Chapter 3

Models for the Recording System in a Forensic Environment

If an audio recording has been properly edited, the edit points may not be detectable by auditory or instrumental means. However, as discussed in chapter 2, editing would mean that the purported original recording is a copy recording. Therefore, when conducting a forensic authenticity examination, establishing that the recording is original is of paramount importance as it eliminates the possibility of editing having taken place. The problem can be stated as follows: '*given a forensic digital audio recording produced on a known or alleged recording system, can it be determined if the recording is an original or a copy*'. The recording itself must provide the data that will reveal the characteristics or signatures of the underlying recording system leading to the identification of its original or copied status. The first stage in finding a solution to the problem was to develop models for the signals and systems used in such recording processes.

In this chapter, generic models to describe portable digital audio recording systems are proposed that incorporate sampling frequencies > 24 kHz and effective number of quantisation bits ≥ 16. The models to be described are original and represent stable, linear time-invariant systems (LTI). An LTI system can be analysed in detail using a set of tools that form the core of signal and system analysis providing insight into the system properties [77 ch2]. The initial model represents a single recording device producing an original recording. A more complex model is developed to describe a recording system that represents an original recording that has been copied through an analogue interface onto another recording device, as discussed in chapter 2. After examining the possibility of simplifying the model, justification is provided for treating the output of the recording system at high frequencies as a white noise input signal modified by a series of low pass filters. This final model provides a basis for recording system discrimination.

As there are analogue and digital processes involved in digitally recording acoustic signals, the models will be described in terms of continuous time and discrete time signals defined as follows:

- A signal described as continuous has a value specified for all points in time. If x is a continuous time-domain signal it is given the notation of $x(t)$.
- A signal described as discrete has a value specified at points equally spaced in time. If x is a discrete time-domain signal it is given the notation of $x[nT]$ where T is the sample spacing, and n is a member of a set of ordinals $\mathbb{Z}$: $n \in \mathbb{Z}$

This notation distinguishes between the continuous time signal $x(nT)$ and the discrete time signal $x[nT]$ at time $t = nT$.

3.1 Recording System Model

When referring to a digital audio recording system, the digitised signal corresponds to a discrete representation of the analogue signal in both the time and amplitude domains, the result of sampling and quantisation respectively. In a high quality digital audio system the bandwidth and dynamic range are representative of the maximum capabilities of the human auditory system. A good comprehensive treatment of the theory and practice relating to the subject of digital audio recording technology is presented by Watkinson [78].

The basic format of a digital audio signal is Pulse Code Modulation (PCM), whereby the data rate increases linearly with sampling frequency and amplitude resolution. The models to be described refer primarily to portable digital audio recording systems that are not based on perceptual encoding techniques. Figure 3.1 shows in its simplest form, a digital audio recording and playback system [14 p28].

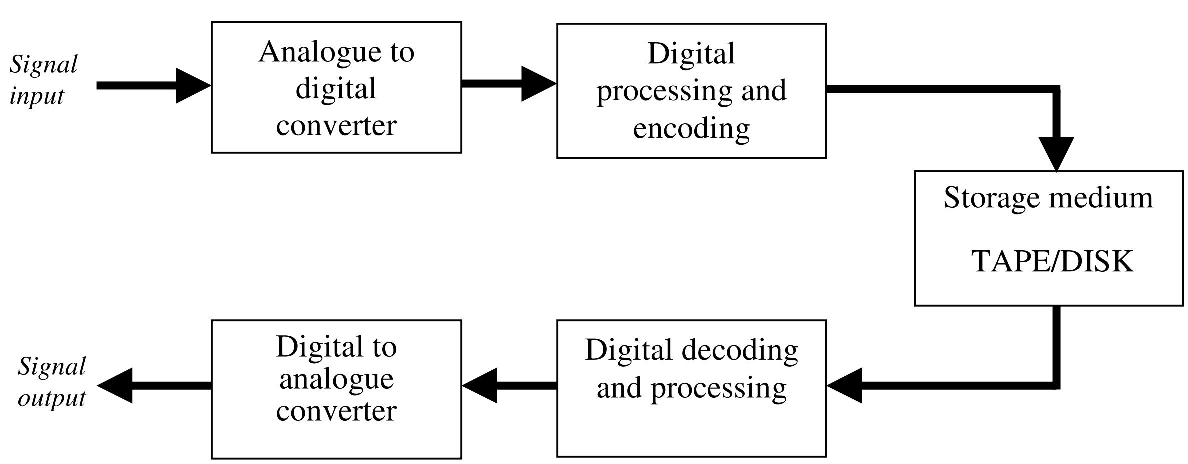

Fig 3.1: Basic digital recording and playback system.

One of the advantages of a digital audio recording system over that of analogue, is that if it has been properly designed, the audio information is largely uncorrupted by the recording medium, and the quality of the audio is therefore dependent on the conversion processes involved in getting signals into and back out of the digital domain.

A well engineered uncompressed audio recording system having a wider bandwidth and a larger signal to noise ratio than a signal that passes through it will result in the complete transfer of the signal information *x(t)*, providing the input signal level is matched to the amplitude level window of such a recording system. However, it is always anticipated that the system will introduce a measure of noise indicated by $\eta(t)$ in fig 3.2.

$$x(t) = IN \longrightarrow \boxed{\text{Optimum system}} \longrightarrow x(t) + \eta(t) = OUT$$

Fig 3.2: An optimum recording system maintains the input signal information.

In practice, even well engineered systems will cause errors to be present in the output created by non-linearities in the conversion processes. The recorder designer will aim to keep these errors as perceptibly small as possible. From a forensic perspective, additional signal components introduced in a copying process would be useful in discriminating between the original recording and a copy recording made of it.

3.1.1 Model for a Recording System Producing an Original Recording

A digital audio system will use a finite precision representation of the analogue input signal, and conversion into the digital domain will impose limitations on the available bandwidth and signal to noise ratio of the recording. The purpose of the ADC is to change a signal from the analogue domain, characterised by one that potentially varies continuously with time, to that of a digital signal representing amplitude snapshots of the original continuous time signal, achieved by a process of sampling and quantisation. The quality of an audio signal passing into and out of a digital recording system is bounded by the limitations imposed on it by the ADC and DAC conversion processes of that system.

A number of common methods are available to convert an analogue audio signal into the digital domain. Blesser's paper [79] provides a comprehensive examination of the digitisation of audio signals; introducing the important concepts of anti-aliasing, quantisation noise and jitter within a framework of PCM and oversampling conversion techniques.

