

Chapter 5

Defining and Analysing the Signal Processing Procedures

In this chapter the signal processing requirements outlined in chapters 3 and 4 are analysed. In summary, the technique should restrict estimates to the low pass roll-off region of the recording, where the signature of the recording process becomes apparent. This ROI should be described by the simplest parametric model and should employ well-established estimation techniques wherever possible. Potentially suitable procedures have been investigated and are reported in the following subsections:

- Estimation of the log of the power spectrum over a ROI. (section 5.1 to 5.3)
- Application of the RDT. (section 5.4 to 5.5)
- Estimation of a slope parameter of the RDT of the log of the power spectrum. (section 5.6)

The statistical properties of the estimates are derived from a time-domain signal that can be characterised over the spectral ROI as low pass filtered Gaussian white noise (GWN).

5.1 Spectral Estimation Method

The first stage of the analysis technique derived in chapter 4, requires the spectral estimation of an ensemble of time-domain data to be evaluated over a limited range of frequencies close to the Nyquist limit.

The subject of spectral analysis is very broad and can be split into two main classes of estimation: parametric and non-parametric. Analytical descriptions of many of the methods of spectral analysis within each class may be found in well-known texts [122-124]. The parametric methods describe random processes using a rational transfer function

model. The overall estimate is a 3-step procedure: selection of model, estimation of the parameters of the assumed model and production of the spectral estimate implied by the model [125 p440]. It is known that parametric methods of analysis may be complicated but can produce a higher resolution estimate than non-parametric methods when the signal duration is short [126]. However, application of the wrong model type or wrong parameter estimate can produce completely erroneous spectral estimates [126]. In general, non-parametric methods based on Fourier transforms are satisfactory when the data size is adequate.

The power spectral density (psd) function describes how the variance of a random time-domain process is distributed in frequency: *density = power per unit bandwidth*. In this section the problem of estimating the psd of a wide-sense stationary (WSS) random process is considered; where a WSS process can be defined as having its first and second order moments unchanging over time [125 p40-42].

The standard non-parametric estimation technique for spectral analysis is based on the periodogram [122 ch2]. The periodogram produces an estimate of the psd from the information contained in the data alone, making no assumptions about the signal characteristics. The periodogram allows the direct estimation of the psd by taking the square of the magnitude of the Fourier transform of the time-domain sequence. It will be assumed that for calculation efficiency, any practical implementation of expressions involving the DFT will be carried out using a fast Fourier version of the transform (FFT).

The ordinates of the periodogram are proportional to the squared magnitude of the DFT of a finite length sequence, and is from a computational point of view, simple to evaluate. Taking N time-domain samples, $\{x(1), x(2), \dots, x(N)\}$, consisting of a WSS Gaussian random process with an unknown spectral density, the spectral estimator for the periodogram is defined as [122 p24]:

$$\hat{\Phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=1}^N x(t) e^{-j\omega t} \right|^2 \quad (5.1)$$

When N represents the total number of time-domain samples and s represents a discrete positive frequency coefficient between $0, 1, \dots, S-1$ $\{S = N/2 + 1\}$, the forward discrete psd of a discrete data sequence (5.1) can be written as:

$$\hat{\Phi}[s] = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-\frac{j2\pi sn}{N}} \right|^2 \quad (5.2)$$

The discrete frequency is defined here as a fraction of sampling frequency given by s/N and the discrete radian frequency ω is defined over equally spaced samples between 0 and π radians, $\omega \in [0, \pi]$, and given by:

$$\omega = \frac{2\pi s}{N}$$

5.1.1 Performance Limitations of the Periodogram

The periodogram approach to spectral estimation is computationally efficient but has inherent performance limitations. Two measures that may be used to characterise the performance of the periodogram estimation, are the bias and the variance [123 app4A]. The bias is a measure of the average deviation **away** from the true value of the psd (5.3) and the variance is the average deviation **about** the true value of the psd (5.4).

$$\text{Bias}\{\hat{\Phi}(\omega)\} = E\{\hat{\Phi}(\omega)\} - \Phi(\omega) \quad (5.3)$$

$$\text{Var}\{\hat{\Phi}(\omega)\} = E\left[\hat{\Phi}(\omega) - \Phi(\omega)\right]^2 \quad (5.4)$$

The bias in the periodogram occurs as the result of having a finite length data sequence. This finite or truncated data sequence is equivalent to windowing (multiplying) an infinite length data sequence with a rectangular (or some other) function and is discussed in detail in chapter 6. The expected value of the periodogram is therefore the convolution of the power spectrum with the power spectrum of the window function. As the length of data sequence increases, the window function spectrum converges to an impulse in the frequency-domain; indicating that the periodogram is an asymptotically unbiased estimator [122 p33]:

$$\lim_{N \rightarrow \infty} E\{\hat{\Phi}_p(\omega)\} = \Phi(\omega) \quad (5.5)$$

The main limitation of the basic periodogram defined by (5.2) is that it has a high statistical variance. Marple [123 app4A], shows that the periodogram variance does not decrease by increasing the data length of the sample. The increased information contained in a sample of larger data length exhibits itself in the frequency-domain, as an increase in the number of independent frequency coefficients and therefore has improved frequency resolution but increases the scale of the computation. Alternatively, averaging a set of periodograms together can improve on the statistical properties of the periodogram [127], [128]. Bartlett proposed a simple ensemble averaging method to reduce the fluctuations of the periodogram [129]. The idea is to split up the available length N data sequence

estimator.

Segmentation of the data as described, results in a non-linear relationship between FFT size and calculation time. Mulgrew et al [130 p287] state the number of multiplication's required for an M point FFT to be $(M/2)\log_2 M$. Given the number of transform points M for a fixed data length N , the number of data segments will be N/M , and the total number of multiply operations for K segments of data is therefore given by (5.8). For a fixed data length, the graph of fig 5.2 shows a change in the number of multiply operations for a change in the number of segments.

$$\text{FFT multiplications} = \left(\frac{N}{2}\right)\log_2\left(\frac{N}{K}\right) \quad (5.8)$$

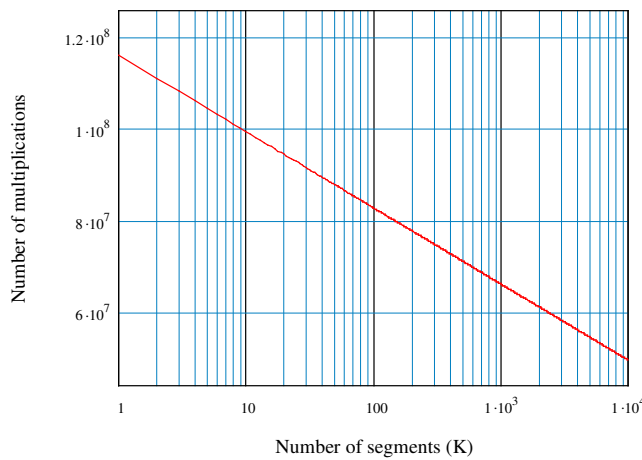


Fig 5.2: The number of segments versus the number of FFT multiply operations are shown for a fixed data length of 10Mb.

It is clear from fig 5.2 that for a fixed data length, increasing the number of segments and therefore the number of periodograms to be estimated, reduces the processing time required for the overall averaged periodogram calculation. However, as the calculation time is not critical for the application, segmentation will be optimised for the variance and resolution requirements of the RDT slope estimate as discussed in the following chapters.

5.2 Frequency Transform Method: Chirp Z-Transform v FFT

In estimating the power spectrum over the ROI there is a requirement for a detailed analysis of the psd over a very limited range of frequencies near to the Nyquist limit. As described in the previous section, an estimation of the power spectrum based on the periodogram is carried out using the Fourier transform. This section examines the chirp z-transform (CZT) as an alternative method of time to frequency domain conversion.

To obtain a frequency resolution of $\leq \Delta f$ using a sampling frequency of $1/T$, would require that the number of sampled data points N be: $\geq 1/(T \times \Delta f)$. Therefore small Δf

would require a large number of data samples, and depending on the bandwidth of interest, would result in the calculation of redundant spectral information. A derivation of this redundancy follows where:

- fs : sample rate
- bw : required bandwidth
- M : transform size

The ratio of the maximum bandwidth to the required bandwidth is given by: $fs/(2 \cdot bw)$.

Dividing this into the total number of spectral points in the power spectrum $M/2$ gives the total number spectral points over the required bandwidth: $M \cdot bw/fs$. Subtracting this value from the total number of spectral points gives the number of spectral points of redundancy: $(M/2) - M \cdot bw/fs$. Dividing by $M/2$ will give the redundancy factor: $1 - 2 \cdot bw/fs$. It can be seen that the redundancy factor does not depend on original transform size, only on the bandwidth of the ROI and sampling rate. The percentage of redundant transform points is given by:

$$\% \text{ redundancy} = R_{bw} = \left(1 - \frac{2 \cdot bw}{fs}\right) \cdot 100 \quad (5.9)$$

As an example, using a sampling rate of 32kHz and a bandwidth of 600Hz the redundancy R_{bw} is:

$$R_{bw} = \left(1 - \frac{2 \cdot bw}{fs}\right) \cdot 100 = 100 - \frac{120}{32} = 96.25\%$$

In order to concentrate all transform points over the required bandwidth and eliminate the redundancy produced by the FFT, the possibility of using the CZT has been investigated.

5.2.1 The Chirp Z-Transform (CZT)

A method of computing an M -point frequency transform giving increased detail of a Fourier spectrum over a limited frequency range is the requirement. Rabiner et al state:

“One very useful application of the CZT is the ability to efficiently evaluate high resolution narrow frequency band spectra.What is often required is high resolution for a limited range of frequencies. The CZT is extremely well suited to such cases since it allows selection of initial frequency spacing, independent of the number of time samples. Hence high resolution data over a narrow frequency range can be obtained at low cost.”[131]

This tends to imply that the CZT can provide superior resolution over a limited frequency band to that of the FFT. This statement will now be examined.

Discrete signals can be described by the z-transform. When a signal is represented

by a discrete spectrum the FFT is used to evaluate the z-transform at discrete points given by:

$$z = e^{\frac{2\pi jm}{N}} \quad \text{for } m = 0, 1, \dots, N-1 \quad (5.10)$$

These points are all positioned uniformly around the unit circle of the z plane.

The CZT can compute the z-transform for an input sequence off the unit circle of the z-plane. If the values of z (5.10) are arranged on circular or spiral contours, then they can be expressed as a discrete convolution. This is the basis of the CZT presented by Rabiner, Schafer and Rader [131],[132] and this subsection is based on their work. The z-transform expressed along a spiral contour is shown in fig 5.3 and is given by:

$$z_m = AW^{-m} \quad m = 0, 1, \dots, M-1$$

where M is an arbitrary integer and A and W are arbitrary complex numbers:

$$A = A_0 e^{j2\pi\theta_0}$$

$$W = W_0 e^{j2\pi\phi_0}$$

Here, A is the starting point of the z-plane contour and W defines the contour itself. The following six statements describe the general behaviour of the independent parameters of the CZT:

1. If $A_0 = 1$ then the contour starts on the unit circle with a displacement from the origin of $2\pi\theta_0$.
2. If $A_0 > 1$ then the contour starts outside the unit circle with a displacement from the origin of $2\pi\theta_0$.
3. If $A_0 < 1$ then the contour starts inside the unit circle with a displacement from the origin of $2\pi\theta_0$.
4. If $W_0 = 1$ then the contour is a concentric circle with an angular spacing between points given by $2\pi\phi_0$.
5. If $W_0 < 1$ then the contour spirals outward with an angular spacing between points given by $2\pi\phi_0$.
6. If $W_0 > 1$ then the contour spirals inward with an angular spacing between points given by $2\pi\phi_0$ (fig 5.3).

