The Application of Dither and Noise-Shaping to Nyquist-Rate Digital Audio: an Introduction.

Christopher Hicks,
Communications and Signal Processing Group,
Cambridge University Engineering Department,
United Kingdom.

cmh@eng.cam.ac.uk

September 12, 1995

Abstract

This paper discusses the theory of dithered and noise-shaped quantisation, and in particular their application to Nyquist rate digital audio. Dither is shown to be essential if quantisation-related distortion is to be avoided. Careful application of noise-shaping is shown to be of audible benefit by increasing the perceived dynamic range of a signal. Various drawbacks of the technique are discussed. In particular the 'fragility' of signals processed in this way is noted, and the ease with which the benefits may be undone is illustrated.

1 Introduction

The importance of dithering in digital audio systems has been recognised, at least on an academic level, for some time. Despite this, the subject remains poorly understood in the audio profession at large, and the unpopularity of many early digital recordings can be attributed to signal distortion at low levels, as a result of unsuitable or non-existent dither.

The technique of noise-shaping has been applied to digital audio in a number of specific areas. The best known of these are analogue to digital and digital to analogue conversion, where it is used to obtain high resolution, low bandwidth performance from a low resolution, high bandwidth converter [1].

A more recent application is CD mastering from high-resolution recordings stored on digital tape or hard disc [2]. To convert a recording of, say, 20 or 24 bit resolution to the CD format requires a re-quantisation to 16 bits. This inevitably introduces noise into the signal, and in a straightforward quantisation this noise is approximately white.

Noise-shaping can be used create a non-white noise spectrum. For example it is possible to lower the noise power spectral density in the frequency bands where the ear is most sensitive, at the expense of higher noise power in other bands (where the ear is less sensitive). This process lowers the perceived quantisation noise floor and increases the subjective dynamic range of the signal.

2 Dithered Quantisation

Consider the system shown in figure 1, whose output $y(n)$ is the sum of a high-resolution input $x(n)$ and a random dither process $d(n)$, linearly quantised by Q. Without the dither, quantisation of highly correlated signals (such as music) results in tonal distortion components being added to the signal.

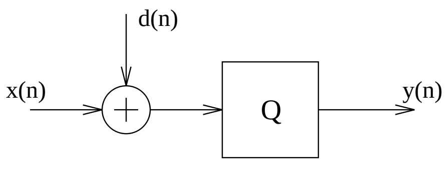

Figure 1: Simple dithered quantiser

The distribution of d is critical; it must effectively decorrelate the quantisation error from the input signal x, while adding a minimum of noise power to the output signal y. It must also linearise the quantiser transfer function, such that the expected output $E[y(n)] = x(n)$, the input. If d is a white noise source with a suitable distribution then the error $e(n) = y(n) - x(n)$ is also white, regardless of the spectrum of x.

In practice a triangular distribution for d of peak deviation $\pm q$ (the quantiser step size) is found to be suitable for high-quality audio [1]; this is often referred to as TPDF[1] dither. In this case the expected error $E[e(n)]$ is zero, the error power $\sigma_e^2 = q^2/4$, and both are independent of the signal x.

Thus the action of the dithered quantiser may be modelled as a simple addition of a stationary white noise source $e(n)$. The SNR is degraded by the TPDF dither by 4.8 dB, compared with the undithered case, but this is far preferable to the distortion that can arise from undithered quantisation.

The dither signal itself need not be white provided that it is sufficiently random that it is not perceived as a tonal component in the output signal. However, the noise added by the quantiser itself is still approximately white (if the dither is of suitable power) and the output noise is the sum of these two noise sources. Therefore, using this scheme the minimum noise power density achievable (at a particular frequency) is approximately that due to the quantiser alone, this being

$$\frac{Tq^2}{12}\ \mathrm{Hz}^{-1},$$

where $1/T$ is the sample rate.

3 Noise-shaped Quantisation

This limit may be lowered by the addition of a feedback loop around the dithered quantiser as shown in figure 2.