In order to optimise the analogue to digital conversion process, the audio signal goes through a number of pre-conditioning stages. A possible configuration of the signal conditioning process required for digital conversion is shown in fig 3.3.

The digitised signal would ideally be a perfect representation of the acoustic signal entering the recorder. However, frequency response limitations, sampling errors, quantisation errors and analogue noise, detract from this ideal. Although each processing stage identified in fig 3.3 is required to optimise the digital conversion process, errors will be introduced into the final digitised signal. Signal conditioning circuits such as microphone preamplifiers and line amplifiers produce signal conditioning noise, the anti-aliasing filter produces frequency response errors and noise, the pre-conditioning ADC circuits produce noise and possibly distortion, the ADC conversion process produces quantisation noise and sampling/timing errors known as jitter [80].

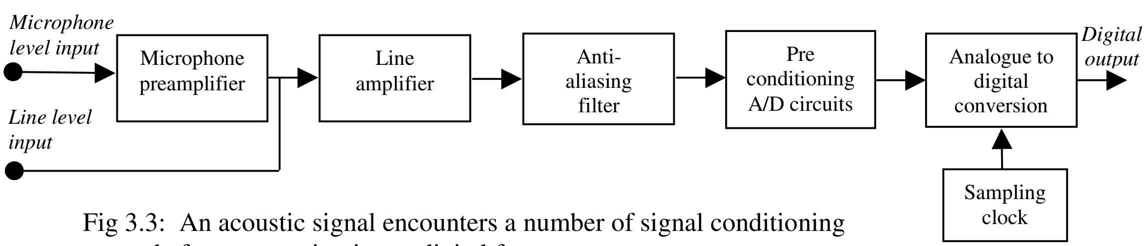

Fig 3.3: An acoustic signal encounters a number of signal conditioning stages before conversion into a digital form.

Conversion Process

As is well known, the sampling process itself introduces images of the audio signal into the frequency-domain. The sampling process can be modelled as a time-domain signal *x(t)* multiplied by a pulse train *p(t)*, with each individual pulse considered as a delta function [77 pp516-520]. From transform theory [81], a multiplication in the time-domain becomes a convolution in the frequency-domain:

$$x(t) \xleftrightarrow{f} X(j\omega)$$

$$p(t) \xleftrightarrow{f} P(j\omega)$$

$$X_p(j\omega) = X(j\omega) * P(j\omega)$$

Where $X_p(j\omega)$ is a periodic function of ω, and consists of shifted replicas of $X(j\omega)$ centred on the sampling frequency *fs* and its harmonics, as shown in fig 3.4.

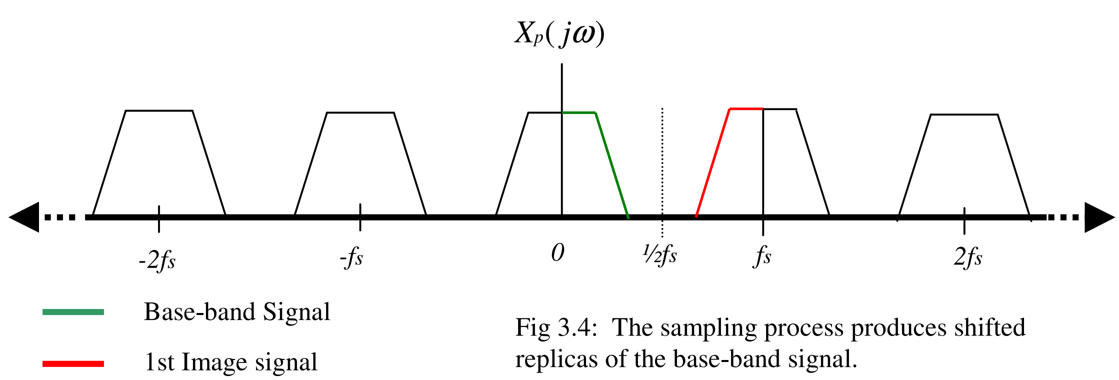

Fig 3.4: The sampling process produces shifted replicas of the base-band signal.

It can be seen from fig 3.4 that if the frequency of the base-band signal were to exceed ½ *fs* then the fold-over signal or 1st image would overlap and superimpose onto the base-band signal. This is known as aliasing, and is perceived as a form of distortion. Once aliasing distortion has taken place, it is not possible to remove it [79]. Aliasing can be prevented or significantly reduced by applying a low-pass filter before the sampling process known as an anti-aliasing filter. The anti-aliasing filter is used to attenuate high frequency input signals in order to stop frequency components greater than ½ *fs* entering the sampling circuitry [79]. The frequency ½ *fs* is known as the Nyquist limit and *fs* is known as the Nyquist rate or sampling frequency. An optimum anti-aliasing filter is very difficult to achieve with PCM conversion because the filter has to work in the analogue domain and the requirement will be for the filter to have a very steep attenuation or "brickwall" characteristic outside of the pass-band. The steep attenuation is necessary so that *fs* can be as low as possible for a given bandwidth without aliasing occurring. However, achieving this goal introduces ringing and phase non-linearity in the pass-band [82]. Sensitivity to deviation of component values from their nominal values caused by temperature and ageing can also be a problem with analogue filters. Accurate, low noise, low distortion analogue anti-aliasing filters only come with a high degree of circuit complexity. They are realised with high tolerance, high stability components that are expensive and usually take up greater space, which is an important consideration for portable recording equipment.

When analysing the behaviour of the quantisation process of an ADC, two approaches are possible:

1. With a known input signal, the spectrum characterisation and behaviour of the errors produced by the quantisation process can be established by deterministic analysis [83].
2. A probabilistic analysis can be carried out where the effect of quantisation is modelled by an additive white noise, uniformly distributed in amplitude and uncorrelated with the input signal.

The second approach is often referred to as the 'quantisation theorem' [84-87]. As the acoustic signals contained on a forensic audio recording are complicated and unpredictable, the probabilistic approach has been taken. The quantisation theorem, models the quantisation error $q[nT]$ of the digitisation process as additive noise, independent of the signal and having a probability density function (pdf) that is uniform and a frequency spectrum that is white. This is a widely used statistical model [84-87] and is also known as the additive independent white noise model (AIWN). However for the model to be valid the following conditions are required [84-88]:

1. The quantisation noise is white.
2. Quantum steps Δ are small compared to variations in the signal level. Consequently, independence will be approximately met since the error $q[nT]$ describes only signal variations relative to the next quantum level and has little to do with the overall value of the signal at the time [86].
3. The error $q[nT]$ has an equal probability of taking on any particular value between $-\Delta/2$ and $+\Delta/2$.

