

Chapter 4

Recording System Identification

The requirement was to develop a method to determine if a recording, allegedly produced on a known recording system, was consistent with an original or a copy. A signature of the recording system needed to be estimated using only data from the recording. Chapter 4 reports on a method to achieve this requirement based on the models developed in the previous chapter, where it was proposed that high frequency attenuation of environmental noise and speech components, leaves a recording containing predominantly signal conditioning noise approaching the Nyquist limit. This chapter shows that these signal and noise characteristics allow useful information relating to the magnitude response of the anti-aliasing and anti-imaging processes of the recording system to be obtained and this forms the basis of recording system identification.

The relationship between the input and output of an LTI system in the frequency domain is provided by the transfer function (TF). Based on linear transform theory, a number of methods for the determination of a systems TF have been developed, including, frequency response analysis using sinewave input signals [110], spectral analysis using stochastic input signals [111] and various sophisticated time domain parametric identification methods [112]. Connections between the various TF estimators are provided by Ljung [111] and are all based on measuring the output signal after exciting the unknown system under test with a known input signal.

Consideration has been given to non-parametric methods of system identification where the spectral response estimation allows the nature of the data and underlying systems to be exposed in a graphic and informative manner [113]. Non-parametric frequency domain estimators of the TF are based on the Fourier transform, having the

advantage of fast and effective processing not requiring any previous knowledge regarding a model structure. The identification problem is split into separate independent sub-populations at each frequency irrespective of the colour of the input signal. The asymptotic Gaussian characteristics of the spectral components allows the use of least square procedures, leading to minimum variance estimators [114], [115].

4.1 Recorder Transfer Function over a Spectral Region of Interest

This section investigates the magnitude squared transfer function of a recorder over a spectral region of interest using system identification techniques. A common model used for system identification when the input and output measurements are disturbed by noise is shown in fig 4.1, [115], [116].

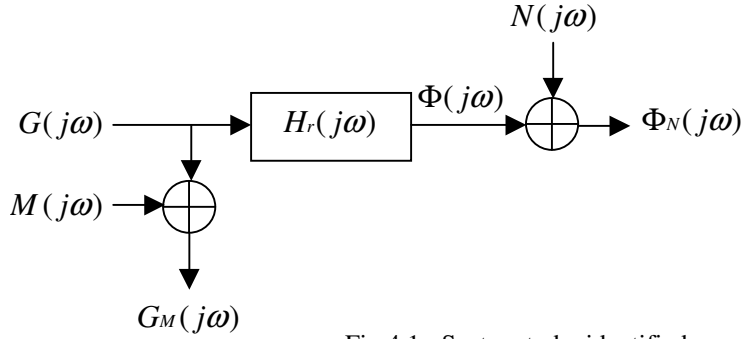


Fig 4.1: System to be identified.

The TF estimate $\hat{H}_r(j\omega)$ is given by the ratio of output spectrum $\Phi(j\omega)$ plus additive output measurement noise $N(j\omega)$ to input spectrum $G(j\omega)$ plus additive input measurement noise $M(j\omega)$:

$$\hat{H}_r(j\omega) = \frac{\Phi(j\omega) + N(j\omega)}{G(j\omega) + M(j\omega)} = \frac{\Phi_N(j\omega)}{G_M(j\omega)}$$

The initial assumptions have been made that the input and output measurements do not contain additive noise and additionally, the effects of signal truncation are zero. The TF then becomes:

$$H_r(j\omega) = \frac{\Phi(j\omega)}{G(j\omega)}$$

An audio recording represents the output of the system under analysis $\Phi(j\omega)$, while the TF, $H_r(j\omega)$ and the input signal $G(j\omega)$ are unknown. However, from the previous chapter it has been assumed that that the acoustic input signal spectral energy

decreases for increasing ω and at frequencies approaching the Nyquist limit, the signal energy will have decayed below the signal conditioning noise level, leaving a region with a constant power spectral density (psd). Therefore, at high frequencies we can model the input signal as white noise, which can be completely described by its variance. This high frequency region will be modified by the response of any low-pass filtering as shown in the model of the recording system, fig 3.12. This suggests that an estimate of the recording's power spectrum or psd for frequencies approaching the Nyquist limit may be used to discriminate between a system producing an original recording or a copy recording.

A power spectrum contains no phase information, and is a real valued non negative quantity [117 p323]. If the DFT, or the more computationally efficient fast Fourier transform (FFT) is multiplied by its complex conjugate the power spectrum can be obtained directly from the time-domain data [118 ch6.5]. The magnitude squared TF, which is the ratio of the output psd to the input psd is given by:

$$|H_r(j\omega)|^2 = \frac{|\Phi(j\omega)|^2}{|G(j\omega)|^2} \quad \text{where } 0 \leq \omega \leq \pi$$

Clearly, the psd of the input of a recording $|G(j\omega)|^2$ is unknown, but at high frequencies will be dominated by the signal conditioning noise $\sigma_{nf(Ar)}^2$ as discussed in the previous chapter. Therefore, the psd over this frequency region ω will be constant (4.1) [118 ch6.5]; where $\omega \in \omega$, and is defined between upper and lower frequency limits of ω_2 and ω_1 respectively and will be known as the region of interest (ROI).

