HIGH DYNAMIC RANGE AUDIO APPLICATIONS FOR DIGITAL SIGNAL PROCESSING

Julian Dunn*
Cambridge, England
*Email: JDunn@iee.org

ABSTRACT

Application of digital signal processing to audio is often limited by available converter technology and recorder datawidth to processes with controlled dynamic range. In digital mixing consoles some analogue gain control is normally required to achieve adequate performance for the input channel ADCs and digital multi-track recorders.

This paper examines such dynamic range bottlenecks and derives specifications for typical applications. Equipment that has been designed to these specifications is described.

1. INTRODUCTION

The digital audio systems discussed in this paper are being used to process high fidelity audio signals for replay with a minimum loss in quality. This is often not possible because the dynamic range of the final product cannot match the source due to limitations at key points in the chain.

This paper will review the dynamic range requirements for digital audio equipment at various points in the chain. Using measurements of real signals and the performance of the human ear it will show that these requirements can be defined, and are not just a performance specification that continuously needs to increase as newer technology becomes available.

Some of these requirements will be shown to far exceed the performance of currently available equipment. In some respects this limitation is a product of designing equipment to `ideal' n-bit performance specifications, when perhaps `non-ideal' n+4 bit performance is actually required.

Noise shaping techniques can improve the apparent dynamic range of some of this equipment to match operational needs. In addition two techniques of `Dynamic Range Enhancement' have been developed at Prism to improve on the performance of the equipment:

A coding algorithm has been developed that allows the use of 16 bit linear PCM recorders for recording the output of analogue to digital converters (ADCs) with outputs longer than 16 bits, and replaying the signal onto high dynamic range media, such as 20-bit disk based digital audio editing systems, with the minimum of degradation in the noise floor.

An ADC system is also being designed specifically for high dynamic range applications. The specification for this unit will be described and related to real applications.

2. DEFINITION OF DYNAMIC RANGE

Throughout this paper the dynamic range of a channel refers to the ratio of the energy of a sine wave of maximum amplitude (0dBFS) to the energy of the error signal produced when the device is modulated at low levels (typically at -60dBFS). This definition is further qualified in this paper when referring to perceived, or perceptually weighted, dynamic range. The weighted dynamic range is defined as the dynamic range of a channel with a white noise floor that appears to have the same dynamic range as the channel in question, when the channels are compared with their noise at the threshold of hearing. This is a rudimentary way of accounting for the variation in the perception of noise at different frequencies. These topics are covered more rigorously elsewhere [6,7,1,2].

3. EXAMPLES OF DYNAMIC RANGE REQUIREMENTS

3.1 The Domestic Listener

The end result of this audio processing is the reproduction of sound to the listener. This may be taking place in an environment in which there is little or no extraneous noise to mask noise in the audio signal chain. Fielder [11] has shown that white noise at 4dB SPL is just audible for people in either a home listening or professional monitoring environment, even with environmental noise levels of 50dB SPL. This compares with maximum sound levels between 105dB and 112dB SPL (corresponding to digital full scale) for listening levels in domestic or review situations [6]. Subtraction of the 4dB figure from the peak level figures implies a dynamic range requirement, for the listener, of from 101dB to 108dB.

3.2 The Source Material

Some of the background noise in a recording is part of the ambience of that recording. This could be from the audience, traffic, or even a ventilation system. This noise can have the effect of masking noise that is due to the audio signal processing and recording, but this masking effect is often very small. An example from Cohen [12] is the Davies Symphony Hall, in San Francisco, with the ventilation and air conditioning on. This shows the noise level up to 8dB above the threshold of audibility below 500Hz and below the threshold in the critically sensitive region between 3 and 4kHz. This part of the spectrum is where white noise is heard if it is only just above the threshold of hearing.