Quantisation noise cannot always be assumed to be white [88], however it can be, when the quantised signal is of a high level and has a high spectral bandwidth. Under these conditions the noise is statistically random from sample to sample. Considering the quantisation process to be non-linear, Maher succinctly describes the process of whitening as follows:

> "*The signal will contain spectral components at the input signal frequencies and at harmonics of the input frequencies, and intermodulation products at sums and differences of multiples of the input frequencies. The presence of the harmonic and intermodulation terms results in a plethora of components spread over a wide bandwidth. These quantisation noise components will alias in the frequency-domain adding to the whitening effect. For an input signal containing many spectral components the numerous distortion products produced by the quantizer becomes effectively continuous across the audible band and takes on a spectrum which is essentially low level white noise."*[89]

The unsampled quantisation noise power spectrum will decay to insignificance at high frequencies outside of the audio band. The area under the curve of this noise spectrum is as shown in fig 3.5. Hawksford states:

> *"Inevitably, when the amplitude quantised signal is sampled, the spectrum of the quantisation distortion is changed. However what is probably surprising on initial encounter is that when the quantised signal is sampled the quantisation noise power remains unchanged and is concentrated into the frequency band 0 to fs/2 Hz, where fs is the sampling frequency, although a replication of this spectrum also occurs about each harmonic of fs Hz."* [90]

The diagram shown identifies the total quantisation distortion spectrum under the curve and how the different areas fold back into the base-band frequency range.

The quantisation model has been based on uniform quantisation, where the quantum step Δ is the same for all conversion levels. The AIWN model for non-uniform quantisers such as floating point types have been considered by Widrow *et al* [84], where the limits of the validity of the model have been pointed out by Gray at [87].

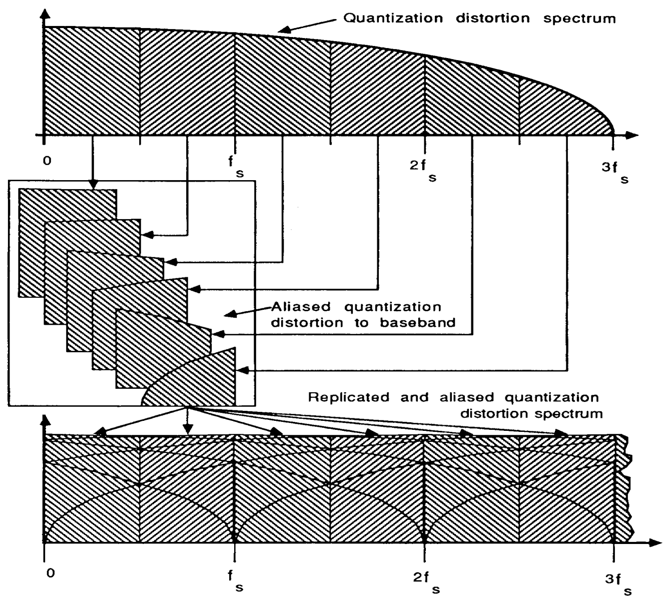

Fig 3.5: Aliasing of quantisation distortion to the base-band together with replication of aliased spectrum about the harmonics of the sampling rate. Taken from [90].

One of the conditions where the AIWN model is known to breakdown is when the input signal is low and the size of the quantum step Δ becomes relatively large [91]. This results in the quantisation error retaining the character of input-dependent distortion or noise modulation known as granulation noise [89]. The effects of granulation noise are demonstrated in fig 3.6 where the spectrum of the output response of a recording system is shown for a 1kHz sinoisoidal input signal recorded at levels between –20 dB full scale (FS) and –90 dBFS. As the signal is reduced, harmonically related components start to appear above the noise floor [92].

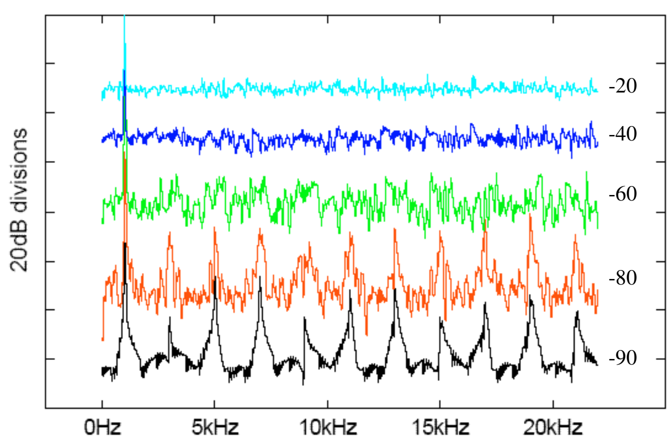

Fig 3.6: FFT analysis of undithered 16 bit quantisation of a 1 kHz tone at –20, -40, -60, -80 and–90 dBFS . Traces offset by 25 dB,[92].

In order to decorrelate the signal with the error and eliminate the audible effects of granulation noise, random noise is often deliberately added to the input of a recorder in a process known as dithering [91], [93], [94]. The amount of additive noise required is small, and is in the region of one to two least significant bits (LSB) peak-to-peak [91],

[79]. It is also recognised that natural thermal noise produced by signal conditioning circuits can act as a dither source in its own right. Vanderkooy and Lipshitz state:

> *" Normal recorded audio signals, such as those from most microphone and mixer setups, would have sufficient noise to adequately dither a 16-bit ADC."*[95]

Generally, microphone circuits incorporated in portable type digital audio recorders produce non-linear error in the form of signal conditioning noise. Signal conditioning noise may contain peak values many times greater than a single quantisation level. An example is shown in fig 3.7 where the noise has a Gaussian distribution with a standard deviation of around 17 quantum levels for a recording level set to maximum. It will therefore be assumed that any audio signals described by the proposed model will be self dithering and the problem of granulation noise effects will not be considered any further.