[1] Triangular Probability Density Function

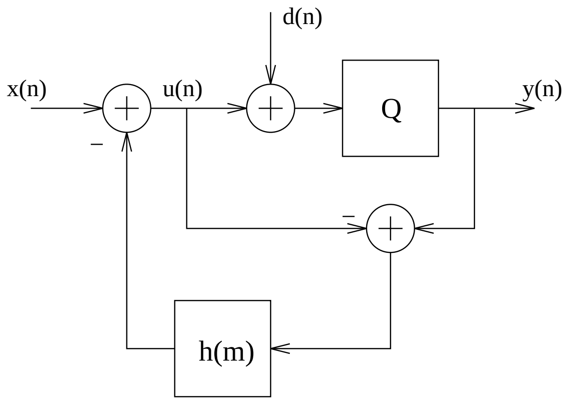

Figure 2: Noise-shaped Quantiser

The (non-linear) difference equations describing this system are

$$u(n) = x(n) + (y(n) - u(n)) \star h(m) \quad (1)$$
$$y(n) = \mathcal{Q}\left[u(n) + d(n)\right] \quad (2)$$

where $h(m)$ is the filter impulse reponse, $\star$ represents the discrete convolution operator and $\mathcal{Q}[.]$ is the quantisation function. It was shown above that if d has suitable statistical properties then the system is linearised and the combined effect of adding $d(n)$ and quantising may be modelled as the addition of an error $e(n)$ which is independent of x. Equation 2 may therefore be rewritten as a linear equation

$$y(n) = u(n) + e(n). \quad (3)$$

Eliminating $u(n)$ and taking z-transforms gives the system transfer function

$$Y(z) = X(z) + E(z)\left(1 - H(z)\right). \quad (4)$$

Note that for the system to be realisable only negative powers of z are permissible in the numerator of $H(z)$; there must be at least one sample delay in the feedback loop for the system to be causal. It follows from this that setting $H(z)$ to unity and eliminating all the noise is impossible. As with any causal recursive filter no positive powers of z are permissible in the denominator.

Substituting $e^{j\theta}$ (where $\theta = \omega T$, $\omega = 2\pi f$) for z gives the output spectrum

$$Y(e^{j\theta}) = X(e^{j\theta}) + E(e^{j\theta})\left(1 - H(e^{j\theta})\right). \quad (5)$$

From equation 5 it can be seen that the filter H changes the error spectrum by the function $(1 - H(e^{j\omega T}))$ but does not affect the signal itself.

Integrating this transfer function over frequency gives the noise power gain G and the total noise power σ_n^2 in the output signal as

$$G = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|1 - H(e^{j\theta})\right|^2 d\theta \quad (6)$$
$$\sigma_n^2 = G\sigma_e^2. \quad (7)$$

Rewriting equation 6 as

$$G = 1 + \frac{1}{2\pi}\int_{-\pi}^{\pi} H(e^{j\theta})H^*(e^{j\theta})d\theta \quad (8)$$

it is clear that the total noise power can never be reduced as the integral term in equation 8 cannot be negative.

In the case of a realisable FIR filter

$$H(z) = \sum_{p=1}^{P} b_p z^{-p} \quad (9)$$

equation 8 becomes

$$G = 1 + \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{p=1}^{P} b_p e^{j\theta p}\sum_{q=1}^{P} b_q e^{-j\theta q} d\theta. \quad (10)$$

Note that all the product terms for which $p \neq q$ integrate to zero by orthogonality. Swapping the order of integration and summation this expression becomes

$$G = 1 + \frac{1}{2\pi}\sum_{p,q=1}^{P}\int_{-\pi}^{\pi} b_p b_q e^{j\theta(p-q)}d\theta, \quad (11)$$
$$G = 1 + \sum_{p=1}^{P} b_p^2. \quad (12)$$

The noise power gain is seen to be a simple function of the filter coefficients, and the form of this expression hints at a law of diminishing returns; as the length of the filter impulse response is increased the noise power gain is also, generally, increased. This, in turn, implies that as we try to exercise more control over the shape of the quantisation noise floor, the actual signal to noise ratio is degraded.