$$|G(j\omega)|^2 \equiv \frac{\sigma_{nf(Ar)}^2}{\pi} \equiv k_G^2 \quad (4.1)$$

The magnitude squared TF over this spectral ROI can be estimated by:

$$|H_r(j\omega)|^2 = \frac{|\Phi(j\omega)|^2}{k_G^2} \quad (4.2)$$

From (4.2), the power spectrum of the recording over the ROI, represents the magnitude squared TF of the unknown recording system multiplied by the constant k_G^2 :

$$|\Phi(\omega)|^2 = |H_r(\omega)|^2 \cdot k_G^2 \quad (4.3)$$

The magnitude of the constant k_G^2 (4.1) is proportional to the signal conditioning noise power, and is therefore a function of the make, model and recording level of the original recording machine.

4.2 Low-Pass Response of a Recording System

Ideally, the frequency response of a digital audio recorder should be flat up to the Nyquist limit. However, to avoid aliasing components appearing within the passband, anti-aliasing low-pass filters are employed to steeply attenuate signals from a cut-off point near to the Nyquist limit. In conjunction with the high frequency acoustic signal attenuation and the spectral constant resulting from the signal conditioning noise, the low pass filter response provides information that may be used to discriminate between original and copied recordings.

A filter's pass-band can in general be defined as the spectral range where the power of the input signal is transferred to the filter's output with unity gain. The stopband can be defined as the spectral region of the filter that eliminates the input spectral power from the output. The spectral region in between the pass and stop bands is known as the transition region. The bandwidth of the transition region is only zero for an ideal filter; in practice a smooth transition from the pass-band to the stopband should occur over a frequency region dependent on the filter characteristics. Fig 4.2 shows how the positions of the pass, stop and transition bands are determined [119 p124].

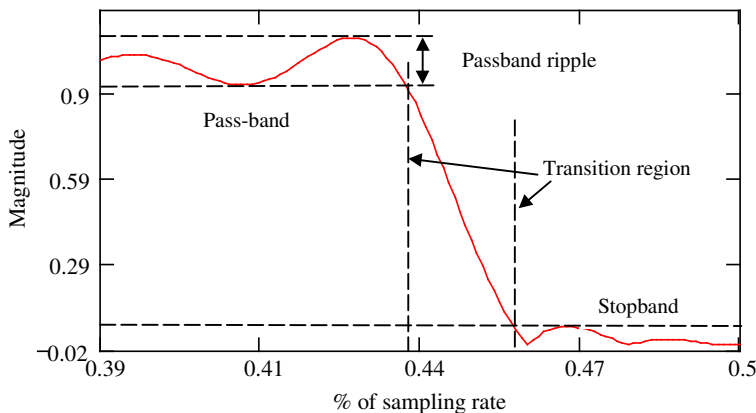


Fig 4.2: Higher frequency region of a low-pass filter magnitude response, identifying the positions of the passband, the stopband and the transition region [119 p124].

From the proposed model developed in chapter 3, the low-pass anti-aliasing filter response will be equivalent to the overall response of the original recorder. When the model for the copying recording system contains N low-pass filters, an overall system response will be equivalent to the response obtained by cascading N low-pass filters. Multiplying the response of each filter together produces the overall frequency response of the system.

The model assumes that a recording will have been produced on a system where signal conditioning noise power has a flat spectrum and dominates towards the end of the pass-band and beyond. Any shaping of the noise spectrum at high frequencies will be entirely due to the low-pass filters within the system. An estimate of the recording's power

spectrum for this region would represent the magnitude squared transfer function of the recording system that it has been produced from, multiplied by a constant that represents the signal conditioning noise power spectral density (4.3). The ability to differentiate between the psd of different numbers of cascaded filters, makes it theoretically possible to detect if a recording is consistent with an original recording or some form of copy.

4.2.1 Cascading Filters

This section investigates how the squared magnitude response of a low pass filter is affected when it is cascaded with other filters having identical specifications.

The magnitude squared response $|H(j\omega)|^2$ of a Butterworth low-pass filter having a cut off frequency of 20kHz is shown in fig 4.3. Also shown on the same graph is the response of this filter cascaded with up to four identical filter responses:

$$F(\omega)^5 = F(\omega) \cdot F(\omega) \cdot F(\omega) \cdot F(\omega) \cdot F(\omega) \quad \text{where} \quad F(\omega) = |H(j\omega)|^2$$

Kaiser and Hamming [120] show that passing the data through an identical filter twice will:

1. Approximately double the ripple in the pass-band.
2. Double the attenuation in decibels in the stopband.
3. Leave the transition bandwidth the same.

Increasing the stopband attenuation, and maintaining the transition bandwidth results in an increase in the rate of attenuation over the transition band. This rate of attenuation increases for each successively cascaded filter as indicated by fig 4.3.