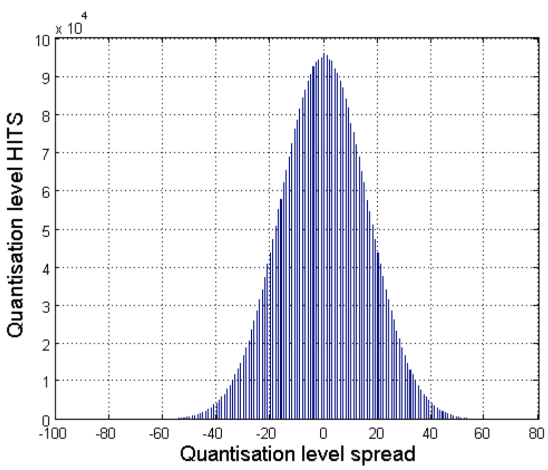

Fig 3.7: Histogram of an audio recording of signal conditioning noise, produced from a Sony NT-2 digital micro cassette system. The recording has a Gaussian distribution with a standard deviation of around 17 quantum levels.

The transfer information capacity of a communications channel can be defined as a function of bandwidth and signal to noise ratio [96]. Given a constant channel capacity, noise and bandwidth ratios may be interchanged. Oversampled and quantisation noise shaped systems are based on this principle [89]. Oversampling and noise shaping techniques as discussed by Hausser [97], allow some of the quantisation noise power to be located outside of the audio pass-band and into the redundant frequency space created by oversampling. Therefore, for a given amplitude resolution, oversampling allows the number of quantisation levels and therefore number of bits to be reduced.

For PCM recording, it is necessary to convert the oversampled data back to the original Nyquist rate of the recording system so that maximum use can be made of the medium's storage space. This is achieved by a process known as decimation, where the oversampled digital signal is filtered to remove signal components above half the original Nyquist rate and is then down-sampled back to the original Nyquist rate [97]. In this overall process, the onus on anti-aliasing filtering has been shifted away from the analogue domain to that of the digital domain, where the low-pass filter characteristics can be made accurate and with absolute phase linearity.

In practice, the conversion processes may therefore include low-pass filters applied in both analogue and digital domains. For example, in conventional PCM conversion, the anti-aliasing low-pass filtering will occur entirely in the analogue domain, whereas an oversampled system may have a low-pass anti-aliasing filter implemented predominantly in the digital domain, requiring only low order low-pass filtering in the analogue domain [97]. Without loss of generality, the models presented describe the anti-aliasing process in the analogue domain before any digital conversion process takes place irrespective of the type of conversion process.

A Generic Model

As detailed mathematical descriptions of individual types of conversion processes proved unnecessary, the AIWN model describing the digital noise produced by the conversion process will be used generically in the models proposed. This uniform noise model of quantisation forms the basis of the forensic recording system model, however for a complete description, the overall output of the ADC needs to account for signal conditioning noise produced by the signal conditioning circuits prior to digital conversion. The signal conditioning noise $\eta(t)$ will be independent of the acoustic signal $x(t)$ allowing a time-domain signal sequence $g(t)$ prior to conversion to be modelled as a sum of the acoustic signal and the signal conditioning noise:

$$g(t) = x(t) + \eta(t) \tag{3.1}$$

The noisy analogue signal will then be presented to the sampling and quantisation processes for digital conversion. By quantising the total analogue input signal given by (3.1), the model of the output y of a quantiser Q which forms part of a sampled data system is shown by fig 3.8.

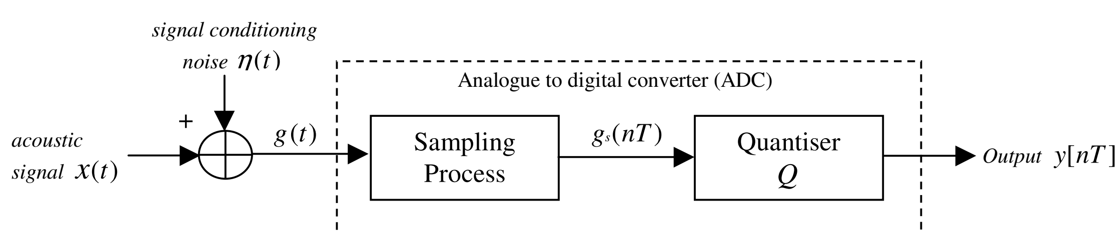

Fig 3.8: Basic input output model of a continuous to discrete time converter.

From fig 3.8 the quantiser output will be given by:

$$y[nT] = Q\{g_s(nT)\} \quad \text{where } g_s(nT) = x_s(nT) + \eta_s(nT)$$

The quantisation error $q[nT]$ will be the difference between the sampled analogue input signal $g_s(nT)$ and the digital output signal $y[nT]$:

$$q[nT] = Q\{x_s(nT) + \eta_s(nT)\} - \{x_s(nT) + \eta_s(nT)\} = y[nT] - g_s(nT)$$

Bertico et al [98] considered a noise model for digitised data that resulted in white Gaussian noise superimposed on an input sequence that accounts for both the wideband signal conditioning and quantisation noise sources. This was based on the results of their research that showed that when the ratio of the standard deviation of signal conditioning noise to quantisation step is greater than 0.2, ($\sigma/\Delta > 0.2$), the overall output noise of the ADC can be considered Gaussian and white. It is clear from this result, that when including the signal conditioning noise in the forensic recording model, the output may also be considered Gaussian and white.

It has been assumed that the total analogue signal is simply the sum of the acoustic signal and the signal conditioning noise (3.1). In practice the acoustic signal will pass through a gain setting stage with a variable gain which can be represented by a gain factor A_{or}. This gain represents the setting of the recording level and is variable in the sense that it is normally specified by the user prior to making the recording or may change dynamically, according to the level of the incoming acoustic signal.

The single noise source η has so far been shown to be additive to the acoustic signal and therefore would also be multiplied by a single gain factor A_{or}. This would be a gross simplification of a real system, which would have many noise sources and gain stages within the analogue circuits. At low frequencies, noise would be dominated by flicker noise also known as $1/f$ noise, generated by most active and some passive devices in the analogue circuits. Flicker noise has an approximately $1/f$ characteristic for frequencies lower than a few kilohertz, above this its power is low and is usually neglected [99 p434]. At frequencies above a few kHz the signal conditioning noise will be primarily the result of thermal and shot noise [99 ch23.17]. Thermal noise will be produced by all resistive elements of the system, while shot noise will be produced from the random injection of charge carriers across the depletion layer of semiconductor junctions. Both thermal and shot noise can be modelled as white noise. Each of the many sources of noise will be subject to various accumulations and amplifications throughout the analogue circuits prior to the final stage of digital conversion. The conclusion is that the signal conditioning noise will not be directly proportional to the recording gain A_{or}, but can be modelled as some function of the recording gain $f(A_{or})$ which would be specific to the make and model of the recorder. The output is also affected by the anti-aliasing filter response $h(t)$, whose effects in the time-domain are modelled as a convolution with the applied input signal. A convolutional process will be indicated by the operator $*$ in all equations that follow.