The derivation of the CZT can be found at [131] and is based on expressing the DFT as a convolution. This convolution can be implemented efficiently by using the FFT algorithm [119 ch 9]. The CZT has advantages over the standard DFT/FFT [132] that include:

- The number of time samples N is not required to be equal to the number of ordinates of the z-transform M .

- The angular spacing of z_m can be specified.
- N or M are not required to be composite integers.
- The contour need not be a circle but can spiral in and out with respect to its origin.
- The point z_0 is arbitrary.

If the requirement is to calculate the z-transform entirely on the unit circle then $A_0=1$ and $W_0=1$, and under these conditions the algorithm is known as the chirp transform (CT) described by Oppenheim and Schaffer [133 pp623-628]. A special case of the CT is where $A=1$, $W = \exp(-j2\pi\phi_0)$ and $N=M$, which is equivalent to the DFT (fig 5.4).

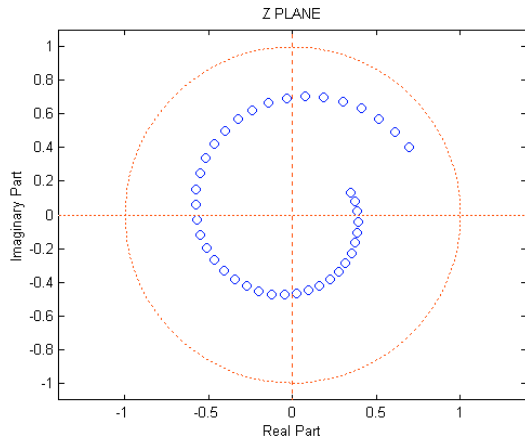


Fig 5.3: Example of a spiral contour :

$$M = 40 \quad A = 0.8e^{\frac{j\pi}{6}} \quad W = 1.02e^{\frac{-j2\pi}{M}}$$

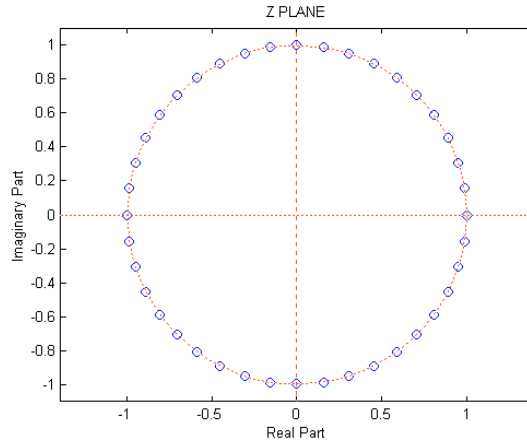


Fig 5.4: Example of the equivalent of a DFT :

$$M = 40 \quad A = 1 \quad W = e^{\frac{-j2\pi}{M}}$$

Fig 5.5 shows a z-transform produced by the CT over a narrow ROI near to the Nyquist limit, typical of that required for the application. After execution of the CT the output spectrum is a complex vector:

$$[X_m] = [X_0, X_1, \dots, X_{M-1}] \tag{5.11}$$

The M complex points in $[X_m]$ are distributed over the frequency-domain in accordance with (5.11). The frequency ROI is given by: $\omega_m = \omega_0 + m\Delta\omega$, $m = 0, 1, \dots, M-1$

$$f_1 \text{ is the starting frequency and given by: } f_1 = \frac{\omega_0}{2\pi}$$

$$f_2 \text{ is the finishing frequency and given by: } f_2 = \frac{\omega_0 + M\Delta\omega}{2\pi}$$

Where the total number of transform points M is determined from the frequency range and resolution. The final output is a set of equally spaced samples of the Discrete Fourier Transform on the unit circle. The CT has been used in this research to calculate a subset of

the DFT. Where samples ω are found on the unit circle of the z-plane: $r=1$, $\pi > \omega > 0$.

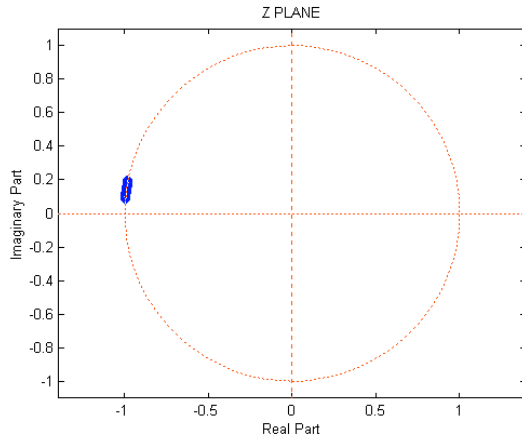


Fig 5.5: A z-transform produced by the CT over a narrow ROI close to the Nyquist limit.

5.2.2 True Resolution

If the time-domain data sequence of length N equals the transform size M , then the highest resolution obtainable from the DFT is given by [134]:

$$\Delta f = \frac{f_s}{N} \quad (5.12)$$

The resolution as stated by (5.12) actually has no direct connection with the DFT, but is a function of the truncation of the data signal sequence to length N . This truncation is equivalent to a multiplication of the data of N samples with a rectangular window of length N [134]. In the frequency-domain this is the same as convolving the frequency response of the rectangular window with the individual frequency coefficients of the data. Now as the bandwidth of the rectangular window is a constant for a given data length N , the ability to resolve detail is limited by this bandwidth. The only way to improve true resolving power is to increase the data length N . Implied in this statement, is an increased window length in the time-domain, which results in the bandwidth of the window becoming narrower in the frequency-domain, improving resolvability. The conclusion must be that as the resolution is a function of the truncated data length N , the CT cannot improve on the resolution of the DFT and equation (5.12) must hold for both the DFT and CT.

5.2.3 Between the Spectral Lines

The requirement was to determine what the CT offers in terms of calculating transforms that have a frequency spacing Δf that is less than the inherent resolution implied by the data and transform size (5.12).

The DFT/FFT and CT produce discrete representations of the spectral data,

producing a gap between each spectral sample, known as the picket fence effect¹, [135 pp671-674]. If the spectral sample spacing is too wide (low transform size) the true variation of the spectrum may not be adequately indicated. The CT can therefore be used as a picket fence reduction technique to decrease the spectral sample spacing to an arbitrary value over the frequency band of interest.

Adding a series of zero valued samples to the end of the time-domain data sequence enables an increase to be made in the transform size of the DFT/FFT. This is known as zero padding [119 ch3.8]. Zero padding results in the same picket fence reduction effect for the FFT as produced by the CT. Picket fence reduction techniques do not change the fundamental resolving power of either the DFT/FFT or the CT. As Kay and Marple state:

“There is a common misconception that zero-padding the data sequence before transforming will improve the resolution of the periodogram. Transforming a data set with zeros only serves to interpolate additional PSD values within the frequency interval $-1/(2\Delta t) \leq f \leq 1/(2\Delta t)$ Hz between those that would be obtained with a non zero padded transform..... In no case of zero padding, however is there an improvement in the fundamental frequency resolution. Zero padding is useful for 1) smoothing the appearance of the periodogram estimate via interpolation, 2) resolving potential ambiguities...3) reducing the “quantization” error in the accuracy of estimating the frequencies of the spectral peaks.”[126]

A comparison will be made between the power spectrum of a spectrally shaped noise signal using both the FFT and CT, where both transform methods are averaged and transformed using identical parameters:

Signal type: low pass filtered Gaussian white noise,
 $f_s = 32 \text{ kHz}$
 $f_{low} = 15 \text{ kHz}$
 $f_{high} = 15.6 \text{ kHz}$
 $bw = 600 \text{ Hz}$
 $M = 16384$
 No of averages = 40 (non overlapped)
 Window type: Rectangular

The power spectrum derived from the CT is a direct 16384-point transform over a 600 Hz ROI. The power spectrum derived from the FFT is a 16384-point transform over the whole region 0 to 15.6 kHz, but only transform points covering the ROI will be required from the FFT. The number of points using the FFT are given by:

$$\text{Points} = \frac{M \cdot bw}{f_s} = \frac{16384 \times 6}{320} = 307$$

¹ The picket fence effect can have an absolute value as given by Harris [134] referred to as ‘scalloping loss’.

In fig 5.6, the FFT power spectrum of 307 points has been over-laid on the CT power spectrum of 16384 points. Both traces show a very similar overall spectral response differing only in the fine detail.

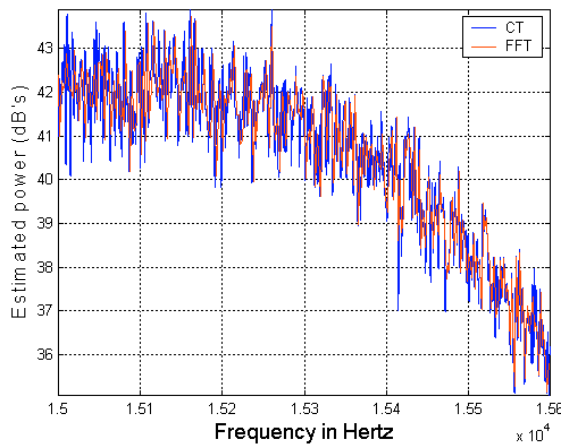


Fig 5.6: Power spectrum from the FFT is overlaid onto CT showing similar response but differing in the finer detail.

The differences in the detail between the two transforms can be seen by examining the first 50 Hz of the data, fig 5.7. In order to show that the zero padding picket fence reduction technique is equivalent to results obtained by using the CT, the data used for the FFT version has been zero padded by a factor of 6, giving the total number of FFT points across the ROI as $6 \times 307 = 1842$. The result of zero padding the data is shown in fig 5.8, where the spectral shape across the fine detail is now identical with the CT.

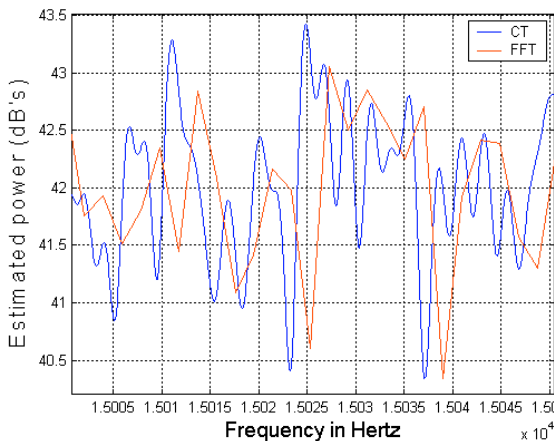


Fig 5.7: Fine detail of CT overlaid by FFT.

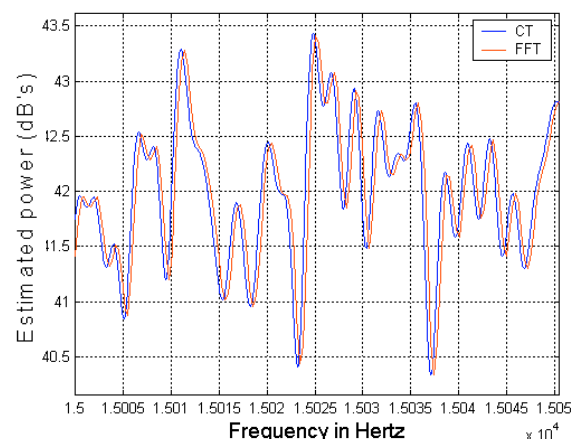


Fig 5.8: Fine detail of CT overlaid by zero padded FFT.

These results confirm that zero padding an FFT can produce the same results as the CT. Of course the FFT has calculated many redundant transform points outside the ROI, as given by (5.9).

It has been shown that the CT produces no fundamental increase in resolution between spectral coefficients. The extra transform points produced by the CT will have a

value based on the already existing information. These extra transform points will therefore produce a positive correlation between closely spaced coefficients. This may be undesirable when the results are to be used in a linear regression application [136 ch12.1]. The additional more closely spaced spectral points distributed over the region resulting from the CT represents spectral energy in the form of leakage from adjacent spectral points (discussed in detail in chapter 6) and does not represent any new information. This can be considered as serial spectral autocorrelation between the transform points.