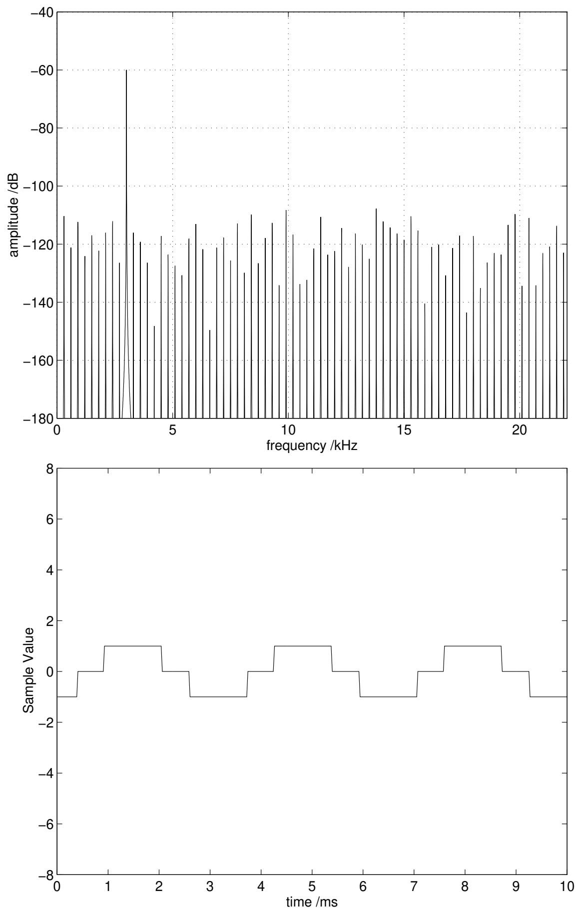

Figure 3: Undithered Quantiser

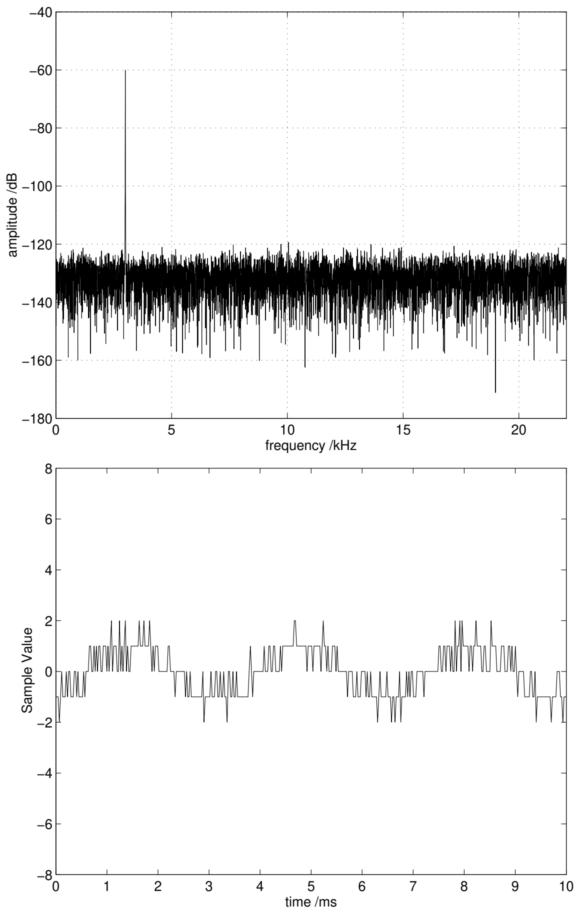

Figure 4: Dithered Quantiser

4 Simulations

A number of simulations are presented here to express graphically the effects of various types of dithered and noise-shaped quantisation. The test signal in for the spectra is a 3 kHz sinusoid, 60 dB below the maximum amplitude available in a 16 bit system. All spectra have been estimated by a 16,384 point windowed DFT. The test signal for the time-domain plots is a 300 Hz sinusoid at -90 dB.

Figure 3 shows the quantiser output if the signal is undithered. The spectrum shows a clear line structure, the spacing of these lines being a function of both the test signal frequency and the sample rate. Their frequencies are therefore modulated by the signal, and the audible result is more akin to distortion than noise.

Compare this with the same signals to which have been added white TPDF dither of peak amplitude q, prior to quantisation (figure 4). It is clear that the noise floor is less structured in this latter case. The SNR is approximately 33 dB for the dithered signal, which implies a best-case SNR of approximately 93 dB for a full-scale signal. This agrees well with the expected 4.8 dB reduction of SNR noted above.

Figures 5 and 6 show the actions of quantisers incorporating two different noise-shaping filters. Figure 5 uses $H(z) = z^{-1}$ and the calculated SNR of 30 dB agrees with the noise power gain of 3 dB calculated from equation 12. The dynamic range at frequencies below 6 kHz has been improved at the expense of dynamic range at higher frequencies. This is a particularly important filter as its implementation requires just one memory element, and two additions per sample period.