Based on the discussion and the noise model of quantisation, a generic model for the recording system is shown in fig 3.9.

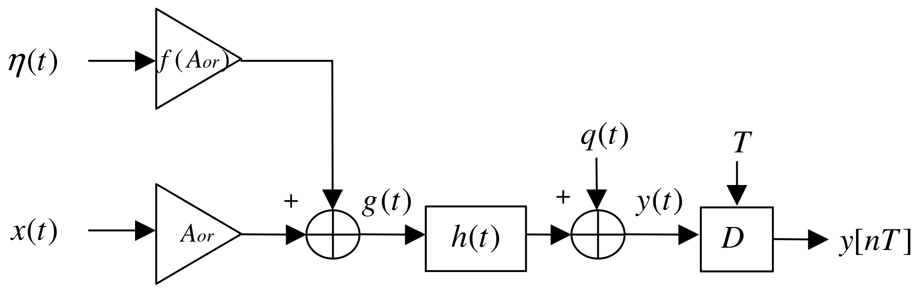

Fig 3.9: Generic model for the digitisation of the signal and associated conditioning noise.

It should be noted that all elements are modelled in the continuous time-domain. The output is then converted to discrete time using an ideal continuous time to discrete time convertion process *D*, where *T* represents the sampling interval. From the model shown in fig 3.9, an expression can be formed:

$$y[nT] = D[(\{A_{or} \cdot x(t) + f(A_{or}) \cdot \eta(t)\} * h(t)) + q(t), T] \tag{3.2}$$

The digital output $y[nT]$ represents the recorded audio data to be analysed. The digitised recorded signal can be considered independent of the recording medium and comprises the sum of two components:

1. The sound pressure variations as converted by the microphone prior to entering the recorder.
2. Noise produced from within the recording system.

It will be assumed that the bandwidth of the analogue conditioning circuits are wider than the bandwidth of the recording system and therefore the spectrum of the signal conditioning noise will be wider than the recording channel.

Using the additive noise model, the power output from the quantiser is greater than its input by an amount $\overline{q^2}$. Jayant and Noll [100 ch 4.7.3] suggest that a model may include a less than unity gain component before the quantiser to compensate for this fact. However, absolute gain is not an issue and the modification would not make a significant difference to the results put forward here and therefore will not be considered any further.

3.1.2 Model for a Recording System Producing a Copied Recording

The model for the recording system producing an original recording has been extended to include a basic copying system. In this revised model the original acoustic signal will be recorded onto a recording system and then copied from this system into a further recording system via a DAC and ADC process. No intermediate editing, signal manipulation or further ADC and DAC stages in the overall process have been considered. The system model proposed will be considered to produce the minimum difference between an original recording and its copied counterpart and will therefore be the most difficult to detect.

In theory, the DAC conversion process is relatively simple. An *N* bit word can be transformed back to an analogue waveform by an *N* bit DAC conversion system. The same impairments encountered by the ADC conversion process are experienced by the DAC process. In a linear PCM system, quantisation levels are spaced apart by 1 LSB and the maximum quantisation error is +/- ½ LSB. This is the ideal, and in practice the DAC introduces linearity errors in the transfer function [101]. In addition to linearity errors, clock jitter results in the same problems as previously identified for the ADC process. As Watkinson points out [78 p722] ADCs and DACs have the same transfer function; they are only distinguished by the direction of operation. Therefore, it is not surprising that noise shaped oversampling DACs are common-place, and used for the same reasons as noise shaped oversampling ADCs.

The model put forward for the copying system is shown in fig 3.10, and consists of the original recording model plus the additional playback and re-recording processes involved in producing a copy. To simplify the model, signal conditioning noise generated by the copying process η_{cr}, represents the combination of analogue noise generated by both the playback and the copying process. The gain stage A_{cr} represents the output gain of the recorder used for playback and the recording gain of the copy recorder. In practice the recording level of the copy recorder is adjusted independently of the playback machine's output which is usually fixed at line level.

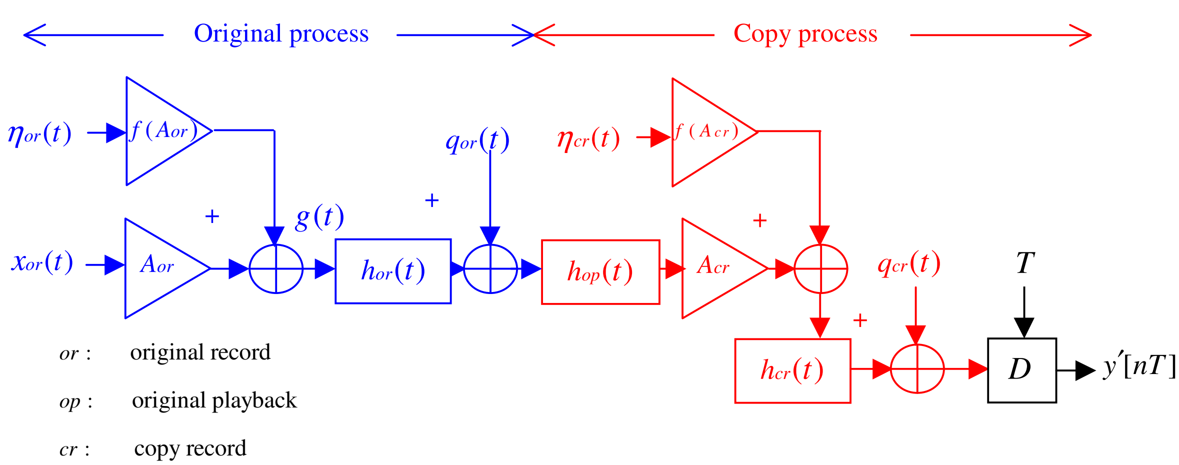

Fig 3.10: Recording system producing a copy recording.