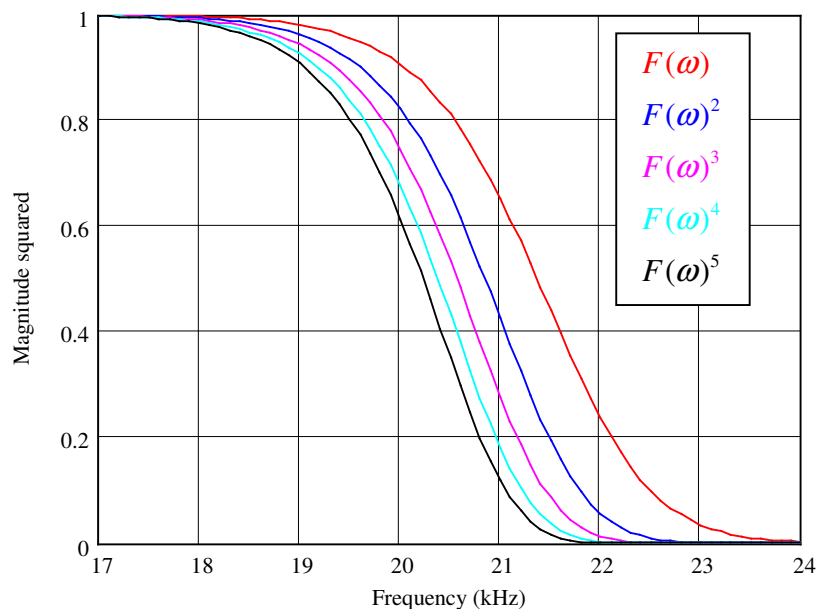


Fig 4.3: The effects at high frequency of multiplying a low-pass response by up to four times with the same filter.

The behaviour of cascaded filters having identical responses can be examined by using an amplitude change function [120]. The amplitude change function shows the output response of a number of cascaded filters versus the output response of a single filter. In fig 4.4, the concept has been modified to show the magnitude squared change function, illustrating how the magnitude squared response of a single filter is transformed when cascaded with one or more identical filters. Importantly, this function is independent of both filter shape and frequency. This means that providing the filters used in cascade have the same specification, the resulting magnitude squared change function is identical for all filter responses and is therefore a universal descriptor of the output transformation of filters used in cascade.

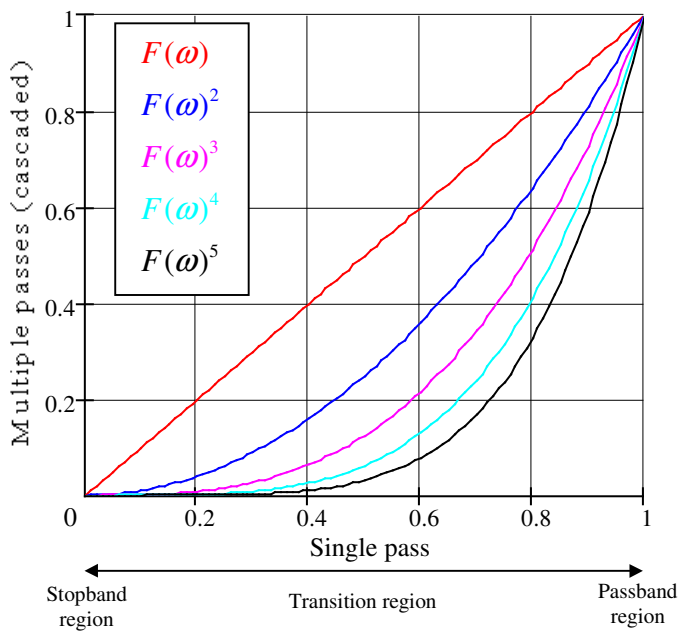


Fig 4.4: Magnitude squared transformation graph showing the output response of cascaded filters versus the output of a single filter.

If the original and copied recording systems are modelled as a series of cascaded filters with the same specification, it would be useful to establish where the change in the output response between a copied and original recording is at its most sensitive. The filter magnitude squared change function [120] can be used to find this sensitivity by establishing the rate of change of difference between the magnitude squared response of a single filter $F(\omega)$ and N multiple cascaded filters:

$$\frac{d}{dF(\omega)} [F(\omega)^N - F(\omega)] = N \cdot F(\omega)^{(N-1)} - 1 \quad (4.4)$$

The rate of change of magnitude squared difference function as given by (4.4), has been plotted for up to five cascaded filters as shown in fig 4.5. Again, this function is independent of both filter shape and frequency. It can be concluded from fig 4.5 that the

greatest rate of change of difference between a single and cascaded filter output is found approaching the pass-band region. This is with the exception of a two-filter configuration ($2F(\omega) - 1$), which has equal positive and negative rates of change of difference and therefore reaches maximum nearing both the pass-band and the stop-bands.

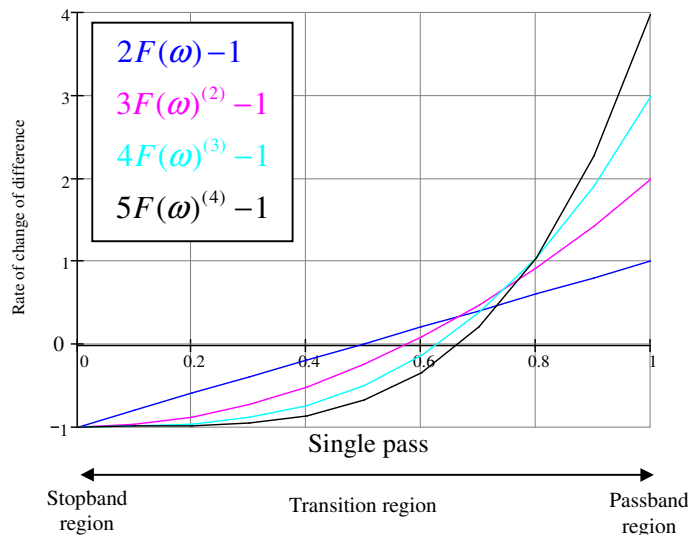


Fig 4.5: Sensitivity to cascaded filter output change in the form of a rate of change of difference between the magnitude squared response of a single filter and up to four further identical cascaded filters are shown.