Maximum sound levels in recordings will obviously vary between material, and they will depend on listening or microphone positions. Fielder [11], also drawing on earlier work [13], has shown that at typical listening positions levels can reach up to 122dB SPL in very percussive passages of orchestral classical music with a solo piano capable of producing over 100dB SPL. These figures can be easily exceeded with close miking techniques, without necessarily exceeding the capability of the microphone.

These figures can be combined to show that musical source material recorded with microphones at a typical listening position can require a dynamic range of up to 118dB from an audio processing system to reproduce it without the latter introducing audible noise. This requirement can be increased further with the use of microphones sited closer to the source.

3.3 Headroom and Footroom

Sound engineers require a certain amount of headroom in order to provide some margin for signals exceeding expected levels. This headroom adds to the dynamic range requirement and its size depends on many factors:

The expense of re-recording if signal clipping is found to be unacceptable
The consistency of sound levels between rehearsals
Sound level metering accuracy

Practical experience at Thames Television in the UK has shown a requirement for a working headroom 10dB above the intended peak level for controlled level signals [9]. Much of this is to avoid clipping transient programme peaks that do not register on the monitoring level meters.

For similar reasons some parts of a recording may be recorded at too low a level. This could be corrected by the application of gain to the channels affected. This can often be done in the analogue domain at the time of the recording by adjustment of microphone amplifier gain, but it may be only discovered after the event. If the gain is applied after part of the processing chain then it will also raise the processing noise level in that channel. An allowance of noise floor below that which is required for the final product is called footroom.

Some footroom is also required to ensure that the sum of the noise powers introduced at each stage of processing is not audible.

3.4 The dynamic range requirement for digital audio equipment

An adequate dynamic range is the minimum required that will not itself produce degradation to the signal quality that would be audible to a listener. The previous sections show that this requires dynamic ranges as follows (allowing 12dB for headroom and footroom):

Dynamic range before level control: *118dB (plus headroom and footroom - 130dB)*

This will maintain the dynamic range of the original signal. The size of the headroom and footroom at this stage will depend on operational practices.

Later in the chain the signal level can be controlled to match the more limited dynamic range of the replay system, but headroom is still required to allow for error, and footroom will be required if any further intermediate processing is to be applied to the signal.

Dynamic range after level control: *108dB (plus any headroom and footroom - 120dB)*

A BBC report [10] concludes that 22 bits of coding range are adequate for carrying signals that have not been controlled in level (including those feeding multi-track recorders for subsequent mix-down) and that 18 bits are required for controlled signals, allowing 2 bits of headroom before producing a 16 bit signal adequate for distribution or disc mastering. (The removal of headroom in the reduction to 16 bits can be done using automatic peak limiting techniques to avoid clipping)

4. THE DYNAMIC RANGE OF AVAILABLE EQUIPMENT

4.1 Microphones

The microphones reported in [11] had overload levels between 120 and 140dB SPL, with noise corresponding to an 8dB SPL white noise threshold. Those in [12] have noise figures between 1dB and 10dB above the threshold of hearing. This suggests that microphones are close to matching the dynamic range requirement mentioned above.

4.2 Analogue to Digital Converters (ADCs)

Recent advances in ADC design have been reported. The most significant advance has been the development of the noise-shaping oversampling converter. At dbx Adams [4] developed a chip set that he reported as having a dynamic range of 114dB when used with two front-end devices. More recently Adams et al [5], at Analog Devices, have developed a single package device that has a dynamic range of 105dB. There are other units available which claim a similar dynamic range to the latter: several of them using an UltraAnalog module set that appears to be based on the dbx parts.

These figures indicate that the required dynamic range of 118dB for conversion of uncontrolled signals is not far beyond the capabilities of available technology. The dynamic range for controlled signals of 108dB could already be digitised without noticeable degradation of the noise floor using the dbx design mentioned above. There would also be 6dB of dynamic range in hand to allow for some limited headroom and footroom.