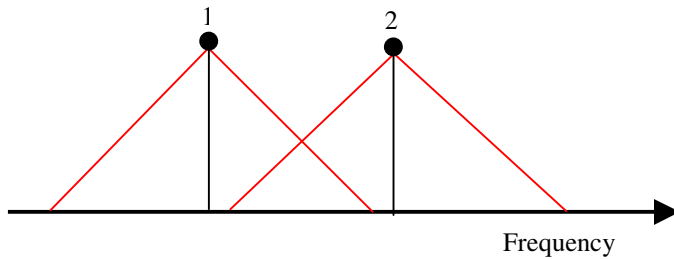


Fig 5.9A: Shows two uncorrelated spectral points (FT).

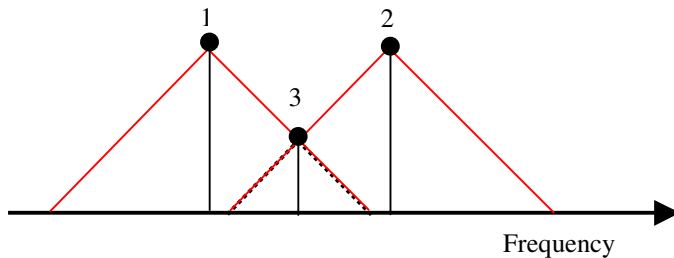


Fig 5.9B: As in A but point 3 is based on the spread of energy from points 1 and 2 not on any additional information (CT).

The simplified principle, ignoring superposition, is demonstrated in fig 5.9. Fig 5.9A represents two spectral transform points 1 and 2 produced by the FFT, the diagonal lines indicate the spread of energy out from each point due to spectral leakage (section 6.2). It can be seen that points 1 and 2 are independent of one another. A CT producing twice the number of transform points over the region is represented by fig 5.9B, where a further transform point is situated between the original two points. The value of this intermediate transform point will be dependent on the magnitude of points 1 and 2, resulting in serial autocorrelation.

It can be concluded that if the number of overall FFT transform points (non-zero padded) are high enough, then over the ROI the shape of the spectral response can be adequately estimated using this transform. The CT can have advantages when the number of points over a region of interest is low for a particular transform size. For example, using the 600 Hz bandwidth and a transform size of 512 points, the CT has all 512 points spread over the region but the FFT will only have:

$$\text{Points} = \frac{M \cdot bw}{f_s} = \frac{512 \times 6}{320} \approx 10$$

It should be remembered that both methods still have the same fundamental resolution:

$$\frac{F_s}{M} = \frac{32 \times 10^3}{512} = 62.5 \text{ Hz}$$

5.2.4 Calculation Time

The CT is a more complex transform than the FFT and incorporates the efficiency of the FFT to perform the convolution required to obtain its output. Rabiner et al state:

“As the number of samples grows large, the computation time for the chirp z-transform grows asymptotically..... This is the same sort of asymptotic dependence of the fast Fourier transform, but the constant of proportionality is bigger for the chirp z-transform because two or three fast Fourier transforms are required instead of one.” [131]

Empirical measurements show that for a transform size of 64 points the FFT calculates the result approximately 2 times the speed of the CT. At a transform size of 65536 points, the FFT is computed approximately 16 times faster than the CT. These measurements are based on the standard algorithms incorporated into the Matlab signal processing toolbox [137]. As discussed, a limited spectral ROI leads to the production of a large amount of redundant information when using a conventional FFT; if this redundancy were exploited by calculating only a subset of output points [138], the FFT speed compared to the CT would have even greater efficiency.

5.2.5 Conclusions from the Comparison

The comparison between the DFT/FFT and CT used for the estimation of the power spectrum of the transition region of a low-pass function leads to the following set of conclusions.

- The CT offers no increase in the fundamental frequency resolution than that of the DFT/FFT.
- If the number of transform points over the ROI are high enough, the DFT/FFT will adequately describe the required spectral low-pass response.
- The CT results in correlation between the closely spaced spectral coefficients.
- The CT can have advantages over the DFT/FFT when the DFT/FFT transform size is low and the frequency range of interest is limited. These advantages however, are only equivalent to that of zero padding an FFT to produce the same number of transform points over the ROI.

- For a given transform size the FFT calculates the transform significantly faster than the CT. The FFT speed advantage could be improved further if any redundancy in the calculation were to be exploited.

The overall conclusion was that the CT provided no advantages over that of using the FFT when the data size of the transform is high enough. It became apparent that there can be distinct disadvantages in using the CT. Therefore the FFT was the method chosen for the periodogram calculations used in the spectral estimation procedure prior to applying the RDT.

5.3 Logarithmic Transformation

In order to be able to identify low level detail in wide dynamic range spectral responses, a logarithmic transformation of the power spectrum is often applied. The magnitude of the spectral coefficients of the Bartlett periodogram given by (5.6) can be transformed logarithmically: $\log_b(\hat{\Phi}_B[\omega])$. A transform can be to any base b of logarithm; for engineering applications it is usual to work with 10 times the logarithm of the signal to base 10, producing a result that is scaled in decibels (dB). For mathematical convenience, unless otherwise stated, in the analysis that follows logarithmic conversion has been carried out using \log to the base e giving the result in Nepers.

It will be shown in subsection 5.5.4 that statistically, after a logarithmic transformation, the population variance for each spectral coefficient will be the same and independent of its mean value. This results in a logarithmically converted signal showing the same resolution at both low and high signal levels.

Aside from the compression in scaling of the amplitude, the homogeneity in variance across the spectral coefficients is an important requirement for applications involving linear regression, and was a strong reason for applying a logarithmic transformation to the power spectrum [139].

5.4 Reverse Difference Transform

An investigation into the behaviour of the change in spectral response over the ROI due to the application of a novel simple transformation to the log-power spectrum, known as the reverse difference transform (RDT) as introduced in section 4.2, will be extended in this section.

The spectral ROI with a centre frequency Ω is selected from a narrow frequency

region of the digital recorder's low-pass response, where the pass-band rolls off towards the stop-band before encountering the Nyquist limit. This spectral region $R(\omega)$ is reflected around Ω by a simple mathematical transformation, to produce a frequency reversed logarithm of the power spectrum. The reversed logarithmic spectrum is subtracted from the non-reversed, fig 5.10.

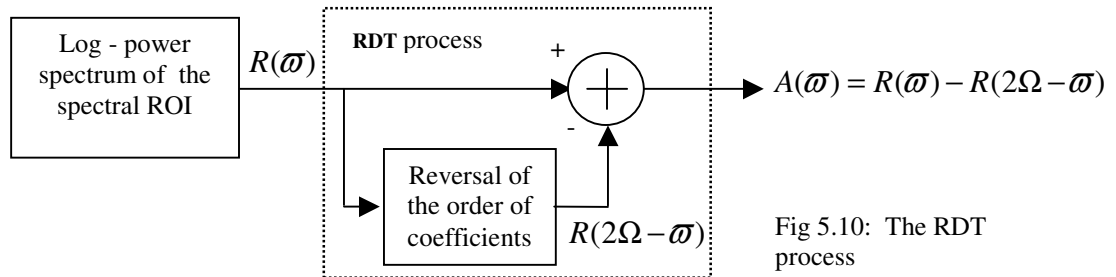


Fig 5.10: The RDT process

Over the ROI, the RDT approximately linearises the spectral response and produces a normalisation of the resulting spectrum that takes the central point of the ROI through 0 dB. The mechanisms by which these changes take place have been examined by modelling the recorder as a low-pass filter having the response characteristics of its anti-alias filter.

5.4.1 Filtered Response

The characteristics of a filter in the frequency-domain can be described by the function representing its complex gain:

$$H(j\omega)$$

Similarly the input signal can be described in the frequency-domain by the function:

$$G(j\omega)$$

When the signal is passed through the filter, the output will be described by the product of these two functions:

$$F(j\omega) = G(j\omega) \cdot H(j\omega) \quad (5.13)$$

If the power spectrum is required then (5.13) can be multiplied by its complex conjugate:

$$F(j\omega)^*$$

resulting in a cancellation of the phase and a squaring of the modulus components:

$$|F(j\omega)|^2 = F(j\omega) \cdot F(j\omega)^*$$

Before the application of the RDT the response of the power spectrum is converted logarithmically. The examples to be shown are scaled in dB:

$$R(\omega) = 10 \cdot \log_{10} |F(j\omega)|^2$$

5.4.2 Power Series Expansion Example Using a Butterworth Response

In general, the type of low-pass anti alias filter applied in the production of a digital recording will be unknown. In this subsection, for illustrative purposes, a Butterworth filter is used to describe the behaviour of the RDT.

The frequency response of the Butterworth filter will be defined in terms of its log-power spectrum. The low-pass Butterworth filter will have the following squared magnitude response [140]:

$$|H(j\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}$$

Expressing this in dB produces a result for the log-power spectrum as a function of ω :

$$R(\omega) = 20\log_{10}|H(j\omega)| = 10\log_{10}\left(\frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}\right) \quad (5.14)$$

Given a cut off frequency ω_c of 20 kHz and a filter order N of 50, produces a log-power spectrum having a relatively high rate of change of attenuation outside of the pass-band as shown by fig 5.11. As discussed in subsection 4.2.2, the ROI will usually incorporate the most non-linear part of the power spectral response known as the ‘knee’.

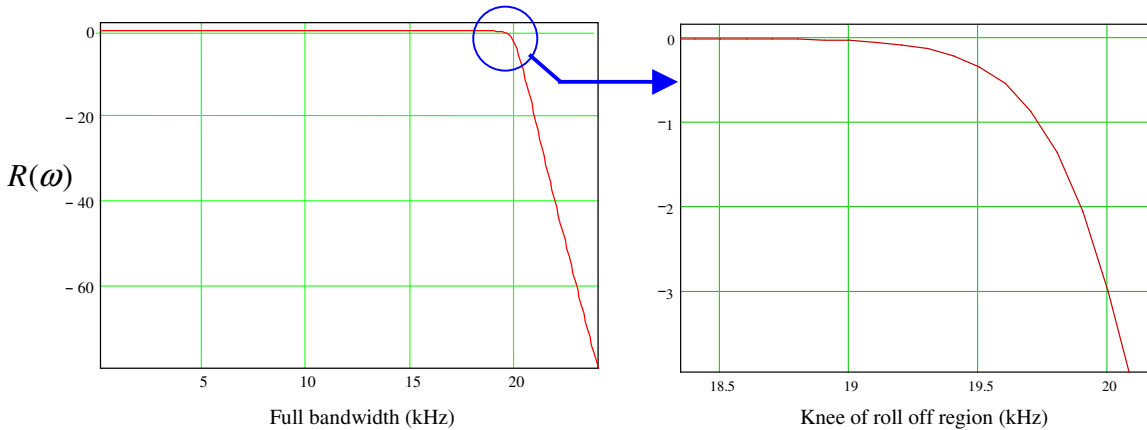


Fig 5.11: Showing traces for full bandwidth and roll off region of the Butterworth filter response.

A Taylor power series expansion [141] will be used to model a Butterworth filter over a spectral ROI. From the model, the RDT will be applied and the results analysed.

A 6th order power series expansion of the Butterworth filter given by (5.14), applied from the centre of the ROI produces the result, where $\Omega = 19.6$ kHz :

$$\tilde{R}(\omega) = 50.3 - 2.59\omega - 5.77(\omega - \Omega)^2 - 7.31(\omega - \Omega)^3 - 4.22(\omega - \Omega)^4 + 2.84(\omega - \Omega)^5 \quad (5.15)$$

Fig 5.12 shows the result of this expansion on the same axis as the original filter response. This shows how close the 6th order series approximates over the ROI.