From the model shown in fig 3.10, an expression may be formed:

$$y'[nT] = D\left[\begin{matrix}\left([A_{cr}\{[(\{Aor \cdot x(t) + f\left(Aor\right) \cdot \eta or(t)\} * hor(t)\right) + q_{or}(t)] * h_{op}(t)\} + \\ f\left(A_{cr}\right) \cdot \eta_{cr}(t)] * h_{cr}(t)\big) + q_{cr}(t), T\end{matrix}\right] \quad (3.3)$$

The blue symbols represent the original recording system, the red symbols the additional processes involved in producing the copy. It can be seen from (3.3) that compared to the original recording model, two further filter convolutions are applied to the combination of the original acoustic signal and original signal processing noise. One convolutional process is derived from the anti-imaging low-pass filter used in playback and the other is from the anti-aliasing filter of the copy recorder. In the copying model two further additive noise sources were also identified:

- η_{cr}, defined as the sum of any electronic noise produced by the DAC playback section and input signal conditioning noise produced by the copying recorder.
- q_{cr}, which is the quantisation noise produced by the digital conversion process of the copy recorder. The results from the analysis of the effects of the additive quantisation noise on the original recording process will hold for the copied recording.

It has been assumed that both new noise sources will be independent of the original noise sources and signal sources and independent of each other.

3.2 Assumptions about the Noise Sources and Acoustic Signals

The second part of developing a final model for a recording system was to examine the noise parameters and the acoustic signal characteristics typical of those encountered in a forensic environment. It was found that by making assumptions about these noise sources and acoustic signals the model derived in the previous section could be simplified. From the simplifications, differences between an original and a copy recording system can be established using the recorded data alone.

Any single forensic audio recording may contain signals from a range of sources independent or coincident in time that could fall into different signal categories. In general, a signal may be categorised as, stationary or non stationary and be further sub divided into: random stationary, deterministic stationary, continuous non stationary and burst non stationary signal types [102] as shown by fig 3.11.

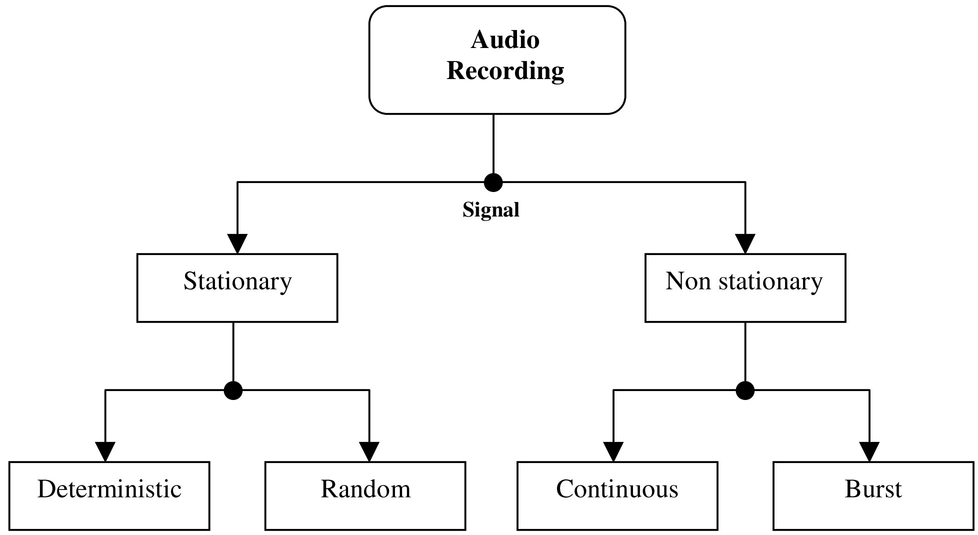

Fig 3.11: Basic signal type breakdown chart.

Definitions for the signal types now follow giving typical forensic recording examples:

- **Stationary signals**: A strict mathematical definition of the stationarity of a signal can be found at [103 ch9.1]. A random stationary signal will for practical purposes be defined here as a signal in which the first and second order moments i.e. mean, variance and covariance do not change over time, where time is limited to the duration of the recording. The random stationary signal will have a spectrum that is continuously distributed over a frequency band and can only be predicted in terms of statistical parameters. An example of stationary noise is the electronic signal conditioning noise generated by the analogue electronics of the recorder. A deterministic stationary signal will be defined as a signal with an instantaneous value that is predictable at all points over time. Induced power line frequency and its harmonics are an example.
- **Non stationary signals**: Non stationary signals are those signals that do not meet the requirements for stationary signals. Non stationary signals can be split between continuous and burst type. A continuous non stationary signal can be defined as a signal that is continuous over the analysis period but is non stationary beyond a timeframe of a few milliseconds. Speech and traffic noise are good examples. Burst type signals are defined as signals that are not continuous but have a definite starting and ending point. Burst noise can therefore take many forms having a widely differing range of signal characteristics. Examples are: car horn, ringing telephone, gunshots, opening or closing of a door.

The following sections show that the original and copy recording system models can be simplified by making some assumptions about the acoustic signals and noise sources.

3.2.1 Significance of Quantisation Noise Power

The signal conditioning noise to quantisation noise ratio (sqr) will be given by:

$$sqr = 10\log_{10}\left[\frac{\text{noise}_{\eta f(A_{or})}}{\text{noise}_q}\right] \tag{3.4}$$

It is well known [94] that for a signal quantised by an *N*-bit ideal linear PCM converter the quantisation noise variance is given by:

$$\overline{q^2} = \frac{\Delta^2}{12} \tag{3.5}$$

Where Δ is the quantisation interval.

The total conditioning noise power represents a white 'zero mean' Gaussian signal and is therefore completely described by its variance:

$$\overline{[\eta \cdot f(A_{or})]^2} = \sigma^2_{\eta f(A_{or})} \tag{3.6}$$

substituting (3.5) and (3.6) into (3.4), for the *N*-bit ideal linear PCM converter the conditioning noise to quantisation noise ratio can be found:

$$sqr = 10\log_{10}\left[\frac{12 \cdot \sigma^2_{\eta f(A_{or})}}{\Delta^2}\right] \tag{3.7}$$

It should be noted that for noise shaped oversampled systems the quantisation noise power over the audio band of the recorder can be significantly less than that indicated by (3.5) [97].

It is found in practice, that the signal conditioning noise power levels for portable audio recorders of 16 bit wordlength set to maximum recording level, can produce $sqr's$ of > 30 dB. As the conditioning noise power levels are much greater than the quantisation noise power at normal recording levels, the effects of quantisation noise on the model's output will be considered negligible.

3.2.2 Conditioning Noise Sources

In the model derived for the copy recording system, there are two sources of signal conditioning noise, the first, is introduced by the original recording process and the second by the copy recording process.