4.3 Multi-track tape recorders

The author is not aware of any multi-track tape recorders that operate with word-lengths of greater than 16 bits. With appropriate dither this corresponds to a dynamic range of 93.3dB. There could be significant operational benefit from operating multi-track machines that do not require input level control. Unfortunately this is not possible with such a restricted dynamic range. The dynamic range of each channel is 15dB less than that required for the final product, and 25dB less than the possible dynamic range of the source - any allowance made for headroom and footroom will further increase this deficit.

It is argued by some that the effective dynamic range of a multi-track recorder is enhanced, as a final mix is a combination of several channels. This effect is small. The most extreme benefit would be if 24 channels of a 48 track machine carried the same signal. Assuming the noise is incoherent, between the channels, then the increase in dynamic range would be:

$$10.\log(24) = 10.8 \text{ dB}$$

This improvement would raise the dynamic range of the system to 104dB, remaining 4 dB worse than the final product requires - without allowing anything for headroom or footroom.

4.4 Mixing consoles

Most professional digital mixing consoles have been designed with digital interfaces of at least 20 bit wordlength. The intermediate processing is often performed to a minimum word length of 24 bits or in a floating point format - which will have a more than adequate dynamic range. (The other problems related to the application of floating point arithmetic in digital audio processing do not directly affect the dynamic range as defined in section 2, and will not be considered here).

An interface wordlength of 20 bits has a dynamic range of 117dB (with flat TPDF dither). This is not adequate for uncontrolled input signals with headroom or footroom. Interfaces implementing the full AES-3 wordlength of 24 bits would be able to allow 23 dB for headroom and footroom.

4.5 Stereo recorders

This type of recorder is normally be used for (re-)recording the mixed-down signal. This will be at controlled levels so that a 20 bit format is adequate. All the new digital video recorder formats can maintain this resolution, as can the newer analogue video machines with audio PCM tracks. There is one audio only machine that the author is aware of that has this resolution - the Mitsubishi X-86 open reel machine.

Nagra have produced a portable 24bit recorder. This coding range is adequate for recording even uncontrolled levels with headroom and footroom.

Most recorders used for the recording of processed audio signals have a word-length of only 16 bits. The dynamic range of these recorders is inadequate for the task without application of some of the techniques described later. At present this is not seen as a problem by many users because the signal dynamic range has already been degraded by earlier digital and analogue dynamic range bottlenecks.

4.6 Distribution media

Compact Discs (CDs) and Digital Audio Tape (DAT) machines both provide 16 bit resolution. With TPDF dither this provides a dynamic range of 93.3 dB. The perceived dynamic range is improved by some disc mastering systems through the use of noise shaping techniques - which can provide enough subjective improvement to achieve the required 108dB [2].

4.7 Domestic replay systems

As there has not been any noise shaped programme material available unsurprisingly the consumer electronics industry has not been producing players of enough dynamic range to reproduce the signals adequately. However there are some DAC devices available that could provide the required performance. For example the delta-sigma TDA1547/SAA7350 combination from Philips has a potential dynamic range of about 105dB. Operated in pairs these devices may provide the 108dB dynamic range required for domestic replay without audible noise.

5. TECHNIQUES FOR INCREASING DYNAMIC RANGE

The dynamic range limitations in the equipment type described in the previous section generate bottlenecks in the signal path. At these points either maximum level or noise floor has to be sacrificed compared with the requirements outlined earlier. The bottlenecks are, in order, the 16 bit multi-track recorder, the 16 bit post-mix recorder, and the 105dB ADC. At these points there are dynamic range shortfalls of 25dB, 15dB and 13dB respectively, before allowing for operational headroom and footroom. Digital processing techniques can reduce the first two of these bottlenecks.