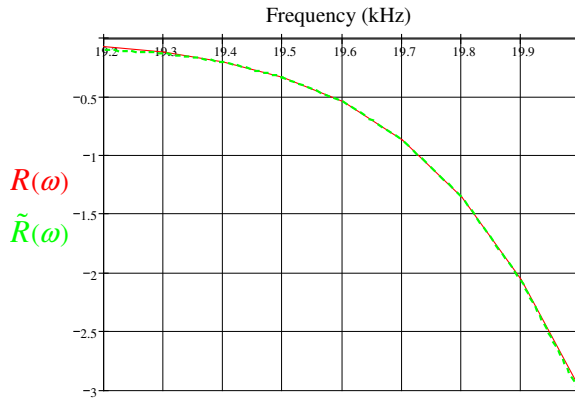


Fig 5.12: The result of a 6th order power series expansion $\tilde{R}(\omega)$ shown overlaid onto the original filter response $R(\omega)$. A close approximation around the frequency ROI of 19.2 kHz to 20 kHz is achieved.

The frequency will be expressed as a function of deviation, $\omega_{\Delta} = \omega - \Omega$ therefore (5.15) becomes:

$$\tilde{R}(\omega_{\Delta}) = 50.3 - 2.59\omega_{\Delta} - 5.77\omega_{\Delta}^2 - 7.31\omega_{\Delta}^3 - 4.22\omega_{\Delta}^4 + 2.84\omega_{\Delta}^5$$

As part of the RDT technique the result is reversed in frequency giving:

$$\tilde{R}(-\omega_{\Delta}) = 50.3 + 2.59\omega_{\Delta} - 5.77\omega_{\Delta}^2 + 7.31\omega_{\Delta}^3 - 4.22\omega_{\Delta}^4 - 2.84\omega_{\Delta}^5$$

Both responses are shown overlaid on the same graph in fig 5.13. Subtracting the reversed result from the original removes the even order terms and doubles the odd order terms from the series to give the result:

$$\tilde{R}(\omega_{\Delta}) - \tilde{R}(-\omega_{\Delta}) = \tilde{A}(\omega_{\Delta}) = -5.18\omega_{\Delta} - 14.62\omega_{\Delta}^3 + 5.68\omega_{\Delta}^5 \quad (5.16)$$

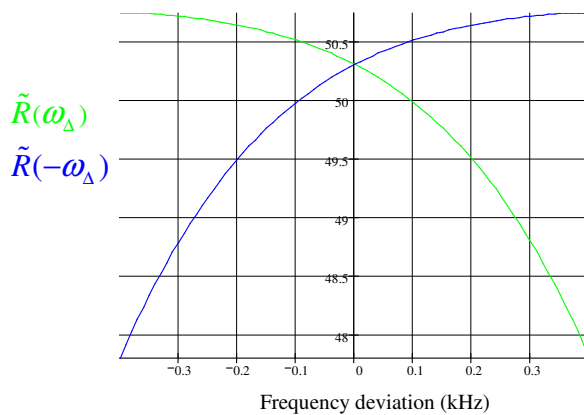


Fig 5.13: Forward and reversed series power expansions.

Any constant term added to the filter response representing the gain of the recording as discussed in chapter 4, would also be removed by the subtraction process. The result is that the RDT response at zero frequency deviation is normalised through zero. Further, the ROI which has a relatively small frequency deviation from its centre frequency, results in the response approximating to a straight line. Note, that if the signal is passed through the system twice as a result of a copying process, both the reversed and original series

coefficients will be doubled. This will also result in the doubling of odd order terms after the subtraction process:

$$2\tilde{A}(\omega_{\Delta}) = -10.36\omega_{\Delta} - 29.24\omega_{\Delta}^3 + 11.36\omega_{\Delta}^5$$

Plotting this result along with that of the single pass (5.16), shows a clear distinction between the two results, fig 5.14.

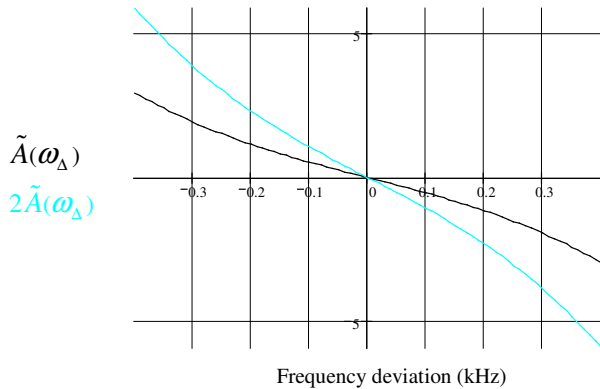


Fig 5.14: Shows the normalising and linearising effects over the spectral ROI of the RDT of the log of the power spectrum using single and double passes through the Butterworth filter.

The shapes of the responses produced by the RDT in fig 5.14 are consistent with the definition of an odd function [135 ch2.2]:

$$f(-\omega) = -f(\omega) \tag{5.17}$$

A Butterworth filter has been used for illustration purposes but the results apply equally for other low-pass filter types. The approximate linearisation produced by the RDT results from the removal of all the even order terms describing the spectral shape, in conjunction with having a relatively small frequency deviation away from the centre frequency of the ROI. Typical deviations in practice would be between ± 300 Hz and ± 500 Hz depending on recorder type and sampling rate. Successive passes of the signal through a low-pass response such as would occur in a copying process, increases the value of the odd order terms, producing a greater slope value in the overall linear trend. This provides discrimination between an original and a copy recording. The RDT also produces a normalisation of the spectral response by removal of any constant terms. This normalisation allows for the direct comparison of responses from recordings produced at different recording levels and from different recording systems.

5.5 The Coefficient Distribution Theory of the RDT

In the previous section, it has been shown that the response across the low-pass transition region of the recording system can be transformed by the RDT into a good approximation of a linear function over the ROI. Regression analysis techniques can be applied to

establish the relationship between the RDT of log-power and frequency (section 5.6). In order to analyse the performance of the regression analysis, the form of the pdf of the RDT is required. The purpose of this section is therefore to show the derivation of the statistical model of the probability density function (pdf) for the individual frequency coefficients of the RDT of the log-power spectrum. A pdf is defined on a continuous interval that has a total area of unity. The derivation of the distribution represents the probability densities of individual frequency coefficients gathered over a number of Fourier transforms K .

The derivation for the pdf has been carried out in stages, following a sequence that reflects the order of signal processing as shown in fig 5.15:

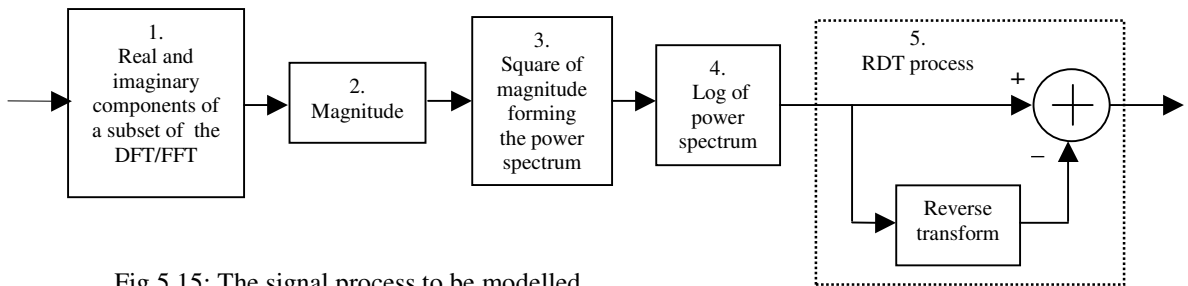


Fig 5.15: The signal process to be modelled.

The five sequential blocks shown in fig 5.15 can be described as follows:

1. The real and imaginary coefficients produced from a subset of frequencies of a DFT/FFT.
2. The magnitude formed by taking the root of the sum of squares of real and imaginary coefficients.
3. The power spectrum produced from the square of magnitude.
4. The log of the power spectrum.
5. The RDT of the log of the power spectrum.

5.5.1 The Real and Imaginary Probability Density Functions of the DFT

The DFT of a stochastic time-domain signal produces a set of coefficients consisting of real and imaginary parts:

$$X(\omega) = X_R(\omega) + jX_I(\omega)$$

As the probability densities prior to the RDT will refer to a single arbitrary frequency coefficient, reference to the frequency of the coefficient has not been included in the mathematical description of any pdf or associated graphs.

The DFT provides a linear transformation between the time and frequency-domains. Assuming the time-domain data comes from a Gaussian pdf, then as every linear

function of a Gaussian process is another Gaussian process [117 p292], the real and imaginary parts of the transform will also be Gaussian. This statement is widened by Scoukens et al [114] who show, based on the Central Limit Theorem, that the Gaussian pdf for the coefficients of the real and imaginary parts hold for random time-domain signals in general, not just time domain signals having a Gaussian pdf. This result applies to both the real and imaginary coefficients of the DFT when derived from the sum of a large number of random time-domain samples.

Considering a white noise time domain sequence, the resulting distribution for the real X_R and imaginary X_I parts of a single transform point will take the form of two independent, zero mean, equal variance, Gaussian distributed random variables [114], [142]. It has been assumed that the correlation for a white noise sequence filtered by a low pass response, having a cut off frequency close to the Nyquist limit, will be low and the probability densities will therefore not significantly change from the white noise condition. This is fully supported by Schoukens et al [114], who also show that when the noise is highly correlated, providing the Fourier transform size is high (≥ 128 points) the form of the pdf does not significantly change. This is confirmed using empirical data in subsection 7.1.3. The probability densities of both real and imaginary coefficients are therefore:

$$\begin{aligned} \text{pdf of a real coefficient :} \quad f_{X_R}(x) &= \frac{1}{\sigma_{ft} \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_{ft}^2}} \\ \text{pdf of an imaginary coefficient :} \quad f_{X_I}(x) &= \frac{1}{\sigma_{ft} \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_{ft}^2}} \end{aligned}$$

For the purpose of illustration unless otherwise stated the variance of the real and imaginary parts σ_{ft}^2 have been given the arbitrary value of 2. This illustrative Gaussian pdf is described for the real coefficient by (5.18) and shown in fig 5.16 and the illustrative imaginary coefficient will be identical.

$$f_{X_R}(x) = \frac{1}{1.414\sqrt{6.28}} e^{-\frac{x^2}{4}} \tag{5.18}$$

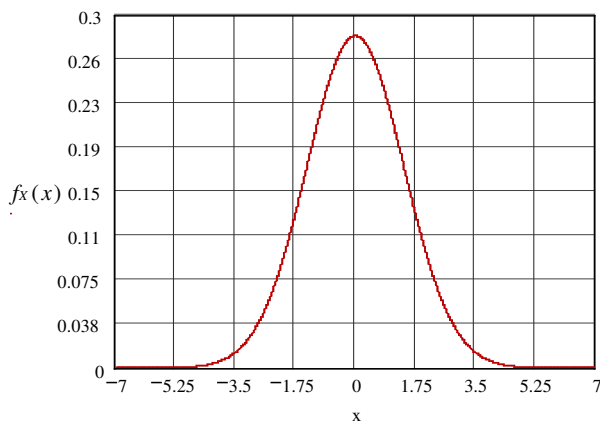


Fig 5.16: The pdf for the real or the imaginary coefficient of a single point of an FFT taken over an infinite number of transforms.

