into its real and imaginary parts. In order to find the minimum of this function, the derivative with respect to both the real and imaginary parts is set to zero, as in (5-13). For the real component one expands the modulus-squared term in (6-7), to obtain (6-8).

From the definition of the GWT (6-6), the derivatives in (6-8) can be evaluated, and are given in (6-9) and (6-10). Substituting (6-9) and (6-10) into (6-8) and simplifying the result leads to (6-10). Via a similar process, (6-11) can be obtained.

$$\frac{\partial D\left[x[n],Y[m,u]\right]}{\partial x_r[n]} = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\left(X[m,u]-Y[m,u]\right)\frac{\partial X^*[m,u]}{\partial x_r[n]} + \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\left(X[m,u]-Y[m,u]\right)^*\frac{\partial X[m,u]}{\partial x_r[n]} \tag{6-8}$$

$$\frac{\partial X[m,u]}{\partial x_r[n]} = h\left[n-g(m),u\right] \qquad \frac{\partial X^*[m,u]}{\partial x_r[n]} = h^*\left[n-g(m),u\right] \tag{6-9}$$

$$\frac{\partial D\left[x[n],Y[m,u]\right]}{\partial x_r[n]} = 2\sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\mathrm{Re}\left[\left(X[m,u]-Y[m,u]\right)h\left[n-g(m),u\right]\right] \tag{6-10}$$

$$\frac{\partial D\left[x[n],Y[m,u]\right]}{\partial x_i[n]} = 2\sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\mathrm{Im}\left[\left(X[m,u]-Y[m,u]\right)h\left[n-g(m),u\right]\right] \tag{6-11}$$

Since for an optimum the derivatives defined in (6-10) and (6-11) will be zero, then it follows that (6-12) must be satisfied.

$$\frac{\partial D\left[x[n],Y[m,u]\right]}{\partial x[n]} = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\left(X[m,u]-Y[m,u]\right)h\left[n-g(m),u\right] = 0 \tag{6-12}$$

Substituting into (6-12) the definition of a GWT, given by (6-6), an expression for the required signal in terms of the desired GWT can be written, see (6-13). Re-arranging (6-13) leads to the linear formulation expressed in (6-14), where $K[l,n]$ is a kernel function dependant only upon the wavelet used in the decomposition of the signal.

$$\sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\sum_{l=0}^{L}x[l]h\left[l-g(m),u\right]h\left[n-g(m),u\right] = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}Y[m,u]h\left[n-g(m),u\right] \tag{6-13}$$

$$\sum_{l=0}^{l_x-1} x[l]K[l,n] = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1} Y[m,u]h[n-g(m),u]$$
$$K[l,n] = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1} h[l-g(m),u]h[n-g(m),u] \qquad (6\text{-}14)$$

Equation (6-14) describes the MLS inversion of the GWT. Such a formulation was described in an analogous, if less formal manner, by Irino [Irino92, 93]. These papers stop their development short of the form given in (6-14). They stated that there was no reason to proceed, since the system of equations was not easily solved due to the size of the matrices used. Such computational concerns are less severe now, and therefore the MLS of a GWT now developed.

The equation (6-14) is linear in terms of the reconstructed signal, and so can be written in a matrix form (6-15).

$$\mathbf{Kx} = \mathbf{Y} \qquad (6\text{-}15)$$

where

$$\mathbf{Y}_{n,1} = \left(\sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1} Y[m,u]h[n-g(m),u] \right) \qquad (6\text{-}16)$$
$$\mathbf{K}_{l,n} = \sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1} h[l-g(m),u]h[n-g(m),u]$$
$$\mathbf{x} = [x[0] \quad x[1] \quad \cdots \quad x[l_x-1]]$$

$$\beta = \sum_{u=0}^{U-1} P(u) \qquad (6\text{-}17)$$

$$\mathbf{x} = \mathbf{K}^{\#}\mathbf{Y} \qquad (6\text{-}18)$$

The matrix $\mathbf{K}$ has dimensions of $\max(P(u)), l_x$ and each element is the result of a double summation, requiring $\beta.l_x.\max(P(u))$ multiplications.

Once the kernel function has been computed, its pseudo-inverse (denoted $\mathbf{K}^{\#}$) must be found in order to compute the signal (6-18). Direct application of a standard pseudo-inverse routine can be used to achieve this. However the dominant computational load resides in the calculation of $\mathbf{K}$. Improvements in computational time are possible by using Fourier based methods [Rabiner78, Oppenheim75]. Each

row of the kernel matrix can be expressed as a convolution. Through the use of DFT based convolution methods, a complete set of points can be computed via one FFT, rather than requiring the computation of each point individually.

Although the calculation of the kernel and its inverse is computationally intensive, the kernel is data independent. This means that, for a given set of wavelets and a particular signal length, the kernel function and its inverse need only be computed once. This makes the technique useful for applications where the data length, and the set of decomposing wavelets are fixed.

The application of (6-18) recovers more points than the signal length, owing to the manner in which the convolution operates. If the signal is reconstructed though direct application of (6-18), a zero-padded version of the signal is recovered, *i.e.* the signal is assumed to have finite support. By using FFT's to compute the correlation there is an implicit assumption that the signal is zero outside the defined points. This assumption can be seen in Figure 52, with the time-axis extending before the start of the signal and after the end. In the case of Figure 52, the linear chirp can be seen as a curve in time-scale, between $t=0$ and $t=1$. By limiting the column space of the kernel so that only the non-zero points are recovered, the computational burden is reduced.

6.2.3.1 Examples of MLS Signal Reconstruction from the WT

Two examples are now given to illustrate MLS signal reconstruction from the WT. As previously, the two signals used are a linear FM chirp sequence, and a Gaussian pseudo random sequence. Both signals are 100 points long, and are assumed to be sampled at 100 Hz.

The decomposing wavelets selected were 20 Morlet wavelets (see Table 5 for definition), ranging in size from 51 to 13 points long, covering a scale range from 0.04 to 0.01.

6.2.3.1.1 Example 1 Linear FM Chirp Sequence

The time series of a Linear FM chirp sequence has been given previously in Chapter 4 (Figure 21), and the amplitude of the GWT is shown in Figure 52.

Through the use of the amplitude and phase of WT and application of (6-18), the signal was reconstructed. Since both the amplitude and phase information are used,

the signal can be reconstructed precisely, with no ambiguity. Figure 53 and Figure 54 depict the errors between the original and reconstructed signal for both the modulus and phase. These errors are small, indicating the success of the algorithm.

6.2.3.1.2 Example 2 Pseudo Random Gaussian White Noise Sequence

The time series of a pseudo random Gaussian white noise sequence has been given previously (Chapter 4, Figure 25), and magnitude of WT using a Morlet wavelet is shown in Figure 55. As before, unlike the linear chirp signal, no structure can be seen in time-scale. The signal was reconstructed using both the modulus and phase of the GWT, and the modulus and phase errors are depicted in Figure 56 and Figure 57 respectively. As with the previous examples, the reconstruction exhibits a high level of fidelity.

It should be noted that this signal is full band, *i.e.* it has significant energy at all frequencies between 0-50 Hz. However the decomposition provided only covers the region between 0.01-0.04 (appropriately 25-100 Hz, aliased). Since the filters are not perfect, there is significant energy both due to the low scale (high frequency) filters aliasing back and the filters having imperfect stop-bands.

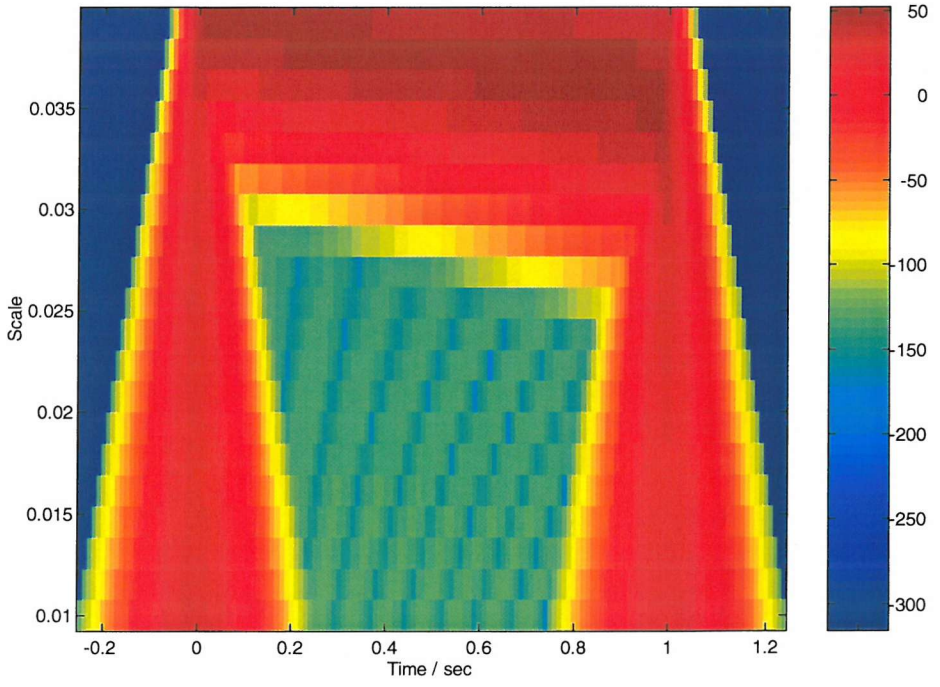

Figure 52 Magnitude of GWT of Linear FM Harmonic Signal (Morlet Window)

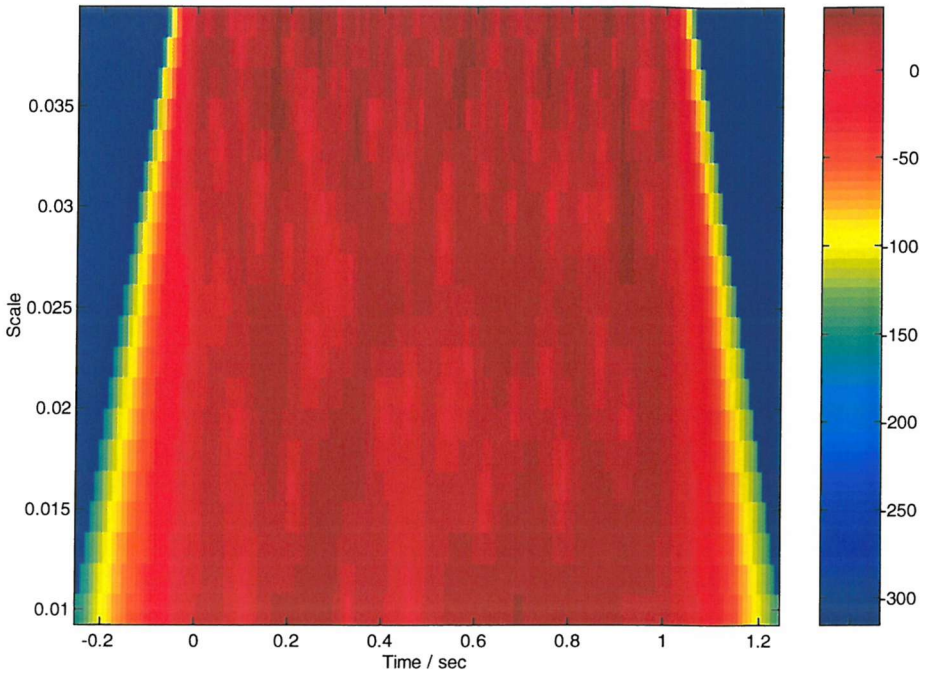

Figure 55 Magnitude of GWT of Gaussian White Noise Signal (Morlet Window)

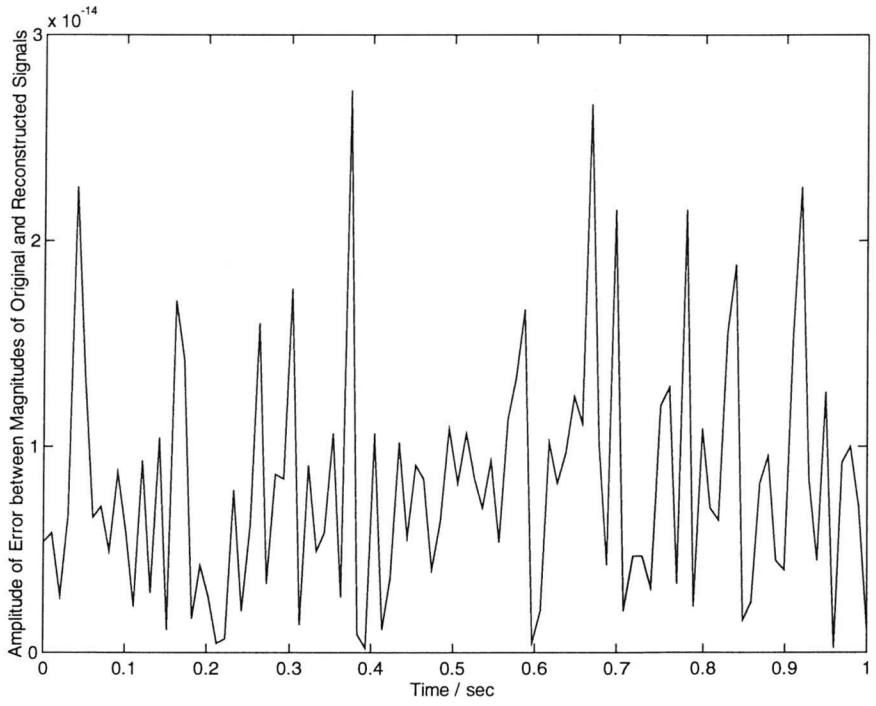

Figure 56 Error Between Amplitude of Original and Reconstructed Signal – Gaussian White Noise

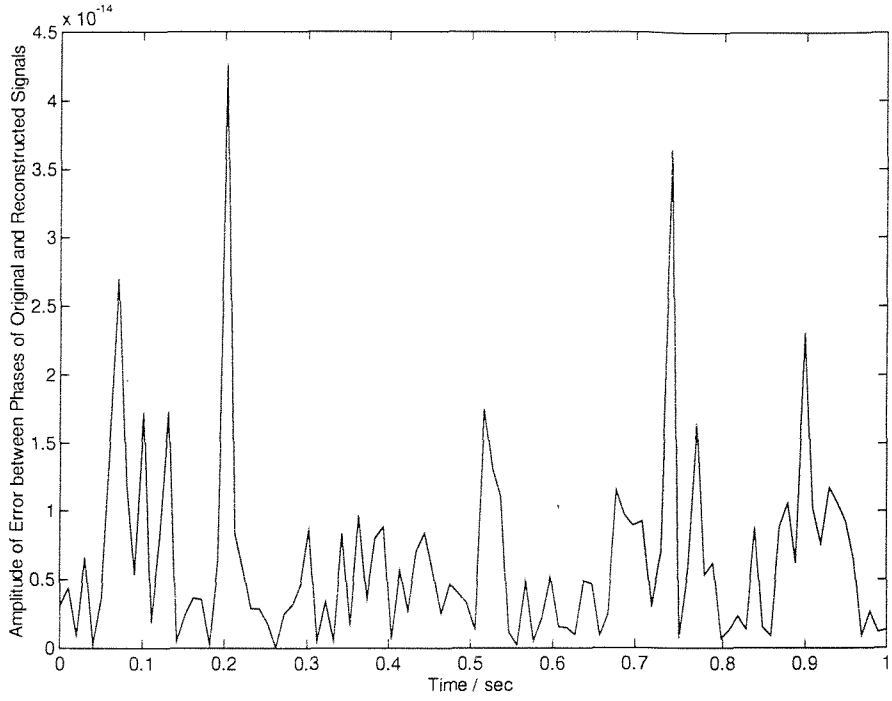

Figure 57 Error Between Phase of Original and Reconstructed Signal – Gaussian White Noise

6.3 Signal Reconstruction from the Phase of the Generalised Wavelet Transform

In Chapter 4, the problem of signal reconstruction from the phase of the ST-FT was discussed. The solution presented used a linear matrix formulation, derived from the definition of the phase of the ST-FT. The ST-FT can be considered to be an example of the GWT. The techniques developed for the ST-FT are now extended to the GWT.

Consider the GWT (6-6), expressed in modulus and phase form in equation (6-19). The phase of the transform can be expressed as given by (6-20).

$$X[m,u] = \left|X[m,u]\right| e^{j\phi[m,u]} \tag{6-19}$$

$$\tan\left(\phi[m,u]\right)=\frac{\mathrm{Im}\left[\sum_{l=0}^{l_x-1}x[l]h\left[l-g(m),u\right]\right]}{\mathrm{Re}\left[\sum_{l=0}^{l_x-1}x[l]h\left[l-g(m),u\right]\right]} \tag{6-20}$$

Unlike the ST-FT, the windowing function used in the GWT is generally complex. Defining the real and imaginary parts of the windowing function via (6-21), the summations in (6-20) can be expanded to give (6-22), which can be written as (6-23).

$$h[n,u]=h_r[n,u]+jh_i[n,u] \tag{6-21}$$

$$\tan\left(\phi[m,u]\right)=\frac{\sum_{l=0}^{l_x-1}x_r[l]h_i\left[l-g(m),u\right]+\sum_{l=0}^{l_x-1}x_i[l]h_r\left[l-g(m),u\right]}{\sum_{l=0}^{l_x-1}x_r[l]h_r\left[l-g(m),u\right]-\sum_{l=0}^{l_x-1}x_i[l]h_i\left[l-g(m),u\right]} \tag{6-22}$$

$$\begin{aligned}&\sin\left(\phi[m,u]\right)\left(\sum_{l=0}^{l_x-1}x_r[l]h_r\left[l-g(m),u\right]-\sum_{l=0}^{l_x-1}x_i[l]h_i\left[l-g(m),u\right]\right)\\&-\cos\left(\phi[m,u]\right)\left(\sum_{l=0}^{l_x-1}x_r[l]h_i\left[l-g(m),u\right]+\sum_{l=0}^{l_x-1}x_i[l]h_r\left[l-g(m),u\right]\right)=0\end{aligned} \tag{6-23}$$

Equation (6-23) can be rearranged in the form shown in (6-24).

$$\begin{aligned}&\sum_{l=0}^{l_x-1}x_r[l]\left(h_r\left[l-g(m),u\right]\sin\left(\phi[m,u]\right)-h_i\left[l-g(m),u\right]\cos\left(\phi[m,u]\right)\right)\\&-\sum_{l=0}^{l_x-1}x_i[l]\left(h_r\left[l-g(m),u\right]\cos\left(\phi[m,u]\right)+h_i\left[l-g(m),u\right]\sin\left(\phi[m,u]\right)\right)=0\end{aligned} \tag{6-24}$$

The modulating wavelet can be written in terms of its modulus and phase form, to finally yield (6-27).

$$h[n,u]=\left|h[n,u]\right|e^{j\eta[n,u]} \tag{6-25}$$

$$\begin{aligned}h_r[n,u]&=\left|h[n,u]\right|\cos\left(\eta[n,u]\right)\\h_i[n,u]&=\left|h[n,u]\right|\sin\left(\eta[n,u]\right)\end{aligned} \tag{6-26}$$

$$\sum_{l=0}^{l_x-1} x_r[l]\left(\left|h\left[l-g(m),u\right]\right|\sin\left(\phi[m,u]-\eta\left[l-g(m),u\right]\right)\right) \qquad (6\text{-}27)$$
$$-\sum_{l=0}^{l_x-1} x_i[l]\left(\left|h\left[l-g(m),u\right]\right|\cos\left(\phi[m,u]-\eta\left[l-g(m),u\right]\right)\right)=0$$

The similarity between the expression for the ST-FT (4-33), and the more general result is (6-27), is evident. Equation (6-27) reduces to (4-33) when the appropriate windowing function (as defined by Table 5) is selected.

In a similar manner to that already described for the ST-FT, equation (6-27) can be written as in a matrix form, given in (6-28).

$$\mathbf{Ax} = \mathbf{0} \qquad (6\text{-}28)$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{Z} & \cdots & \mathbf{Z} & \mathbf{C}_1 & \mathbf{Z} & \cdots & \mathbf{Z} \\ \mathbf{Z} & \mathbf{S}_2 & \cdots & \vdots & \mathbf{Z} & \mathbf{C}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z} & \mathbf{Z} & \cdots & \mathbf{S}_P & \mathbf{Z} & \mathbf{Z} & \cdots & \mathbf{C}_P \end{bmatrix} \qquad (6\text{-}29)$$

$$\mathbf{S}_S = \begin{bmatrix} |h[0,0]|\sin(\phi[s,0]-\eta(0,0)) & |h[1,0]|\sin(\phi[s,0]-\eta(1,0)) & \cdots & |h[N-1,0]|\sin(\phi[s,0]-\eta(N-1,0)) \\ |h[0,1]|\sin(\phi[s,1]-\eta(0,1)) & |h[1,1]|\sin(\phi[s,1]-\eta(1,1)) & \cdots & |h[N-1,1]|\sin(\phi[s,1]-\eta(N-1,1)) \\ \vdots & \vdots & \ddots & \vdots \\ |h[0,U]|\sin(\phi[s,U]-\eta(0,U)) & |h[1,U]|\sin(\phi[s,U]-\eta(1,U)) & \cdots & |h[N-1,U]|\sin(\phi[s,U]-\eta(N-1,U)) \end{bmatrix}$$

$$\mathbf{C}_S = \begin{bmatrix} |h[0,0]|\cos(\phi[s,0]-\eta(0,0)) & |h[1,0]|\cos(\phi[s,0]-\eta(1,0)) & \cdots & |h[N-1,0]|\cos(\phi[s,0]-\eta(N-1,0)) \\ |h[0,1]|\cos(\phi[s,1]-\eta(0,1)) & |h[1,1]|\cos(\phi[s,1]-\eta(1,1)) & \cdots & |h[N-1,1]|\cos(\phi[s,1]-\eta(N-1,1)) \\ \vdots & \vdots & \ddots & \vdots \\ |h[0,U]|\cos(\phi[s,U]-\eta(0,U)) & |h[1,U]|\cos(\phi[s,U]-\eta(1,U)) & \cdots & |h[N-1,U]|\cos(\phi[s,U]-\eta(N-1,U)) \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} x_r[0] & x_r[1] & \cdots & x_r[l_x] & x_i[0] & x_i[1] & \cdots & x_i[l_x] \end{bmatrix}^T$$

In the above formulation, the matrix $\mathbf{Z}$ has U columns and $g(m)$ rows (the amount of overlap), where U is the number of decomposing wavelets, and P is the number of time-slices. In this expression it is assumed that all wavelets have the same length. This can be achieved by symmetrically zero padding the smaller wavelets.

The homogeneous system of equations (6-29) can be solved using the methods discussed in Section 4.4.2 for the ST-FT.

In the ST-FT case, whether the matrix $\mathbf{A}$ has sufficient rank is dependant upon the amount of overlap used between successive time-slices. Since typically different wavelets are of different sizes, no simple rule can be derived as to the amount of overlap required in order to ensure reconstruction. Furthermore, as in the ST-FT, it is assumed that the first signal point is non-zero.

6.3.1 Examples of Signal Reconstruction from the Phase of the GWT

As before, two example signals are used to demonstrate this technique, a linear chirp, and a pseudo random Gaussian sequence. The same decomposing wavelets used in the examples of MLS signal reconstruction from the GWT were also used in these examples. Both signals are 100 points long and assumed to be sampled at 100Hz.

6.3.1.1 Example 1 Linear FM Chirp Sequence

The linear FM sequence used previously in Chapters 4 and 5 is again used here. In the construction of the GWT, the signal is not assumed to be causal, or zero past the end of the signal. The magnitude of the GWT is given in Figure 58. Note that this representation does not contain the start and end transients that appear in Figure 52. This is due to the different methods used when computing the GWT near the signal's start and end.

The phase from which the signal is reconstructed is given in Figure 59. The difference in amplitude between the original (normalised with respect to the maximum sample), and the reconstructed signal, is given in Figure 60, and the error in phase is given in Figure 61. As before, the errors between the original and reconstructed signals are low. These errors can be seen to increase toward the start and end of the signal. This is because these locations contribute to fewer time-slices and so are subject to greater uncertainty.

6.3.1.2 Example 2 Pseudo Random Gaussian Sequence

The magnitude and phase of the GWT for a pseudo-random Gaussian sequence are shown in Figure 62 and Figure 63. The error between the normalised original and reconstructed signal is shown in Figure 64 and Figure 65. Once more the error between the two signals is low, increasing towards the start and end of the signal.

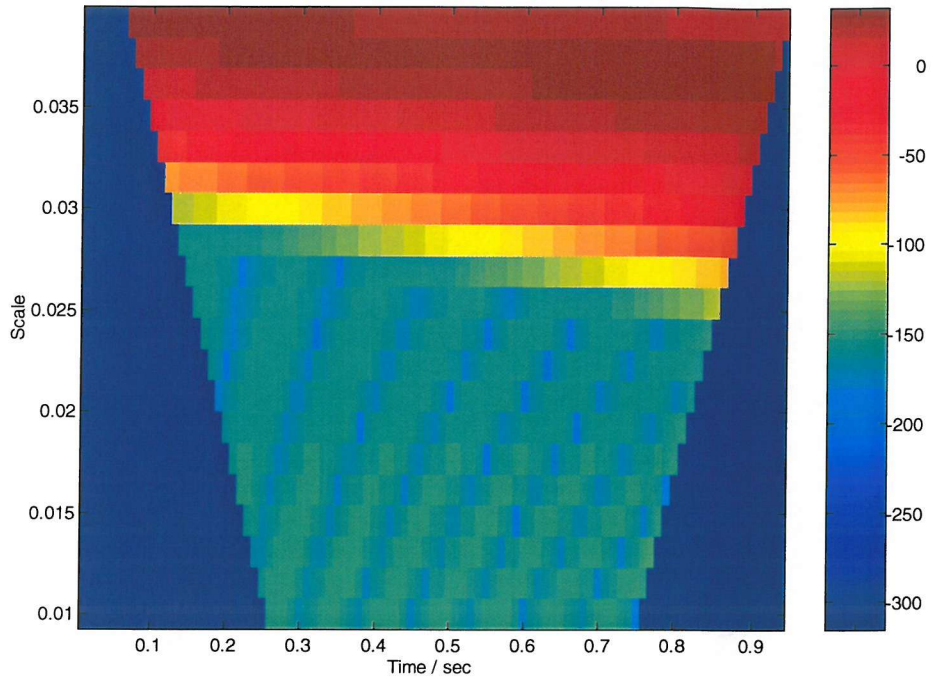

Figure 58 Magnitude of GWT of Linear FM Harmonic Signal (Morlet Window) – No Causal Assumption.