5.1 High pass dither

Spectrally flat triangular probability density function (PDF) dither is required to eliminate signal related quantization distortion products: producing a natural noise floor with a flat spectrum. This is the optimal PDF for eliminating both input-dependent distortion and noise modulation [3], and increases the noise power over re-quantization alone by a factor of 3. Use of this form of dither provides the following dynamic range in a linear PCM channel quantized to n bits:

$$\text{Dynamic Range} = 6.02n - 3.01 \text{ dB}$$

The spectrum of this noise can be shaped, without changing its PDF, to have more of its energy at the higher audio frequencies by using a simple algorithm [8]. This high pass dither results in lower noise levels at frequencies below about 8kHz, with a perceived dynamic range improvement of approximately 3.4 dB [1]. This modifies the previous formula to:

$$\text{Weighted dynamic range} = 6.02n + 0.4 \text{ dB}$$

5.2 Noise shaping through error feedback

At the point of requantization the error signal (noise) being introduced is known. If this signal is filtered and fed back to the input of the quantizer the spectrum of this noise is shaped. With simple feedback the weighted dynamic range increase over high-pass TPDF dither alone is a further 3.5dB - while the total unweighted noise increases by about 2 dB [2]. Noise shaped to be minimally audible can improve dynamic range by 16dB compared with high pass TPDF dither alone [1] but at the cost of a large unweighted noise penalty as the noise levels at the less audible high frequencies rise. For example, the N(2,9) shaper quoted in [1] increases the perceived dynamic range of the quantizer by 13.5dB (over that achieved from high-pass TPDF dither alone) with a reduction in the unweighted dynamic range of 23.5dB.

This noise, if it becomes audible, has an unnatural spectrum, and an increased amount of footroom may be required to ensure that it does not rise above the threshold of hearing in any processing. For minimally audible shaped noise any equalisation providing boost at high frequencies will reduce this footroom immediately. Lipshitz et al [3] recommend that this amount of noise-shaping is reserved only for the final requantization before replay, as the white re-quantization noise floor added by any intermediate processing will rapidly `fill in' the carefully shaped spectrum and degrade the dynamic range.

The two digital recorder bottlenecks can clearly be reduced using this technique. It has been shown that a 16 bit post-processing recorder could gain over 19dB of weighted dynamic range using this technique [2], giving it a dynamic range of:

$$\text{Weighted DER} = 6.02n - 3.01 + 19 \text{ dB}$$

$$= 112.31 \text{ dB} \quad (\text{for } n = 16)$$

This gives it enough dynamic range for handling level controlled signals that do not require headroom or footroom. This amount of shaping causes the unweighted noise floor to rise by more than 20dB above the unshaped noise level. Authors have also suggested that it may be prudent to limit the degree of shaping [1,2] either before any editing or before the last quantization point, and this could reduce the benefit so derived to less than 12dB.

The same technique could be used at the initial, uncontrolled level, recorder. This would also reduce the bottleneck at this point. The improvement may not be as great as greater allowance will have to be made for footroom to avoid the audio processing altering the spectrum of the shaped noise and thus making it audible. The shortfall may be reduced from 25dB (plus headroom and footroom) to 15dB. This still leaves the 16 bit multi-track recorder, working with uncontrolled signals, restricting the dynamic range of the digital audio system, but it now can have approximately the same apparent dynamic range as the ADCs driving it.

5.3 Schemes that require decoding

The shaped dither and noise techniques have the advantage that they do not have to be undone. The signal remains a linear PCM signal with the same level as before. There are other techniques for improving perceived dynamic range that can be applied that require `decoding' to reproduce the original signal. These range from 15/50us emphasis (which requires de-emphasis) to the data compression schemes used for digital audio broadcasting.

Emphasis can make apparent improvements in dynamic range if the higher amplitude signals are only at low frequencies. In order to exploit it properly programme level meters have to operate on the emphasised signal, and the operators need to be familiar with a frequency dependent maximum sound level. For these reasons it is not being considered here.