5.5.2 Magnitude pdf

The magnitude Z of independent random variables X_R and X_I representing the real and imaginary coefficients of the frequency transform is given by the square root of the sum of the squares of the real and imaginary coefficients:

$$Z = \sqrt{X_R^2 + X_I^2} \quad (5.19)$$

Where $Z = Z_\omega[k]$ and represents a random variable of the magnitude of a single Fourier coefficient having a value ω taken over K Fourier transforms. The random variable Z identified by (5.19) will have a Rayleigh pdf described by (5.20) as formulated for narrow band random noise in well-known papers by Rice [142] and [143]. The Rayleigh pdf is an asymmetrical function and is shown in fig 5.17.

$$m_z(z) = \begin{cases} \frac{z}{\sigma_{ft}^2} e^{\left(\frac{-z^2}{2\sigma_{ft}^2}\right)} & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (5.20)$$

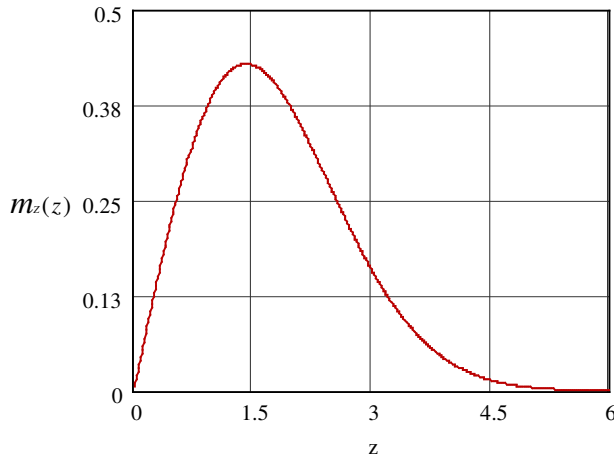


Fig.5.17: The magnitude of the FFT has a Rayleigh pdf .

From appendix A.1 it has been shown that the mean, variance and mode values for this probability density are:

$$\text{mean: } \bar{Z} = \sqrt{\frac{\pi}{2}} \cdot \sigma_{ft} \quad \text{variance: } \sigma_Z^2 = 0.429\sigma_{ft}^2 \quad \text{mode: } \hat{Z} = \sigma_{ft}$$

5.5.3 Power pdf

The power spectrum has been formed by taking the square of the magnitude of the Fourier transform, therefore (5.19) becomes:

$$Z^2 = X^2 + Y^2 \quad (5.21)$$

It is desired to determine the pdf of the power variable from the known pdf of the

magnitude described by (5.20). Having obtained the relationship between the random variables of magnitude and power, a transformation of variables is required, [144 ch2.3]. A new random variable Q is created and is derived from the magnitude variable Z , where $Q = Q_\omega[k]$ is a random variable representing the power of a single Fourier coefficient, having a frequency value ω taken over K Fourier transforms. Using a function $q = g_z(z)$ the pdf of Q will be given by:

$$p_Q(q) = \frac{m_z(g_Q^{-1}(q))}{\left| \frac{dq}{dz} \right|} \quad (5.22)$$

Now if

$$q = g_z(z) = z^2 \quad \text{hence} \quad z = \pm q^{1/2}$$

Then

$$\frac{d}{dz} z^2 = 2z = 2q^{1/2}$$

The Rayleigh pdf $m_z(z)$ represents the magnitude of the Fourier transform and has no meaning for negative values of z , therefore substituting the distribution of q for the positive root (5.22) can be written as:

$$p_Q(q) = \frac{m_z(q^{1/2})}{2q^{1/2}} \quad (5.23)$$

Combining (5.23) with (5.20) the form of the distribution for the power variable is found to be exponential:

$$p_Q(q) = \frac{1}{2\sigma_{f_i}^2} e^{-\frac{q}{2\sigma_{f_i}^2}} \quad q \geq 0$$

$$= 0 \quad q < 0 \quad (5.24)$$

This is a Chi-square function [103 p93] given by (5.25) with n equal to two degrees of freedom and Γ is the Gamma function [145]:

$$p_Q(q; n) = \frac{1}{2^{\frac{n}{2}} \cdot \sigma_{f_i}^2 \cdot \Gamma\left(\frac{n}{2}\right)} \cdot q^{\left(\frac{n}{2}-1\right)} \cdot e^{-\frac{q}{2\sigma_{f_i}^2}} \quad q > 0$$

$$= 0 \quad q \leq 0 \quad (5.25)$$

The Chi-square function χ^2 is the sum of squares of independent normally distributed random variables with zero means and common variance. The number of degrees of

freedom n equates to the number of independent random variables concerned. The form of the pdf for (5.24) & (5.25) is shown in fig 5.18. From appendix A.2 it has been shown that the mean, variance and mode values for this function are:

$$\text{mean: } \bar{Q} = 2\sigma_{ft}^2 \quad \text{variance: } \sigma_Q^2 = 4\sigma_{ft}^4 \quad \text{mode: } \bar{Q} = 0$$

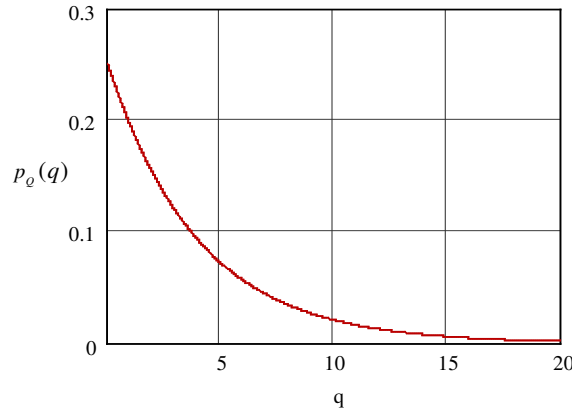


Fig.5.18: The power spectrum has an exponential pdf .

As a point of note, the power spectrum could have been formed directly from the sum of squares of the real and imaginary parts of the DFT (5.21), therefore, the Chi-square distribution may have been assumed without reference to the magnitude/Rayleigh pdf.

5.5.4 Logarithm of the Power pdf

The logarithm of the power spectrum is given by:

$$R = \log_e Q$$

It is desired to determine the pdf of the logarithm of the power variable from the known pdf of the power variable described by (5.24). Having obtained the relationship between the random variables of power and log of power, a transformation of variables is required, [144 ch2.3]. A new random variable R is created and is derived from the power variable Q , where $R = R_\omega[k]$ is a random variable representing the power of a single Fourier coefficient, having a value ω taken over K Fourier transforms. Using a function $r = h_Q(q)$ the pdf of R will be given by:

$$l_R(r) = \frac{p_Q(h_R^{-1}(r))}{\left| \frac{dr}{dq} \right|}$$

Now if :

$$r = h_Q(q) = \log_e(q)$$

Then

$$\frac{d}{dq} \log_e(q) = \frac{1}{q} = \frac{1}{e^r} \quad \text{hence } q = e^r$$

Substituting the distribution for r :

$$l_R(r) = \frac{p_Q(e^r)}{1/e^r} \quad (5.26)$$

Combining (5.24) with (5.26) the distribution becomes:

$$l_R(r) = \frac{e^r}{2\sigma_{ft}^2} e^{-\frac{e^r}{2\sigma_{ft}^2}} \quad -\infty \geq r \leq \infty \quad (5.27)$$

The pdf described by (5.27) is illustrated in fig 5.19.

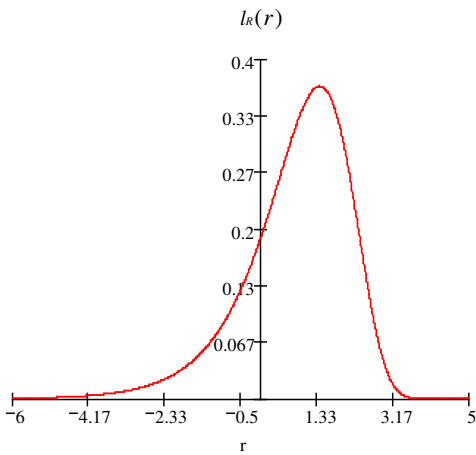


Fig. 5.19: The pdf for the log of power spectrum.

From appendix A.3 the variance and mode values for this function have been shown to be:

$$\text{variance: } \sigma_R^2 = \frac{\pi^2}{6} = 1.645 \quad \text{mode: } \hat{R} = \ln(2\sigma_{ft}^2)$$

For a change in original variance of the real and imaginary coefficients σ_{ft}^2 , the log of the power pdf has a variance and shape that remains constant, but shifts in location along the r scale. Therefore, the mean of the log of the power spectrum is totally dependent on the original variance and this is demonstrated in fig 5.20 where the probability densities for different original variance values are displayed. It is also worth noting that the mean does not coincide with the mode and as the shape of the function is a constant, the difference between the mode and the mean will also be a constant. It has been found from the pdf (appendix A.3) that the difference between the mean and the mode is equal to the well-known Eulers constant, Υ , having a value of 0.577 [146]. From this a closed form equation for the mean of the pdf is obtained:

$$\bar{R} = \ln(2\sigma_{ft}^2) - \Upsilon \quad (5.28)$$

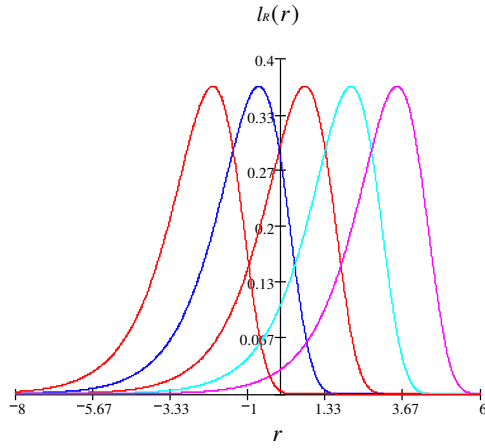


Fig 5.20: Probability densities of the log-power spectrum, shown for 5 different original variance values. The pdf's have a constant shape but different mean values.

5.5.5 Reverse Difference Transform (RDT) pdf

From the pdf of the log-power spectrum, the form of the pdf of the RDT is required. The RDT pdf $\varphi_A(a)$ of a random variable A , may be obtained by establishing the pdf of the difference between two random variables R_{ω_1} and R_{ω_2} . Both random variables are described by the probability densities for the logarithmic power (5.27), each taken from a different spectral coefficient, with frequencies ω_1 and ω_2 .

Let R_{ω_1} and R_{ω_2} consisting of values taken over K Fourier transforms, be statistically independent random variables, having log-power density functions $l_{R_{\omega_1}}(r)$ and $l_{R_{\omega_2}}(r)$, and let the difference between these two variables for each transform represent the RDT random variable A :

$$A[k] = R_{\omega_1}[k] - R_{\omega_2}[k]$$

It is known that when two random variables are statistically independent, the probability density of a combination of the two variables will be the convolution of their probability densities [144 ch3.5]. Given the two log of power probability densities, $l_{R_{\omega_1}}(r)$ and $l_{R_{\omega_2}}(r)$, it is therefore necessary to calculate the convolution to obtain the pdf $\varphi_A(a)$ of the RDT variable A . The required convolution is:

$$\varphi_A(a) = \int_{-\infty}^{\infty} l_{R_{\omega_2}}(r) \cdot l_{R_{\omega_1}}(a+r) dr$$

The two log-power probability densities will be identical except for their mean values \bar{R} (5.28). The mean value for each log-power pdf is determined by the original variance value σ_{ft}^2 of the real and imaginary parts of the Fourier transform. Each pdf will come from a different Fourier coefficient within the spectral ROI and will have a different

original variance value σ_{ft}^2 , designated $\sigma_{\omega_1}^2$ and $\sigma_{\omega_2}^2$. The RDT can then be found from:

$$\varphi_A(a) = \int_{-\infty}^{\infty} \left[\frac{e^r}{2 \cdot \sigma_{\omega_1}^2} \cdot e\left(\frac{-e^r}{2 \cdot \sigma_{\omega_1}^2}\right) \right] \cdot \left[\frac{e^{r+a}}{2 \cdot \sigma_{\omega_2}^2} \cdot e\left(\frac{-e^{r+a}}{2 \cdot \sigma_{\omega_2}^2}\right) \right] dr \quad (5.29)$$

After integrating (5.29) becomes:

$$\varphi_A(a) = \frac{\sigma_{\omega_1}^2 \cdot \sigma_{\omega_2}^2 \cdot e^a}{\left(\sigma_{\omega_2}^2 + \sigma_{\omega_1}^2 \cdot e^a\right)^2} \quad (5.30)$$

Equation (5.30) is a form of **logistic** probability density and will be termed the RDT pdf.