4.2.2 The Spectral Region of Interest

It has been shown how the psd of an unknown recording system may be estimated by using a recording produced on that system. This is achieved using the signal conditioning noise to provide an estimate of the recording systems low-pass spectral response between the end of the pass-band, through the transition band and into the stopband. In practice two aspects of the digital conversion process may limit the spectral region available for analysis:

1. Low level signal conditioning noise will be close to the overall digital system noise floor (≤ 15 dB) predominantly comprised of quantisation noise and will limit the available ROI, illustrated by fig 4.6A.
2. The attenuation outside the pass-band is limited by the relative position of the Nyquist limit, illustrated by fig 4.6B.

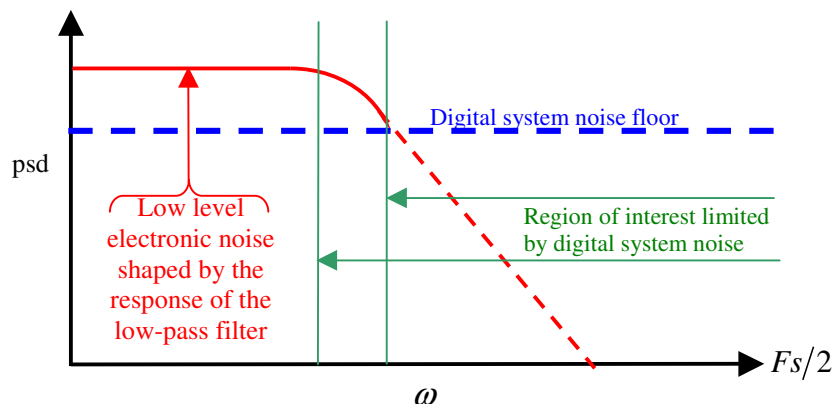


Fig 4.6A: The spectral ROI is limited by the digital system noise floor.

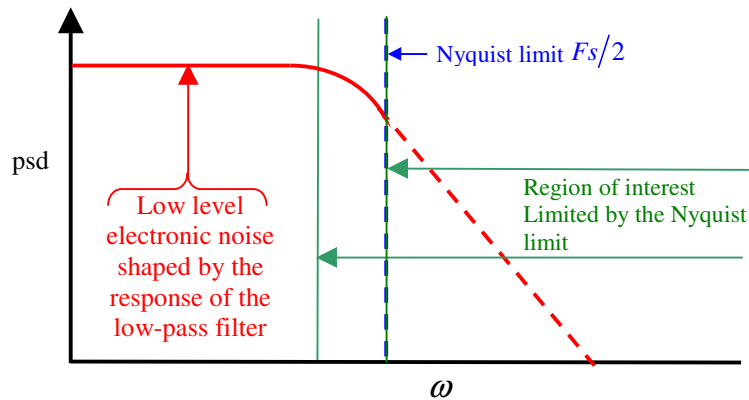


Fig 4.6B: The spectral ROI is limited by the Nyquist limit.

Therefore, the ROI will predominantly include the knee of the curve of the low-pass response. The knee of the curve is the most non-linear part of the response but as shown by fig 4.5, has the greatest difference in the rate of change of attenuation between a single filter and cascaded filters. This suggests that the discrimination between an original recording and a copy recording would be greatest over this non-linear region.

As the cut off frequencies and possibly attenuation characteristics of anti-aliasing and anti-imaging filters will be different for different sampling rates, the ROI for different sampling rates will also be different. For statistical reasons discussed in subsection 5.6.5, the chosen ROI should be as wide as possible, in practice this is limited to the confines of the end of the pass band and digital noise floor/Nyquist limit. It is not proposed to discuss the characteristics of any particular filter type used for anti-aliasing or anti-imaging purposes¹.

4.3 Transforming the Data over the Spectral ROI

The proposal is to use an estimation of the power spectrum of a recording over a spectral ROI that will include the knee of the curve originating from the low-pass response of the recording system. This spectral estimate will be used to determine attenuation characteristics across the ROI in order to establish if the data is from an original or a copied recording. Ideally, the requirement will be to produce a single parameter estimation that is both robust and statistically tractable.

There are two initial problems to be considered: firstly, the spectral shape over the ROI is non-linear, making a single parameter descriptor likely to be poor and secondly there is a constant of unknown value associated with each recording, given by (4.2), that makes comparisons of recordings produced at different recording levels difficult. As an

¹ Legadec et al discuss the various filter specifications and design criteria for anti-alias and anti-image analogue filters used in the ADC and DAC conversion processes at [121].

example, two power spectrums were produced of similar recordings made at different recording levels over a ROI between the end of the pass-band and the Nyquist limit; fig 4.7 shows their averaged power spectrums. To be able to estimate the response over the ROI using a single parameter descriptor, it was necessary to produce a method of analysis which normalises the power spectrum and linearises its response.