The more sophisticated techniques use forms of floating point, instead of linear, coding. These code a signal to a resolution that varies with level. Take, for example, a 16 bit data channel. A 14+2 bit scheme uses 14 bits to record the signal, with a 14 bit mantissa being ranged in level by the 2 bit exponent so that the quantization stepsize varies in four 6dB steps. This has the benefit of increasing the dynamic range by 12dB (as this is the noise improvement at low signal levels) but the signal to noise ratio for signals above -6dBFS is reduced by 12dB, compared with linear coding, and the noise floor will fluctuate with signal levels between full scale and -24dB.

This noise modulation will be inaudible if the original signal masks it. The required mantissa length is determined by this masking effect, which depends on the spectral content of the signal. Lee and Lipschutz have examined the effect with various computer generated signals and found that a 13bit mantissa was required for the modulation to be inaudible [15]. Fielder [14] has also examined the audibility of modulation noise with more forms of signal and found that with signals at the extremes of the audio spectrum the mantissa has to have even greater resolution. The signal to noise ratio required to mask the modulation noise varies from a maximum of over 100dB with a maximum level 40 Hz tone to a minimum of 60dB at 2-3 kHz. Emphasis can attenuate these lower frequencies and Fielder has shown in tests with real music that using a form of emphasis a signal to noise ratio of 75 dB is satisfactory for the modulation noise to be inaudible in a floating point system.

6. DYNAMIC RANGE ENHANCEMENT OF DIGITAL RECORDER CHANNELS

The aim of the work described later, and referred to as `dynamic range enhancement' (DRE) is to improve the dynamic range of those elements in the programme chain that are limiting the dynamic range of the finished product. As has been shown above these elements are the ADC, the input signal recorder and a 16 bit post processing recorder. The dynamic range of the recorder channels can be improved using the noise shaping techniques described above. These could allow a 16 bit channel to have a weighted dynamic range of 108dB - which matches the requirement for the finished product, but without provision for headroom or footroom. Prism have produced the DRE coding algorithm to increase the dynamic range of a 16 bit channel even further in order to provide this headroom and footroom.

For some applications the large amount of headroom for intermediate processing stages is required to eliminate the possibility of clipping. Evans [9] and Manson [10] allow 10 and 12dB respectively to cope with signals that have exceeded the expected maximum signal level. Clipping is less acceptable than the more gentle overload behaviour of analogue systems - where the effect is a gradual increase in low order distortion rather than a wide band discontinuity. For linear PCM the top 6dB of dynamic range uses half the codes available. The DRE algorithm reduces the resolution with which these signals are coded so that the headroom does not use up such a large proportion of the data space. This behaviour is already typical of most converter architectures, as shown in figure 1. This plot of THD+N versus signal level for a high quality ADC/DAC pair shows the THD+N increasing rapidly above -6dBFS - and is particularly appropriate in situations where this part of the dynamic range is headroom reserved for transient overlevel signals.

Below the headroom area any reduction in the resolution of the signal would cause noise modulation for normal programme signals, not just transient overloads. This noise modulation is audible if it is not masked by the original signal. The DRE algorithm only slightly increases the signal resolution for high level signals, with the effect increasing below this level allowing the noise level for lower level signals to be significantly reduced.

The algorithm was developed last year to allow an ADC/DAC unit with a dynamic range greater than 100 dB to be used with a DAT machine. The programme runs in a DSP56001 running at 20MHz, and code operates on the signal in both directions simultaneously to allow for simultaneous decoding of the recorder monitoring output through the DAC while the ADC is encoding the data. This unit has been used for professional recordings - allowing the transfer of a digital audio recording with a dynamic range of over 100dB from the recording location to a 20 bit disc based audio editor, without loss in resolution, using an ordinary 16bit DAT recorder.

Figure 2 shows a plot of the THD+N floor versus signal level for this system. The top line shows the performance of the codec when operated back to back via the 16 bit truncation of a DAT machine. (The truncation is dithered, using TPDF dither, in the codec). The bottom line shows the unit operating in 20 bit mode and bypassing the DAT machine. The DRE line was plotted via the DAT machine which truncated the data to 16 bits. It can be seen from the graph that below -20dBFS the noise of the DRE path is the same as the 20 bit path. Above this level the noise is allowed to increase until at -8dBFS it is greater than that of the 16 bit path. Even at high levels the DRE THD+N performance is within 4 dB of the dynamic range of the original signal.