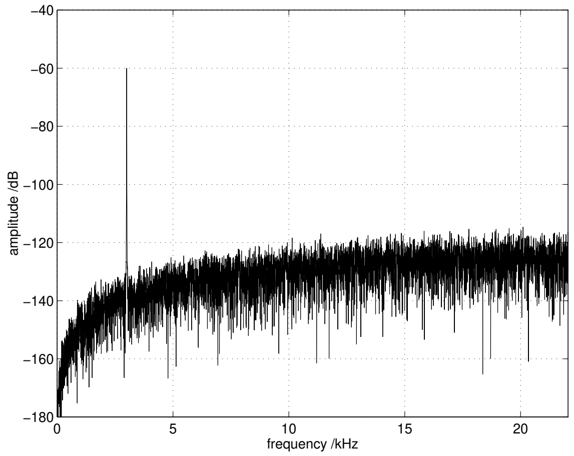

Figure 5: Quantised and noise shaped, $H(z) = z^{-1}$.

Even greater perceptual benefit can be had by tailoring the noise-shaper response to the response of the ear, which is most sensitive to broadband noise at around 3 kHz. There is a second, smaller peak around 12 kHz, above which its sensitivity decreases rapidly[3].

Equation 12 shows that the more terms are included in the filter, the greater the overall noise gain. Experimental evidence suggests that perceptually optimal performance is achieved with filter lengths between 5 and 15 for the FIR case [2]. The higher order filters allow more precise shaping of the noise transfer function at the cost of increased computation and SNR degradation.

Figure 6 shows the output of a quantiser incorporating a seventh-order FIR noise-shaping function, designed with these perceptual criteria in mind. The dynamic range at 3 kHz has been increased by about 15 dB, and again the measured SNR matches closely that predicted by equation 12, being about 14 dB worse than the non-noise-shaped signal. The high level of high-frequency noise is obvious in the time-domain plot, but the form of the original sinusoid is also clearly evident.

The subjective dynamic range increase in this case is around 7 dB. This is to say that the audible noise

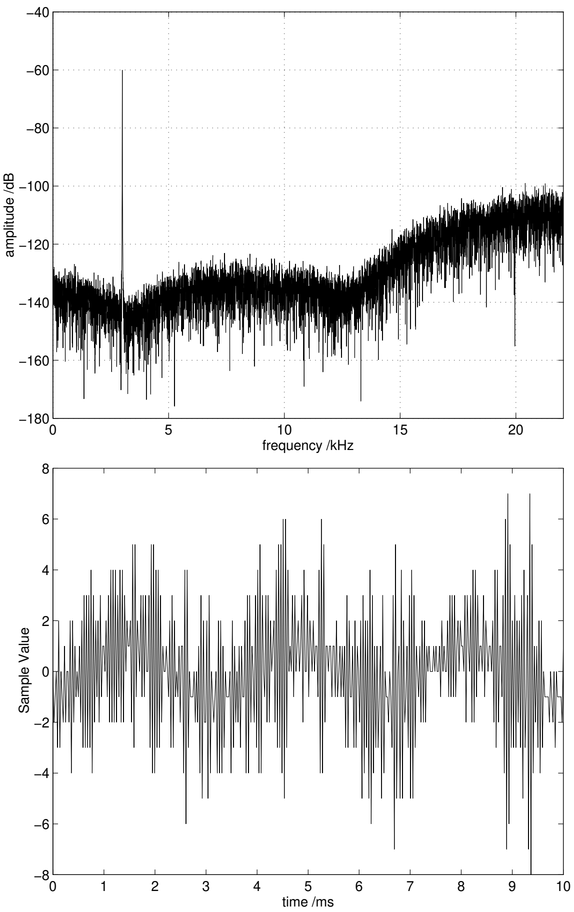

Figure 6: Perceptually noise-shaped.

level for the noise-shaper is 7 dB lower than for white quantisation using the same wordlength. Alternatively, an extra bit may be dropped from the wordlength by using the noise-shaper, without affecting the perceived noise level.

5 Discussion

We have seen that noise-shaping can be used to increase the perceived dynamic range of a digital audio signal. This is accomplished at the expense of overall SNR and high frequency headroom. Its use has a number of further implications, the most important of which are discussed here.

Subsequent Digital Processing

One major limitation of noise-shaped signals is that they are extremely 'fragile'—that is to say it is very easy to completely undo all the benefits of the noise-shaping. For example, an operation as simple as applying a gain to the signal in the digital domain may contain an implicit quantisation that adds noise to the signal. Such noise sources can easily dominate the original noise-shaped floor at those frequencies at which the noise shaper has most attenuation.