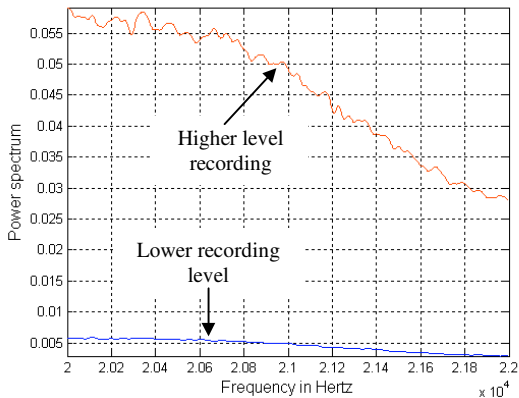


Fig 4.7: 1024 point, averaged power spectrums of two similar recordings produced by the same recorder but at different recording levels. The recordings were produced on a 44.1 kHz sampled system and the spectrums are shown over the transition region of the response up to the Nyquist limit.

4.3.1 The Reverse Difference Transform

For the purposes of comparing recordings made at different recording levels, the spectrum has been used to effectively normalise itself. The method developed, utilises the fact that the pre-filtered noise spectrum is symmetrical about the ROI and the roll off region of the low-pass filter response is asymmetric. We define the following parameters:

Modulus squared of recorder transfer function over ROI: $|H_r(j\omega)|^2$

Noise Power Spectral Density over ROI: $|\Phi_{nn}(j\omega)|^2$

Signal Power Spectral Density over ROI: $|\Phi_{ss}(j\omega)|^2$

Provided the signal is independent of the noise, the power spectrum of an original recording will be the sum of the noise and signal power spectral densities multiplied by the square of the magnitude of the recorder transfer function:

$$|\Phi_{rr}(j\omega)|^2 = |H_r(j\omega)|^2 \cdot (|\Phi_{nn}(j\omega)|^2 + |\Phi_{ss}(j\omega)|^2)$$

The upper (ω_1) and lower (ω_2) frequencies form the boundaries of the ROI which is considered to be characteristic of the recording process and provides the data to be analysed. If the centre frequency of the ROI is Ω , where $\Omega = (\omega_1 + \omega_2)/2$, then in the high frequency roll off region the recording spectrum can be reflected about Ω by a simple mathematical transformation to give a new spectrum reversed in frequency:

$$|\Phi_{rr}(j(2\Omega - \omega))|^2 = |H_r(j(2\Omega - \omega))|^2 \cdot (|\Phi_{nn}(j(2\Omega - \omega))|^2 + |\Phi_{ss}(j(2\Omega - \omega))|^2)$$

Dividing the recording spectrum by the reversed recording spectrum:

$$\frac{|\Phi_{rr}(j\omega)|^2}{|\Phi_{rr}(j(2\Omega - \omega))|^2} = \frac{|H_r(j\omega)|^2 \cdot (|\Phi_{mm}(j\omega)|^2 + |\Phi_{ss}(j\omega)|^2)}{|H_r(j(2\Omega - \omega))|^2 \cdot (|\Phi_{mm}(j(2\Omega - \omega))|^2 + |\Phi_{ss}(j(2\Omega - \omega))|^2)} \quad (4.5)$$

We assume that over the spectral ROI the signal spectrum will be negligible. That is:

$$|\Phi_{ss}(j\omega)|^2 = |\Phi_{ss}(j(2\Omega - \omega))|^2 = 0$$

Also assuming the noise to be a constant k_G^2 , the noise spectrum will be symmetrical about the frequency Ω so that:

$$|\Phi_{mm}(j\omega)|^2 = |\Phi_{mm}(j(2\Omega - \omega))|^2 = k_G^2 \quad \text{Then (4.5) reduces to:}$$

$$\frac{|\Phi_{rr}(j\omega)|^2}{|\Phi_{rr}(j(2\Omega - \omega))|^2} = \frac{|H_r(j\omega)|^2}{|H_r(j(2\Omega - \omega))|^2} \quad (4.6)$$

Therefore it can be concluded that (4.6) relates solely to the transfer function of the recording system and is not dependent on recording levels.

As will be discussed in Chapter 5, converting to a logarithmic form produces the same amplitude resolution at both low and high signal levels and one benefit of this is that it aids the comparison of recordings having a different psd. Converting (4.6) to dB:

$$10 \log_{10} \frac{|\Phi_{rr}(j\omega)|^2}{|\Phi_{rr}(j(2\Omega - \omega))|^2} = 10 \log_{10} \frac{|H_r(j\omega)|^2}{|H_r(j(2\Omega - \omega))|^2}$$

Which can also be written as:

$$20 \log_{10} |\Phi_{rr}(j\omega)| - 20 \log_{10} |\Phi_{rr}(j(2\Omega - \omega))| = 20 \log_{10} |H_r(j\omega)| - 20 \log_{10} |H_r(j(2\Omega - \omega))|$$

at $\omega = \Omega$ the calculation will always normalise to 0 dB. The process of reflecting a log-power spectrum around a centre frequency of a spectral ROI and subtracting it from the original log-power spectrum will be known as the **Reverse Difference Transform (RDT)**.