From appendix A.4 it has been shown that the mean and variance values for this pdf are:

$$\text{mean: } \bar{A} = 2 \ln\left(\frac{\sigma_{\omega_2}}{\sigma_{\omega_1}}\right) \quad \text{variance: } \sigma_A^2 = \frac{\pi^2}{3} = 3.29$$

The mean is therefore dependent on the ratio of original variance values σ_{ω_1} and σ_{ω_2} and the variance for the RDT is found to have a constant value of 3.29 for all values of original variance except when $\sigma_{\omega_1} = 0$ and/or $\sigma_{\omega_2} = 0$ which does not occur in practice.

The pdf of the RDT (5.30) can also be expressed in trigonometric form. For an RDT probability density having a mean value \bar{A} of zero, where $\sigma_{\omega_1}^2 = \sigma_{\omega_2}^2$, the numerator and denominator can be multiplied by $e^{-a/2}$ and trigonometric substitutions made, where after simplification the new function becomes a **hyperbolic secant squared** function:

$$\frac{1}{4} \text{sech}^2\left(\frac{a}{2}\right)$$

For a function where $\sigma_{\omega_1}^2 \neq \sigma_{\omega_2}^2$ the mean \bar{A} as derived for the exponential form is non zero and has to be incorporated:

$$\tilde{\varphi}(a) = \frac{1}{4} \text{sech}^2\left(\frac{a - \bar{A}}{2}\right) \quad (5.31)$$

This sech squared function confirms that the pdf of the RDT is symmetrical. The mode for this function can be found by differentiating (5.31) with respect to the mean and equating to zero:

$$\frac{d}{d\bar{A}} \frac{1}{4} \text{sech}^2\left(\frac{a - \bar{A}}{2}\right) = \frac{1}{2} \text{sech}^2(a - \bar{A}) \cdot \tanh(a - \bar{A})$$

As expected, after equating to zero the value for the mode is found to be the same as the

mean. The RDT pdf (5.30) or (5.31) having a mean of zero is shown by the blue trace in fig 5.21. Also shown by the red trace for comparison purposes, is a Gaussian pdf matched by selecting a variance value to produce the same probability density value at the mean/mode as that of the RDT pdf.

To compare the RDT pdf to a Gaussian pdf the additional higher order central moments of skewness and kurtosis are required. The Gaussian pdf has a skewness of zero and a kurtosis of 3. The skewness and kurtosis for RDT pdf have been calculated in appendix A.4 and are found to be:

skewness: $\gamma_A = 0$ kurtosis: $K_A = 4.2$

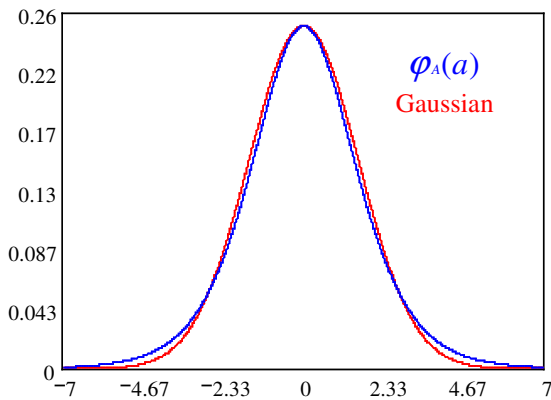


Fig. 5.21: The pdf of the RDT $\varphi_A(a)$. Also shown for comparison purposes is a Gaussian pdf matched to have the same probability density value at the mean/mode as that of the RDT by adjustment of its variance value. It can be seen that the tails decay at different rates.

The RDT pdf has a skewness of zero but has a higher kurtosis than that of Gaussian, making it supergaussian. The RDT pdf therefore has larger tails than the Gaussian pdf, and this has been highlighted in fig 5.22 by overlaying the upper tails of $\varphi_A(a)$ with a Gaussian pdf having the same variance.

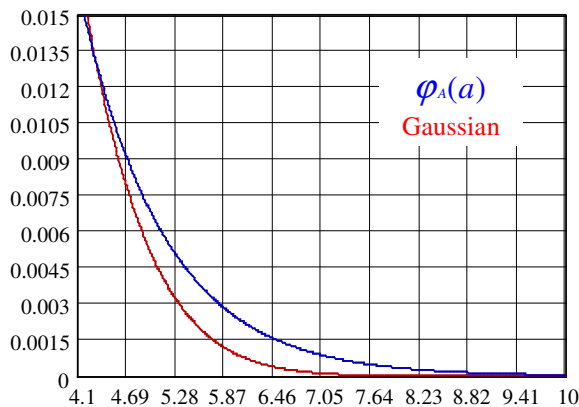


Fig 5.22: Upper tails of $\varphi_A(a)$ and Gaussian pdf's. The Gaussian having the same variance value as $\varphi_A(a)$. A larger tail is seen for $\varphi_A(a)$.

5.5.6 Statistical Attributes of the RDT

In summary, the Fourier coefficients are comprised of independent real and imaginary

parts that have a Gaussian pdf. Finding the magnitude of the Fourier coefficient requires a combination of these real and imaginary parts and results in a Rayleigh pdf. A power transformation is produced by squaring the magnitude of the random variable and changes the Rayleigh into an exponential pdf. A further transformation occurs by taking the logarithm of the power spectrum, producing an asymmetrical density that has the important property of having a constant variance and shape, changing only its mean value for a change in original variance of the real and imaginary parts of the Fourier coefficient. The RDT probability density is produced from two log-power probability densities, each formed from a different Fourier coefficient. The resultant pdf is symmetrical and similar to Gaussian, having a constant variance and shape, changing only in mean value for a change in variance of either of the two log-power density functions. The mean is a biased estimator for the log-power spectrum but unbiased for the RDT of the log-power spectrum. The final pdf of the RDT can be summarised as:

1. Having a constant variance that is independent of the original variance of the real and imaginary parts of the Fourier transform and is therefore independent of the variance of the original time-domain signal.
2. Is symmetrical and similar in shape to a Gaussian pdf. As discussed in the next 2 chapters, after carrying out a spectral averaging process the RDT pdf will converge very closely to that of Gaussian.

Both of these attributes are important in the application of regression techniques examined in the next section.

5.6 Slope Estimation Using Regression Techniques

The RDT has been found to produce an approximately linear vector of frequency related coefficients over the spectral ROI. This section, investigates the requirement to fit a linear regression model to the coefficients to ascertain the rate of change of spectral level to frequency, in the form of a single parameter which will be referred to as the slope [136_{ch1.3}]. The value of the slope parameter cannot be known with complete certainty and investigation into the optimisation of the slope estimate has been undertaken using regression analysis. Regression analysis is a collection of statistical techniques for determining estimates for the parameters of the regression model [147].

5.6.1 Linear Regression Analysis of the RDT of the Log-Power Spectrum

The application calls for a fixed single regressor ω_{Δ} associated with the frequency

deviation over the ROI of the RDT of the log-power spectrum, and a random variable A related to the magnitude of the RDT of the log-power spectrum. A is a set of values referred to as the population and $A(\varpi_{\Delta})$ is a member of a partition of A , known as a sub-population. So for each distinct value of ϖ_{Δ} there is a sub-population of values from the entire population of variable A . Making the assumption that the mean values of these sub-populations fall on a straight line, the regression model will relate the means of the set of values $A(\varpi_{\Delta})$ by the relationship [136 ch1.3]:

$$E(A(\varpi_{\Delta})) = \beta_0 + \beta_1 \varpi_{\Delta}$$

Where β_0 and β_1 are the regression coefficients: β_0 is the intercept and β_1 is the slope.

The mean values of the sub-populations $A(\varpi_{\Delta})$ are a linear function of ϖ_{Δ} but the measured values $a(\varpi_{\Delta})$ of sub-population $A(\varpi_{\Delta})$ do not fall on an exact straight line. This can be generalised to a probabilistic linear model with the assumption that the expected value of $A(\varpi_{\Delta})$ is a linear function of ϖ_{Δ} . For a fixed value ϖ_{Δ_j} , the value $a(\varpi_{\Delta_j})$ will be represented by a linear model plus its error ε and observations in the sample may be expressed as [136 ch1.3]:

$$a(\varpi_{\Delta_j}) = \beta_0 + \beta_1 \varpi_{\Delta_j} + \varepsilon_j \quad j = 0, 1, \dots, J-1 \quad (5.32)$$

Where ε_j is a random error term having a mean $E(\varepsilon_j) = 0$, and a variance $\sigma^2(\varepsilon_j)$ and j is the j^{th} RDT transform point taken from a total of J points of the RDT transform.

Therefore, a distribution of values exist for each fixed value of ϖ_{Δ_j} and as β_0 and β_1 are constants the variance is solely determined by the error ε_j . The pdf of each variable $A(\varpi_{\Delta})$ is that derived for the pdf of the RDT in the previous section, $\varphi_A(a)$, having a constant variance equal to 3.29.

When only sample data are available, averaging produces the best predicted value of a randomly chosen observation from the sub-population and is equal to the estimated sub-population mean. Since $E(\varepsilon_j) = 0$:

$$E(a(\varpi_{\Delta_j})) = E(\beta_0 + \beta_1 \varpi_{\Delta_j} + \varepsilon_j) = \beta_0 + \beta_1 \varpi_{\Delta_j} + E(\varepsilon_j) = \beta_0 + \beta_1 \varpi_{\Delta_j}$$

Figure 5.23 represents the regression model passing through the means of the RDT probability densities, shown for 3 different values of ϖ_{Δ} .

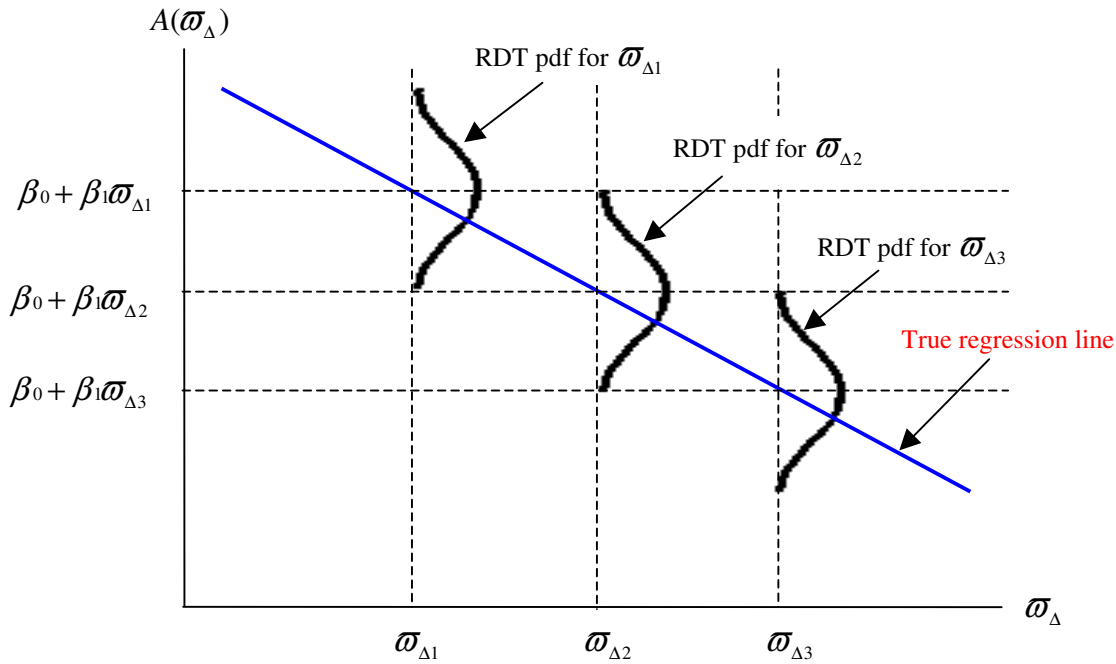


Fig 5.23: Linear regression model showing the RDT distribution as a function of ϖ_{Δ}

In summary the slope of the regression can be calculated from the mean change in A for a unit increase in ϖ_{Δ} and is the basis for determining a single parameter value for the RDT of the log-power spectrum, indicating original or copied status of a recording.