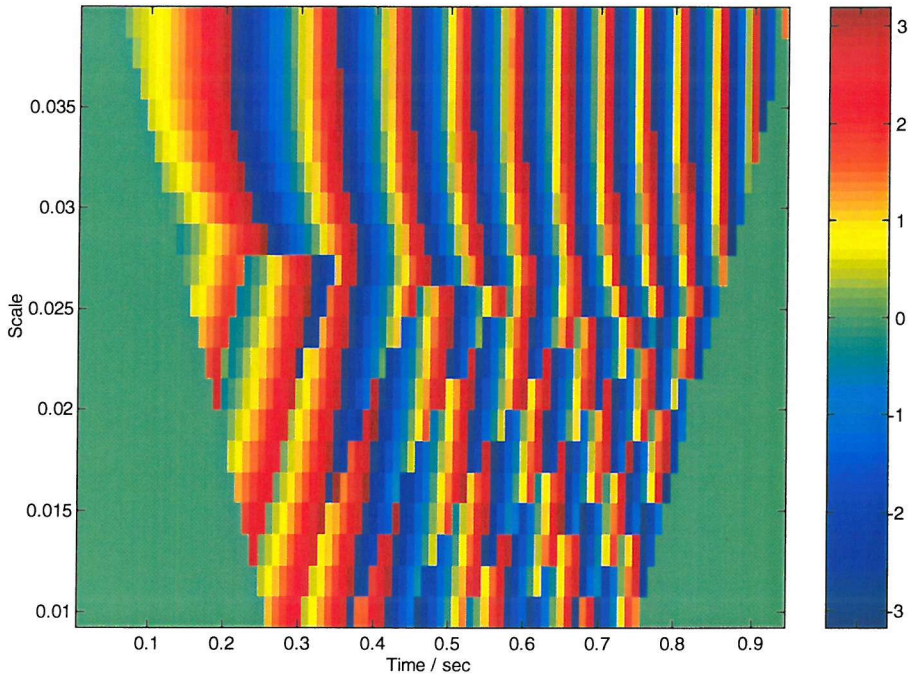

Figure 59 Phase of GWT of Linear FM Harmonic Signal (Morlet Window) – No Causal Assumption.

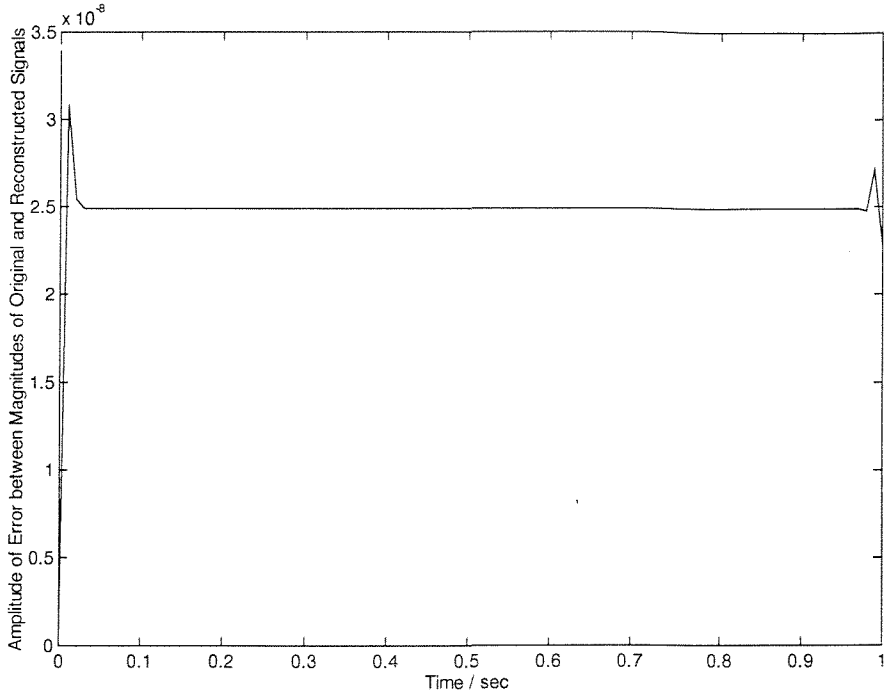

Figure 60 Error Between Amplitude of Original and Reconstructed Signal - Linear Chirp

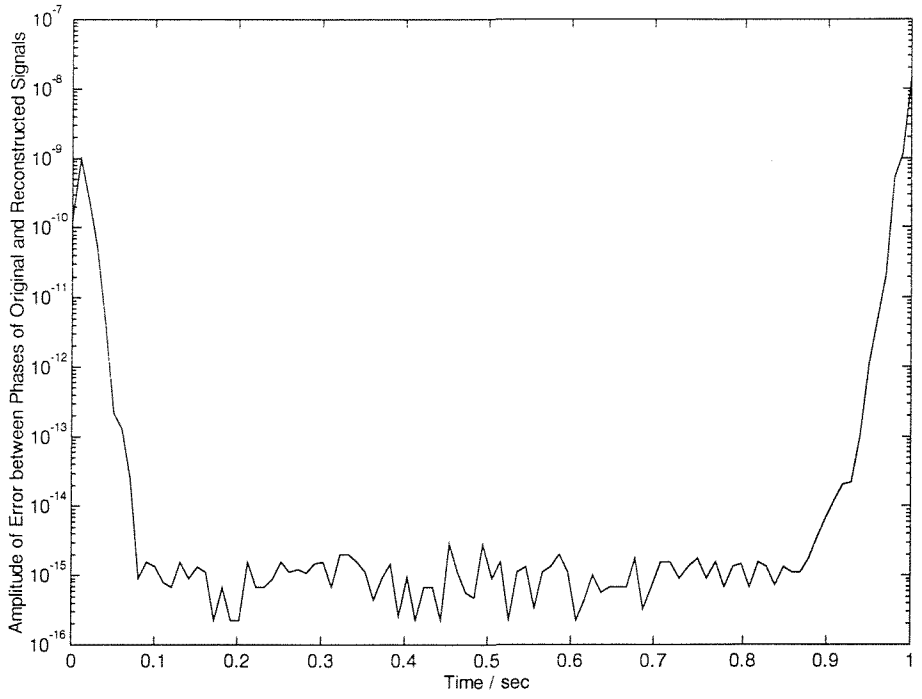

Figure 61 Error Between Phase of Original and Reconstructed Signal - Linear Chirp

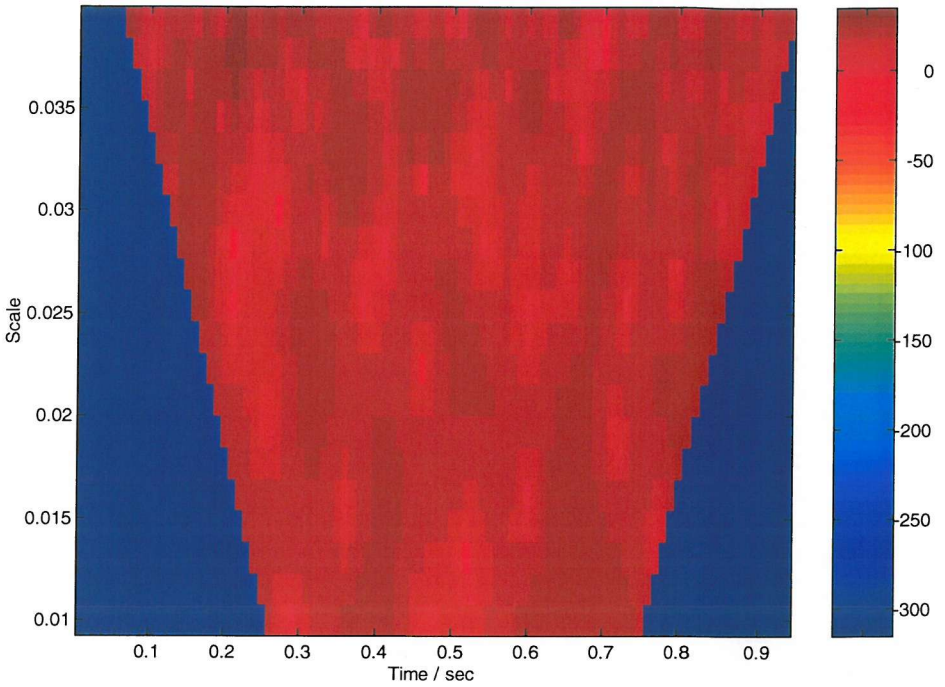

Figure 62 Magnitude of GWT of Gaussian White Noise Signal (Morlet Window) – Non - Causal

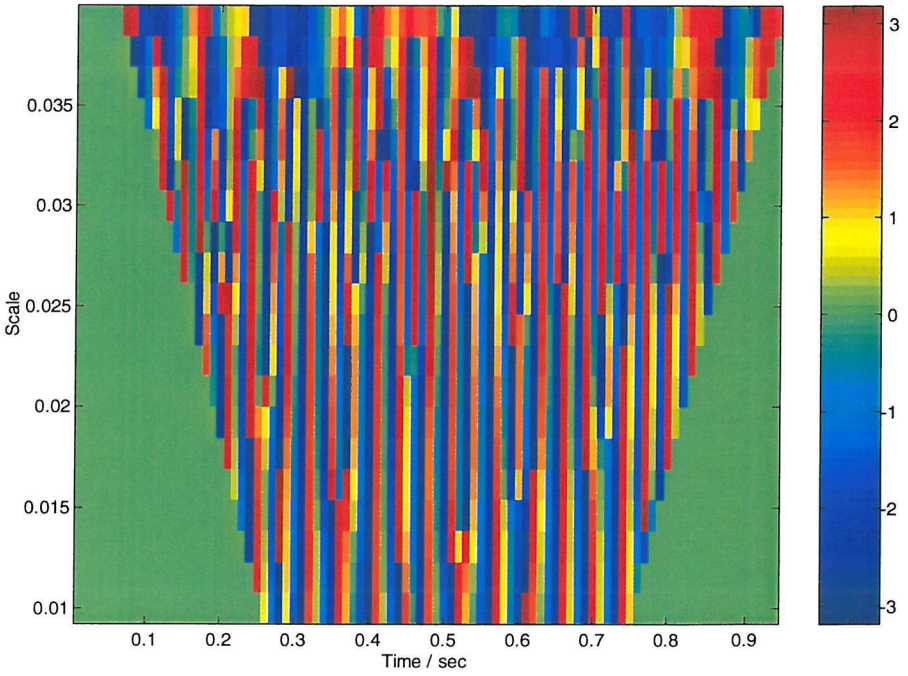

Figure 63 Phase of GWT of Gaussian White Noise (Morlet Window) – No Causal Assumption

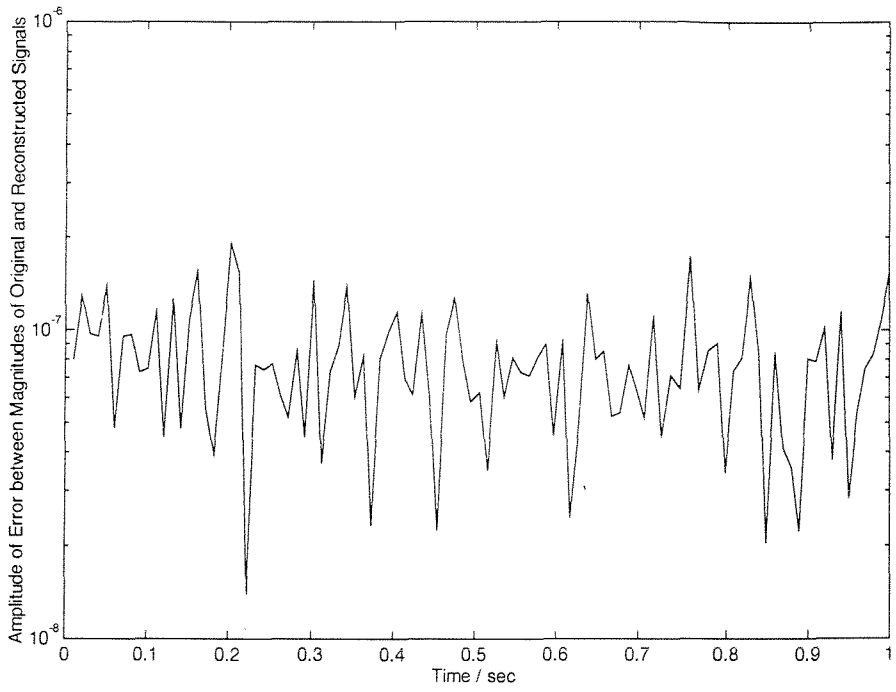

Figure 64 Error Between Amplitude of Original and Reconstructed Signal – Gaussian White Noise

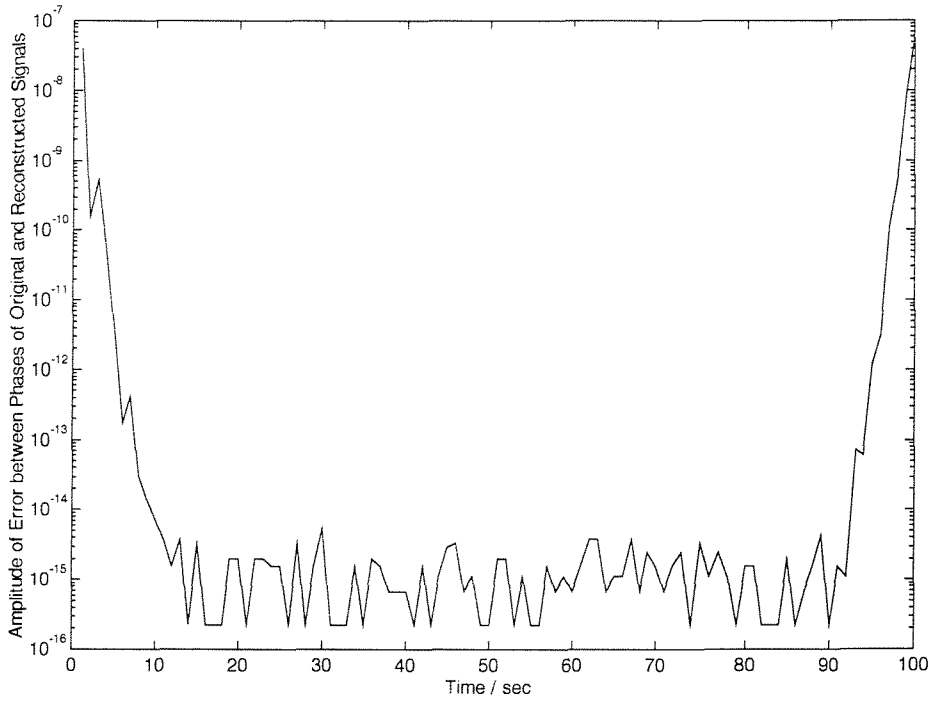

Figure 65 Error Between Phase of Original and Reconstructed Signal – Gaussian White Noise

6.4Signal Reconstruction from the Modulus of the WT Transform

6.4.1.1 Inverse WT based method

In the approach to signal reconstruction from the magnitude of the ST-FT, Griffin [Griffin83, 84] developed an algorithm based upon repeated application of the MLS signal reconstruction from ST-FT. Irino [Irino92, 93] followed the same methodology; however they found that the MLS inversion of the WT was too computationally intensive for the decomposing wavelets they wished to use. Therefore, rather than repeatedly applying a MLS inversion formula for the WT, Irino choose to use a faster, but non-MLS inversion formula. This inversion formula used was given previously in (6-3).

Based upon this inversion formulation, Irino [Irino92, 93] defined the 'Replace and Inverse Wavelet Transform – Modified Wavelet Transform Magnitude' or RIWT-MWMT algorithm as expressed in (6-30) and (6-31).

In this equation, $X^p[m,a]$ is the WT of the p^{th} signal estimate, $x^p[n]$, $|Y[m,a]|$ is the desired modulus, $h[n]$ is the decomposing wavelet.

$$x^p[n]=\sum_{a=0}^{U-1}\sum_{m=0}^{P(a)-1}\frac{1}{\sqrt{a}}\hat{X}^p[m,a]h\left[\frac{n-m}{a}\right] \tag{6-30}$$

$$\hat{X}^p[m,a]=|Y[m,a]|\frac{X^{p-1}[m,a]}{|X^{p-1}[m,a]|} \tag{6-31}$$

If the MLS inversion formula, defined for the GWT in Section 6.2.3, were used instead of the simple inversion formula given (6-3) then a matrix inversion would be required at each iteration. Thus this method requires less computational resource per iteration. This approach does however suffer from two limitations. Firstly, unlike the MLS based method for the GWT (of which the WT is one example) presented in Section 6.5.1.3, the distance measure between the WT of the current estimate of the signal and the desired WT is not monotonically decreasing. Secondly, for an invalid input WTs, this algorithm will not tend toward the MLS solution. The MLS GWT based approach does not suffer from this limitation.

6.5 Signal Reconstruction from Magnitude of the GWT Transform

Sections 6.2 and 6.3 describe signal reconstruction from the complete GWT, and only using its phase. The final extension is to describe signal reconstruction from solely magnitude information. This section begins by reviewing previous techniques for signal reconstruction from the magnitude of the WT and then presents the MLS based inversion for the GWT.

6.5.1.1 Piece-wise Method

As stated in the introduction to Chapter 5 Section 5.4 Anderson [Anderson93] generalised the ST-FT approach of sequential reconstruction from the magnitude of the ST-FT, to the GWT. If the GWT is maximally overlapped such that $g(m)=m$, (6-32) can be written.

$$X[m,u]=\sum_{l=-m}^{l_x-m-1} x[l+m]h[l,u] \tag{6-32}$$

Writing the decomposing wavelet in terms of modulus and phase as given in (6-33), where $h_m[n]$ is low pass filter (*i.e.* its amplitude varies very slowly). If $\phi[n,u]$ is of the form $\phi[n,u]=f(u)n$ then windowing provides a decomposition at the frequency $f(u)$. Making this assumption and substituting (6-33) into (6-32) results in (6-34). This equation is of finite length, defined between the points $0\le l\le \lambda(u)$, where $\lambda(u)$ is the length of the wavelet for the scale, u.

$$h[n,u]=h_m[n]e^{-j\phi[n,u]} \tag{6-33}$$

$$X_m[u]=\sum_{l=0}^{\lambda(u)-1} x[l+m]h_m[l]e^{-j(f(u)l)} \tag{6-34}$$

Deriving the magnitude of (6-34), and writing the signal in terms of real and imaginary parts as defined in (6-35), results in (6-36).

$$x[n] = x_r[n] + jx_i[n] \tag{6-35}$$

$$\left|X_m[u]\right| = h_m[0]^2\left(x_r[n]^2 + x_i[n]^2\right) + a_u[n]x_r[n] + b_u x_i[n] + c_u[n] \tag{6-36}$$

where

$$a_u[n] = 2h[0]\sum_{l=0}^{\lambda(u)-1} h_m[l]\left(x_r[l+m]\cos\left(f(u)l\right) - x_i[l+m]\sin\left(f(u)l\right)\right)$$

$$b_u[n] = 2h[0]\sum_{l=0}^{\lambda(u)-1} h_m[l]\left(x_r[l+m]\sin\left(f(u)l\right) + x_i[l+m]\cos\left(f(u)l\right)\right)$$

$$c_u[n] = \left|\sum_{l=0}^{\lambda(u)-1} x(l+m)h(l)e^{jf(u)m}\right|^2$$

Working from a similar formulation, Anderson [Anderson92, 94] stated that evaluation of these equations at two distinct frequencies (scales), results in a system of two non-linear equations. When the two equations are subtracted from one another, the squared terms are eliminated, yielding a linear relationship between $x_r[n]$ and $x_i[n]$. Solving for $x_i[n]$ in terms of $x_r[n]$ and substituting into either of the original non-linear equations, yields a quadratic in terms of one unknown $x_r[n]$. Solution of the quadratic equation gives two candidate values. Computation of these values at two distinct frequency (scale) points resolves this ambiguity by selecting the common value present in both solutions. Once the value of $x_r[n]$ is known, $x_i[n]$ can be found directly. Anderson [Anderson92] noted that signal reconstruction in this manner only take places to within a phase constant, as expected though use of only modulus information. Anderson [Anderson94] described a more advanced version of this algorithm, based upon the use of an error criterion. Once one signal point has been constructed, the next can be recovered in an analogous manner.

Although requiring very few scale points (but more than two), this technique does require the signal to be causal and maximally overlapped. The causality constraint is another manifestation of the requirement for some degree of signal knowledge prior

to reconstruction, a limitation of all piece-wise approaches shown throughout this thesis.

6.5.1.2 Quadratic Form based method

Consider the definition of the GWT, (6-6). The squared modulus of this function, at a given point in time-scale, can be defined as given in (6-37), which can be expanded as (6-38).

$$\left|X\left[m,u\right]\right|^2 = \mathrm{Re}\left[\sum_{l=0}^{\lambda(u)-1} x[l]h\left[l-g(m),u\right]\right]^2 + \mathrm{Im}\left[\sum_{l=0}^{\lambda(u)-1} x[l]h\left[l-g(m),u\right]\right]^2 \tag{6-37}$$

$$\begin{aligned}\left|X\left[m,u\right]\right|^2 = &\left[\sum_{l=0}^{\lambda(u)-1} x_r[l]h_r\left[l-g(m),u\right] - x_i[l]h_i\left[l-g(m),u\right]\right]^2 \\ &+\left[\sum_{l=0}^{\lambda(u)-1} x_r[l]h_i\left[l-g(m),u\right] + x_i[l]h_r\left[l-g(m),u\right]\right]^2\end{aligned} \tag{6-38}$$

This equation can be expressed in matrix form (6-39).

$$\mathbf{x}^T\mathbf{A}_u\mathbf{x} = \left|X\left[m,u\right]\right|^2 \tag{6-39}$$

where

$$\mathbf{A}_u = \mathbf{A}_r + \mathbf{A}_i$$

$$\mathbf{A}_r = \left[\mathbf{h}_r \;\vdots\; -\mathbf{h}_i\right]^T\left[\mathbf{h}_r \;\vdots\; -\mathbf{h}_i\right]$$
$$\mathbf{A}_i = \left[\mathbf{h}_r \;\vdots\; \mathbf{h}_i\right]^T\left[\mathbf{h}_r \;\vdots\; \mathbf{h}_i\right]$$

$$\mathbf{h}_r = \left[h_r(0,u) \;\; h_r(1,u) \;\; \cdots \;\; h_r((l_x-1)-g(m),u)\right]$$
$$\mathbf{h}_i = \left[h_i(0,u) \;\; h_i(1,u) \;\; \cdots \;\; h_i((l_x-1)-g(m),u)\right]$$

$$\mathbf{x} = \left[x_r[0] \;\; x_r[1] \;\; \cdots \;\; x_r[l_x-1] \;\; x_i[0] \;\; x_i[1] \;\; \cdots \;\; x_i[l_x-1]\right]^T$$

The quadratic form in (6-39) defines the squared magnitude for the GWT for one point in time-scale. As previously, the generalisation of this equation from one point in time-scale to multiple points in time, and then to multiple points in time-scale, is non-trivial. The full extension is described by (6-40).

$$\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x} = \mathbf{X} \tag{6-40}$$

where

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}^T \end{bmatrix} \qquad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{B}_{0,0} \\ \mathbf{B}_{1,0} \\ \vdots \\ \mathbf{B}_{P(0)-1,0} \\ \mathbf{B}_{0,1} \\ \vdots \\ \mathbf{B}_{P(U)-1,U-1} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} \left|X[0,0]\right|^2 \\ \left|X[1,0]\right|^2 \\ \vdots \\ \left|X\left[P(0),0\right]\right|^2 \\ \left|X[0,1]\right|^2 \\ \vdots \\ \left|X\left[P(U)-1,U-1\right]\right|^2 \end{bmatrix}$$

$$\mathbf{B}_{m,u} = \begin{bmatrix} \mathbf{RR}_{m,u} & \mathbf{RI}_{mn,u} \\ \mathbf{IR}_{m,u} & \mathbf{II}_{mn,u} \end{bmatrix} \qquad \begin{bmatrix} \mathbf{ARR}_u & \mathbf{ARI}_u \\ \mathbf{AIR}_u & \mathbf{AII}_u \end{bmatrix} = \mathbf{A}_u$$

$$\mathbf{RR}_{m,u} = \begin{bmatrix} \mathbf{Z}_{1-g(m),1-g(m)} & \mathbf{Z}_{1-g(m),\lambda(u)} & \mathbf{Z}_{1-g(m),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{\lambda(u),1-g(m)} & \mathbf{ARR}_u & \mathbf{Z}_{\lambda(u),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{l_x-g(m)-\lambda(u),1-g(m)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),\lambda(u)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),l_x-g(m)-\lambda(u)} \end{bmatrix}$$

$$\mathbf{RI}_{m,u} = \begin{bmatrix} \mathbf{Z}_{1-g(m),1-g(m)} & \mathbf{Z}_{1-g(m),\lambda(u)} & \mathbf{Z}_{1-g(m),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{\lambda(u),1-g(m)} & \mathbf{ARI}_u & \mathbf{Z}_{\lambda(u),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{l_x-g(m)-\lambda(u),1-g(m)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),\lambda(u)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),l_x-g(m)-\lambda(u)} \end{bmatrix}$$

$$\mathbf{IR}_{m,u} = \begin{bmatrix} \mathbf{Z}_{1-g(m),1-g(m)} & \mathbf{Z}_{1-g(m),\lambda(u)} & \mathbf{Z}_{1-g(m),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{\lambda(u),1-g(m)} & \mathbf{AIR}_u & \mathbf{Z}_{\lambda(u),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{l_x-g(m)-\lambda(u),1-g(m)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),\lambda(u)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),l_x-g(m)-\lambda(u)} \end{bmatrix}$$

$$\mathbf{II}_{m,u} = \begin{bmatrix} \mathbf{Z}_{1-g(m),1-g(m)} & \mathbf{Z}_{1-g(m),\lambda(u)} & \mathbf{Z}_{1-g(m),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{\lambda(u),1-g(m)} & \mathbf{AII}_u & \mathbf{Z}_{\lambda(u),l_x-g(m)-\lambda(u)} \\ \mathbf{Z}_{l_x-g(m)-\lambda(u),1-g(m)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),\lambda(u)} & \mathbf{Z}_{l_x-g(m)-\lambda(u),l_x-g(m)-\lambda(u)} \end{bmatrix}$$

and as before, $P(u)$ is the number of time-slice for a given wavelet decomposition, $\lambda(u)$ is the length of the wavelet, $\mathbf{Z}_{m,n}$ is a zero matrix of size $m \times n$. The matrices $\mathbf{ARR}_u, \mathbf{ARI}_u, \mathbf{AIR}_u, \mathbf{AII}_u$ are defined by symmetrically splitting up $\mathbf{A}_u$ as defined by (6-39).