Neglecting the acoustic signal, and removing the quantisation noise sources from both original and copied recording models, the output will consist of analogue signal conditioning noise only. The dominant signal conditioning noise in the original recording system is considered to be produced by the microphone pre-amplification circuits [104], [105] in conjunction with source noise generated by the microphone itself.

When producing a copy recording from the original recording, the copy may be made by coupling into the copy recorder usually via the line input or less likely via the microphone input. Therefore, as shown by fig 3.10, the copy recording will introduce a further additive source of analogue conditioning noise $\eta_{cr} \cdot f(A_{cr})$. This additional noise source when introduced by the process of copying via the line level input should be small when compared to the high levels of signal conditioning noise generated and amplified by the microphone input circuits of the original recorder. When producing the copy via the microphone input of the copying recorder instead of the line input, the input sensitivity will be high, and the recording level of the copy recorder will therefore need to be low in order to reduce the incoming signal to an acceptable recording level. This low copy recording level will also keep any additional signal conditioning noise introduced by the copying process low.

For both copy recording scenarios the original signal conditioning noise power is usually much greater than the conditioning noise power introduced by the copy process. Therefore:

$$\sigma^2_{\eta_{or} \cdot f(A_{or})} \gg \sigma^2_{\eta_{cr} \cdot f(A_{cr})}$$

The effects of signal conditioning noise introduced from the playback and copying process on the models overall output will be considered negligible.

3.2.3 Revised Model Based on Assumptions

A new simplified model for the recording system can be formed, based on the assumptions that:

1. Quantisation noise is much smaller than the signal conditioning noise and can be neglected.
2. The additional signal conditioning noise introduced by a copying process is much smaller than the original recording signal conditioning noise and can be neglected.

Taking both assumptions into account a revised model is shown in fig 3.12.

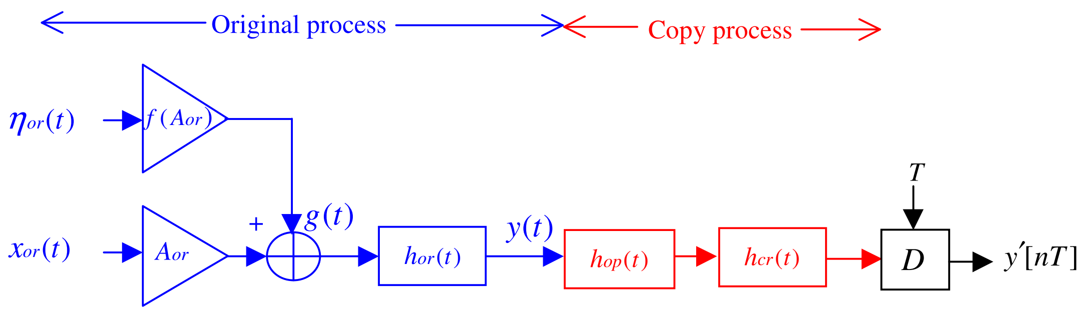

Fig 3.12: Revised copy recording system.

In this simplified model the original recording process is modified by the addition of the anti-imaging, $h_{op}(t)$ and anti-aliasing, $h_{cr}(t)$, low-pass filter convolutions resulting from the copy process:

$$y'[nT] = D\left[(\{A_{or} \cdot x(t) + f(A_{or}) \cdot \eta_{or}(T)\} * h_{or}(t) * h_{op}(t) * h_{cr}(t)), T\right] \qquad (3.8)$$

Therefore the model described by fig 3.10 reduces to the original recording model fig 3.9 plus two additional low pass filters.

A number of different recording scenarios may be involved in a copying/tampering process. For example, audio data may have been transferred to the copy recorder via some form of computer based editing system as discussed in chapter 2. If this is carried out entirely by analogue interfacing then as illustrated in fig 3.13, a total of four further low-pass filters will be cascaded in the copying process, which is in addition to the anti-aliasing filter used to produce the original recording.

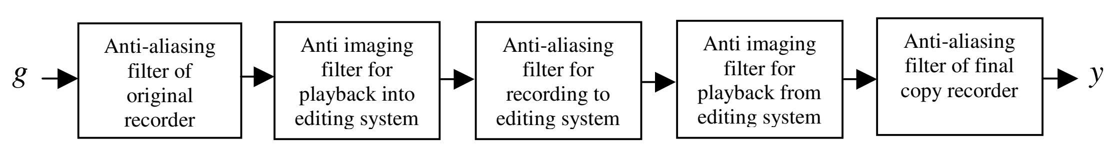

Fig 3.13: Making an edited copy recording using analogue interfacing may result in a total of four additional low-pass filters being applied to the originally recorded signal.

In its simplified form, the digital output of a recording system can be described as the total analogue input signal $g(t)$, convolved with a total of N low-pass filters:

$$y[nT] = D\left[g(t) * h_1(t) * h_2(t), * \cdots\cdots h_N(t), T\right] \qquad (3.9)$$

The time-domain model of (3.9) may be equivalently expressed in the discrete frequency-domain as the total input signal $G(j\omega)$ multiplied by the total recording system transfer function $H_r(j\omega)$, where the transfer function of the recording system is simply the overall response of the cascaded low-pass filters:

$$\Phi_{rr}(j\omega) = G(j\omega)\cdot H_r(j\omega) \qquad \text{where} \quad H_r(j\omega) = H_1(j\omega)\cdot H_2(j\omega),\cdots\cdots H_N(j\omega)$$

Conversion from discrete time to discrete frequency can be carried out using the Discrete Fourier Transform (DFT): $y[nT]\xrightarrow{fd}Y[j\omega]$.

3.2.4 Acoustic Signal Behaviour in a Forensic Environment

Understanding the overall trends and features of the acoustic signals that are to be recorded was paramount in establishing parameters from the overall recording that led to the possible identification of a recording system.

In a forensic context, the acoustic signals of interest are primarily speech recorded under various conditions and environments. Speech consists of complex sound patterns formed by the articulatory apparatus that are rich in harmonics and decrease in energy for increasing frequency [73]. Pinson and Denes report that on average, speech energy is greatest in the 100 Hz to 600 Hz region: above these frequencies the energy decreases, until at around 10 kHz it has fallen by 50 dB below the peak level [106]. This is supported by the work of Dunn and White, who report that for conversational speech 30cm from the mouth, only very small levels of energy exist for frequencies above 12 kHz [107].