This algorithm, which is still under development, has been designed to behave in an appropriate manner with digital recorder error concealment schemes and metering. The effect of interpolations on the DRE signal is between a zero order hold and a linear interpolation, with the concealment as effective as with linear coding at low levels.

7. A DYNAMIC RANGE ENHANCED ADC

Some of the principles used in the development of the DRE algorithm have been applied to the design of a high dynamic range ADC. The prototype performance indicates that an unweighted dynamic range in excess of 117 dB will be achieved, and with combinations of units improvement on this is expected. The unit will produce an output word-length of 24 bits that can be reduced to match recorder wordlengths using noise-shaping and/or DRE techniques. Using both techniques will allow its use with 16 bit recorders without reducing the perceived dynamic range of the ADC.

8. CONCLUSION

The dynamic range requirements for audio processing have been summarised. Comparisons with digital audio equipment show that the performance of the digital multi-track recorders is 25dB short of the dynamic range of some source material. Noise shaping techniques can give 16 bit distribution media, such as CDs and DAT, enough dynamic range to be able to eliminate audible noise at replay, and 20 bit dynamic range recorders are required to handle the level controlled signals, prior to distribution, without degradation of this. Some noise shaping techniques can to be applied to recording devices earlier in the chain, but noise shaping on its own is not enough to overcome the dynamic range bottleneck that occurs if uncontrolled signals are recorded onto 16 bit multi-track machines.

A coding scheme has been developed that can be applied to 16 bit recorders to allow them to be used with controlled signals without compromising the dynamic range of the final product. It can also be used with digital multi-track tape machines to significantly reduce the noise that they add to the final signal.

9. REFERENCES

[1] Stuart, J. Robert and Wilson, Rhonda J. - `A Search for Efficient Dither for DSP Applications', Preprint 3334, presented at the 92nd AES Convention (March 1992)

[2] Lipshitz, Stanley P., Vanderkooy, John and Wannamaker, Robert A. - `Minimally Audible Noise Shaping', *J. Audio Eng. Soc.*, vol 39 No. 11, pp 836-852 (November 1991)

[3] Lipshitz, Stanley P., Wannamaker, Robert A. and Vanderkooy, John - `Quantization and Dither: A Theoretical Survey', *J. Audio Eng. Soc.*, vol 40 No. 5, pp 355-375 (May 1992)

[4] Adams, Robert W. - `An IC Chip Set for 20-bit A/D Conversion', *J. Audio Eng. Soc.*, vol 38 No. 6, pp 440-458 (June 1990)

[5] Adams, R.W., Ferguson, P.F., Ganesan, A., Vincelette, A., Volpe, A. and Libert, R - `Theory and Practical Implementation of a Fifth-Order Sigma-Delta A/D Converter', *J. Audio Eng. Soc.*, vol 39 No. 7/8, pp 515-528 (July/August 1991)

[6] Stuart, J. Robert - `Estimating the Significance of Errors in Audio Systems', Preprint 3208, presented at the 91st AES Convention (October 1991).

[7] Stuart, J. Robert - `Predicting the Audibility, Detectability, and Loudness of Errors in Audio Systems', Preprint 3209, presented at the 91st AES Convention (October 1991)

[8] Skritek, Paul- `Prospective techniques for improved signal-to-noise ratio in digital audio conversion systems', Full length version of paper presented at the 82nd AES Convention, London (March 13 1987)

[9] Evans, P. - `Digital audio in the broadcast centre', *EBU Review - Tech.* Nos. 241-242 , pp 118-125 (June/August 1990)

[10]Manson, W.I. - `Digital Coding Ranges for use in an all-digital studio-sound complex', BBC Research Department Report No. 1985/16 (December 1985)