Just as in the ST-FT case, the formulation is non-linear in terms of the signal. As before, an iterative solution is proposed in order to solve this equation. This is given in (6-41). In general the matrix $\left(\hat{\mathbf{x}}\tilde{\mathbf{A}}\right)$ is not square, and thus pseudo inverse methods must be used.

$$\mathbf{x}_{p+1} = \left(\hat{\mathbf{x}}_p \tilde{\mathbf{A}}\right)^{\#} \mathbf{X} \tag{6-41}$$

As in the ST-FT case, this iteration is not guaranteed to converge even for an valid input GWT. Indeed trials have shown that, unlike the ST-FT case, this approach fails to converge even for valid GWT, unless careful selection of the decomposing wavelets is made. As before, a Newton based technique could also be used, solving an alternative form of equation (6-40) given in (6-42). As stated in Chapter 5 Section 5.1.6 Newton's algorithm is only guaranteed to converge when the starting point is close to the root. A modified version of Newton's algorithm [Press92], also consistently failed to converge to the correct solution for an arbitrary starting point. Furthermore this method required the multiple computations of the function and its Jacobian, making it computationally intensive.

$$\hat{\mathbf{x}}\tilde{\mathbf{A}}\mathbf{x} - \mathbf{X} = 0 \tag{6-42}$$

The computational load and the problems with convergence of the solution mean that the quadratic form based method is computationally problematic.

6.5.1.3 MLS Based Method

In Section 6.2.3 a MLS approach to signal reconstruction was developed and computational issues considered. Thus, in the same manner as described by Griffin [Griffin83] for the ST-FT, a MLS based signal reconstruction from the magnitude of the GWT can be now described. An iterative algorithm for signal reconstruction is described in (6-43) and (6-44), where $X^p[m,u]$ is the GWT of the p^{th} estimate of the signal $\mathbf{x}^{p+1}$, and $Y[m,u]$ is the desired GWT.

$$\mathbf{x}^{p+1} = \mathbf{K}^{\#}\mathbf{Y}^p \tag{6-43}$$

$$\mathbf{Y}^p = \left(\sum_{u=0}^{U-1} \sum_{m=0}^{P(u)-1} \hat{Y}^p[m,u] h\left[n - g(m), u\right] \right) \tag{6-44}$$

$$\hat{Y}^{p}[m,u]=|Y[m,u]|\frac{X^{p}[m,u]}{|X^{p}[m,u]|}$$

This iterative algorithm is found to be monotonically decreasing with regards to the distance measure given in (6-45). A proof of this is given in Appendix 2. It should be noted that, although the computation of kernel may take some time, it need only be computed once at the start of the algorithm. Computation of the vector $\mathbf{Y}^{p}$ can be achieved through one outer product and one summation, and thus can be computed quickly even for large signals.

This approach is very similar in definition to the non-MLS inversion of the WT described in the previous section and given in equations (6-30) and (6-31). As illustrated in Table 5, the windowing function can be selected such that the GWT provides a WT decomposition of the signal. Therefore the two algorithms can be considered as attempting to achieve the same goal. However since this algorithm uses the MLS inversion formula, it will yield an optimum solution. This is not true of the approach presented by Irino [Irino92, 93]. Thus the inversion described by (6-43) and (6-44) is expected to be more robust to invalid GWT, in analogy with the ST-FT case described by Griffin [Griffin84].

$$D_{M}[x[n],Y[m,u]]=\sum_{u=0}^{U-1}\sum_{m=0}^{P(u)-1}\left(|X[m,u]|-|Y[m,u]|\right)^{2} \tag{6-45}$$

The algorithm is stopped when the distance measure or reconstructed signal changes less than a given tolerance, or if a set number of iterations is exceeded.

6.5.1.3.1 Examples of Signal Reconstruction from Modulus of GWT based on MLS GWT Reconstruction

As previously, two example signals are now used to demonstrate this algorithm, a linear FM chirp and pseudo random sequence. Both sequences are 100 points long and are assumed to be sampled at 100 Hz. The Morlet wavelet as used to decompose the signal into time-scale. The algorithms were stopped after 500 iterations.

6.5.1.3.2 Example 1 Linear FM Chirp Sequence

The linear FM sequence used throughout this chapter was used an input, and a GWT computed. The magnitude of the GWT used here was given previously in Figure 52.

Figure 66 shows the error between the GWT of the current best estimate of the signal, and the desired GWT. The error given in these plots is the sum of the square error across the whole of the TF plot used for reconstruction. The error between the desired and reconstructed GWTs never reaches zero, however in the time-domain the error between the magnitude of the desired and the reconstructed signal is still small (*i.e.* error is less than 10% of the peak value of 10), as given in Figure 67. The error between the signals can also be seen in the differences between the phases, shown in Figure 68. Allowing the procedure to run for a larger number of iterations results in reduced errors between the original and reconstructed signals.

6.5.1.3.3 Example 2 Pseudo Random Gaussian Sequence

Similar performance between can also be seen when using a pseudo-random input sequence. The error of the converging system is shown in Figure 69, with the final error between the original and reconstructed signals shown in Figure 70 and Figure 71. The error at the last iteration was lower in the case of the Gaussian pseudo-random noise than for the linear chirp. This is reflected in the error between the desired and reconstructed signals being much lower. The maximum error in the magnitude of the signal is less than 7% of the max (maximum value of 1), compared to 10% in the linear chirp case. Once again, allowing the procedure to run for a larger number of iterations results in reduced errors between the original and reconstructed signals.

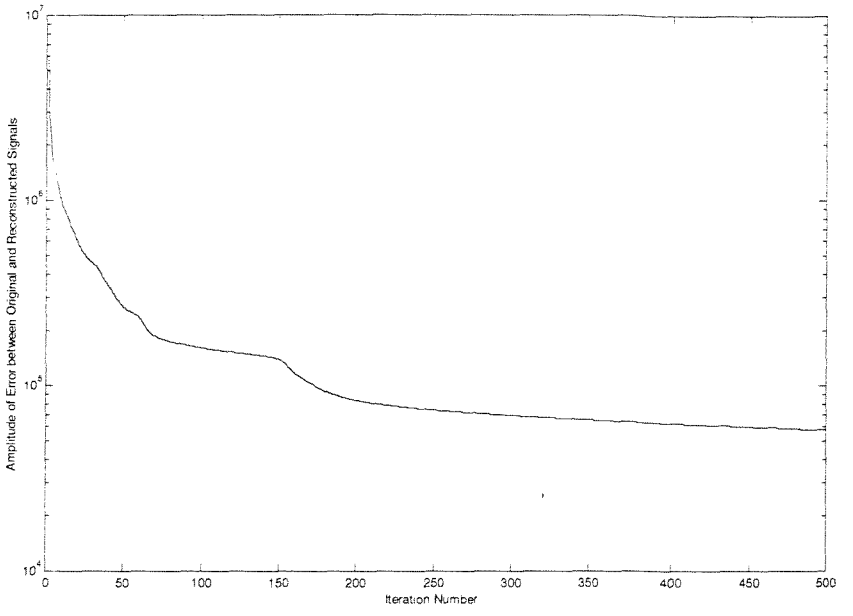

Figure 66 Error between GWT of Signal Estimate and Desired Signal – MLS based Reconstruction from Magnitude, Linear Chirp

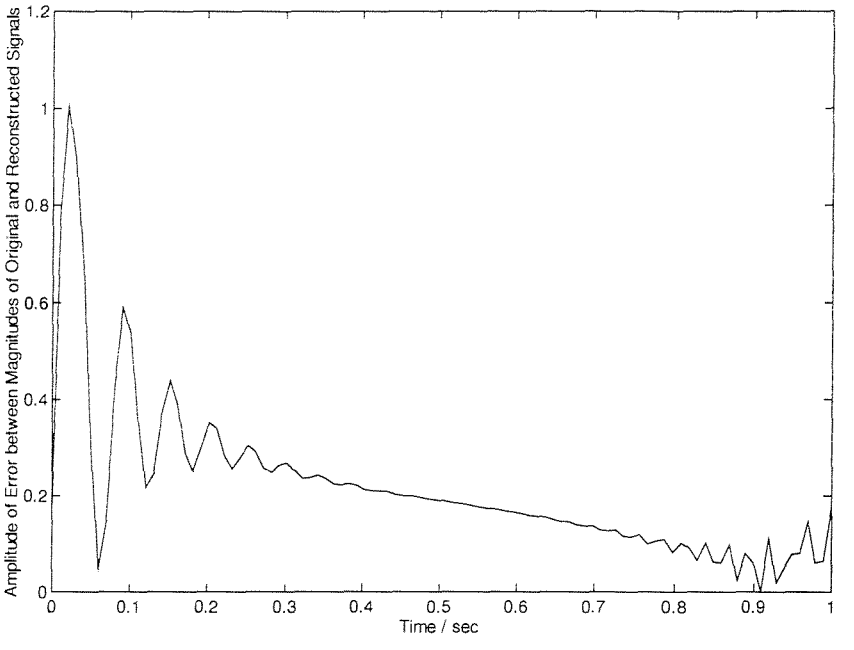

Figure 67 Error in Amplitude between Original and Reconstructed Signals – MLS based Reconstruction from Magnitude, Linear Chirp

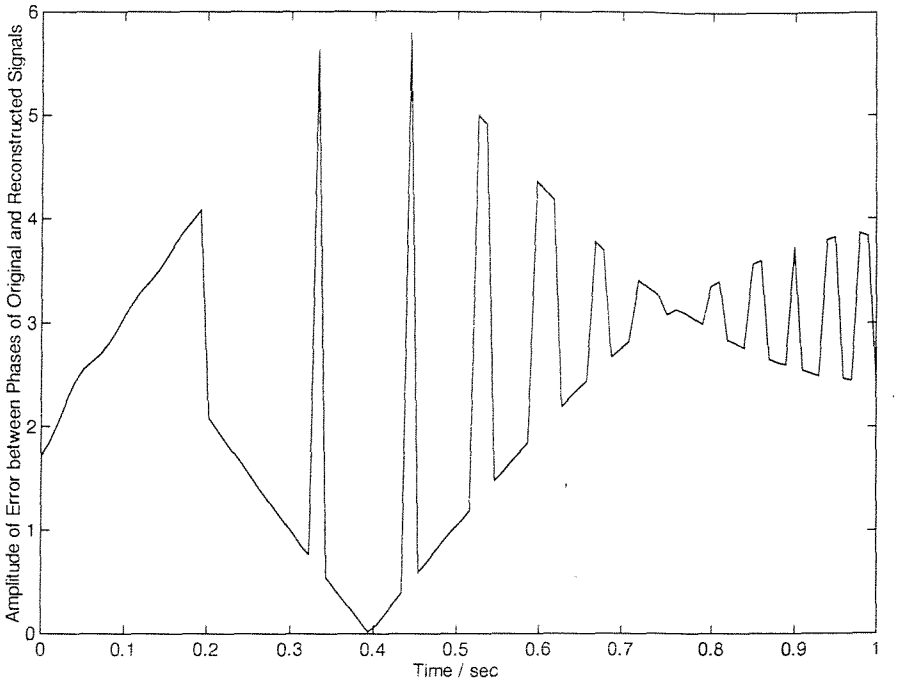

Figure 68 Error in Phase between Original and Reconstructed Signals – MLS Based Reconstruction from Magnitude, Linear Chirp

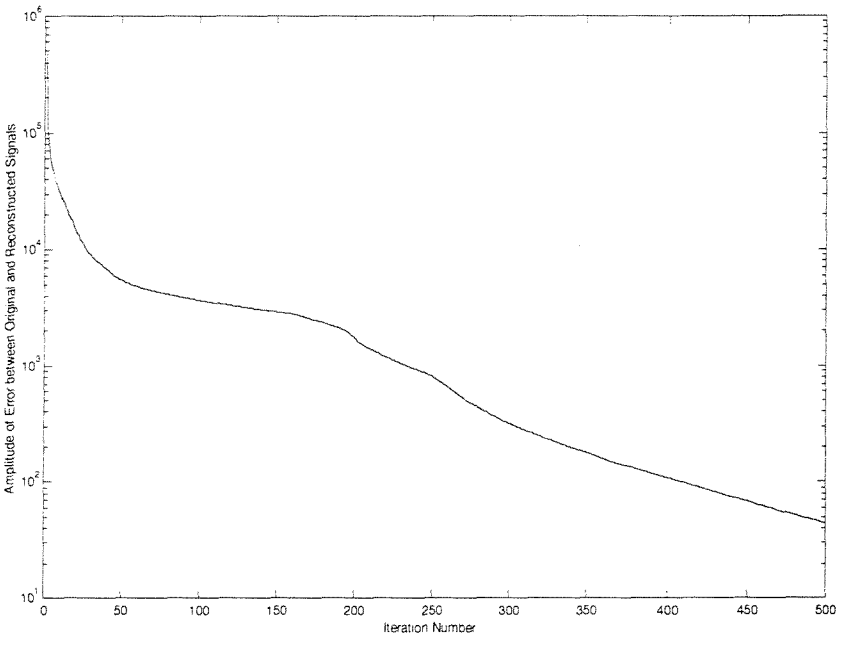

Figure 69 Error between GWT of Signal Estimate and Desired Signal – MLS Based Reconstruction from Magnitude, Pseudo Random Gaussian Sequence

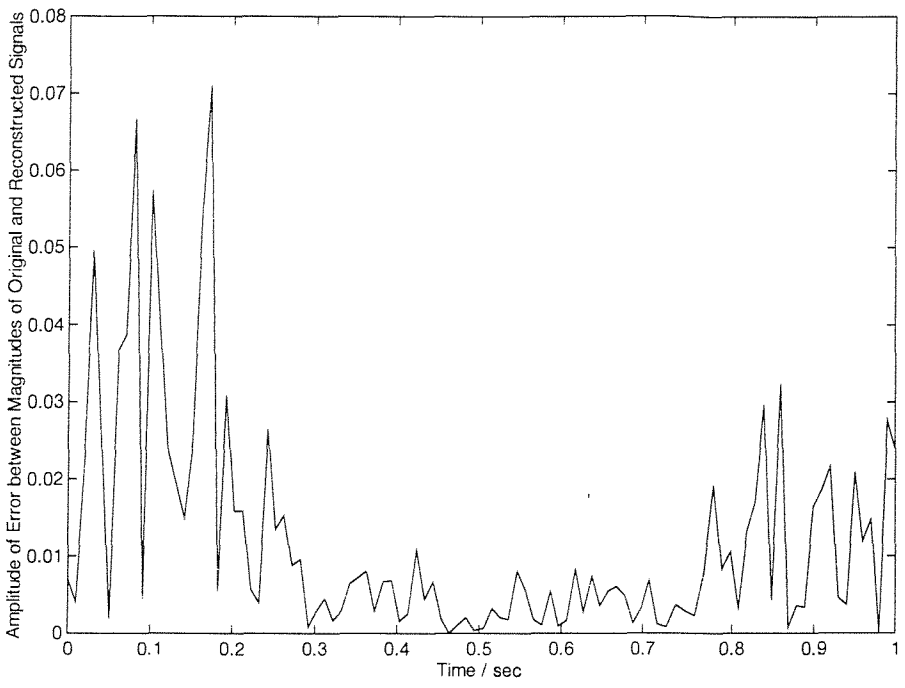

Figure 70 Error in Amplitude between Original and Reconstructed Signals – MLS based Reconstruction from Magnitude, Pseudo Random Gaussian Sequence

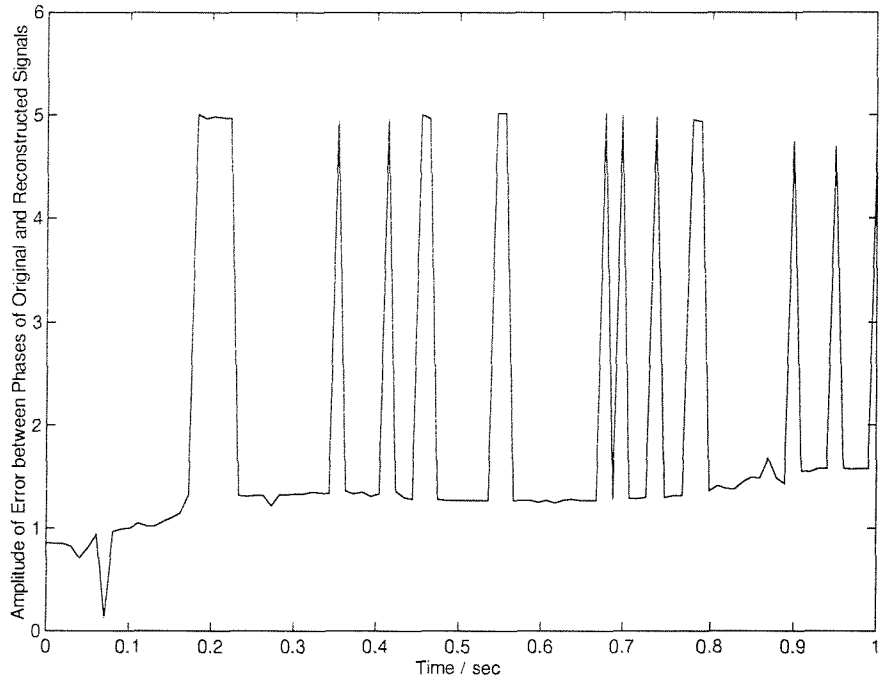

Figure 71 Error in Phase between Original and Reconstructed Signals – MLS based Reconstruction from Magnitude, Pseudo Random Gaussian Sequence

6.6 Conclusions

This chapter has generalised the theory of signal reconstruction from magnitude or phase presented previously for the ST-FT to cover a wide range of decomposition methods, the so called GWT.

This chapter began by reviewing present methods of signal reconstruction from the WT. These methods, although quick to implement, produce sub-optimal (in a MLS sense) solutions when invalid WTs are used. Developing the work presented by Griffin [Griffin83, 84] for the ST-FT, the MLS inversion of the GWT was defined and demonstrated. Although possessing qualities present in the ST-FT case (*i.e.* non-causality), the MLS signal reconstruction from the GWT requires much more computer resources than the ST-FT, requiring a large matrix inversion.

In Chapter 4 signal reconstruction from the phase of the ST-FT was presented. Herein signal reconstruction from the phase of the GWT was developed. In a direct analogy with the ST-FT case, signal reconstruction takes place via the inversion of a signal dependant matrix. Examples showing the algorithm working for a set Morlet decomposing wavelet functions were presented.

Signal reconstruction from the magnitude of the WT (as with the ST-FT) have been presented previously in the literature. Two previously defined methods were reviewed, one based upon the piece-wise reconstruction of a causal signal [Anderson94], and the other based by the algorithm described by Griffin [Griffin83, 84] for signal reconstruction, without using a MLS based inversion procedure. The former of these two requires some prior signal knowledge and the latter, despite being fast to implement, cannot be shown to converge to the optimal solution.

Two novel alternatives to these approaches were presented for signal reconstruction from the magnitude of the GWT. The first is based on the solution to a quadratic form. This technique suffers all the limitations present in the ST-FT case, requiring large computational resources with no guarantee of convergence. Trials of two different methods to solve this problem have shown that convergence may not occur even if the GWT is valid.

The second technique follows the methodology presented in Griffin [Griffin83, 84] but using the MLS signal reconstruction from GWT. This solution to the problem has a number of benefits over those defined previously. However since the algorithm is based upon the MLS inversion of the GWT, it suffers from the poor computational performance of that algorithm, making it currently unsuitable for reasonable length signals. This method is guaranteed to converge toward the MLS solution even if the input is an invalid as the magnitude of a GWT.

Chapter 7 Enhancement of Heart Sounds via Time-Frequency Modification

7.1 Introduction

The traditional method for physicians to hear to the sounds created by the heart beating has been through the use of a stethoscope. By listening to the sound created, a skilled physician can identify some structural defects present in the heart of a patient. Abnormal sounds, such as a heart 'murmur', may reflect some pathological defect in the heart. The pitch, duration and location of the murmur in the cardiac cycle are all important parameters used in diagnosis [Leung99].

7.2 Time Domain Form of Heart Sounds

As heard through a traditional or an electronic stethoscope, the heart sounds like 'lub-dup lub-dup'. These sounds correspond to the two dominant sounds in the cardiac cycle. The first sound, the 'lub', is known as the first heart sound (S1) and the second, the 'dup', as the second heart sound (S2). An explanation of the mechanisms creating these sounds is not offered here, and the reader is referred to [Leung99, Rangayyan88, Durand95].

In addition to S1 and S2, other sounds can also occasionally be heard when listening to the heart. These range from ejection sounds (high-pitched clicks) to filling sounds (low-pitched thumps). The sounds of interest in this thesis are heart 'murmurs'; these sounds are created by the turbulent flow of blood in the cardiovascular system. There are a number of potential causes for this turbulence.

In this research the murmurs which are investigated are: Atrial Septal Defects (ASD), Ventricular Septal Defects (VSD) and non-pathological innocent murmurs. Temporally these murmurs all exist in the region between S1 and S2 in the heart sound.

ASD refers to a defect (hole) in the wall separating the left and right atria. The sound created is due to the additional flow of blood being pumped trough the pulmonary artery.

VSDs are due to a defect in the wall separating the left and right ventricles. The pumping action of the heart creates a high pressure, which results in the flow of blood through the defect becoming turbulent. VSDs are typically easier to diagnose due to their long duration and loudness.

Innocent murmurs can occasionally be found in young children. These murmurs are created due to a more dynamic rather normal flow of blood in the heart becoming turbulent, and thus creating flow noise [Harris76, Leung99]. These murmurs can be difficult to distinguish from ASDs.

7.3 Extension of the Duration of a Signal via Spectrogram

A typical cardiac cycle in an adult lasts for around a second (less in children), making diagnosis of a murmur difficult. The murmur is typically lower in amplitude compared to S1 and S2. Both of these factors make accurate diagnosis a complicated task.

With a view to improving the ability of the physician to diagnose the type of murmur present, the cardiac cycle can be recorded and replayed at a slower rate. This helps to separate S1 and S2, but it has an additional effect; the contraction and expansion of a signal in time alters its spectral content. This relation was stated mathematically in Chapter 2 Section 2.2.3.2, and is used by the WT to alter the shape in time and frequency of a particular wavelet. This effect can be heard when playing a record at the wrong speed: there is a shift in pitch as well as an alteration of the temporal support of the sound.

In a series of papers Zhang *et al.* [Zhang88a-b] describe a matching pursuit algorithm applied to the expansion of heart sounds. The matching pursuit algorithm decomposes a signal in terms of known time-frequency 'atoms'. Each atom can be scaled, dilated, and shifted in time or frequency to provide the best local match to the signal. An iterative algorithm is described to decompose the signal: at each iteration the atom that best matches the signal is found. The remainder of the signal, after this

atom has been subtracted, is the starting point for the next iteration. This method works well when the signal being analysed is efficiently modelled by a sum of decomposing functions; Zhang *et al.* use exponentially decaying sinusoids. However, the number of decomposing functions required increases rapidly when the signal is not well matched to the decomposing functions. As a result this procedure has a problem synthesising murmurs in heart sounds due to their random nature, since these murmurs can not be efficiently represented using simple TF atoms. In the paper by Zhang *et al.* [Zhang88b] this parametric approach is used to temporally expand and frequency shift heart sounds, in order to improve audio discrimination and thus aid diagnosis and training.

Here an alternative non-parametric method is proposed. Altering the temporal characteristics of the signal, whilst retaining its frequency characteristics, is possible if the process is performed in the TF domain rather than the time domain. In this manner, the signal can be manipulated to increase its duration, without a contraction in frequency. The non-parametric nature of the GWTs makes it more widely applicable than the matching pursuit method.

The process of altering the duration of the signal, whilst retaining the frequency characteristics, is presented. The sounds from the heart are recorded at a sufficiently high sampling rate (in this case 2000 Hz). The signal is then transformed into the TF domain via a GWT. In order to reduce the computation time, the ST-FT is used to provide the TF decomposition.

Once the signal is in the TF domain, this representation can be stretched in time so that the required signal duration is achieved. This procedure can be achieved by interpolating addition time-slices between known time slices. The greater the number of interpolated slices inserted, the longer the temporal duration of the TFR. It follows therefore that the greater the number of interpolated slices in the TFR, the longer the duration of a signal reconstructed from it. This procedure can be difficult to achieve, if the complex ST-FT is used, due to phase wrapping ambiguities between successive time-slices. In order to improve the quality of the interpolated TFR, the squared magnitude of the ST-FT (spectrogram) is used. Use of the magnitude removes the problems of phase unwrapping that are present with a linear TFR.