5.6.2 Assumptions for the Model

Linear models derived using regression analysis are based on a number of assumptions about the population and sample parameters [148 ch3.3]:

1. The order of the model is correct i.e. linear first order. This is approximately true for the RDT of the log-power spectrum providing the frequency deviation away from the centre frequency of the ROI is relatively small, as shown in section 5.4. The model order will be considered further in subsection 7.3.5.
2. The error terms ε_j are uncorrelated random variables with zero mean.
3. The standard deviations of the sub-populations are independent of the value of ϖ_{Δ} , and are the same for each sub-population, this is termed the homogeneity of standard deviations, and was shown to be true for the pdf of the RDT in section 5.5.5.
4. Tests of hypotheses and interval estimation also require the additional assumption that the errors have a Gaussian pdf.

In practice departures from the assumptions or validations of the theory for the model can be tested by analysing the observed errors using residual analysis (section 7.3).

5.6.3 Regression Through the Origin

The random variable A produced by the RDT has been considered with respect to the general linear regression model. The following two factors require a change to the model:

1. The RDT frequency is shown in terms of deviation away from the ROI centre frequency Ω , and therefore at the centre frequency ϖ_{Δ} is zero.
2. It has been shown that the spectrum power is normalised by the RDT process so that at Ω the power is 0 dB.

Taking into account 1 and 2, the regression function will go through the origin (0,0), β_0 is therefore zero and the general linear model of (5.32) becomes:

$$a_j = \beta_1 \varpi_{\Delta j} + \varepsilon_j \quad (5.33)$$

The regression function for (5.33) is:

$$E(A(\varpi_{\Delta})) = \beta_1 \varpi_{\Delta}$$

which is a straight line through the origin, with a slope of β_1 [136 ch4.4].

5.6.4 Point Estimators for Linear Function β_1

There are two common approaches for estimating the slope value β_1 :

1. The method of least squares
2. The method of maximum likelihood

The Method of Least Squares

The method of least squares is based on minimising the sum of the squares of vertical deviations Λ , where:

$$\Lambda = \sum_{j=0}^{J-1} \varepsilon_j^2$$

The least squares estimator of β_1 in the regression model of (5.33) can be obtained by minimising Λ given by:

$$\Lambda = \sum_{j=0}^{J-1} (a_j - \beta_1 \varpi_{\Delta j})^2 \quad (5.34)$$

According to the method of least squares, the estimator of slope coefficient β_1 is the value of $\hat{\beta}_1$ that will minimise Λ . A value for $\hat{\beta}_1$ can be found using analytical techniques [136 ch1.6], where considering the regression through the origin model, the value of the slope estimate is given by:

$$\hat{\beta}_1 = \frac{\sum_{j=0}^{J-1} \varpi_{\Delta j} a_j}{\sum_{j=0}^{J-1} \varpi_{\Delta j}^2} \quad (5.35)$$

If the first three assumptions regarding the model, made in section 5.6.2 are met, then the single strongest justification for linear least squares estimation is the Gauss-Markov theorem. This theorem states that the least squares estimator for regression coefficient $\hat{\beta}_1$ will be unbiased and have minimum variance among all unbiased linear estimates [136 p20].

The Method of Maximum Likelihood

The method of maximum likelihood for the regression model [136 pp30-32] is based on choosing estimators that have values of the parameters that are most consistent with the sampled data and may be used to estimate the slope β_1 of the coefficients of the RDT of the log of the power spectrum that is most likely to have occurred given the corresponding frequency values ϖ_{Δ} . It follows, that the maximum likelihood requires the functional form of the probability distribution of the error terms to be known. From section 5.5.5 it was shown that the error distribution of the RDT of the log of the power pdf for a particular frequency ϖ_{Δ} is a hyperbolic secant squared function:

$$\tilde{\varphi}_A(a) = \frac{1}{4} \operatorname{sech}^2 \left(\frac{a - \bar{A}}{2} \right)$$

Maximum likelihood estimation produces a slope value that makes the calculated values of the spectral coefficients A most likely to have occurred given the corresponding frequency values ϖ_{Δ} . The conditional probability of a value a_j given a value $\varpi_{\Delta j}$, $P(a_j | \varpi_{\Delta j})$ is:

$$\frac{1}{4} \operatorname{sech}^2 \left(\frac{a_j - b_1 \varpi_{\Delta j}}{2} \right) \text{ where } b_1 \text{ is the slope estimate}$$

For independent values, the joint probability for all the coefficients can be found:

$$L(b_1) = \prod_{j=0}^{J-1} \frac{1}{4} \operatorname{sech}^2 \left(\frac{a_j - b_1 \varpi_{\Delta j}}{2} \right)$$

In order to simplify the calculation it is usual practice with maximum likelihood calculations to take the log:

$$\log_e L(b_1) = \sum_{j=0}^{J-1} \log_e \left[\frac{1}{4} \operatorname{sech}^2 \left(\frac{a_j - b_1 \varpi_{\Delta j}}{2} \right) \right]$$

The turning point coincides with the maximum value, and is found by differentiating:

$$\frac{d \log_e L(b_1)}{db_1} = \sum_{j=0}^{J-1} \varpi_{\Delta j} \tanh\left(\frac{a_j - b_1 \varpi_{\Delta j}}{2}\right) \quad (5.36)$$

The optimum maximum likelihood slope estimate of b is where the value of b results in (5.36) becoming zero:

$$\sum_{j=0}^{J-1} \varpi_{\Delta j} \tanh\left(\frac{a_j - b_1 \varpi_{\Delta j}}{2}\right) \Big|_{\beta_1=b_1} = 0 \quad (5.37)$$

The maximum likelihood estimator is approximately a minimum value unbiased estimator [149 p278].

When the pdf of the error terms is Gaussian then the Method of Least Squares and Maximum Likelihood are equivalent [136 p34]:

$$\hat{\beta}_1 = b_1$$

It will be demonstrated empirically in chapter 7, that the difference between least squares and maximum likelihood methods of estimation of the slope β_1 of the RDT of the log-power spectrum are insignificant. The least squares method, will be that chosen for slope estimation. This is due to the estimator $\hat{\beta}_1$ having a near optimal value, providing assumptions 1 to 3 of section 5.6.2 are met and the ease of calculation of the least squares regression parameter $\hat{\beta}_1$ (5.35) compared to that of solving equation (5.37) for the maximum likelihood regression parameter b_1 .

Averaging and Least Squares

The estimator for slope β_1 is unbiased, and this indicates that arithmetic averaging K transforms at J frequency coefficients and then fitting a least squares line, produces the same result as fitting the line directly to all J by K coefficients. Rewriting (5.35) to include K measurements for each j frequency coefficient gives:

$$\beta_1 = \frac{1}{K} \sum_{k=0}^{K-1} \frac{\sum_{j=0}^{J-1} \varpi_{\Delta j} a_{jk}}{\sum_{j=0}^{J-1} \varpi_{\Delta j}^2} = \frac{\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \varpi_{\Delta j} a_{jk}}{K \sum_{j=0}^{J-1} \varpi_{\Delta j}^2}$$

Expanding and rearranging:

$$= \frac{\varpi_{\Delta 1} (a_{00} + a_{01} + \dots + a_{0K-1}) + \varpi_{\Delta 2} (a_{10} + a_{11} + \dots + a_{1K-1}) + \dots + \varpi_{\Delta J} (a_{J0} + a_{J1} + \dots + a_{JK-1})}{\varpi_{\Delta 0}^2 + \varpi_{\Delta 1}^2 + \dots + \varpi_{\Delta J-1}^2}$$

It can be seen that the least squares estimate of all the data is equivalent to averaging first then applying a least squares estimate to the result.

5.6.5 Variance in the Slope Estimate

The value of the slope estimate cannot be known with complete certainty. The variance in the slope estimate represents the uncertainty in the overall analysis procedure and it is shown in this section that the slope variance is inversely proportional to the data size.

From section 5.1, given a fixed overall data size N , an increase in frequency transform size M results in a proportionally reduced number of segments K available for averaging. Assuming second order stationarity of the noise signal [125 pp81-84], each segment k will be identically distributed and independent. Therefore, averaging K transforms will reduce the variance of a single spectral estimate, σ_k^2 , by a factor of K giving the final variance, σ_{avg}^2 , as:

$$\sigma_{avg}^2 = \frac{\sigma_k^2}{K} \quad (5.38)$$

The slope estimate $\hat{\beta}_1$ refers to the different values of slope that would be obtained with repeated sampling when the levels of predictor variable ϖ_Δ are held constant from sample to sample. Using the assumptions made for the regression model in section 5.6.2, the mean and variance of the distribution of the slope are defined as [136 pp44-48]:

$$\begin{aligned} \text{mean: } E\{\hat{\beta}_1\} &= \beta_1 \\ \text{variance: } \sigma^2\{\beta_1\} &= \frac{\sigma^2}{\sum_{j=0}^{J-1} (\varpi_{\Delta j} - \bar{\varpi})^2} \end{aligned} \quad (5.39)$$

If the errors have a Gaussian pdf then the distribution of the slope will also be Gaussian. The σ^2 in the numerator of (5.39) is the constant variance previously derived for the spectral coefficients of the RDT pdf, σ_A^2 . Two factors need to be taken into account and incorporated into (5.39) :

1. The variance of the spectrum is reduced as a result of averaging the number of segments K together (5.38).
2. Due to data redundancy discussed in the next section, the slope of the RDT only needs to be calculated for half of the vector, therefore the maximum number of index points J needs to be halved.

Therefore the measurement variance in the slope estimate becomes:

$$\sigma_M^2\{\beta_1\} = \frac{\sigma_A^2}{K \cdot \sum_{j=0}^{(J-1)/2} (\omega_{\Delta j} - \bar{\omega})^2} \quad (5.40)$$

The overall transform size dictates how many spectral coefficients are found over the ROI. From (5.40) the magnitude of the slope-variance can be reduced by increasing the number of transform points over the ROI. This produces an increase in the value of the sum of the squares of the frequency coefficients. Increasing the number of averages K also reduces the slope-variance. However, for a fixed data size N there is a trade-off between increasing the transform size and increasing the number of averages. It is shown in the graph of fig 5.24 that for a fixed data size the slope-variance remains relatively constant for changing ratios of transform points over the ROI to the number of averages. The slope-variance will therefore be dependent on the overall size of the data and the bandwidth of the ROI. When the number of transform points over the ROI is fixed, $\sigma_M^2\{\beta_1\}$ is inversely proportional to K that in turn will be directly proportional to N . Therefore:

$$\sigma_M^2\{\beta_1\} \propto N^{-1}$$

The product of the slope-variance and data size will be a constant:

$$\sigma_M^2\{\beta_1\} \cdot N = C \quad (5.41)$$

Therefore, the slope-variance can be found from:

$$\frac{C}{N} = \sigma_M^2\{\beta_1\} \quad (5.42)$$

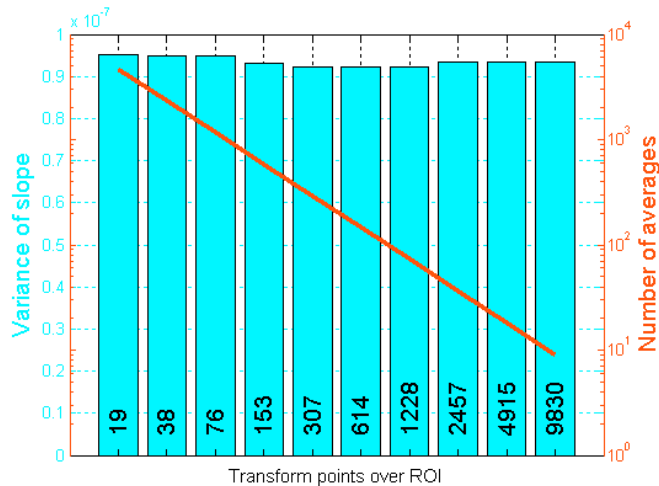


Fig 5.24: For a fixed data size the variance of the slope remains relatively constant for a changing ratio of transform size to number of averages.