Figure 6 shows the spectrum of a sixteen-bit signal (as may be stored on a CD); figure 7 shows the spectrum of the same signal after three small amplitude changes (totalling 0 dB), each with 16-bit accuracy. The dynamic range increase resulting from the noise-shaping has been completely undone, but the excess high-frequency noise remains.

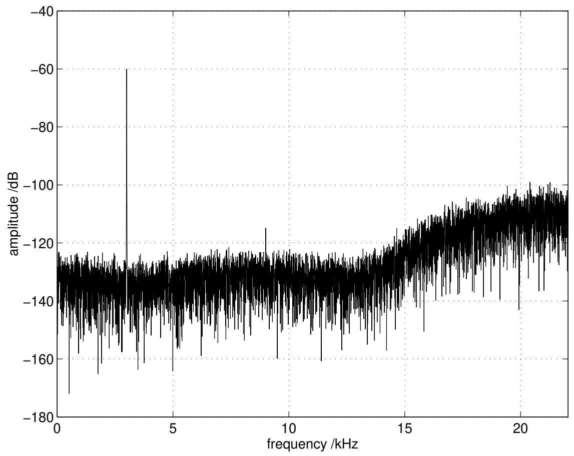

Figure 7: Spectrum after 16-bit gain change.

Notice also the quantisation-related *spuriae* (the most prominent at 9 kHz) which could have been avoided by dithering the (implicit) quantiser. Of course, this dither and quantisation could also incorporate noise-shaping, but this must be approached with caution as successive applications of severe noise-shaping can cause excessive high-frequency noise to be added to the signal.

If simply changing the amplitude of a signal is enough to upset its delicate balance of bits, then more complex operations (such as equalisation or compression) spell instant death!

Sample Rate Conversion

There is a number of different sample rates in common use for digital audio, those most often encountered being 48 kHz, 44.1 kHz and 32 kHz. This range of common sample rates has a number of implications for the design and use of noise-shapers.

Firstly, it is important to note that the frequency response of a given digital filter changes with sample rate, but the response of the ear is fixed absolutely in frequency. Thus different sets of filter coefficients are required in the noise-shaper for each sample rate that is encountered.

Secondly, since a sample rate converter (SRC) is essentially a digital filter, care must be taken to ensure that quantisations implicit in the filter do not adversely affect the signal in a manner similar to the gain change mentioned previously. This includes any quantisation implicit at the output of the converter, for example when rate-converting one 16-bit signal direct to a second 16-bit recording medium.

Finally it has been suggested [4] that asynchronous sample rate conversion is an effective technique to combat the jitter inherent in digital audio interconnections. In particular the use of a single-chip SRC immediately before a DAC enables the DAC itself to run off a clean, local clock, rather than running from a potentially noisy clock recovered (for example) from an S/P-DIF[2] channel. It is important, once again, to retain sufficient wordlength at the SRC output, and to use a DAC with sufficiently high resolution and low internal noise to preserve the dynamic range.

Numerical Headroom

Up to this point we have ignored the effect of saturation of the output word; in short, if the input signal peaks at (or close to) full-scale then addition of dither (and particularly noise-shaped dither) can cause the output word to saturate.

Of course this may be avoided by reducing the amplitude of the input signal, but the effect of this

[2] Sony/Philips Digital Interface Format

varies greatly with output wordlength. The reduction required for the perceptual noise shaper above is approximately 10 LSB's which corresponds (in a 16-bit system) to around 0.003dB, which is, of course, insignificant.

However, in (for example) a 6-bit system (perhaps an audio coding application) this same headroom requirement corresponds to a 1.5 dB loss of dynamic range, which must be traded against the perceived dynamic range increase that results from the noise shaping.

Implications for the Studio

It is important that noise-shaping is the very last operation in the mastering process, and that no further manipulation of the samples occurs with wordlengths shorter than that of the original source tape.

Unfortunately this is not the universal industry practice. Many stand-alone analogue to digital converters of 18 or 20 bit resolution incorporate noise-shaping algorithms so that critical mid-band dynamic range is preserved when recording to a sixteen bit medium such as DAT or CD-R[3]. If the signal is replayed directly then (almost) all is well; if, however, the recording is loaded to an editing system then problems begin to emerge.

Since it is a sixteen bit signal that has been recorded, the temptation when loading to a hard-disc-based workstation for editing is to store the data as sixteen-bit words, since this is efficient in terms of space, speed and therefore also of cost. As shown above it only requires the level of the signal to be changed during editing to undo the benefit of the noise-shaping, but leave behind undesirable high-frequency noise.