As discussed previously, if the original recording is copied then the response is modified by further low-pass filters. Assuming all filters are identical, then for each generation of copy n we will have two further filters applied giving the total number of filters as $2n + 1$, therefore the overall response is given by:

$$|H_r(j\omega)^{2n+1}|^2$$

Making the same assumptions as before the response for any number of re recordings are given by a general expression for the RDT:

$$20(2n + 1) \left[\log_{10} |\Phi_{rr}(j\omega)| - \log_{10} |\Phi_{rr}(j(2\Omega - \omega))| \right] \quad (4.7)$$

Theoretical example

Using the same assumptions, the magnitude-squared response of a Hamming widowed

$\sin x/x$ low-pass filter is shown in fig 4.8, along with the response of the filter after having been cascaded with an identical filter. Each response has been multiplied by a different value of constant k_G to represent the signal conditioning noise levels from recorders having different recording gains. Using (4.3) the filter responses are described by:

$$\text{Original recording: } |\Phi_{r_1}(j\omega)|^2 = |H_r(j\omega)|^2 \cdot k_{G_1}^2$$

$$\text{Copied recording: } |\Phi_{r_2}(j\omega)|^2 = |H_r(j\omega)|^2 \cdot k_{G_2}^2$$

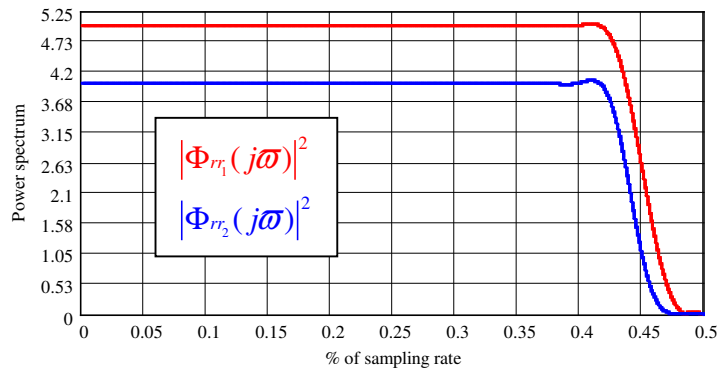


Fig 4.8: Magnitude squared responses representing original and copied recordings produced at different recording levels.

The spectral ROI has been selected to be 0.415 to 0.440 as a percentage of the sampling rate and has a value for Ω of 0.427. Both responses over this ROI are shown in fig 4.9.

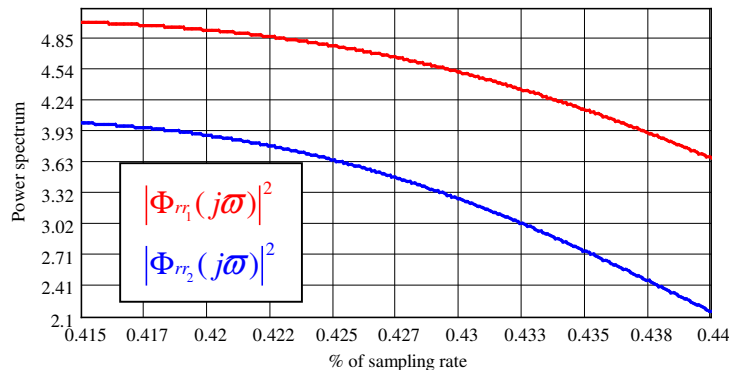


Fig 4.9: As fig 4.8 but limited to the spectral ROI selected to be on the knee of the curve as would be required in practice.

Using (4.7), the RDT is calculated for both responses and as shown in fig 4.10, Ω has normalised to 0 dB and the RDT rate of attenuation for the copy is much greater than the original. It should also be noted that the RDT has approximately linearised the responses.

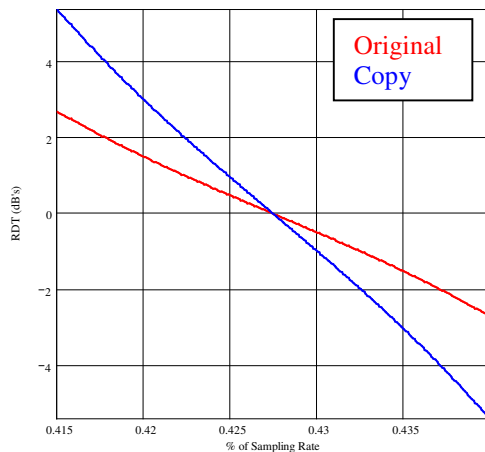


Fig 4.10: The result of applying the RDT to the ROI for both original and copied responses. Both responses are normalised at Ω and approximately linearised.

Increasing the number of cascaded filters results in the RDT response increasing the rate of attenuation proportionally as shown for 5 cascaded filters in fig 4.11.