The result of this procedure is to create an invalid spectrogram. Through the application of the MLS based technique described in Chapter 5 a signal can now be

reconstructed to create a temporally extended signal. Griffin *et al.* [Griffin84] used a similar procedure to alter the duration of speech. Since the resultant spectrogram created is unlikely to be valid, the error between the spectrogram created from the current estimate of the signal and the desired spectrogram will not converge to zero.

Using this approach, four recorded signals were extended. The four signals used are a normal heart sound, and heart sounds containing: an ASD murmur, a VSD murmur and a non-pathological innocent murmur[3]. Three seconds of data were acquired for each. This was sufficient to span three or four cardiac cycles. These time series are shown in Figure 72 to Figure 75. Visual inspection of these time-series yields little information about the presence or type of murmur. Only the VSD murmur can be clearly seen as a rapidly changing signal component between S1 and S2 (Figure 74). This is as expected, since typically a VSD murmur has a higher amplitude than the other types of murmur investigated here. The approximate times for the occurrence of S1 and S2 are given in Table 6.

The ST-FT for each of these signals was computed using a 128 pt Hanning window, with an overlap of 96 samples between successive time slices. The spectrograms for the four signals are given in Figure 76 to Figure 79. The impulse-like nature of S1 and S2 can be seen from these plots, covering at least the range from 0-500 Hz. Again, of the three murmur signals, the VSD murmur can be most clearly seen, as a broad-band (0-900Hz before being cut-off by the anti-aliasing filter) sound lasting almost all the time between S1 and S2. The innocent murmur can also be seen between S1 and S2 as increased signal energy at low frequencies. The ASD murmur cannot be identified as there as little difference between the spectrogram of the ASD and of the normal (healthy) heart beat. Using linear interpolation, the temporal dimension of the spectrogram was extended so that the resultant spectrogram was three times longer than the original. The modified spectrograms for the four signals are shown in Figure 80 to Figure 83.

Using these synthetic spectrograms as an input, application of the MLS based signal reconstruction formula (Chapter 5 Section 5.4.2) enabled the construction of a time series for each spectrogram. The converging error of the algorithm is shown in Figure 84 to Figure 87. The error given in these plots is the sum of the square error across the whole of the TF plot used for reconstruction. A fixed number of iterations

[3] These diagnoses have been made by a skilled physician

(2500) were used. The resulting time-series had the required extended temporal support. The differences between the desired and reconstructed spectrograms are depicted in Figure 88 to Figure 91, and the reconstructed spectrograms are shown in Figure 92 to Figure 95. The error remains very low (below −200dB) until the last few time-slices and then increases to approximately −100dB. This is due to the limited information available for these time-slices. The error between the two solutions can be seen to be very much lower than the maximum amplitude present in the desired spectrogram. Indeed, visually there is little difference between the desired and reconstructed spectrograms. The reconstructed signals are shown in Figure 96 to Figure 99.

Although the time-series and TFR provide a great deal of information regarding the signal, the transformation is most clear when the signal is listened to. The sound files are provided on the disc accompanying this thesis (Standard PC Format). The original versions of the heart sounds are given in the directory '/heart/original/', and the temporal extended versions given in '/heart/extended/'. In this form the difference between the three different murmurs can be more easily heard. The improvement over simply changing the sampling rate can be heard by comparing the temporally extended versions in '/heart/extended' with the re-sampled versions in '/heart/Lower Sampling Rate'. The difference in the fidelity in terms of the spectral content of the two different methods can be clear heard.

With the extended version of the normal heart sound, there is little or no noise between S1 and S2 of each cardiac cycle. With the innocent murmur, some narrow-band flow noise can be heard, but this decays quickly. In contrast the VSD murmur can be heard prominently between beats, lasting almost the whole interval between S1 and S2. In the ASD case, flow noise can also be heard between S1 and S2, but at lower amplitude than with the VSD. In this signal, S2 can be heard to be split into two parts.

As a training aid for physicians, these extended signals allow for easier discrimination between murmurs. Without the extension it was hard to hear the difference between the ASD and innocent murmur, whereas with the extension the difference can clearly be heard. As compared to playing a tape at a slower speed, this method offers the benefit of not altering the spectral content of the signal. Unlike the matching pursuit algorithm, the quality of the results is not affect by the

nature of the signal being analysed, and therefore this technique can be used to extend both normal heart sounds as well as those with murmurs.

Heart Sound	Time of S1	Time of S2
Normal Heart Sound (Figures 72,76,80,88,92 and 96)	0.5	0.1
	1.3	0.8
	2.4	1.6
		2.7
ASD Murmur (Figures 73, 77, 81, 89, 93 and 97)	0.5	0.2
	1.2	0.75
	1.9	1.4
	2.6	2.1
		2.9
VSD Murmur (Figures 74, 78, 82, 90, 94 and 98)	0.2	0.45
	0.75	1.0
	1.4	1.6
	1.9	2.2
	2.5	2.75
Innocent Murmur (Figures 74, 78, 82, 90, 94 and 98)	0.05	0.35
	0.6	0.9
	1.25	1.55
	1.95	2.25
	2.6	2.9

Table 6 Starting Times of S1 and S2 in the Analysed Signals

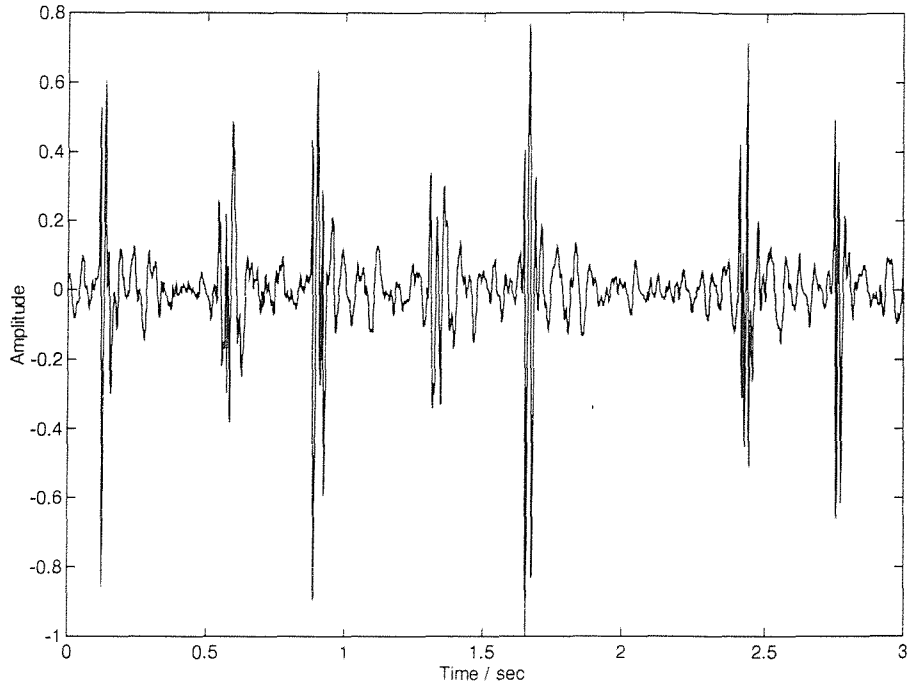

Figure 72 Time Series of a Normal Heart Beat

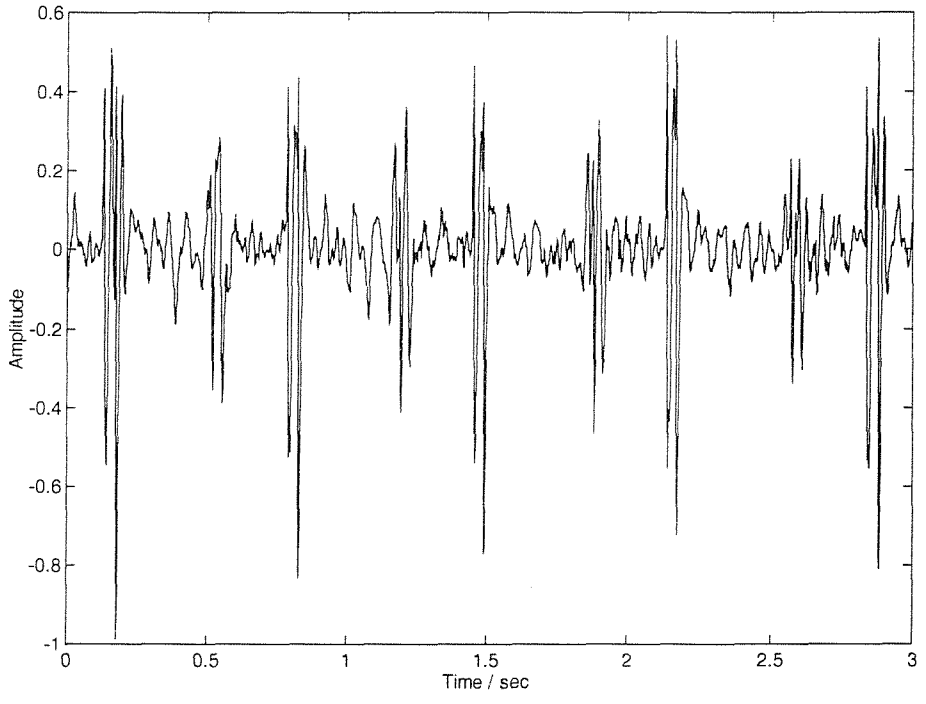

Figure 73 Time Series of a Heart Beat with an ASD

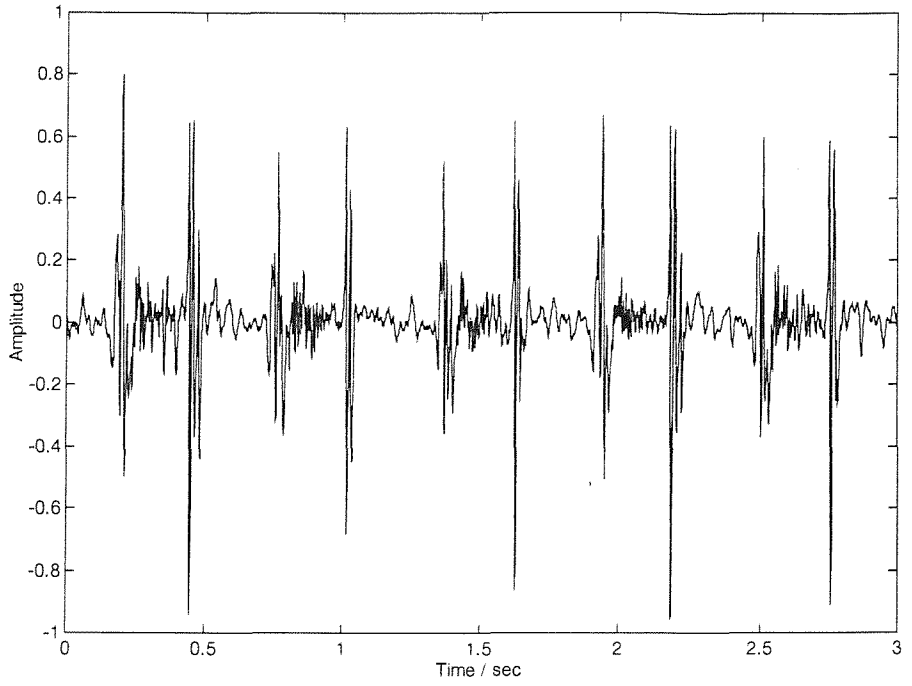

Figure 74 Time Series of a Heart Beat with a VSD

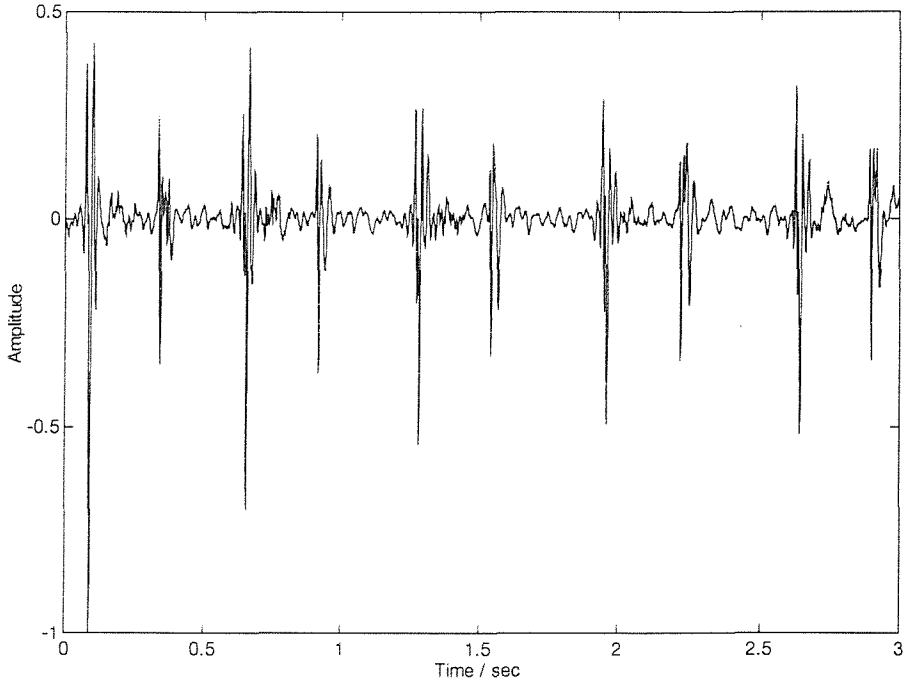

Figure 75 Time Series of a Heart Beat with an Innocent Murmur

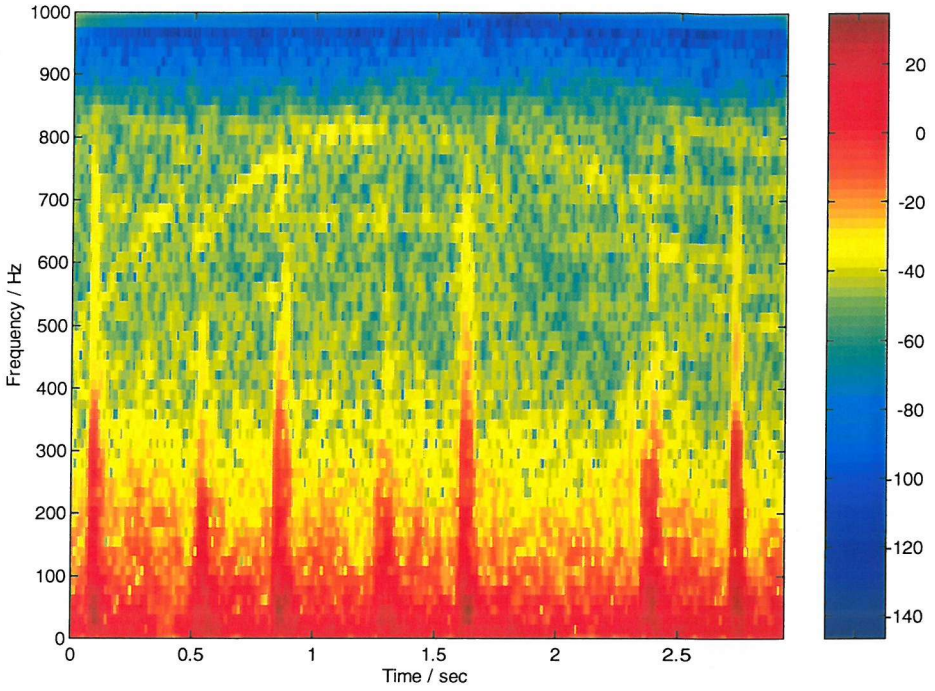

Figure 76 Spectrogram of a Normal Heart Beat

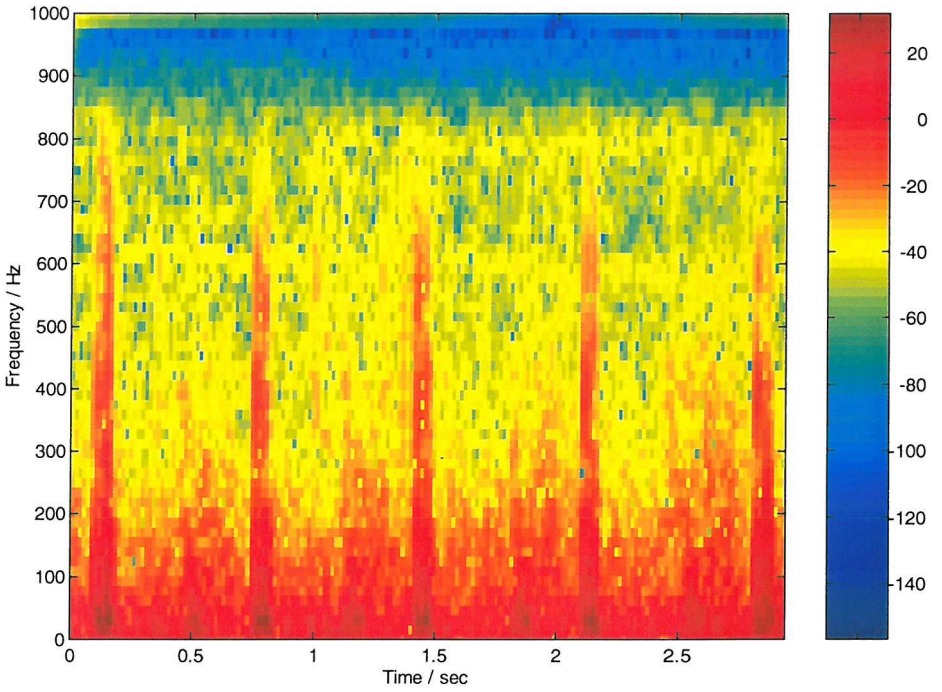

Figure 77 Spectrogram of a Heart Beat with an ASD

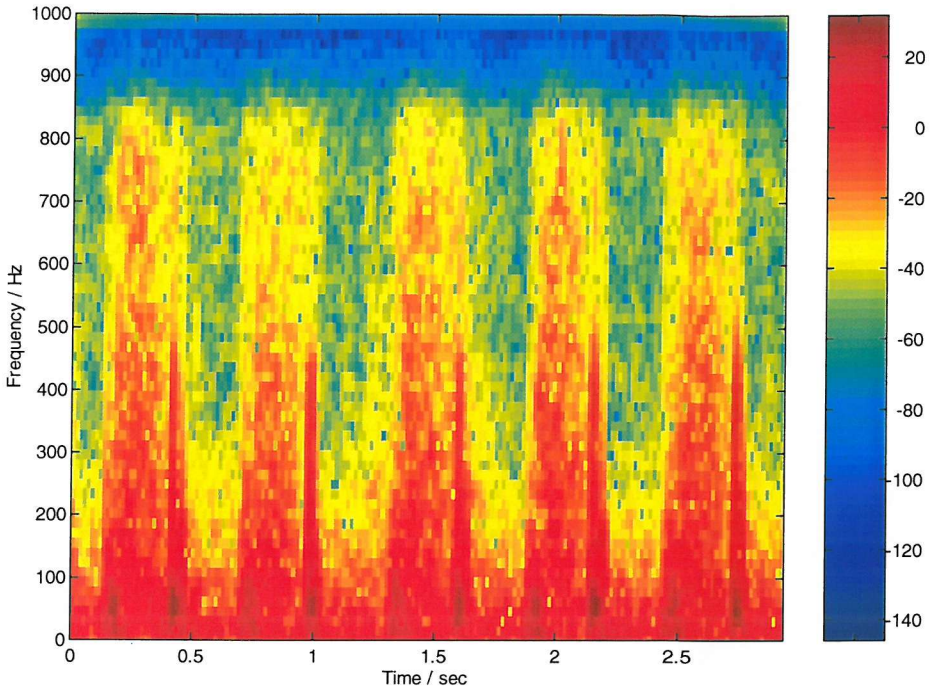

Figure 78 Spectrogram of a Heart Beat with a VSD

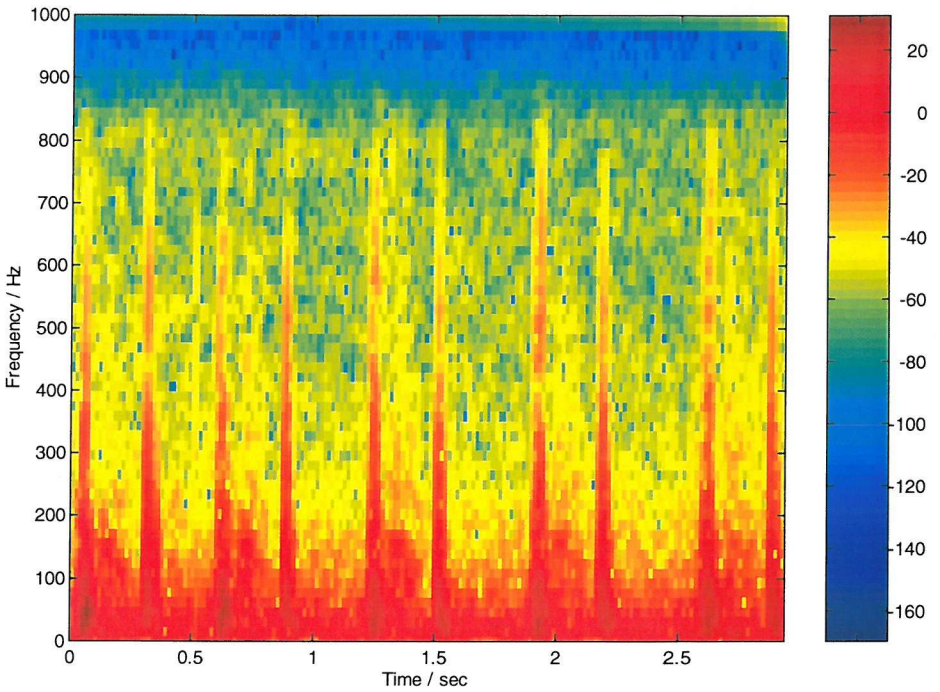

Figure 79 Spectrogram of a Heart Beat with an Innocent Murmur

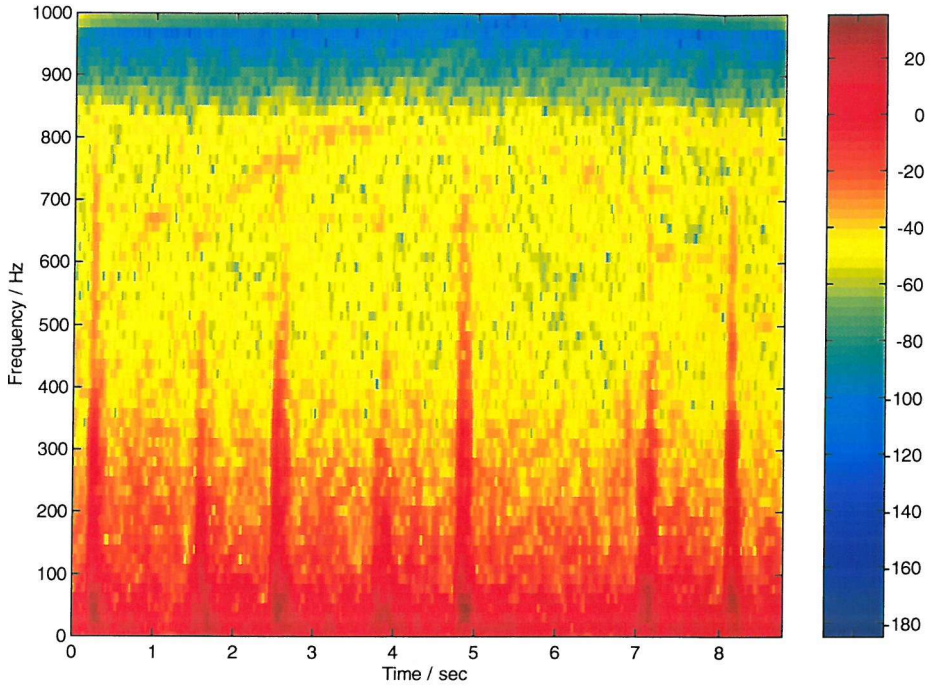

Figure 80 Modified Spectrogram of a Normal Heart Beat with Extended Temporal Support

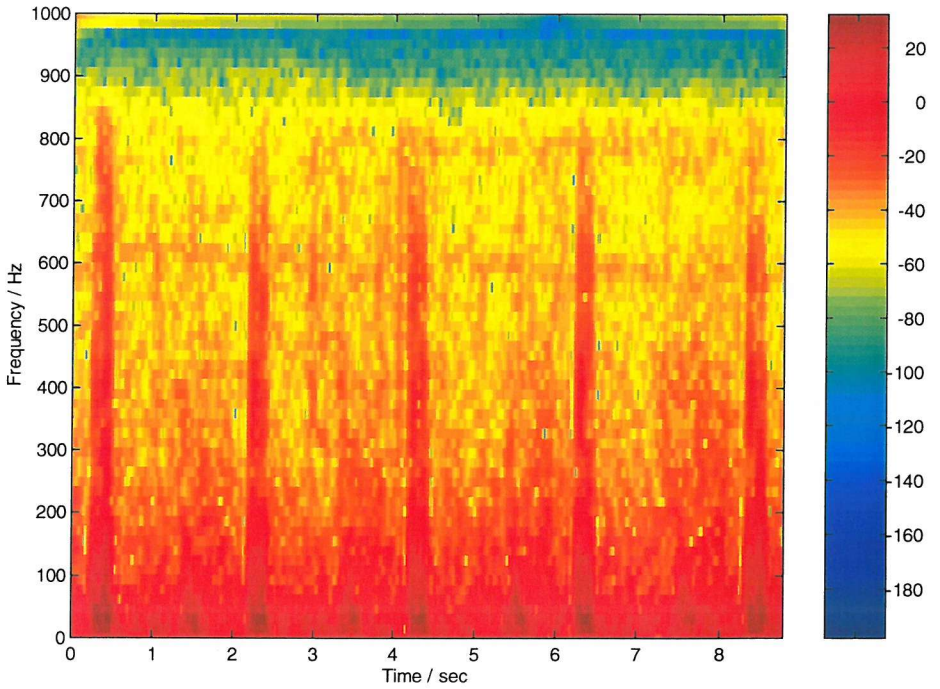

Figure 81 Modified Spectrogram of a Heart Beat with an ASD with Extended Temporal Support