[11]Fielder, Louis D. - `Dynamic-Range Requirement for Subjectively Noise-Free Reproduction of Music', *J. Audio Eng. Soc.*, vol 30 No. 7/8, pp 504-510 (July/August 1982)

[12]Cohen, Elizabeth A. and Fielder, Louis D. - `Determining Noise Criteria for Recording Environments', *J. Audio Eng. Soc.*, vol 40 No. 5, pp 384-402 (May 1992)

[13]Sivian, L.J., Dunn, H.K. and White, S.D. - `Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras', *IRE Trans. Audio*, vol AU-7, pp 1-29 (May/June 1959)

[14]Fielder, L. D. - `The Audibility of Modulation Noise in Floating-Point Conversion Systems', *J. Audio Eng. Soc.*, vol 33 No. 10, pp 770-781 (October 1985)

[15]Lee, Francis F. and Lipschutz, David - `Floating-Point Encoding for Transcription of High-Fidelity Audio Signals' *J. Audio Eng. Soc.*, vol 75 No. 5, pp 266-272 (May 1977)

FIGURES

1. THD versus level for a leading `20 bit' ADC/DAC codec pair.

2. THD versus level for another ADC/DAC pair, showing the effect of limiting the wordwidth to 16 bits, with and without `dynamic range enhancement'.

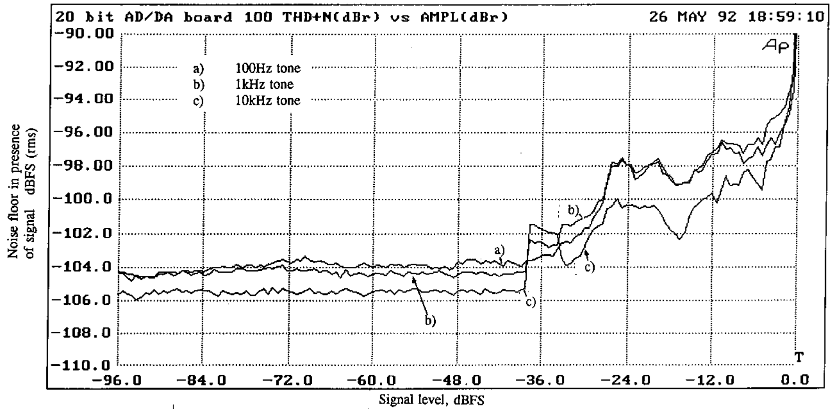

Figure 1 - THD versus level for a leading `20 bit' ADC/DAC codec pair.

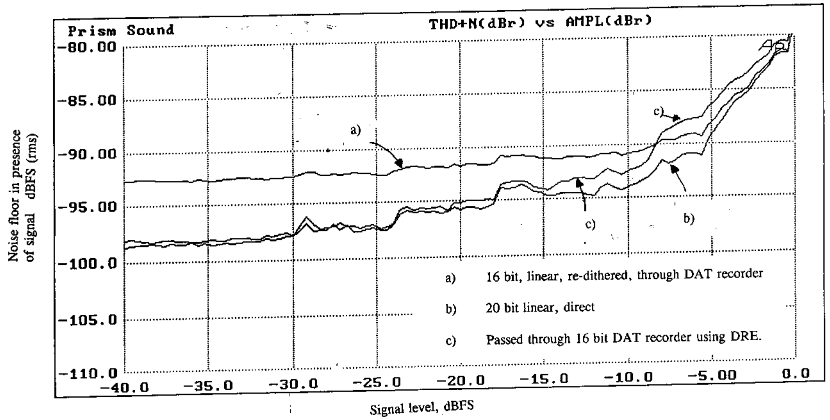

Figure 2 - THD versus level for another ADC/DAC pair, showing the effect of limiting the wordwidth to 16 bits, with and without `dynamic range enhancement'.