Using (5.40) to obtain a value for $\sigma_M^2\{\beta_1\}$ at an arbitrary data size N and a bandwidth of 600 Hz, a value for the constant C can be calculated from (5.41). Using this value for the constant, a graph has been produced from (5.42) showing the variance of the slope

$\sigma_M^2\{\beta_1\}$ for a changing data length N , fig 5.25. Using this graph, the level of slope-variance introduced by the measurement system can be ascertained for varying lengths of recorded audio material.

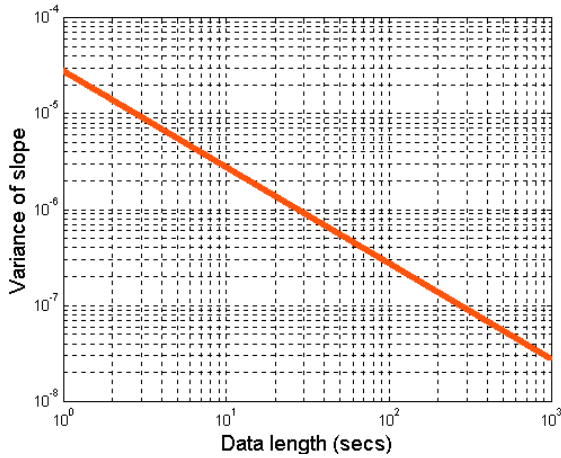


Fig 5.25: Change of measurement slope-variance $\sigma_M^2\{\beta_1\}$ for a change in data length. Data length shown in seconds for a 32 kHz sampling rate.

From the result shown in fig 5.25; given a null hypothesis that a measured slope value represents the true value of an original recording and the alternative hypothesis that the slope value represents the true value of a copied recording, and also knowing the variance of the distribution (assumed Gaussian) for a given data length, critical slope values over the probability density can be defined where the null hypothesis will or will not be rejected. Rejecting the null hypothesis when it is true, produces a type I error, and accepting it when it is false, produces a type II error [149 pp296-298]. It can be seen from fig 5.25 that the variance of a distribution could be proportionally reduced by increasing the data length. Providing the critical values are held constant, a reduction in variance would reduce both type I and type II errors and discrimination between original and copied recordings could be made arbitrarily high. However, as discussed in following chapters, account has to be taken of other factors that will also introduce uncertainties into the estimate. These further uncertainties reduce the benefits of producing very low slope-variance from the measurement system.

5.6.6 Data Redundancy Resulting from the RDT Process and its Significance on β_1

It will be shown in this section that for the purposes of calculating the slope value, 50% of the data produced by the RDT is redundant, requiring only one half of the data for the estimate.

The application of the RDT results in a vector that can be considered in two halves. The half of the vector containing positive data values, has co-ordinates of $(+A, +\sigma_\Delta)$ or

($+A, -\varpi_{\Delta}$) and the other half of the vector, containing negative data values having co-ordinates of ($-A, -\varpi_{\Delta}$) or ($-A, +\varpi_{\Delta}$). The spectral response of the RDT has been shown to produce an odd function (5.17). The negative data values exhibit obvious symmetry and are identical to that of the positive data values except that they been multiplied by -1 . A typical sample data set illustrates these points in fig 5.26.

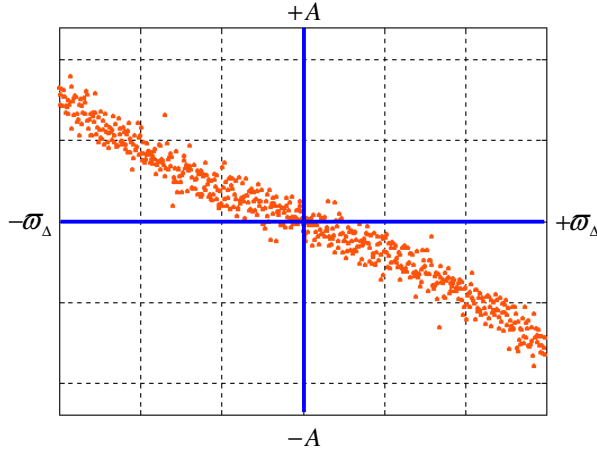


Fig 5.26: Negative half of data vector has obvious symmetry with that of the positive half.

What effect does this have on fitting a regression line to the data in the least squares sense? Intuitively, if both halves of the data vector are identical but 180° apart then calculated individually both halves will have identical slope values. If both halves of the data are left as one vector then the slope value will also be the same as that of half of the vector. The least square estimate of the slope β_1 where the regression is not through the origin is given by (5.43) [150 p48].

$$\beta_1 = \frac{\sum_{j=0}^{J-1} (\varpi_{\Delta j} - \bar{\varpi}) (a_j - \bar{A})}{\sum_{j=0}^{J-1} \varpi_{\Delta j}^2} \quad (5.43)$$

Where \bar{A} is the mean of the averaged RDT values \bar{A} . Analysing the slope as two separate halves β_{1+} and β_{1-} :

for β_{1-} : $\varpi_{\Delta} \leq 0$ when $j = 0, 1, 2, \dots, (J-1)/2$

for β_{1+} : $\varpi_{\Delta} \geq 0$ when $j = (J-1)/2, J/2, (J/2)+1, (J/2)+2, \dots, J-1$

If the two halves of the vector have the same slope value then:

$$\beta_{1-} = \frac{\sum_{j=0}^{(J-1)/2} (\varpi_{\Delta j} - \varpi_{\Delta-}) (a_j - \bar{A}_-)}{\sum_{j=0}^{(J-1)/2} \varpi_{\Delta j}^2} = \beta_{1+} = \frac{\sum_{j=(J-1)/2}^{J-1} (\varpi_{\Delta j} - \varpi_{\Delta+}) (a_j - \bar{A}_+)}{\sum_{j=(J-1)/2}^{J-1} \varpi_{\Delta j}^2}$$

The sum of the squares of the Fourier coefficients makes the value of the denominators of

both half vectors the same. Subtraction of the individual means of each half vector, \bar{A} and $\bar{\omega}_\Delta$, normalises the ω_Δ and A values about zero, therefore β_{1-} and β_{1+} are the same. For the complete vector incorporating both halves of the data, the mean values are zero, therefore (5.43) becomes:

$$\beta_1 = \frac{\sum_{j=0}^{J-1} \omega_\Delta \cdot a_j}{\sum_{j=0}^{J-1} \omega_\Delta^2} \quad (5.44)$$

The slope of the RDT is now represented by (5.44) and produces values for the numerator and denominator twice that of one half of the vector. Therefore, the slope value is the same whether calculated over both halves of the vector or over one half.

$$\beta_1 = \beta_{1+} = \beta_{1-}$$

It may be concluded that the slope-variance (5.40) as discussed in the previous section must take into account this redundancy by calculating the sum of squares for only the positive or negative range of frequency values.

5.7 Summary and Conclusions

This chapter has established theoretical relationships and performance parameters that will be used when assessing the empirical performance of the processes in chapters 6 and 7.

The signal process starts with a basic spectral estimation based on Bartlett's periodogram. This estimate is only required over a limited spectral region near to the Nyquist limit. With this in mind the capabilities of the chirp transform have been compared with that of the Fast Fourier transform. The results showed that the FFT was a better choice for the application.

An analysis of the novel transform named the RDT introduced in chapter 4 has been carried out by modelling a low-pass frequency response, using a 6th order power series expansion around the central frequency of interest. The results showed that the RDT of the log-power spectrum eliminates even order terms in the spectral envelope of a low-pass response. This leaves an approximately linear function over the ROI. Further, any constant terms were also removed by the RDT and resulted in a magnitude normalisation through zero. The normalisation process allows simple visual comparisons to be made of RDT log-power spectra even though each spectrum may have been produced from recordings with widely varying recording levels.

The pdf of the RDT of the log-power spectrum was then derived and from the derivation, it was shown that the sub-population variance for any particular coefficient is constant as a result of a log transformation of the power spectrum. The variance of the RDT pdf is independent of the variance of the time-domain noise under analysis and it follows that the slope-variance will also be independent of the variance of the time-domain noise. When the FFT size is relatively high, only very weak assumptions about the noise in the time domain are required. This makes the derived frequency domain probability densities very robust against changes in time domain noise characteristics. The derived probability densities along with closed form solutions for the means, modes and variances derived in section 5.5 and in appendix A are summarised in table 5.1.

	pdf	Mean	Mode	Variance	Symmetrical
Normal (Real or imaginary parts)	$\frac{1}{\sigma_{ft}\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_{ft}^2}}$	\bar{x}	\bar{x}	σ_{ft}^2	Yes
Raleigh (Magnitude)	$\frac{z}{\sigma_{ft}^2} e^{-\frac{z^2}{2\sigma_{ft}^2}}$	$\sqrt{\frac{\pi}{2}} \cdot \sigma_{ft}$	σ_{ft}	$0.429\sigma_{ft}^2$	No
Exponential (Power)	$\frac{1}{2\sigma_{ft}^2} e^{-\frac{q}{2\sigma_{ft}^2}}$	$2\sigma_{ft}^2$	0	$4\sigma_{ft}^4$	No
Log of exponential (log of power)	$\frac{e^r}{2\sigma_{ft}^2} e^{-\frac{e^r}{2\sigma_{ft}^2}}$	$\ln(2\sigma_{ft}^2) - \gamma$	$\ln(2\sigma_{ft}^2)$	$\frac{\pi^2}{6}$	No
RDT	$\frac{\sigma_{\omega_1}^2 \cdot \sigma_{\omega_2}^2 \cdot e^a}{(\sigma_{\omega_2}^2 + \sigma_{\omega_1}^2 \cdot e^a)^2}$	$2 \ln\left(\frac{\sigma_{\omega_2}}{\sigma_{\omega_1}}\right)$	$2 \ln\left(\frac{\sigma_{\omega_2}}{\sigma_{\omega_1}}\right)$	$\frac{\pi^2}{3}$	Yes

Table 5.1: Summary of the derived probability densities and associated statistics.

The spectral signal conditioning achieved by the RDT satisfies the requirements called for when using simple linear regression techniques. The application of linear regression models to the RDT of the log of the power spectrum has been introduced in order to produce an unbiased single parameter estimation of the response, indicating the generation status of an unknown recording. The methods of least squares and maximum likelihood were discussed as estimators for the slope parameter, similar performance but different complexity of calculation, led to the least squares method to be the chosen estimator.

For a fixed data size, the variance of the periodogram can be reduced by splitting the data up and averaging a number of individual periodograms. This produces a

proportional reduction in the frequency transform size, resulting in a trade off between spectral variance and spectral resolution. However, the parameter of interest is the slope estimate of the RDT of the log-power spectrum, and it is the variance in this slope estimate, not the variance in the spectral estimate that is important. The variance in the slope estimate has been shown to be inversely proportional to the overall data length and is not influenced by the ratio of transform size to number of averages. It has also been shown that the slope-variance can be reduced by having a wider ROI (5.40), and the choice of ROI should be as wide as possible within the confines of the pass-band and digital noise floor / Nyquist limit, as discussed in subsection 4.2.2.