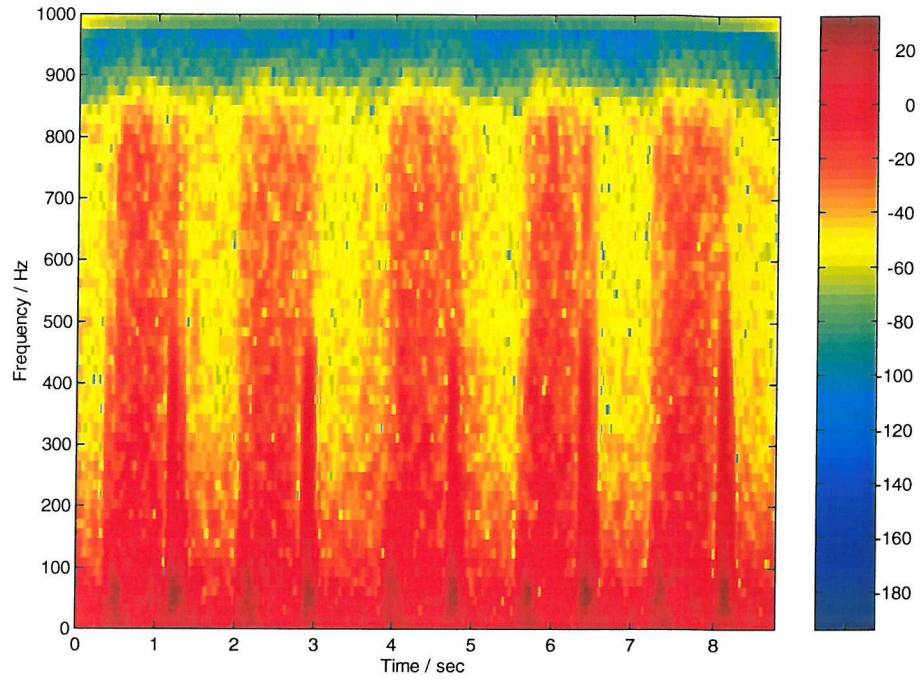

Figure 82 Modified Spectrogram of a Heart Beat with a VSD with Extended Temporal Support

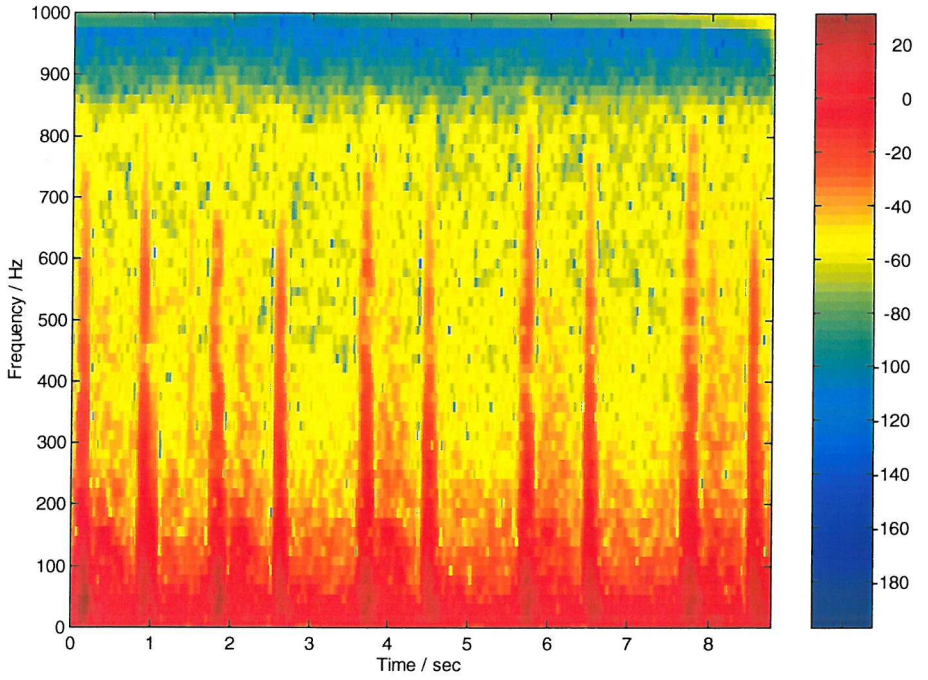

Figure 83 Modified Spectrogram of a Heart Beat with an Innocent Murmur with Extended Temporal Support

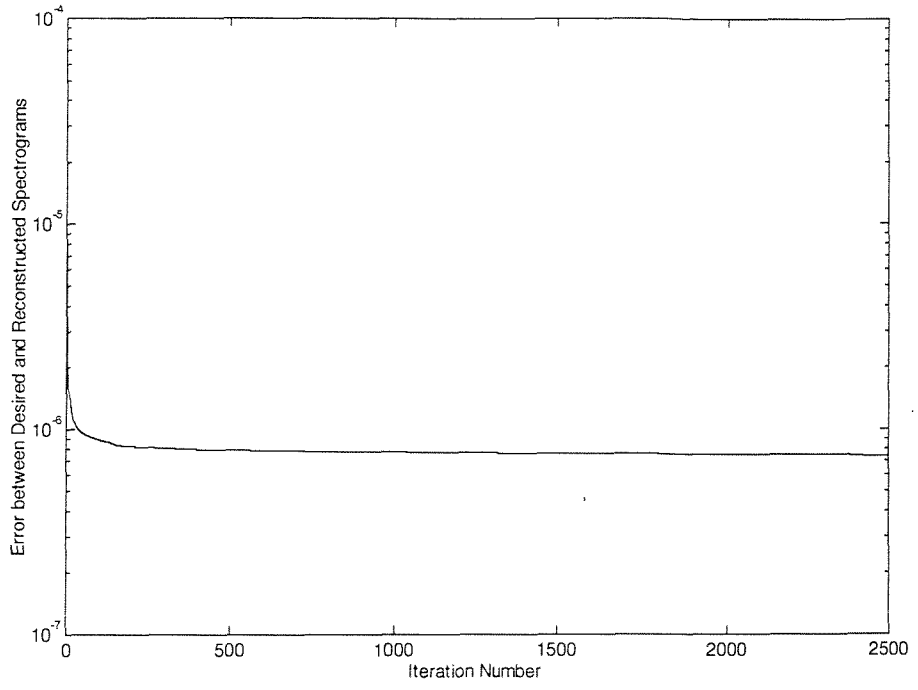

Figure 84 Convergence of the Distance Measure between the Desired and Reconstructed Spectrograms – Normal Heart Sound

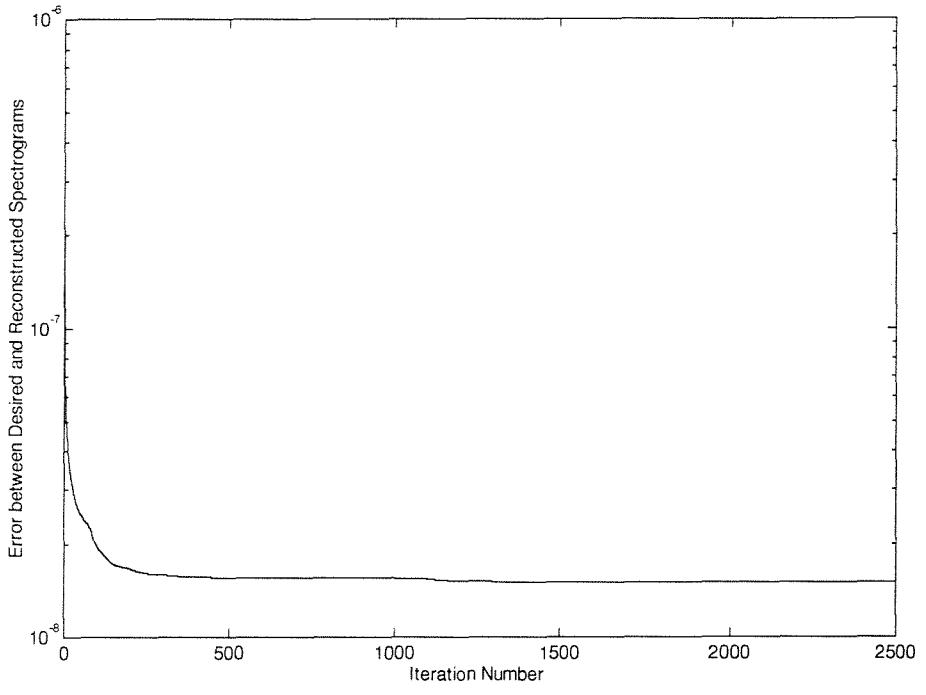

Figure 85 Convergence of the Distance Measure between the Desired and Reconstructed Spectrograms – Heart Beat with an ASD

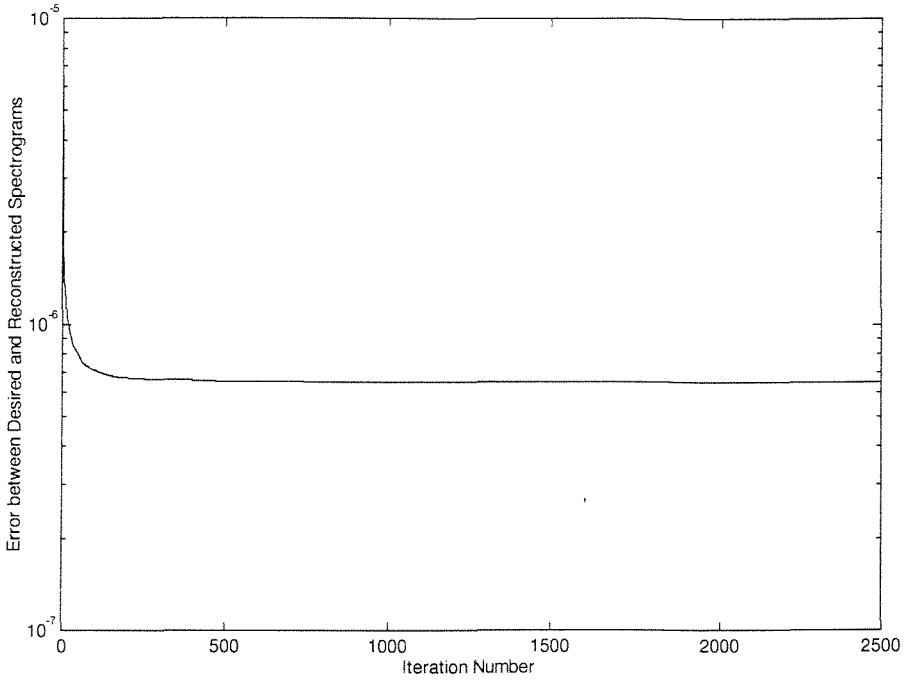

Figure 86 Convergence of the Distance Measure between the Desired and Reconstructed Spectrograms – Heart Beat with a VSD

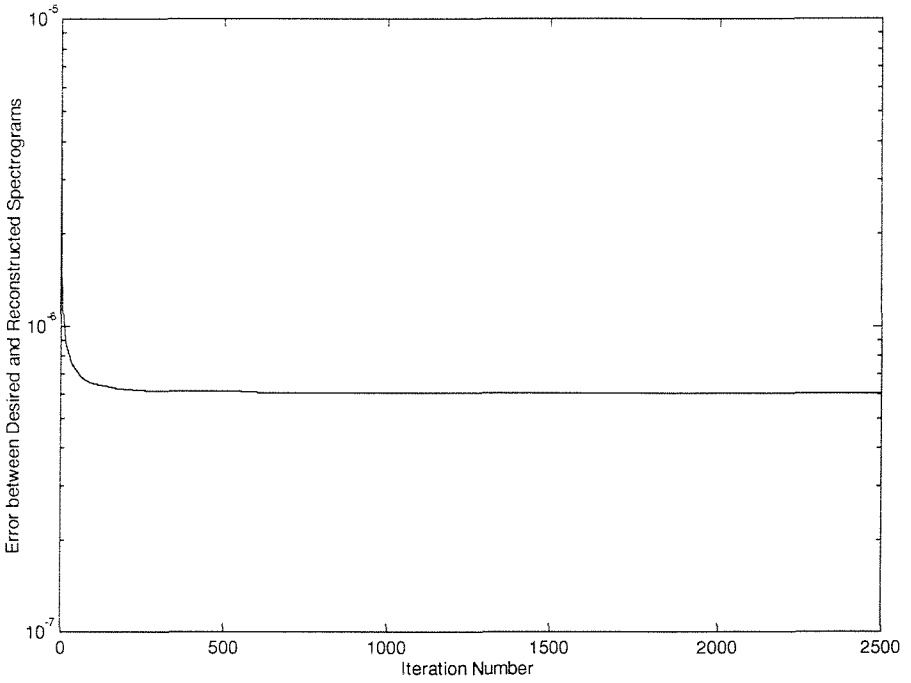

Figure 87 Convergence of the Distance Measure between the Desired and Reconstructed Spectrograms – Heart Beat with an Innocent Murmur

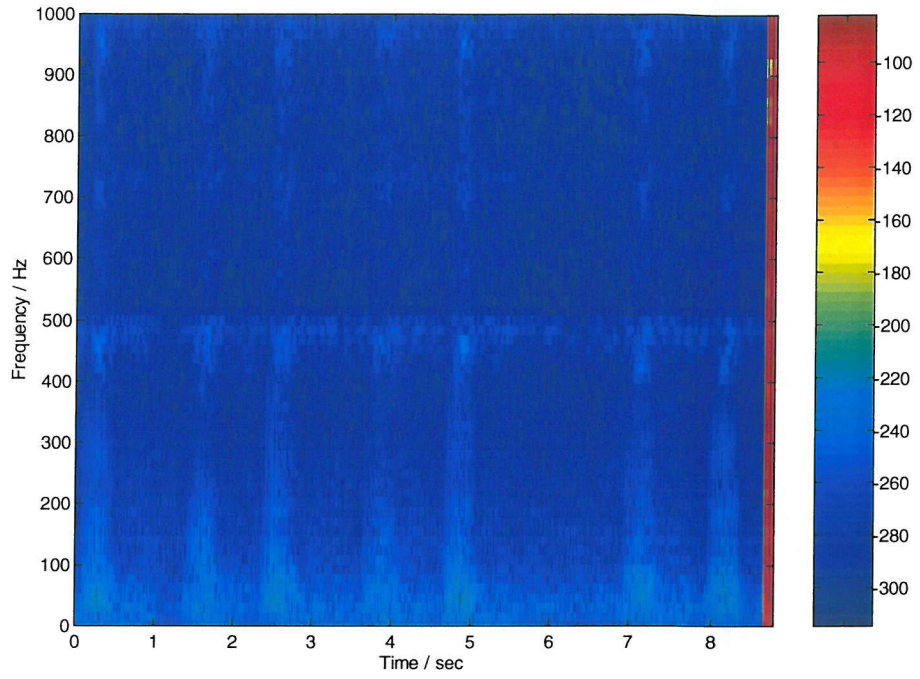

Figure 88 The Error between the Desired and Reconstructed Spectrograms – Normal Heart Sound

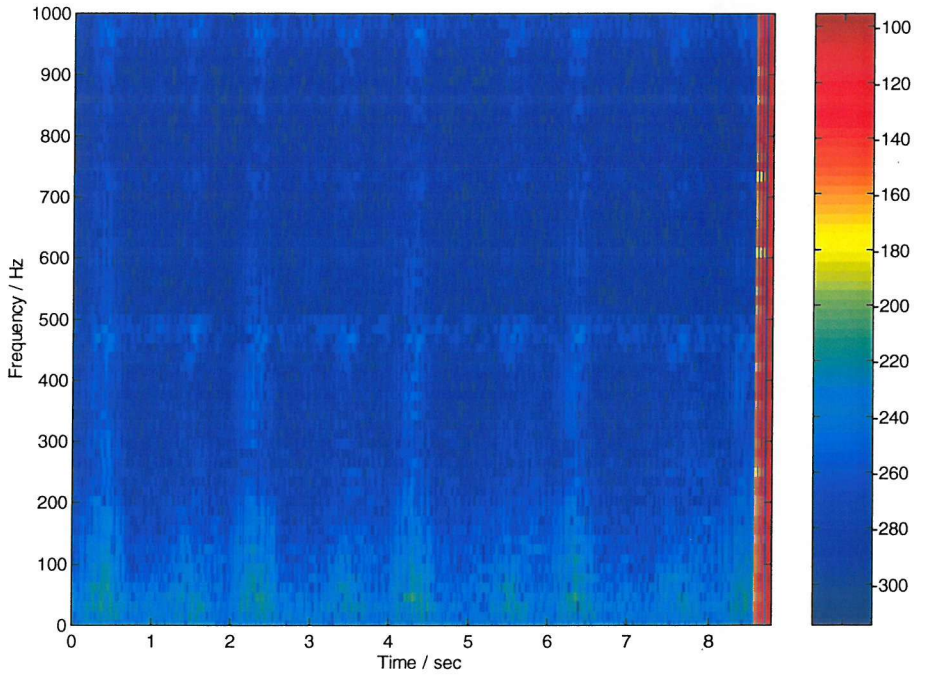

Figure 89 The Error between the Desired and Reconstructed Spectrogram – Heart Beat with ASD

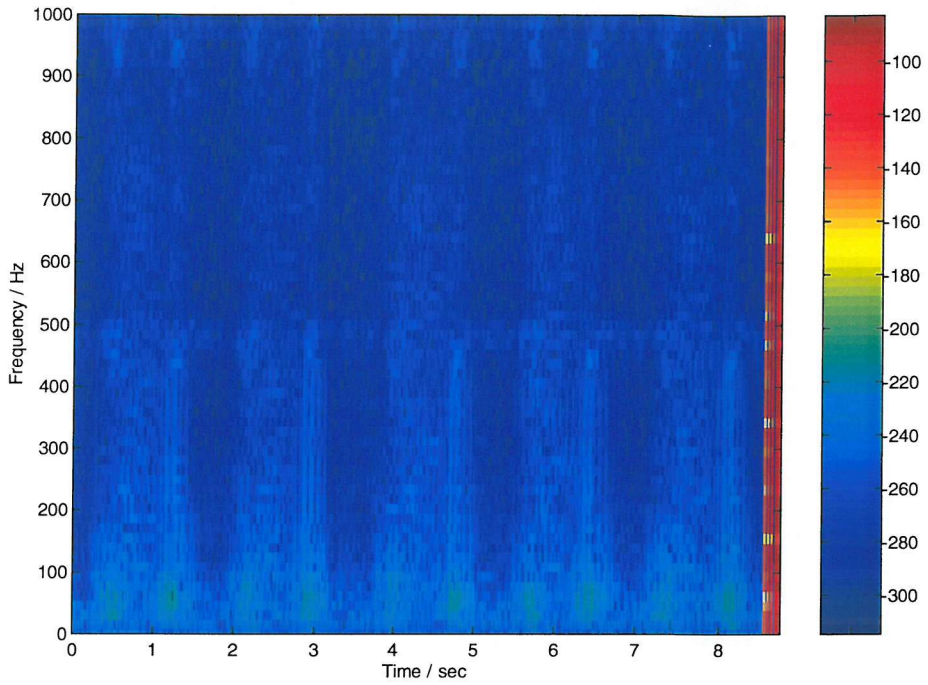

Figure 90 The Error between the Desired and Reconstructed Spectrogram – Heart Beat with VSD

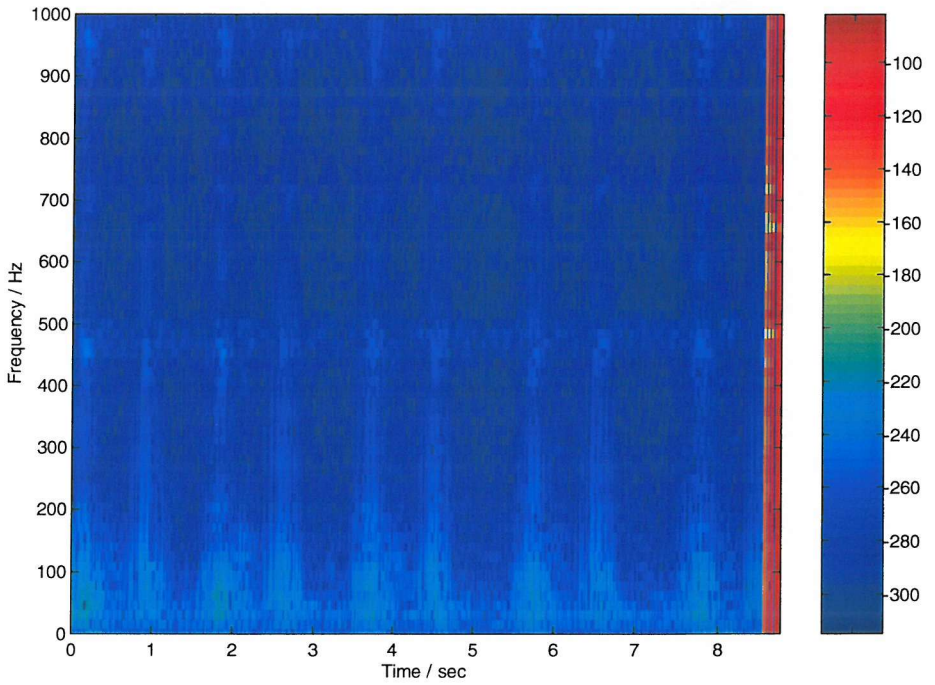

Figure 91 The Error between the Desired and Reconstructed Spectrogram – Heart Beat with Innocent Murmur

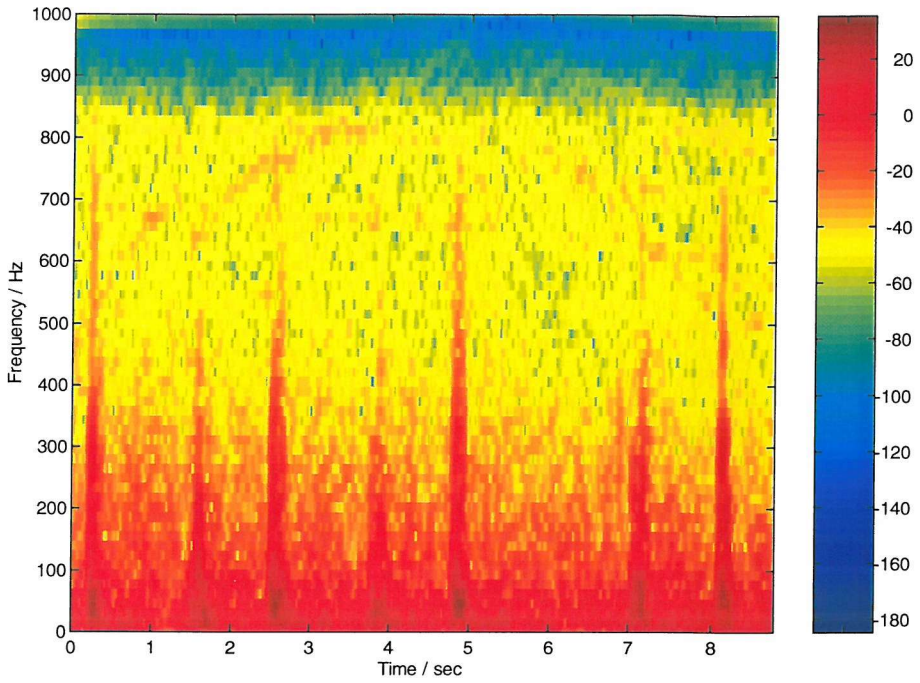

Figure 92 Reconstructed Spectrogram of a Normal Heart Beat whose Temporal Support has been Extended

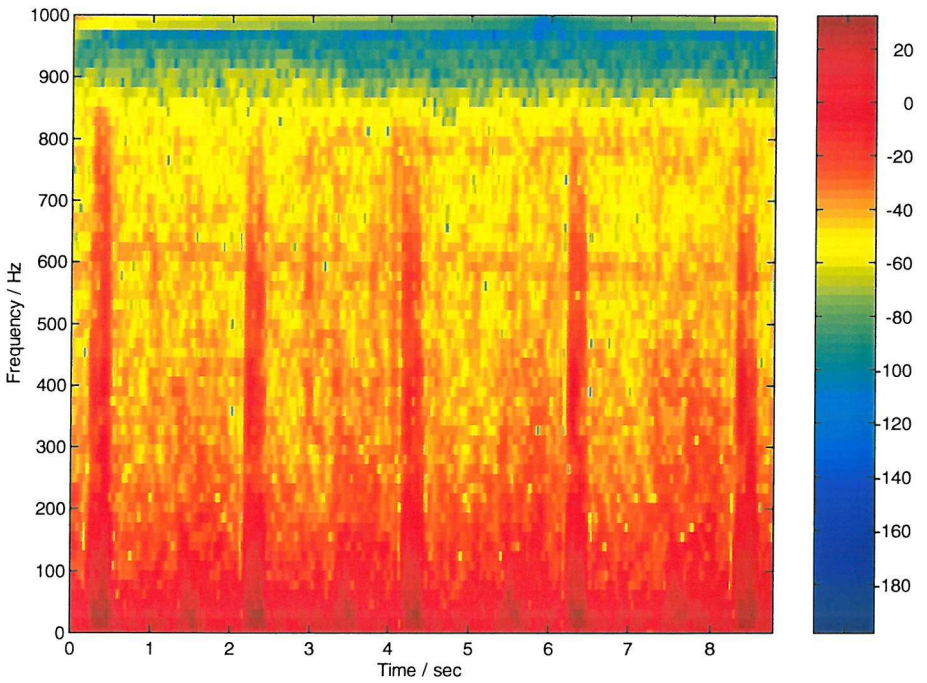

Figure 93 Reconstructed Spectrogram of a Heart Beat with an ASD whose Temporal Support has been Extended