If the application of noise-shaping becomes widespread then the use of 24-bit signal paths throughout the studio is to be encouraged.

Consumer Equipment

If now we assume that a correctly noise-shaped signal has been encoded on a CD, the problems have still not gone away. Availability of cheap DSP power makes it a tempting proposition to do, for example, digital equalisation or sample-rate buffering in consumer replay equipment. This too must be done to high accuracy if the noise-shaped signal spectrum is to be accurately preserved, implying long wordlengths and higher costs.

Furthermore, the operation of the DAC itself must be considered. The vast majority of converters used for digital audio use digital oversampling filters to relax the requirements of the analogue anti-image filter. For noise-shaping to be of benefit, these digital filters also must preserve the full dynamic range of the source tape in the audio band. Many of the standard chipsets do not fulfil this requirement, but rather contain significant arithmetic noise sources that can significantly degrade the perceived dynamic range of noise-shaped signals.

[3]Recordable Compact Disc.

6 Conclusions

The careful application of dither is essential when requantising digital audio samples. Addition of suitable dither is effective at eliminating quantisation-related distortion, with minimal degradation of the signal to noise ratio.

Noise-shaping is a powerful technique that can be used to enhance the perceived dynamic range of a digital recording. It relies on the use of a digital filter to alter the spectrum of the dither and quantisation noise applied to a digital audio signal at the final stage of mastering. However, to deliver this extended dynamic range to the domestic listener requires great care.

In particular, arithmetical operations such as digital filtering, can easily undo all of the benefits of noise-shaping. To this end it is not desirable to apply noise-shaping techniques to the Nyquist rate audio until all digital editing operations are completed; it should be the very final stage of mastering.

Additionally, digital processing in domestic equipment (including oversampling in a DAC) has to be approached with care if the benefits of a noise-shaped signal are to be realised.

References

[1] John Watkinson. *The Art of Digital Audio.* Focal Press, 1994.

[2] Sony Corporation. Super Bit Mapping; technical overview.

[3] B.C.J. Moore. *An Introduction to the Psychology of Hearing.* Academic Press, 1989.

[4] Analog Devices. Technical data for the AD1890 asynchronous sample rate converter, 1993.

Appendix: Sony SBM

Super Bit Mapping[4] is a noise-shaping algorithm developed by Sony for mastering of Compact Discs from 20-bit source tapes. It uses an FIR filter of order 12 for $H(z)$ and a little reverse engineering from [2] enables calculation of a set of filter coefficients (for a 44.1 kHz sample rate) which duplicates the SBM noise-shaping curve. The magnitude of the noise transfer function is plotted in figure 8.

Coefficients for Transversal FIR			
b_1	$1{\cdot}47933$	b_2	$-1{\cdot}59032$
b_3	$1{\cdot}64436$	b_4	$-1{\cdot}36613$
b_5	$9{\cdot}26704\times10^{-1}$	b_6	$-5{\cdot}57931\times10^{-1}$
b_7	$2{\cdot}67859\times10^{-1}$	b_8	$-1{\cdot}06726\times10^{-1}$
b_9	$2{\cdot}85161\times10^{-2}$	b_{10}	$1{\cdot}23066\times10^{-3}$
b_{11}	$-6{\cdot}16555\times10^{-3}$	b_{12}	$3{\cdot}06700\times10^{-3}$

Table 1: Filter to mimic SBM.

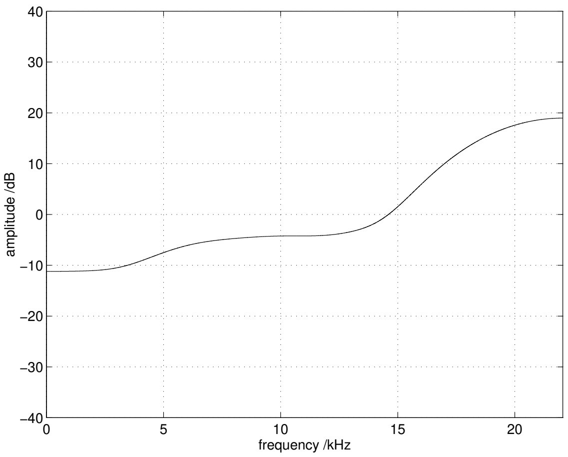

Figure 8: Quantisation noise gain.

[4] *Super Bit Mapping* and *SBM* are Sony trademarks.