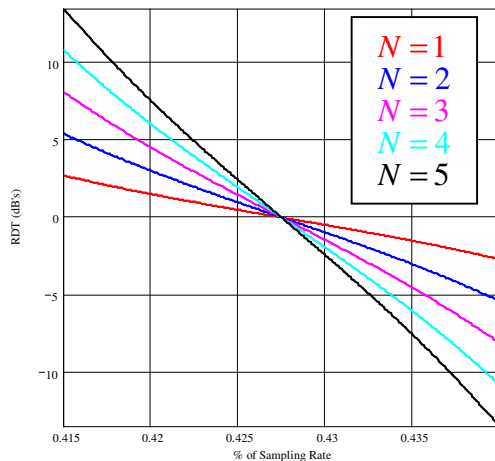


Fig 4.11: The result of applying the RDT to the ROI for up to 5 cascaded filters. After each filter application an increase in slope value occurs.

Frequency has been presented on a linear scale but may also be displayed logarithmically, however the high frequency narrow ROI produces very little change to the spectral shape after log transformation and will not be considered further.

Real data example

An original recording was produced on a Sony NT-2 digital micro cassette recorder sampling at 32 kHz. A copy recording was produced from it by coupling the line output of a playback machine to the line input of another NT-2 recorder. An estimation of the power spectrum of the original and copied recording was then produced for a spectral ROI of 15 kHz to 15.6 kHz. This ROI was chosen based on examination of a number of recordings produced on this recording system, typified by the response above 14 kHz identified in fig 3.16 of chapter 3. The estimated power spectrum over the ROI was produced using 308 transform points over the ROI and to reduce variance between the spectral coefficients the spectrum has been averaged over a total of 234 transforms.

The traces of the averaged power spectrums shown in fig 4.12, indicate that the copy recording has been recorded at a lower level than the original recording and also indicates increasing attenuation with frequency for both recordings. However, it offers no clear evidence of the relative rate of change of attenuation between the original and copied recordings. The log of the averaged power spectrum for the original and copied recordings are shown in fig 4.13, revealing that the copy recording has a greater rate of change of attenuation over the ROI than the original recording. The RDT of the log of power spectrum shown in fig 4.14, approximately linearises the spectral response of the recordings and also normalises the response so that at the centre of the ROI the power is 0 dB removing the effects of different recording levels.

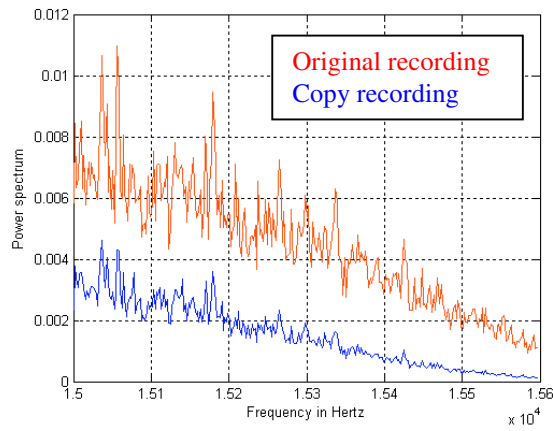


Fig 4.12 Averaged power spectrum over the ROI of the original and copied audio data.

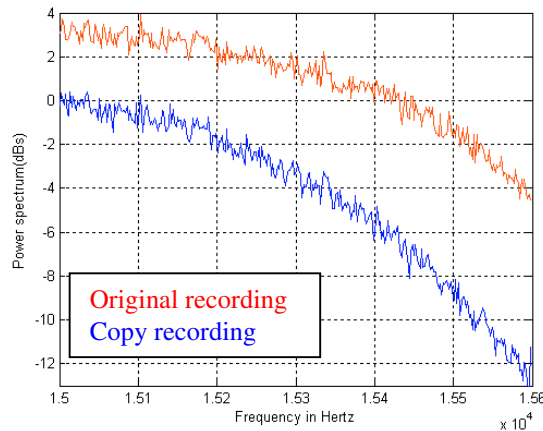


Fig 4.13 Logarithmically converted averaged power spectrum over the ROI of original and copied audio data.

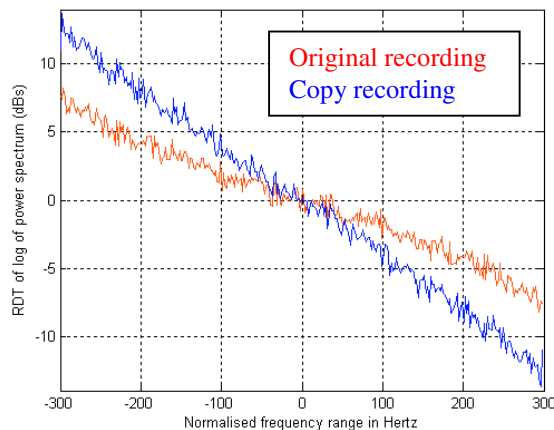


Fig 4.14 RDT of log of averaged power spectrum over the ROI of original and copied audio data.

4.4 The Proposed Signal Processing Procedures

The proposed signal processing procedures are based on estimating the averaged log-power spectrum restricted to a high frequency ROI of the recording, where discrimination between recordings of different generation may be found. Using regression techniques it is proposed that the approximately linear relationship produced by the RDT of the averaged log-power spectrum to frequency will allow a single parameter in the form of a slope

estimate to be given. Inference procedures will then allow discrimination between original and copied recordings to be made. A summary of the proposed signal processing procedures are shown in fig 4.15, and are discussed and analysed in detail in chapter 5 and chapter 6.