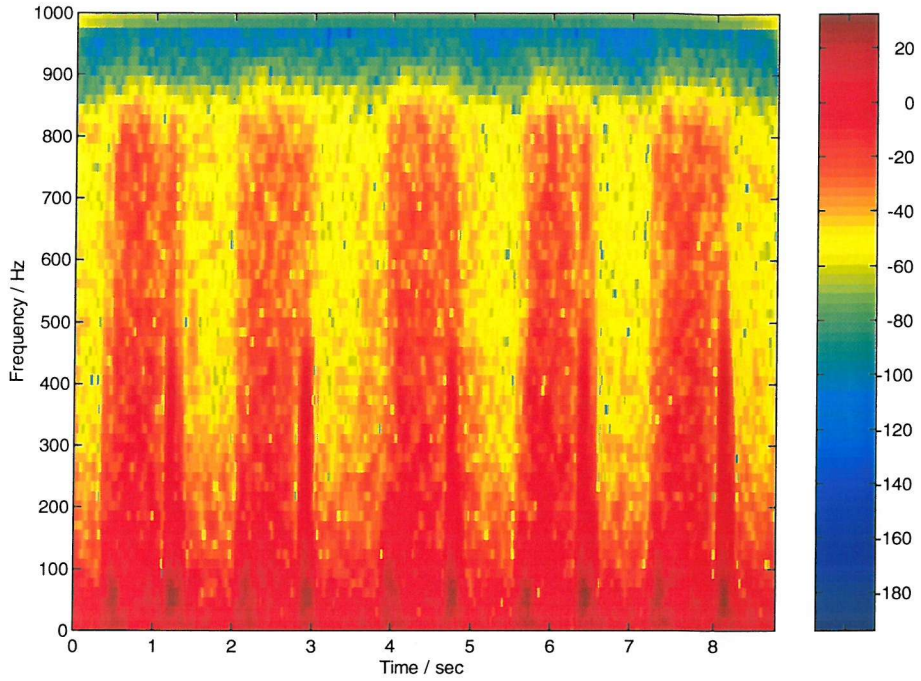

Figure 94 Reconstructed Spectrogram of a Heart Beat with a VSD with Extended Temporal Support

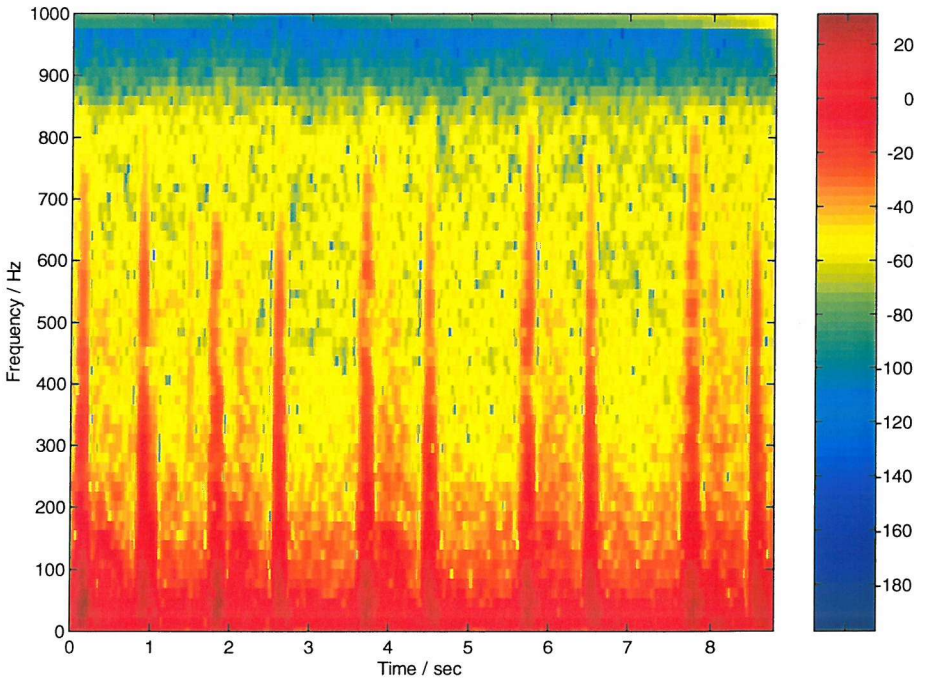

Figure 95 Reconstructed Spectrogram of a Heart Beat with an Innocent Murmur whose Temporal Support has been Extended

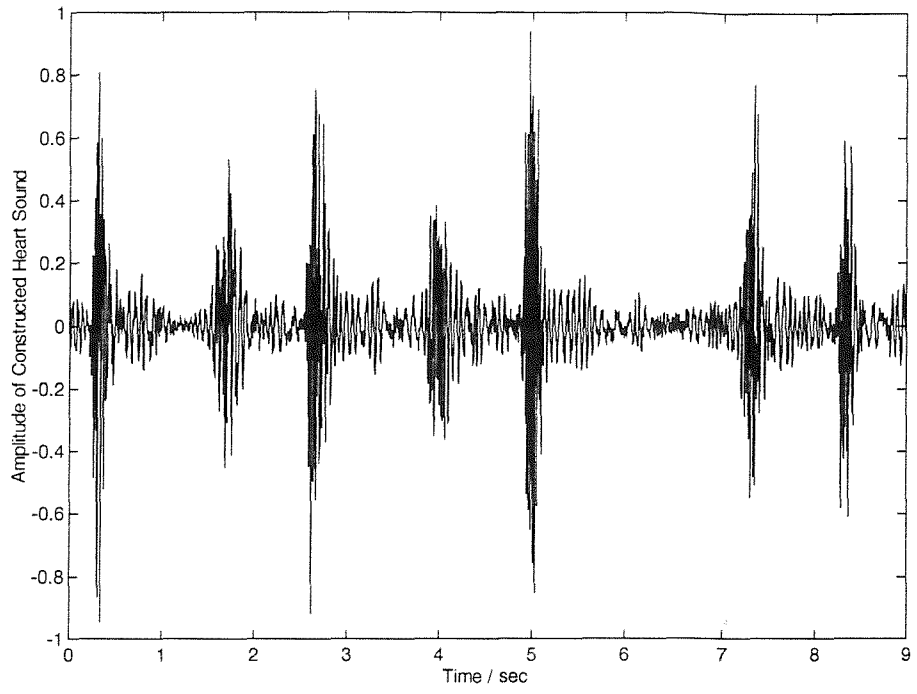

Figure 96 Time Series of Temporally Extended Normal Heart Beat

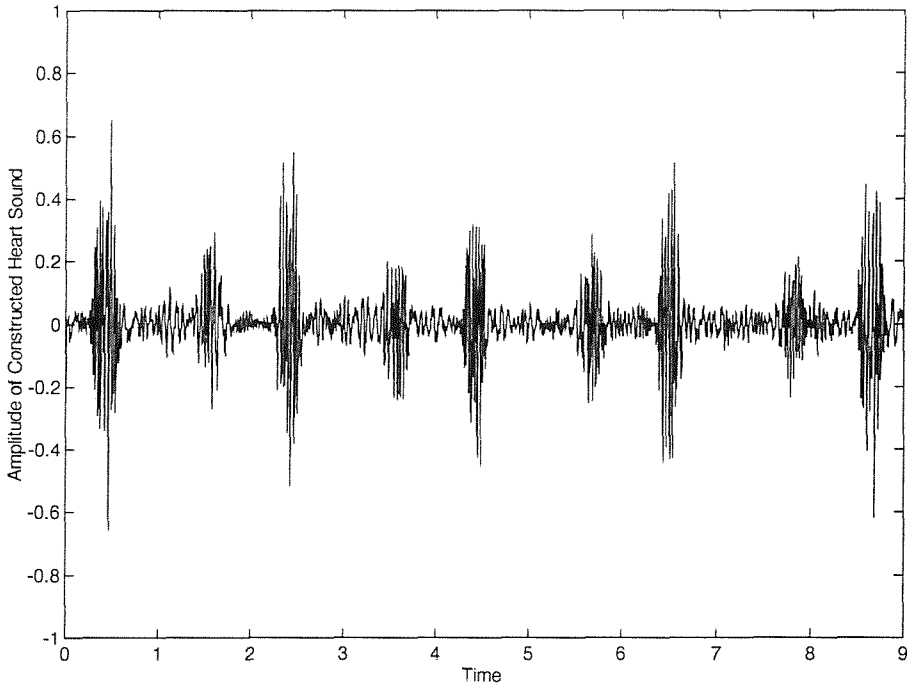

Figure 97 Time Series of Temporally Extended Heart Beat with an ASD

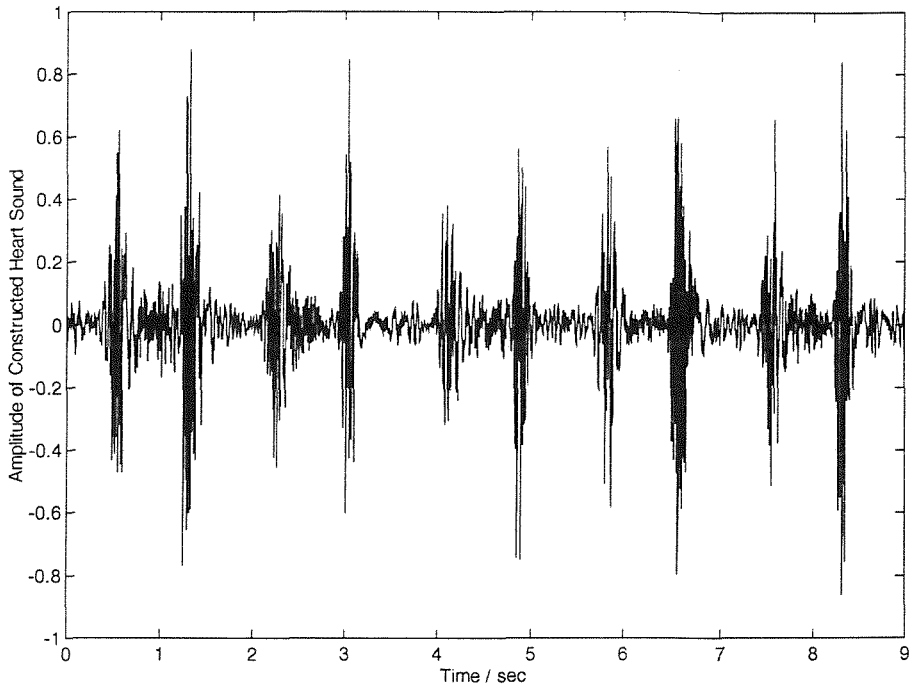

Figure 98 Time Series of Temporally Extended Heart Beat with a VSD

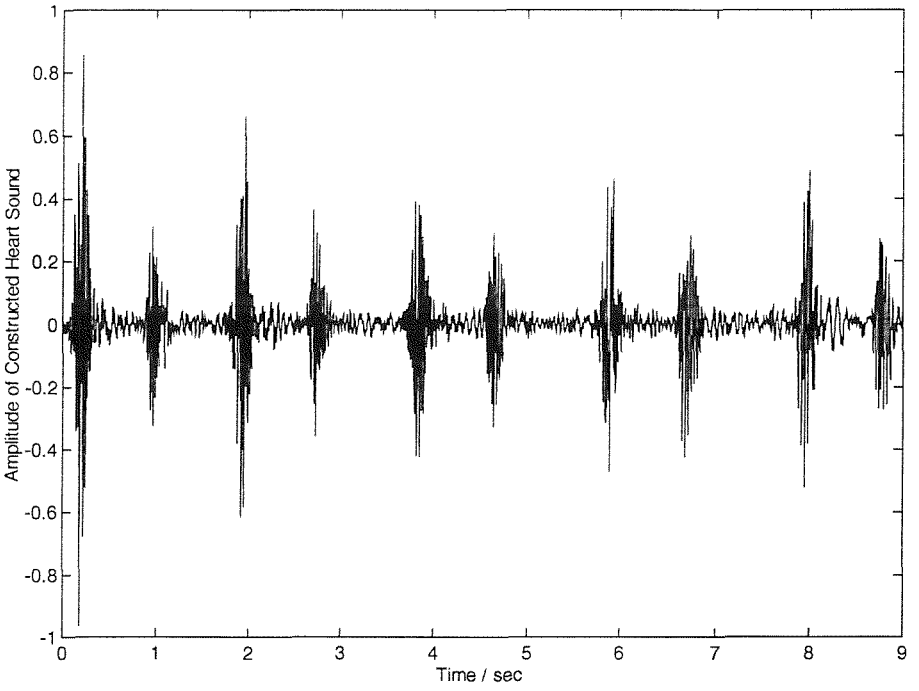

Figure 99 Time Series of Temporally Extended Heart Beat with an Innocent Murmur

7.4 Reconstruction of an Averaged Murmur via the Spectrogram

The human body is a complex structure. As a result, the biological signals generated vary both between patients and over relatively short time scales, for example heart beat to heart beat. The dynamics of any heart murmur present alters in each cardiac cycle and thus any TFR generated will also change according to cardiac cycle being processed. With a view to improving diagnosis of the presence and type of a murmur, a technique to reduce the variability of the signal is desired.

One technique used to reduce the variability is to average in the TF domain over a number of cardiac cycles [White97, Leung98, Leung99]. Averaging in the time-domain is not possible because the signals are random in nature [Beyar84]. Each murmur is a different realisation of a zero mean random process. It follows that simply averaging in the time-domain, the different initial phase of each murmur will result in destructive interference, cancelling out the murmur waveform. A quadratic (or energetic) TFR is not affected by the random nature of the signal and therefore averaging in the TF domain can be performed without cancellation of signal terms. Energetic TFRs are not affected by the random nature of the signals because although each realisation of the murmur has a different starting phase, it has the same energy content in TF. Leung [Leung99] demonstrates this process for a range of different averaging methods and TFRs.

A recording of a series of heart beats from a patient is split in order to separate complete cardiac cycles. The spectrogram is then computed for each of these signals, although other quadratic TFRs have been shown to offer improved visualisation characteristics [Leung99]. The spectrogram for each cardiac cycle is then synchronised so that the first heart sound (S1) is at time zero. The averaged spectrogram is then computed by taking the mean of the spectrograms of each cardiac cycle. Alternative smoothing operations can also be used [Leung99]. The averaging operation is given in (7-1) where the averaged spectrogram $\left|S_a(t,\omega)\right|^2$ is given in terms of spectrograms computed for each cardiac cycle $\left|S_n(t,\omega)\right|^2$. It is unlikely that the averaged spectrogram will be valid spectrogram.

$$\left|S_a(t,\omega)\right|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\left|S_n(t,\omega)\right|^2 \tag{7-1}$$

Using averaged spectrogram, the time series of a typical example of the cardiac cycle can be reconstructed. Examining the average spectrogram between S1 and S2, allows the structure of the murmur to be displayed with greater contrast. Furthermore use of only this part of the signal enables construction of a typical murmur signal, without S1 and S2 being present in the reconstruction.

The averaged spectrogram (area between S1 and S2) computed over 20 cardiac cycles, for the three types of murmurs previously discussed are presented in Figure 100 to Figure 102. In these images the effects of the anti-aliasing filter used in the sampling process can clearly be seen, as the energy content present in the signal fades as it reaches 1000Hz (half the sampling rate). The characteristics of each of the murmur can be seen to be different from one another, with the innocent murmur being almost tonal in nature compared to the ASD and VSD murmurs. The VSD murmur seems to possess a greater amount of energy above 300 Hz compared to the ASD. These averaged spectrograms were then used as the input to the signal reconstruction algorithm based upon the MLS inversion of the ST-FT [Griffin83, 84]. The converging total error is shown in Figure 103 to Figure 105. Since the averaged spectrogram was a synthetic input, a large number of iterations (2500) was used to attempt to get a good reconstruction. That said, the total error is still much higher than those achieved for the temporal extended spectrograms used in the previous section, with a typical sum of the squared error above 1×10^4 for the ASD and innocent murmurs and above 1×10^6 for the VSD murmur. Since the signals are of unequal length, more revealing statistics may be the mean, maximum and minimum squared errors, these are given for the three murmurs in Table 7. The reason for the high errors may be due to the fact that the reconstruction is using an averaged version of the murmur, and thus further removed from a valid spectrogram.

Error Statistic	ASD Murmur	VSD Murmur	Innocent Murmur
Sum of Squared Error	15900	762650	29700
Mean Squared Error	53	1885	107
Maximum Squared Error	656	2357	1040
Minimum Squared Error	7.6e-9	2.5e-7	1.8e-8

Table 7 Error values for the three Synthetic Murmur Types after the 2500 Iterations

Figure 106 to Figure 108 show the spectrogram of the reconstructed signals. There is little visual difference between the desired and reconstructed spectrograms. The time series used to create these plots can be thought of as being the average murmur, and these are given in Figure 109 to Figure 111. The difference in the power of the murmurs can be seen in the different maximum amplitudes present for each of the signals, the VSD (Figure 110) being the loudest as expected, since prior to any signal processing it is the most audible. The VSD murmur has a peak value of ~40dB across a broad range of frequencies, as compared to the ASD which only reaches ~25dB and the innocent murmur which ranges ~35dB but in a narrow range of frequencies. In the innocent murmur case (Figure 111), its tonal nature can also be seen in the time series, as a periodic component starting at 0.2sec and lasting approximately 0.5sec.

As before, although the time series show the signal, the differences between the three signals can be more easily heard than visualised in the time series. These files are found on the enclosed disc in the '\heart\averaged murmur' directory. Judgement of the aural quality of these signals is hindered because of the short duration of the murmur signals. To alleviate this, five repetitions of each signal, each separated by a short-break, are given on each file. The aural differences in the spectral content between the VSD (high frequency), ASD and innocent murmurs (low frequency) can now be clearly heard.

7.5 Conclusions

With a view to improving the diagnostic ability of the physician by offering various audio representations of the cardiac cycle, two different approaches based upon the reconstruction of a signal from a synthetic spectrogram have been presented. In terms of improving the diagnosis, the temporally extended version of the cardiac cycle overall prove to be more useful overall, owing to the physician familiarity with the original signal. The processing power required to compute the extended time series is reasonably small, such that a small personal computer can perform in a few seconds. However the averaged spectrogram proved to be a useful tool in the automatic classification of heart murmurs [Leung99], and could be used for the purposes of training.

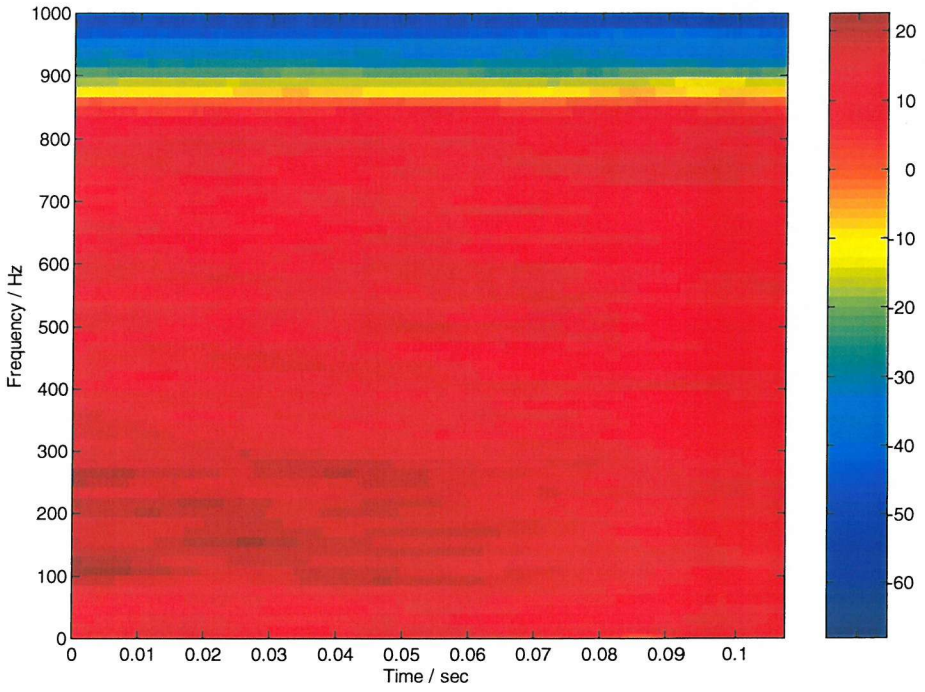

Figure 100 Averaged Spectrogram for the Cardiac Cycle (between S1 and S2) – ASD Murmur

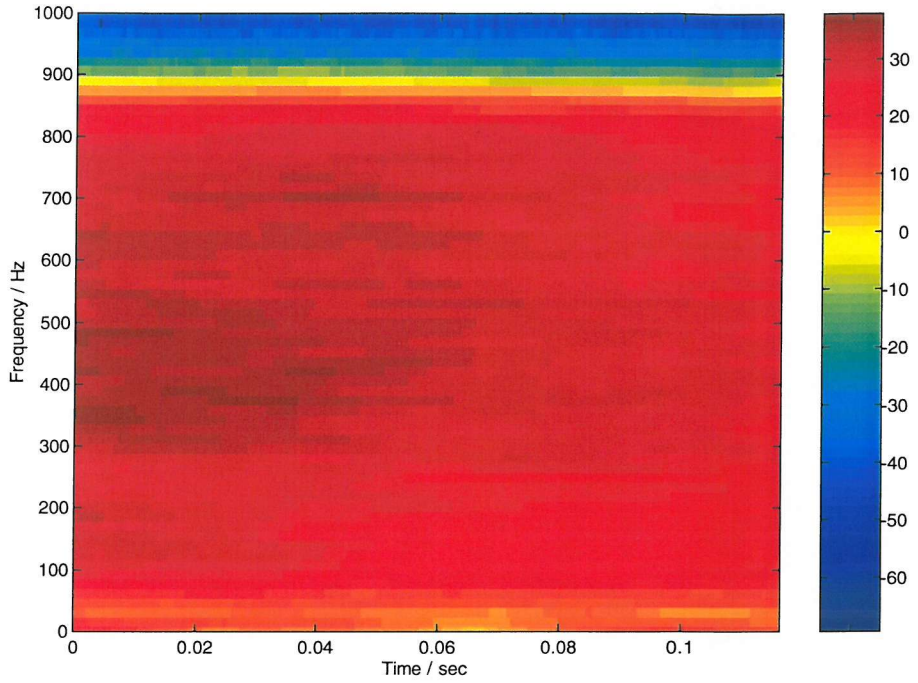

Figure 101 Averaged Spectrogram for the Cardiac Cycle (between S1 and S2) – VSD Murmur

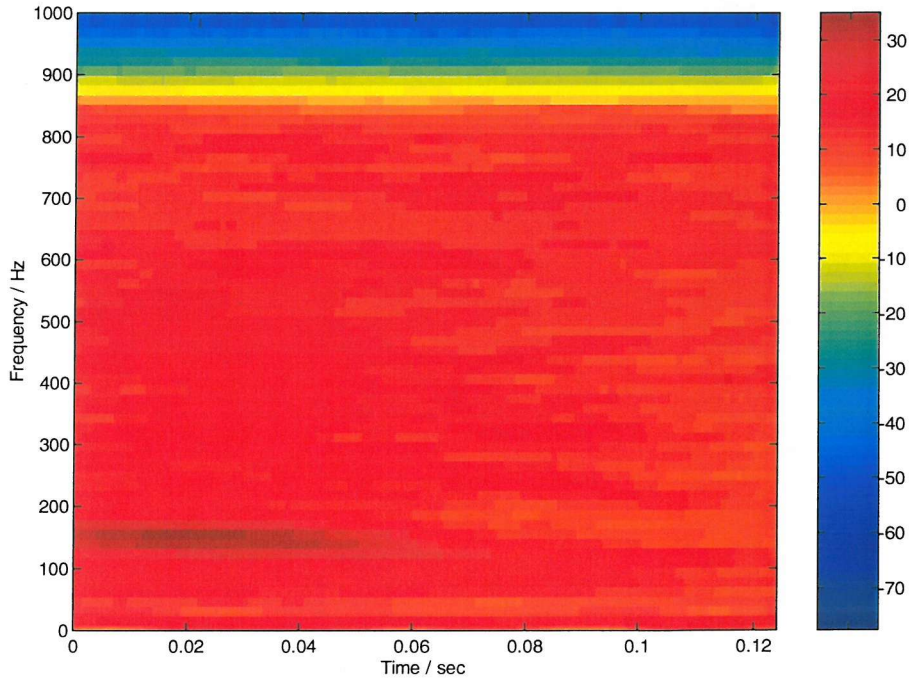

Figure 102 Averaged Spectrogram for the Cardiac Cycle (between S1 and S2) – Innocent Murmur