Microphones are acoustically tailored to suit specific applications, and when used for covert speech recording are often of the electret type, having a frequency response tailored specifically for speech frequencies, thereby attenuating signals above the speech band. A typical frequency response for a microphone used for covert recording is shown in fig 3.14 [108].

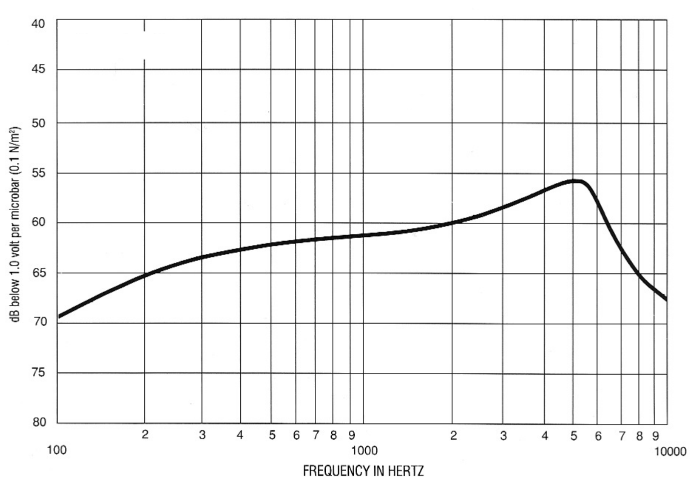

Fig 3.14: Typical microphone frequency response used for covert recording applications [109].

When a recorder captures speech, a range of other acoustic sounds will also be recorded. Vehicle and road traffic noise are often found on forensic audio recordings and fig 3.15 shows sound pressure levels of an unweighted 1/3rd octave plot of urban road traffic noise [109 p22]; significant attenuation occurs towards the higher frequencies.

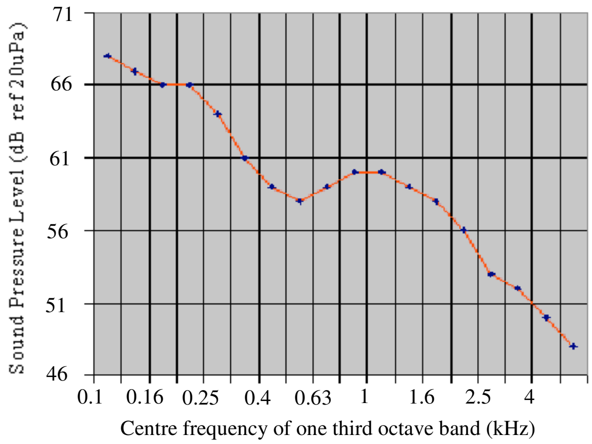

Fig 3.15: One third octave band unweighted frequency analysis of urban road traffic noise [109 p22].

Because of acoustic high frequency attenuation, it is common to find that at the high end of the recording spectrum above approximately 12 kHz, there will be very little acoustic energy. For a forensic recording having a minimum Nyquist limit of ≥ 12 kHz, the assumption is made that near to this limit, the acoustic signal energy will be negligible and therefore signal conditioning noise will predominate at these high frequencies.

To demonstrate this attenuation, a power spectrum was estimated from a two-minute sample recording of conversational speech, using a 16384 point FFT, averaged over 234 transforms. The speech recording was produced using a microphone with a response similar to that identified in fig 3.14 and the result is displayed in fig 3.16. The speech energy can be seen to drop away from its maximum power level around the 500 Hz region, completely disappearing into the signal conditioning noise after 13 kHz. For this 16 bit recording sampled at 32 kHz, a relatively flat smooth spectrum can be identified between 10 kHz and the low pass transition region due to the anti-aliasing filter starting at 15 kHz.

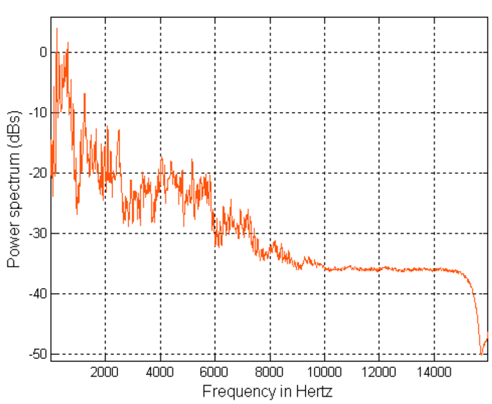

Fig 3.16: 16 bit 32 kHz sampled speech recording showing attenuation at high frequencies. Above 10 kHz the signal conditioning noise is predominant.

Examination of signals typically found in a forensic environment lead to the important conclusion that acoustic signals at high frequencies will have negligible energy leaving signal conditioning noise to predominate over this region.

3.3 Summary

A recording system model has been developed that is suitable to describe high fidelity portable audio recorders used for forensic recording purposes, having sampling frequencies > 24 kHz and effective number of quantisation bits ≥ 16. The initial model described the overall input signal as the sum of an acoustic signal generated from outside the recording system and the signal conditioning noise introduced by the analogue electronics of the recording system before ADC conversion. This combined signal is then modified by an anti-aliasing low-pass filter response. The recording system model was extended for a recording that has been directly copied using an analogue interface. This copied recording model was then simplified based on the following assumptions:

1. Signal conditioning noise of an original recording is the predominant noise source.
2. The signal conditioning noise is white and therefore has a flat psd extending beyond the frequency response of the recording channel.
3. Additive analogue noise sources introduced in a copy process are negligible compared to the signal conditioning noise produced by the original recording.
4. For recording systems of ≥ 16 effective bit resolution, quantisation noise is negligible compared to the signal conditioning noise.
5. Other sources of non-linearities produced by the digital conversion processes are negligible.
6. The acoustic signals are likely to have decayed to negligible levels near to the Nyquist limit when the sampling rate is greater than circa 24 kHz.

The output of the overall recording system is shown from the simplified model to reduce to an input signal plus signal conditioning noise modified by a number of low-pass filter functions. In general, the original acoustic signals will be unknown, but acoustic signals typically recorded in a forensic environment decay at high frequencies. Near to the transition region of the low-pass response of the recording system, the acoustic signal will have decayed to levels below the signal conditioning noise of the recording. Over this spectral region the acoustic signal can be considered negligible and only signal conditioning noise will remain.

It has been found that the response of this spectral region to the signal conditioning noise provides the basis for recording system identification and forms the subject of the following chapter.