4.5 Summary

Approaching the problem as one of system identification and using the assumptions about the acoustic and electronic analogue signals and subsequent digitised signals introduced in chapter 3, a simple recording model in the frequency-domain has been proposed. The recording model does not require detailed knowledge of the recorder's digital conversion process and in its simplest form is represented by a series of low-pass filters and a white noise source. It has been shown that the rate of change of attenuation over the low-pass transition region increases for an increase in the number of overall cascaded low-pass filters and this offers the opportunity to discriminate between original and copied recordings.

In order to exploit the differences in cascaded filter attenuation characteristics, the logarithm of an averaged power spectrum has been used in conjunction with a novel transform given the name RDT. The RDT normalises the data in order to remove differences between recordings that are due to different recording levels. Further, the RDT approximately linearises the normally non-linear response of the log-power spectrum over the ROI, allowing it to be described by a single parameter. Considering the model to be ideal, the overall procedure produces a result that is totally dependent on the spectral response of the recording system and is independent of the acoustic signal and recording levels. It is proposed that linear regression techniques and inference procedures are applied to the result in order to ascertain the original or copied status of the recording under examination.

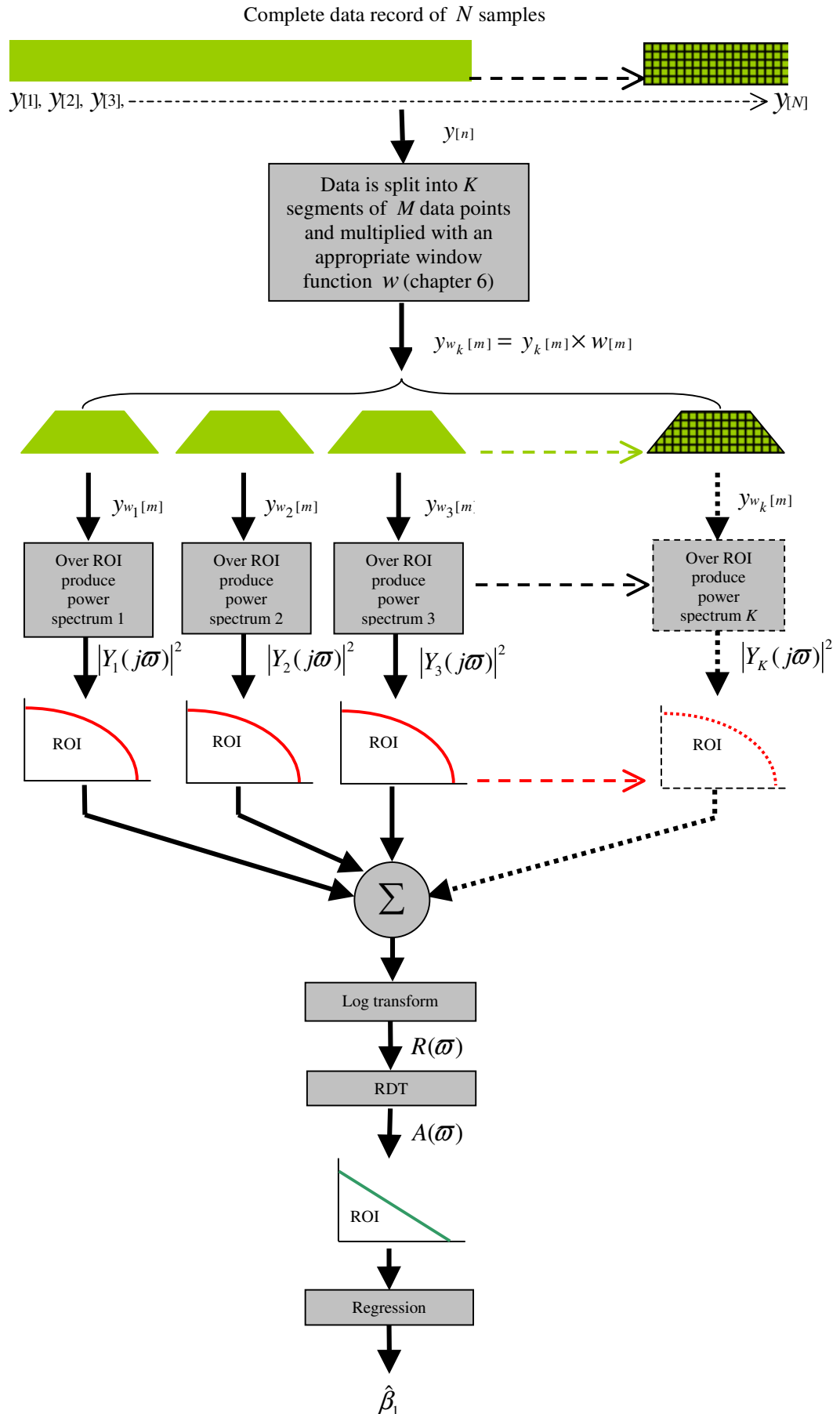


Fig 4.15: Summary of the overall processing sequence. The signal processing procedures are discussed and analysed in detail in chapters 5 and 6.