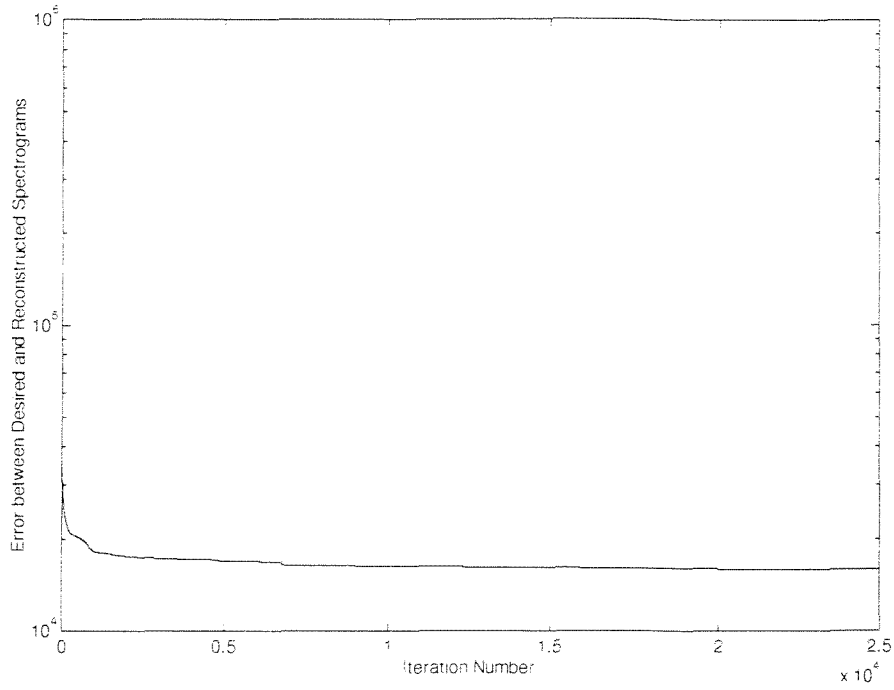

Figure 103 Convergence of Distance Measure between the Desired and the Reconstructed Averaged Spectrograms – ASD Murmur

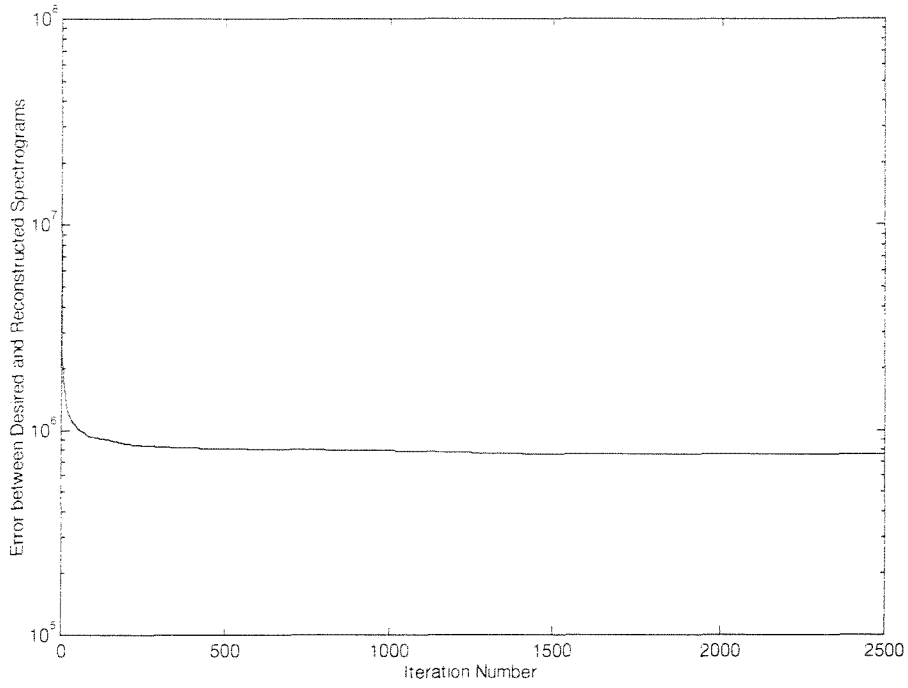

Figure 104 Convergence of the Distance Measure between the Desired and Reconstructed Averaged Spectrograms – VSD Murmur

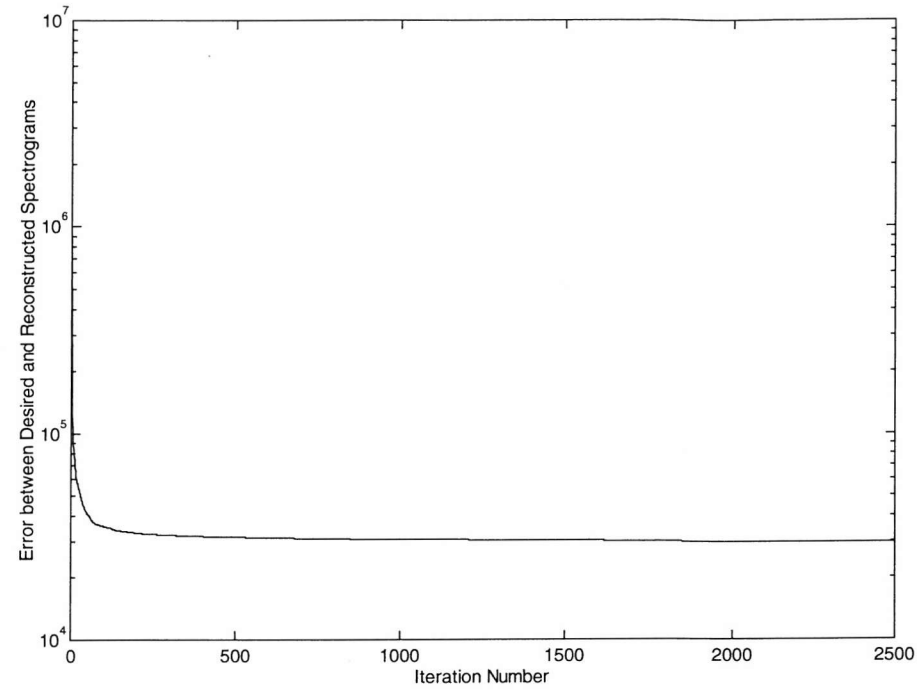

Figure 105 Convergence of the Distance Measure between the Desired and Reconstructed Averaged Spectrograms –Innocent Murmur

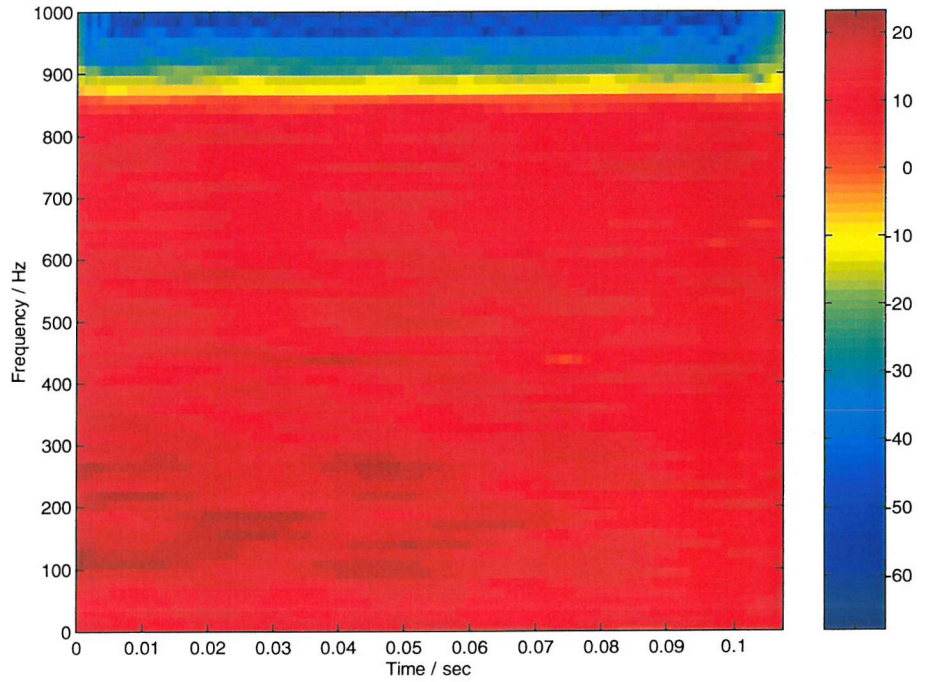

Figure 106 Reconstructed Spectrogram of the Cardiac Cycle (between S1 and S2) – ASD Murmur

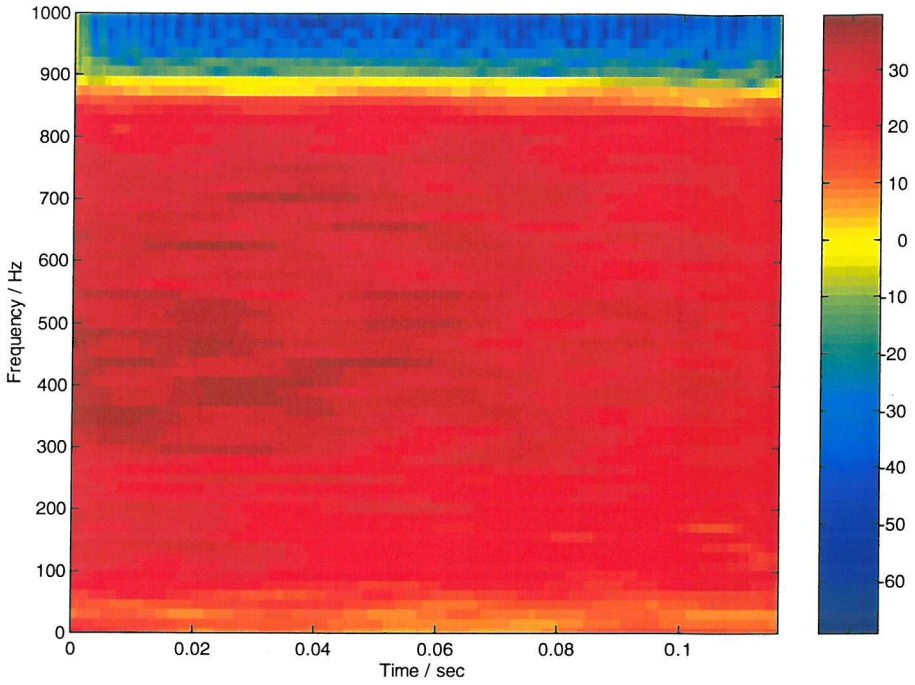

Figure 107 Reconstructed Spectrogram of the Cardiac Cycle (between S1 and S2) – VSD Murmur

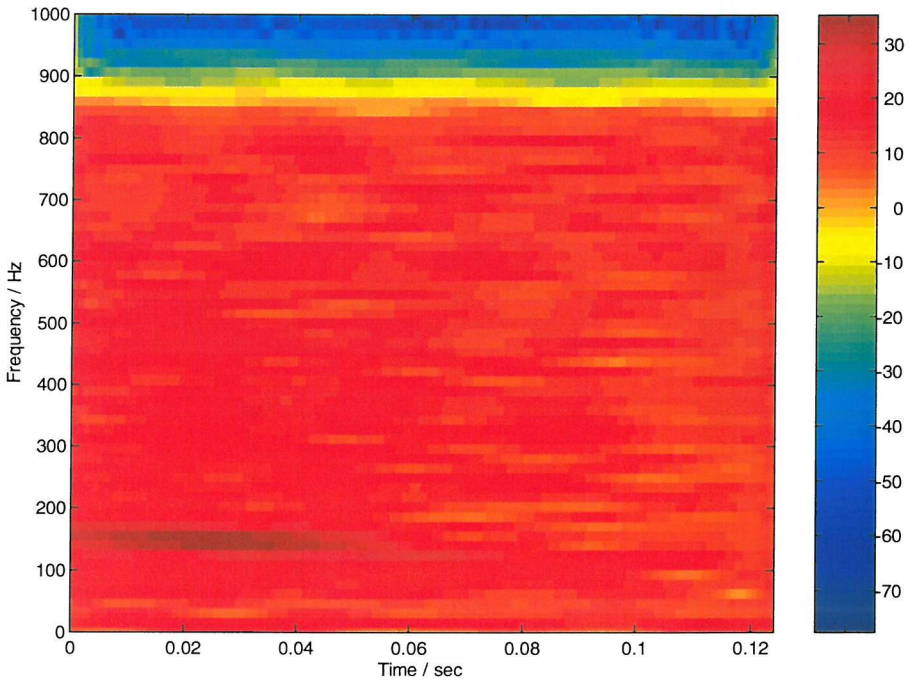

Figure 108 Reconstructed Spectrogram of the Cardiac Cycle (between S1 and S2) – Innocent Murmur

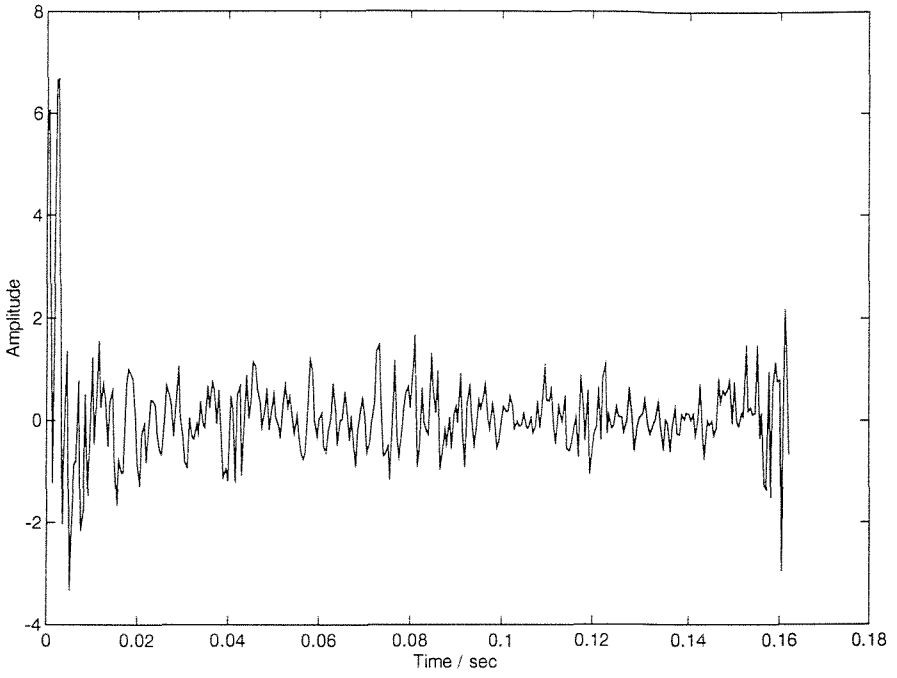

Figure 109 Reconstructed "Average" ASD Murmur

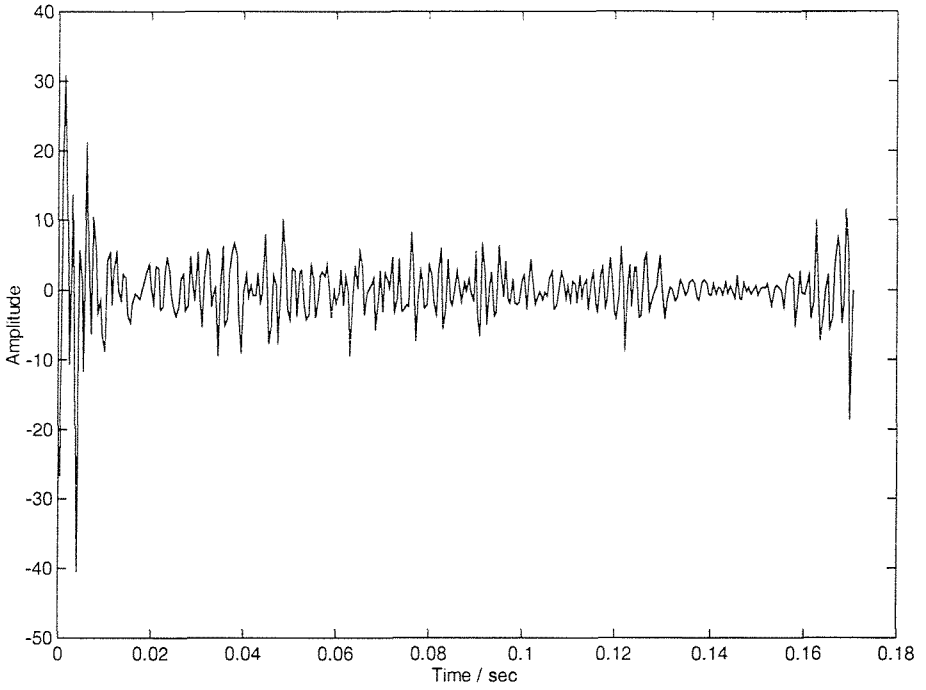

Figure 110 Reconstructed "Average" VSD Murmur

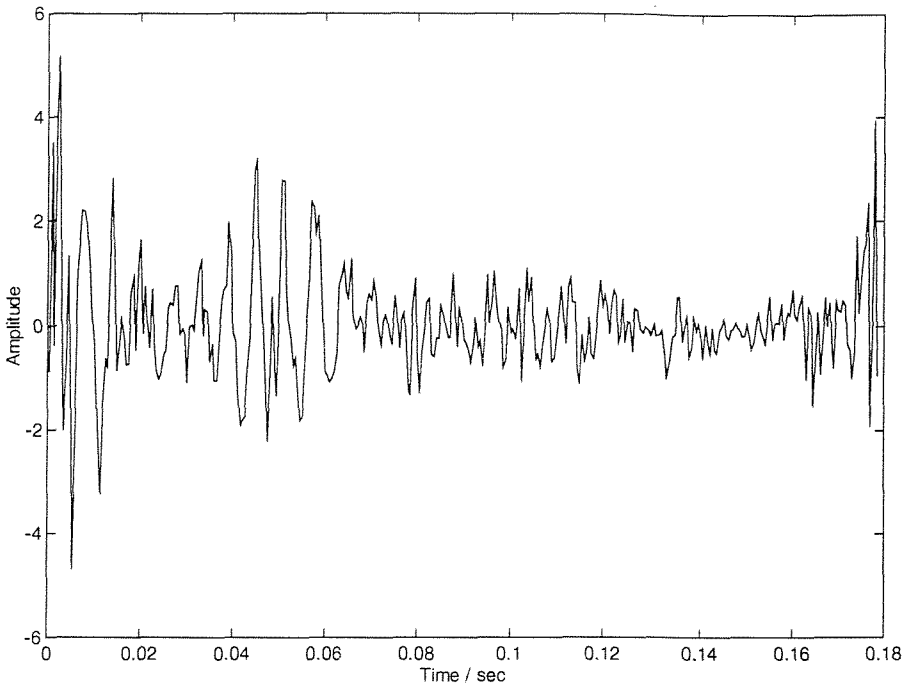

Figure 111 Reconstructed "Average" Innocent Murmur

Chapter 8 Conclusions and Future Work

8.1 Summaries and Discussions

In Chapter 2 the concepts of spectral analysis and TF representations for 1-D signals were reviewed. The differences between stationary and non-stationary, and stochastic and deterministic signals, were outlined. TFRs are one way in which the time-varying spectrum of a non-stationary signal can be displayed. The first TFR presented was the ST-FT, a temporally limited version of the FT. The ST-FT is a complex distribution, but has the benefit of being linear. The other commonly used linear TFR is the WT. Since the linear TFRs are complex valued their (squared) magnitudes are often plotted. The squared magnitudes of the ST-FT and WT are considered to be TFRs in there own right and are called the Spectrogram and Scalogram respectively. The Spectrogram can be considered to be a member of a wider set of bilinear TFRs called Cohen's Class. All these TFRs are bilinear with respect to the signal, and all suffer from a compromise between good local resolution and the presence of cross-terms, visually distracting components present due to the bilinearity.

In Chapter 3 signal dependant TFRs were presented. Three distinct methods were considered: signal dependant kernel design, Cohen-Posch and Reassigned TFRs. Although easily defined in the Ambiguity Plane (AP) and invertible, data adaptive kernel design was not explored further since it required user intervention for the specification of the volume of the kernel. Cohen-Posch TFRs conform to the intuitive requirements for an energetic TFR, specifically positivity and correct marginals. However their performance for many classes of signals is poor. This is due to CP TFRs being sensitive to the orientation of the signals in the TF plane. The reassigned TFR approach does not suffer from this limitation, and the reassigned spectrogram is guaranteed to be positive. However the reassigned TFR does not, in general, conform to the marginal properties. The computation of a reassigned distribution is a two-stage process. It requires first, the computation of the 'reassigned co-ordinates', a set of new co-ordinates dependant upon the local moments in time and

frequency of the distribution. These new co-ordinates reallocate the energy of the spectrogram to a position which is less affected by the time-limiting windowing function used in its computation. The second stage is the reallocation of the energy within the bandwidth of the window to this new location. The reassigned spectrogram was shown to be linked to an instantaneous frequency extraction algorithm called the 'Ridge and Skeleton' approach. It is also related to the partial derivatives of the phase of the ST-FT with respect to time and frequency.

With a view to constructing an inverse to the reassigned spectrogram, the link between the phase of the ST-FT and the reassigned co-ordinates was explored in Chapter 4. It was shown that, using the reassigned group-delay parameter (computed using finite differences), the phase of the ST-FT can be recovered to within a constant per time-slice. The chapter proceeds to describe signal reconstruction from phase, first reviewing existing techniques for signal reconstruction from the phase of the DFT, and then deriving techniques for signal reconstruction from the phase of the ST-FT. Two approaches to achieving this goal are presented, one reconstructing the signal in a piece-wise iterative manner, the other recovering the complete signal in a single step. The global approach was found to be superior to the piece-wise approach, since it did not require any *a priori* signal knowledge and was robust to numerical errors. Through knowledge of the reassigned group-delay parameter, and application of the global reconstruction algorithm it was shown that a real signal can be reconstructed. In the case of a complex signal, the reassigned group-delay does not provide sufficient information for global reconstruction. However a piece-wise algorithm can be determined.

The dual of signal reconstruction from the phase of the ST-FT is reconstruction from the modulus, as described in Chapter 5. Unlike the phase case, reconstruction from modulus (or squared modulus) of the ST-FT has been previously described in the literature. Chapter 5 reviews those approaches, including a Minimum Least Squares (MLS) based approach. The MLS algorithm reconstructs the signal to within an overall phase constant. A new approach to the problem is also presented. This takes inspiration from the method used to formulate reconstruction from phase, defining the squared modulus of a single point in TF in terms of a quadric form. The extension from one point in TF to all points is not simple, and the resultant formula is non-linear in terms of the signal. A fixed point and Newton-Raphson algorithms are used to solve the system of equations. Although Newton's algorithm is

guaranteed to converge close to the solution, it offers no global convergence characteristics. The fixed-point algorithm is not guaranteed to converge, but did so for valid spectrograms in fewer iterations than Newton's algorithm. However, both approaches require much greater computational resources, in terms of both processing power and memory than the MLS algorithm. Therefore for all but the shortest signals, the MLS algorithm is considered to be superior.

The work for signal reconstruction from partial TF information given in Chapters 4 and 5, is generalised in Chapter 6 to cover a wide range of different linear TFRs. A generalised version of the WT (GWT) is presented, which contains both the ST-FT and the WT as members. A MLS reconstruction equation is defined to recover a signal from full (modulus and phase) GWT information. This algorithm requires the computation of a signal-independent kernel matrix and its inverse, a potentially time consuming procedure for long signals. Generalising the algorithms in Chapter 4, signal reconstruction from the phase of the GWT is presented. As with the ST-FT case, signal reconstruction takes place to within an overall amplitude constant. As in the ST-FT case given in Chapter 5, the algorithm for signal reconstruction from the modulus of the GWT is based upon the MLS inversion. Therefore, the procedure suffers from the same limitation as the MLS inversion for the GWT, requiring the computation of the inverse of the kernel matrix. However this matrix need only be computed once for a given length of signal and set of decomposing wavelets.

Two applications of signal construction from synthetic spectrograms are presented in Chapter 7 both involve processing heart sounds. In the first application, heart sounds are temporally extended while maintaining their spectral characteristics, using modifications of the spectrogram of the heart sound. The temporal extension aids in the diagnosis of heart murmurs. Using the MLS based algorithm, a signal was reconstructed from a modified spectrogram. Unlike a parametric based technique described in the literature, the computation required is not dependant upon the type of signal present. The second application reconstructs an average murmur from the averaged spectrogram, an operation not possible purely in the time domain due to the random nature of the flow noise generating the murmur.

8.2 Further Work

A number of further areas for research have presented themselves over the course of this thesis. An overview of three of these is given here.

8.2.1 Robust Signal Reconstruction from the Phase of the GWT

Although a method for signal reconstruction from the phase of the ST-FT (Chapter 4) and GWT (Chapter 6) has been described, these algorithms assume a valid (non-modified or synthetic) phase distribution. If an invalid phase is used, then these algorithms yield poor performance. In order to improve the robustness of signal reconstruction from the magnitude of a linear TFR (Chapter 5 for ST-FT, Chapter 6 for GWT) the solution was found which minimised the squared error between the desired and reconstructed signal. An analogous expression can be written for signal reconstruction from phase, as given in (8-1), where m is the time slice number, u is the frequency (or scale) variable, $T(u)$ is the number of time-slices present for a given frequency, U is the number of decomposing frequencies, $\hat{\phi}[m,u]$ is the desired phase and $\phi[m,u]$ is the phase of the GWT for the signal x.

$$\min_{x}\left(\sum_{u=0}^{U}\sum_{mn=0}^{T(u)}\left(\phi[m,u]-\hat{\phi}[m,u]\right)^2\right) \tag{8-1}$$

The minimisation of (8-1) with respect to the input signal allows for synthetic or modified phase distributions to be used, producing a signal that has a phase distribution closest to that desired.

The reassigned co-ordinates for the spectrogram can be viewed as the derivative of the phase with respect to the time and frequency variables used. Modifications of the reassigned co-ordinates could now take place to construct a signal with desired characteristics. Alternatively a signal could be constructed which possess the qualities defined by a purely synthetic set of reassigned co-ordinates.

8.2.2 Signal Reconstruction from the Reassigned Co-ordinates

In Chapter 4 the link between the phase of the ST-FT and the reassigned co-ordinates (as defined by a numerical difference) was presented. Although simply defined, this method of computation of the reassigned co-ordinates suffers from a number of limitations. These include the need for phase unwrapping, and poor reconstruction when a non-maximally overlap ST-FT is used. Two alternatives for computing the reassigned co-ordinates have been presented in the literature, an analytical [Auger95] and a recursive [Richard97] method.

Although simply defined in terms of the ST-FT, neither of these methods posses a simple inversion. Unlike the difference method employed upon the phase of the discrete ST-FT, these approaches compute a sampled version of the derivative of the phase of the continuous ST-FT. The link between this sampled continuous version and the fully discrete version (as derived by numerical differences) should be explored. If no relationship exists, then an alternative is to attempt to derive a method of allowing signal reconstruction from the discrete samples of continuous partial derivative.

8.2.3 Fast Techniques for Inverse Kernel Computation

In Chapter 6 a MLS method of signal reconstruction from the GWT was presented. This method relied on the computation of a kernel function, which was dependant upon the number and type of decomposing wavelets used. The time taken to compute this kernel can be reduced by using Fourier based methods as described in Chapter 6. The inverse of the matrix was computed numerically using a pseudo-inverse algorithm. No attempt was made to compute this inverse matrix directly, rather than via numerical inversion. An examination of the structure of the kernel may allow analytical computation of its inverse, if there is sufficient structure. Direct computation of the kernel inverse would allow this method to be used for longer signals than is currently possible.

Appendix 1 A Proof of Convexity

In order to prove that a Projection onto Convex Sets (POCS) algorithm will converge to a Cohen-Posch TFR, it sufficient to prove that the sets defined by all positive TFRs, and TFRs conforming to the time marginal and TFRs conforming to the frequency marginal are convex.

A set C is defined to be convex if for any two points $x_1, x_2 \in C$, the point $x_3 = \sigma x_1 + (1-\sigma) x_2$ also lies in C for any value of the parameter σ in the range $0 \leq \sigma \leq 1$.

An arbitrary TFR for a given input signal $x(t)$ is defined as $P_x(t,\omega)$. The set of TFRs which conform to the time marginal for a given input signal are given in [A1]. The set of TFRs which conform to the frequency marginals are given in [A2], where the FT of the input signal is denoted as $X(\omega)$ (Chapter 2 (2-1a)).

$$T_m = \left\{ P_x \mid \int P_x(t,\omega)\, d\omega = \left| x(t) \right|^2 \right\} \qquad \text{[A1]}$$

$$\Omega_m = \left\{ P_x \mid \int P_x(t,\omega)\, dt = \left| X(\omega) \right|^2 \right\} \qquad \text{[A2]}$$

A proof that the sets defined by the time and frequency marginals and positivity are all convex [Clarkson95].

Time Marginal

Define two arbitrary TFRs, P_{x_1} and P_{x_2} that are members of the set of TFRs conforming to the time-marginal as defined in [A1]. In order to prove that this set is convex, it is sufficient to prove that a new TFR, P_{x_3} expressed in [A3] conforms to the time marginal. Simplifying [A3] in terms of the energy per unit time of the signal

leads directly to [A4]. Therefore the TFR $P_3(t,\omega)$ conforms to the time-marginal and thus the set is convex.

$$\int P_{x_3}(t,\omega)d\omega = \sigma\int P_{x_1}(t,\omega)d\omega + (1-\sigma)\int P_{x_2}(t,\omega)d\omega \qquad [A3]$$

$$\therefore \int P_{x_3}(t,\omega)d\omega = \sigma|x(t)|^2 + (1-\sigma)|x(t)|^2 = |x(t)|^2 \qquad [A4]$$

Frequency Marginal

The proof that the frequency marginal also defines a convex set follows the proof for the time marginal. Defining the frequency marginal equivalent of [A3], in [A5], equation [A6] can be shown. Thus, since the TFR $P_3(t,\omega)$ also conforms to the required frequency-marginal and the set is convex.

$$\int P_{x_3}(t,\omega)dt = \sigma\int P_{x_1}(t,\omega)dt + (1-\sigma)\int P_{x_2}(t,\omega)dt \qquad [A5]$$

$$\therefore \int P_{x_3}(t,\omega)dt = \sigma|X(\omega)|^2 + (1-\sigma)|X(\omega)|^2 = |X(\omega)|^2 \qquad [A6]$$

Positivity

Define the following variables: $P_{x_1}, P_{x_2} \in R^+$, two arbitrary positive TFRs. In order to prove convexity we need to prove that $P_{x_3} \in R^+$ as defined in [A7] where $0 \le \sigma \le 1$.

$$P_{x_3} = \sigma P_{x_1} + (1-\sigma)P_{x_2} \qquad [A7]$$

Due to the limits placed upon σ, it is clear that $(1-\sigma) \in R^+$ and thus $P_{x_3} \in R^+$. Thus the set defined by positivity is convex.

Therefore all three sets are convex and thus an POCS algorithm using membership of these three sets as constraints will converge.

Appendix 2 Proof of Monotonically Decreasing Error of Iterative Signal Reconstruction from Magnitude of Wavelet Transform

This appendix shows the following:

1. The signal recovered from the modulus of its Generalised Wavelet transform (GWT) is closed and bounded

2. That the error between successive iterations is decreases

The proof stops short of the showing that the algorithm will converge to a global minimum, merely showing that the error between successive errors decreases. The proof closely follows that of Griffin [Griffin84].

In the following derivation the input signal is assumed to be real. The algorithm is developed for the real and imaginary parts of the signal independently, this results in no loss of generality since the procedure can be repeated for the imaginary part of the signal.

Previously (Chapter 6) reconstruction of a signal from the modulus GWT used formulated in terms of the repeated application of a matrix multiplication. For the definitions of the matrices, see Chapter 6, Section 6.5.1.3. In simplified form the reconstruction procedure can be seen as repeated application of [A8], where $\mathbf{X}^p$ is the GWT matrix of the vector $\mathbf{x}^i$, and $Y[m,n]$ is the given or desired GWT modulus.

$$\mathbf{x}^{p+1} = \mathbf{K}^{-1}\mathbf{Y}\mathbf{h}^p \qquad [A8]$$

$$\mathbf{Yh}^p = \sum_{u=0}^{U-1} \sum_{m=0}^{P(u)-1} \hat{Y}^p[m,u] h[n - g(m), u]$$

$$\hat{Y}^p[m,u] = |Y[m,u]| \frac{X^p[m,u]}{|X^p[m,u]|}$$

Let L be an algorithm on $\square^N$, and suppose that, given an initial signal estimate $\mathbf{x}^0$, the sequence $\{\mathbf{x}^p\}_{p=0}^{\infty}$ is created satisfying

$$\mathbf{x}^{p+1} = L[\mathbf{x}^p] \qquad [A9]$$

Given that the algorithm can be written in terms of the matrix formulation [A9], the system is both closed and bounded for if $\mathbf{K}^{-1}$ exists, and if it, and $\mathbf{Yh}^p$ are finite.

The discrete distance measure between the GWT and the desired GWT minimised in the MLS case, is given in [A10].

$$D[\mathbf{x}^p, Y[m,u]] = \sum_u \sum_m |X^p[m,u] - Y^p[m,u]|^2 \qquad [A10]$$

In the case of reconstruction from modulus, as expressed in [A8] at each iteration the recovered GWT $\mathbf{X}^p$ is replaced with $\hat{\mathbf{Y}}^p$. Using the distance measure [A10] $\hat{\mathbf{Y}}^p$ minimises $D[\mathbf{x}^p, \hat{Y}^p[m,u]]$ for $\mathbf{x}^p$ fixed and $\hat{\mathbf{Y}}^p$ constrained to have magnitude $|Y[m,u]|$. Therefore

$$D[\mathbf{x}^p, \hat{Y}^p[m,u]] \leq D[\mathbf{x}^p, \hat{Y}^{p-1}[m,u]] \qquad [A11]$$

and

$$D[\mathbf{x}^{p+1}, \hat{Y}^{p+1}[m,u]] \leq D[\mathbf{x}^{p+1}, \hat{Y}^p[m,u]] \qquad [A12]$$

Through the application of [A8] $\hat{\mathbf{Y}}^p$ allows estimation of the next signal $\mathbf{x}^{p+1}$. Since the mathematics was derived to minimise $D[\mathbf{x}, \hat{Y}[m,u]]$ for $\hat{\mathbf{Y}}^p$ fixed, it follows that

$$D\left[\mathbf{x}^{p+1},\hat{Y}^{p}\left[m,u\right]\right]\leq D\left[\mathbf{x}^{p},\hat{Y}^{p}\left[m,u\right]\right] \quad [A13]$$

With equality holding if and only if $\mathbf{x}^{p+1}=\mathbf{x}^{p}$. Combining [A12] and [A13] together produces [A14], again with equality holding if and only if $\mathbf{x}^{p+1}=\mathbf{x}^{p}$.

$$D\left[\mathbf{x}^{p+1},\hat{Y}^{p+1}\left[m,u\right]\right]\leq D\left[\mathbf{x}^{p},\hat{Y}^{p}\left[m,u\right]\right] \quad [A14]$$

Writing the expression for $D\left[\mathbf{x},\hat{Y}\left[m,u\right]\right]$ explicitly as given in [A15], reduces to [A16].

$$D\left[\mathbf{x}^{p},\hat{Y}^{p}\left[m,u\right]\right]=\sum_{u}\sum_{m}\left|X^{p}\left[m,u\right]-\left|Y\left[m,u\right]\right|\frac{X^{p}\left[m,u\right]}{\left|X^{p}\left[m,u\right]\right|}\right|^{2} \quad [A15]$$

$$D\left[\mathbf{x}^{p},\hat{Y}^{p}\left[m,u\right]\right]=\sum_{u}\sum_{m}\left(\left|X^{p}\left[m,u\right]\right|-\left|Y\left[m,u\right]\right|\right)^{2}=D_{M}\left[\mathbf{x}^{p},Y\left[m,u\right]\right] \quad [A16]$$

This leads to the distance measure based upon the GWT magnitude, $D_{M}\left[\mathbf{x}^{p},Y\left[m,u\right]\right]$ given in Chapter 6 Section 6.5.1.3. Since $D\left[\mathbf{x}^{p},\hat{Y}^{p}\left[m,u\right]\right]=D_{M}\left[\mathbf{x}^{p},Y\left[m,u\right]\right]$ reapplication of [A14] yields [A17].

$$D_{M}\left[\mathbf{x}^{p+1},Y\left[m,u\right]\right]\leq D_{M}\left[\mathbf{x}^{p},Y\left[m,u\right]\right] \quad [A17]$$

This expression shows that distance measure based upon the GWT magnitude decreases at each iteration.

References

[Ackroyd71] M. Ackroyd 1971 *Journal of Acoustical Society of America* **50**, 1229-1231. Short-Time Spectra and Time-Frequency Energy Distributions

[Allen77] J. Allen 1977 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **25**(3), 235-238. Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform

[Anderson85] J. Anderson 1985 *ICASSP-85*, 533-536. A Filter/Detector Interpretation of the Short-Time Fourier Transform Magnitude

[Anderson92] J. Anderson 1992 *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* 277-280. A Spectral Magnitude Analysis Theorem and Applications

[Anderson93] J. Anderson 1993 *IEEE Transactions on Signal Processing* **41**(12) 3541-3543. A Wavelet Magnitude Analysis Theorem

[Anderson94a] J. Anderson 1984 *IEEE-SP DSP Workshop, Chatham, MA.* Speech Analysis/Synthesis based on Perception

[Anderson94b] J. Anderson 1994 *Sixth IEEE Digital Signal Processing Workshop, New York,* 297-300. Complex Signal Reconstruction from, Time-Frequency Magnitude

[Anton91] H. Anton, 1991 *Elementary Linear Algebra*, Wiley, London

[Auger95] F. Auger and P. Flandrin 1995 *IEEE Transactions on Signal Processing* **43**(5), 1068-1089. Time-Frequency and Time-Scale Representations by the Reassignment Method

[Baraniuk93] R. Baraniuk and D. Jones 1993 *IEEE Transactions of Signal Processing* **41**(4), 1589-1602. A Signal-Dependant Time-Frequency Representation: Optimal Kernal Design

[Baraniuk94] R. Baraniuk and D. Jones 1994 *IEEE Transactions on Signal Processing* **42**(1), 134-146. A Signal-Dependent Time-Frequency Representation: Fast Algorithm for Optimal Kernal Design

[Beyar84] R. Beyar, S. Levkovitz, S. Braun, and Y. Palti 1984 *IEEE Transactions in Biomedical Engineering* **31**(9), 591-596. Heart-Sound Processing by Averaging and Variance Calculation – Physical Basic and Clinical Implications

[Boashash92] edited by B. Boashash 1992 *Time-Frequency Signal Analysis Methods and Applications*, Longman and Cheshire, New York

[Bendat86] J. Bendat and A. Piersol 1986 *Random Data*, John Wiley & Sons, New York.

[Chassande97] E. Chassande-Mottin, I. Daubechies, F. Auger and P. Flandrin 1997 *IEEE Signal Processing Letters* **4**(10), 293-294. Differential Reassignment

[Clarkson95] Edited by P. M. Clarkson and H. Stark 1995 *Signal Processing Methods for Audio, Images and Telecommunications*, Academic Press, London

[Cohen79] L. Cohen and Y. Zaparovanny 1979 *Journal Mathematical Physics* **21**(4), 794-796. Positive Quantum Joint Distributions

[Cohen85] L. Cohen and T. Posch 1985 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **33**(1), 31-37. Postive Time-Frequency Distribution Functions

[Cohen95] L.Cohen 1995 *Time-Frequency Analysis*, Prentice-Hall, New Jersey

[Collar87] A. Collar and A. Simpson 1987 *Matrices and Engineering Dynamics*, Ellis Horwood Limited

[Crochiere80] R. Crochiere 1980 *IEEE Transaction on Acoustics, Speech, and Signal Processing* **28**(2), 99-102. A Weighted Overlap-Add Method of Short-Time Fourier Analysis/Synthesis

[Delprat92] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian and B. Torresani 1992 *IEEE Transactions on Information Theory* **38**(2), 644-664. Asymptoic Wavelet and Gabor Analysis: Extraction of Instantaneous Frequencies

[Durand95] L. Durand and P. Pibarot 1995 *Critical Reviews in Biomedical Engineering* **23**(3/4), 163-219. Digital Signal Processing of the Phonocardiogram: Review of the Most Recent Advancements

[Francos93] A. Francos and M. Porat 1996 *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* 321-324. Parametric Estimation of Multicomponent Signals using Minimum Cross Enropy Time-Frequency Distributions

[Fonollosa96] J. Fonollosa 1996 *IEEE Transactions on Signal Processing* **48**(8), 2086-2091. Positive Time-Frequency Distributions Based on Joint Marginal Constraints

[Gabor46] D. Gabor 1946 *Journal of the IEE* **93**(3), 429-457. Theory of Communication

[Golub90] G. Golub and C. Van Loan 1990 *Matrix Computations*, John Hopkins, Oxford

[Griffin83] D. Griffin and J. Lim 1983 *IEEE ICASSP-83* 804-807. Signal Estimation from Modified Short-Time Fourier Transform

[Griffin84] D. Griffin and J. Lim 1984 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **32**(2), 236-243. Signal Estimation from Modified Short-Time Fourier Transform

[Guillemain91] P. Guillemain, R. Kronland-Martinet and B. Torresani, *Wavelets and Applications*, pp. 38-60. Ed. Meyer 1991 Masson: Springer-Verlag, New York

[Guillemain96] P. Guillemain and R. Kronland-Martinet 1996 *Proceedings of the IEEE* **84**(4), 561-585. Characterization of Acoustics Signal through Continuous Linear Time-Frequency Representations

[Hammond96] J. K. Hammond and P. R. White 1996 *Journal of Sound and Vibration* **190**(3), 419-447. The Analysis of Non-Stationary Signals using Time-Frequency Methods

[Harris76] A. Harris, G. Sutton and M. Towers 1976 *Pysiological and clinical aspects of cardiac auscultation*, Medi-Care Ltd., London, UK

[Hayes80a] M. Hayes, J. Lim and A. Oppenheim 1980 presented at *ICASSP Conference*. Phase-only signal reconstruction

[Hayes80b] M. Hayes, J. Lim and A. Oppenheim 1980 *IEEE Transactions on Acoustics, Speech and Signal Processing* **38**(6), 672-680. Signal Reconstruction from Phase or Magnitude

[Herlufsen84] H. Herlufsen 1984 Bruel and Kjaer *Technical Review* No. 1. Dual Channel FFT Analysis (Part 1)

[Hlawatsch91] F. Hlawatsch and W. Krattenthaler 1991 *IEEE Transactions on Signal Processing* **39**(3), 612-619. Phase Matching Algorithms for Wigner-Distribution Signal Synthesis

[Hlawatsch92] F. Hlawatsch and G. Boudreaux-Bartels 1992 *IEEE Signal Processing Magazine* April 21-66. Linear and Quadratic Time-Frequency Signal Representations

[Hlawatsch94] F. Hlawatsch, A. H. Costa and W. Krattenthaler 1994 *IEEE Transactions on Signal Processing* **42**(9), 2513-2520. Time-Frequency Signal Synthesis with Time-Frequency Extrapolation and Don't-Care Regions

[Irino92] T. Irino and H. Kawahara 1992 *IEEE ICASSP-92*, **1**, 85-88. Signal Reconstruction from Modified Wavelet Transform – An Application to Auditory Signal Processing

[Irino93] T. Irino and H. Kawahara 1993 *IEEE Transactions on Signal Processing* **41**(12), 3549-3554. Signal Reconstruction from Modified Auditory Wavelet Transform

[Janssen87] A. Janssen 1987 *IEEE Transactions on Acoustics, Speech and Signal Processing* **35**(5), 701-705. A Note on "Positive Time-Frequency Distributions"

[James95] J. F. James 1995 *A student's guide to Fourier Transforms*, Cambridge University Press, Cambridge

[Jones95] D. Jones and R. Baraniuk 1995 *IEEE Transactions on Signal Processing* **43**(10), 2361-2371. An Adaptive Optimal-Kernal Time-Frequency Representation

[Kincaid91] D. Kincaid and E. Ward Cheney 1991 *Numerical Analysis : Mathematics of Scientific Computing*, Brooks/Cole Publishing Company, London

[Kodera76] K. Kodera, C. Villedary and R. Gendrin 1976 *Physics of the Earth and Planetary Interiors* **12**, 142-150. A New Method for the Numerical Analysis of Non-Stationary Signals

[Kodera78] K. Kodera, R. Gendrin and C. De Villedary 1978 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(1), 64-76. Analysis of Time-Varying Signals with Small BT Values

[Krattenthaler93] W. Krattenthaler and F. Hlawatsch 1993 *IEEE Transactions on Signal Processing* **41**(1), 278-287. Time-Frequency Design and Processing of Signals via Smoothed Wigner Distributions

[Kronland88] R. Kronland-Martinet 1988 *Computer Music Journal* **12**(4), 11-20. The Wavelet Transform and Analysis, Synthesis, and Processing of Speech and Music Sounds

[Leung98] T. Leung, P. White, W. Collis, E. Brown and A. Salmon 1998 *Applied Signal Processing* 4, 154-167. Time-Frequency Methods for Analysing Paediatric Heart Murmurs

[Leung99] T. Leung 1999 Ph.D. Thesis, University of Southampton.

[Lim88] J. Lim and A. Oppenheim 1998 *Advanced Topics in Signal Processing*, Prentice-Hall, New York

[Lopes97] D. M. Lopes and P. R. White 1997 *IEEE UK Symposium on Applications of Time-Frequency and Time-Scale Methods*, 137-140. A System for Constructing Time-Frequency Representations based upon a Detection and Feedback Technique

[Lopes98] D. Lopes and P. White 1998 *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* 21-24. Signal Reconstruction from the Reassigned Spectrogram

[Loughlin94a] P. Loughlin, J. Pitton and L. Atlas 1994 *IEEE Transactions on Signal Processing* 42(10), 2697-2705. Construction of Positive Time-Frequency Distributions

[Loughlin94b] P. Loughlin, J. Pitton and B. Hannaford 1994 *IEEE Signal Processing Letters* 1(12), 199-202. Approximating Time-Frequency Density Functions via Optimal Combinations of Spectrograms

[Marven96] C. Marven and G. Ewers 1996 *A Simple Approach to Digital Signal Processing*, John Wiley and Sons

[Moss94] J. Moss and J. Hammond 1994 *Mechanical Systems and Signal Processing* 8(3) 243-258. A Comparison between the Modified Spectrogram and the Pseudo- Wigner-Ville Distribution with and without Modification

[Nawab82] S. Hamid Nawab, T. Quatieri and J. Lim 1982 *IEEE ICASSP-82* 1046-1048. Signal Reconstruction from the Short-Time Fourier Transform Magnitude

[Nawab83a] S. Hamid Nawab, T. Quatieri and J. Lim *IEEE Transactions on Acoustics, Speech, and Signal Processing* **31**(4), 986-998. Signal Reconstruction from Short-Time Fourier Transform Magnitude

[Nawab83b] S. Hamid Nawab, T. Quatieri and J. Lim *ICASSP-83 Boston* 800-803. Algorithms for Signal Reconstruction from Short-Time Fourier Transform Magnitude

[Nelson94] P. Nelson and S. Elliot 1994 *Active Control of Sound*, Academic Press, London

[Oppenheim75] A. Oppenheim and R. Schafer 1974 *Digital Signal Processing*, Prentice Hall International, New York

[Owens93] F. J. Owens 1993 *Signal Processing of Speech* Macmillan Press Ltd, London

[Papoulis91] A. Papoulis 1991 *Probability, Random Variables and Stochastic Processess*, Third Edition, McGraw-Hill, New York

[Pinter96] I. Pinter 1993 *Computer Speech and Language* **10**, 1-22. Perceptual Wavelet-Representation of Speech Signals and it Application to Speech Enhancement

[Pitton93] J. Pitton, P. Loughlin and L. Atlas 1993 *ICASSP-93* Vol. 4 436-439. Positive Time-Frequency Distributions via Maximum Entropy Deconvolution of the Evolutionary Spectrum

[Pitton94] J. Pitton, L. Atlas, P. Loughlin 1994 *IEEE Transactions on Speech and Audio Processing* **2**(4), 554-566. Applications of Positive Time-Frequency Distributions to Speech Processing

[Plante98] F. Plane, G. Meyer and W. Ainsworth 1998 *IEEE Transactions on Speech and Audio Processing* **6**(3), 282-287. Improvement of Speech Spectrogram Accuracy by the Method of Reassignment

[Portnoff80] M. Portnoff 1980 *IEEE Transaction on Acoustics, Speech, and Signal Processing* **28**(2), 55-69. Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier analysis

[Portnoff81] M. Portnoff 1981 *IEEE Tractions on Acoustics, Speech, and Signal Processing* **29**(3), 374-390. Time-Scale Modification of Speech Based on Short-Time Fourier Analysis

[Press92] W. Press, W. Vetterling, S. Teukolsky and B. Flannery 1992 *Numerical Recipes in C – The Art of Scientific Computing*, Cambridge University Press, Cambridge

[Priestley81] M. Priestley 1981 *Spectral Analysis and Time Series*, Academic Press, London

[Qian96] S. Qian and D. Chen 1996 *Joint Time-Frequency Analysis: Methods and Applications*, Prentice Hall, Upper Saddle River, New Jersey

[Quatieri81] T. Quatieri and A. Oppenhiem 1981 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **29**(6), 1187-1193. Iterative Techniques for Minimum Phase Signal Reconstruction from Phase or Magnitude

[Quatieri83] T. Quatieri 1983 *Journal of the Optical Society of America* **73**(11), 1523-1526. Frequency Sampling of the Short-Time Fourier-Transform Magnitude for Signal Reconstruction

[Rabiner78] L. Rabiner and R. Schafer 1978 Digital Processing of Speech Signals, Prentice-Hall, New York

[Rangayyan88] R. Rangayyan 1988 *Critical Reviews in Biomedical Engineering* **15**(3), 211-236. Phonocardiogram Signal Analysis : A Review

[Richard97] C. Richard and R. Lengelle 1997 *Signal Processing* **60**, 163-179. Joint Recursive Implementation of Time-Frequency Representations and their Modified version by the Reassignment Method

[Rioul91] O. Rioul and M. Vetterli 1991 *IEEE Signal Processing Magazine*, October 14-37. Wavelets and Signal Processing.

[Shiavi91] R. S. Shiavi 1991 *Introduction to Applied Statistical Signal Analysis*, Aksen Associates, New York

[Stoer93] J. Stoer and R. Bulirsch 1993 *Introduction to Numerical Analysis*, Springer-Verlag, New York

[Tchamitchian91] P. Tchamitchian and B. Torresani 1991 in *Wavelets and Applications, Ed. Beykin et al* Springer-Verlag, New York

[Wittaker35] J. M. Wittaker 1935 *Interpolary Function Theory*, Cambridge University Press, Cambridge

[White82] R. G. White and J. G. Walker (Editors) 1982 *Noise and Vibration* Ellis Horwood Limited, London

[White97] P. White, W. Collis and A. Salmon 1997 *Proceedings IEE Colloquium on Time-Frequency Analysis of Biomedical Signals, London*, 2/1-2/3. Time-Frequency Analysis of Heart Murmurs in Children

[Yegnanarayana83] B. Yegnanarayana and A. Dhayalan 1983 *ICASSP 83 – Boston* 639-642. Noniterative Techniques for Minimum Phase signal Reconstruction from Phase or Magnitude

[Yegnanarayana84] B. Yegnanarayana, D. Saikia and T. Krishnan 1984 *IEEE Transactions on Acoustics, Speech, and Signal Processing* **32**(3), 610-622. Significance of Group Delay Function in Signal Reconstruction from Spectral Magnitude or Phase

[Zhang88a] X. Zhang, L. Durand, L. Senhadji, H. Lee and J. Coatrieux 1998 *IEEE Transactions on Biomedical Engineering* **45**(8), 962-971. Analysis-Synthesis of the Phonocardiogram Based on to Matching Pursuit Method

[Zhang88b] X. Zhang, L. Durand, L. Senhadji, H. Lee and J. Coatrieux 1998 *IEEE Transactions on Biomedical Engineering* **45**(8), 972-

979. Time-Frequency Scaling Transformation of Phonocardiogram Based of the Matching Pursuit Method