HIGH DEFINITION AUDIO: WHAT HAS IT DELIVERED?

K Howard Consultant Technical Editor, Hi-Fi News
Contributing Editor, Stereophile

1 INTRODUCTION

High resolution music carriers in the form of DVD-Audio and Super Audio CD have been available for some years. This paper reports on analyses of the content of 15 tracks from 15 different discs (eight DVD-As and seven SACDs) which represent a variety of source material (analogue and digital, legacy and modern). Spectral content, dynamic power requirement and rate of change measurements for these recordings expose a wide range of content quality and are revealing of the amplifier performance required to reproduce them. Further investigation into the signal-to-noise performance of ten 24-bit PCM recordings, assessed by whether truncation to 16 bits results in correlated quantisation error, indicates that most do not exceed 16-bit S/N performance because of microphone self-noise.

2 ANALYSIS

2.1 Source Discs

The discs and tracks used for the first part of this study are listed in Table 1. Where stereo and multichannel mixes were available on the same disc, the stereo mix was used. On DVDs offering both LPCM tracks and data compressed tracks within a video area on the disc, the LPCM tracks were selected. Analysis was conducted by recording each track to hard disk and applying custom software utilities for the various analyses. Pioneer DV-939A and Philips SACD1000 players were used to replay the DVD-A and SACD material respectively, with the Philips on its highest (50kHz) output filter setting. Wherever 24/96 material on DVD-A was available without downsampling on the player's S/PDIF output, the transfer to hard disk was accomplished digitally via a Lynx Studio Technology L22 sound card. In all other cases the transfer was performed via the L22's analogue inputs at 192kHz sampling rate. Note that this causes the spectral graphs to have different frequency axes, some to 48kHz and others to 96kHz.

2.2 Spectral Content

Figure 1 contains nine example FFT spectra that illustrate various aspects of the frequency content variation found on the tested discs. Figs 1a to 1e were obtained from DVD-As and the remainder from SACDs. Both channels were tested but as the results were broadly similar in most cases only the left channel results are shown here. In each graph the upper, red trace represents the peak or envelope spectrum compiled from the highest amplitude recorded in each FFT bin. The lower, blue trace represents the average spectrum across the whole track. Generally the envelope spectrum is a better indicator of the track's audio bandwidth since the presence of frequencies above 20kHz is often episodic.

Fig 1a clearly shows that this disc either has 48kHz sampling rate or is upsampled from a 48kHz master – it is not possible to say which since the disc packaging contains no information on sampling rates. Since the material is digitally remastered from analogue tape and therefore could easily benefit from an extended HF bandwidth, this seems an odd technical decision.

Fig 1b is from an Fs=192kHz track also derived from an analogue master. Clearly some poorly realised upsampling has been applied to create the 192kHz track, rather than it being recorded from the analogue master at 192kHz.

Fig 1c shows what can be achieved when an analogue master is correctly captured as 24/96 LPCM. Recorded in 1984 – later than to two previous examples – it contains musical information out to about 30kHz.

Fig 1d is from a modern 24/96 audiophile recording and has musical content out to over 40kHz.

Fig 1e is also a modern recording, this time 24/192 and from a mainstream record label. Musical content to well in excess of 40kHz is clearly visible.

Fig 1f is instantly recognisable as being measured from SACD because of the hump of ultrasonic noise appearing above 20kHz. This recording is obviously derived from a 48kHz PCM master, which makes it a strange candidate for release on SACD. Nothing on the disc packaging describes this origin.

Fig 1g is the SACD equivalent of Fig 1c, the two being derived from the same analogue recording. Although they are on different frequency scales, the two spectra look very similar. A closer inspection on a log frequency scale would be advisable before using these two discs to compare DVD-A to SACD.

Fig 1h is from a modern audiophile recording. Two features are of interest here. First, the extended HF musical content disappears into the shaped quantisation noise just below 40kHz. Second, there is a peak in the quantisation noise at about 55kHz which appears to be a characteristic of the A/D converter used for this recording (Philips DSD recorder).

Fig 1i again shows this telltale peak but the spectral content of this recording begins to drop off before 20kHz, despite the presence of cymbals. This suggests that large-diaphragm microphones were used, with restricted ultrasonic capability.

2.3 Dynamics

This part of the investigation was inspired by NAD's Power Envelope research[1,2] in the 1980s which attempted to define a more relevant way of specifying amplifier power output by measuring the burst power requirements of music signals over time intervals down to 20ms. NAD's work was conducted by means of oscilloscope observations since at the time the tools to rip tracks from CD and analyse the content in software did not exist. Today we can do much better by analysing the digital data directly, although in the case of DVD-As and SACDs we often have to record the material to hard disk first as the audio data cannot be accessed on the source disc itself.

The method used to expand on NAD's pioneering work is as follows. First, the RMS signal level throughout the track is determined in block sizes of 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048 and 4096ms. Second, the level of the highest-amplitude 4096ms block is set to be 0dB. Third, the relative levels of the highest-amplitude shorter blocks are plotted on a graph. Figure 2 shows how data from the 2ms blocks is combined to calculate levels for the longer blocks, using 2ms block offsets. To ensure that the highest possible levels are recorded, further single-sample offsetting of the 2ms blocks should also be used but this makes the analysis unacceptably long.

Figure 3 shows a result obtained from an orchestral recording on CD (blue trace) with NAD's Power Envelope plotted for comparison (red trace). It is evident even from this one example that the burst power requirements suggested by NAD are often exceeded, and that the power requirement continues to rise below 20ms. The investigation can be extended using histograms like that shown in Figure 4 (for a piece of piano music) which indicate how often burst power in excess of 0dB is required.

Although the following analyses, as already suggested, use 2ms block offsetting without further single sample offsetting, Figure 5 suggests that the errors are small. Here the blue trace represents the results obtained by 2ms block offsetting, while the red trace shows the results for full single-sample offsetting. Only for time intervals below 16ms are any differences visible.

Figure 6 shows the results obtained from four of the tested tracks, the blue trace here representing the left channel and the red trace the right channel. Fig 6a shows the least challenging result obtained, where the burst power requirement at 2ms reaches about 7.5dB (5.6x). Fig 6b shows the most challenging result obtained from DVD-A, where the 2ms burst power requirement reaches 16dB (40x). Although the left and right channel plots are often very similar, as in these two examples, this is not always the case, as shown by Fig 6c. In this case the presence of various percussion instruments in the right channel makes its burst power requirement significantly more challenging than the left channel's. Fig 6d shows the most challenging result obtained overall, from an audiophile SACD, where the burst power requirement at 2ms exceeds 16dB on both channels.

To convert these findings into burst power requirements for *inaudible* clipping would require knowledge of the relevant audibility thresholds, which I don't believe exist. Moreover, clipping recovery behaviour and 100Hz modulation of the clipping point with unregulated power supplies would likely complicate the issue. If clipping is to be avoided altogether – and there is an argument which holds that any clipping is a negation of high fidelity – then it is clear that a burst power capability of approximately +16dB is required. This can be provided either by an amplifier with continuous capability to spare, or an amplifier – *eg* a commutating amplifier – which is designed to deliver additional power over short timescales.

2.4 Rate of Change

In 1977, Peter Baxandall measured the maximum slew rate obtained when playing music from LP using a differentiating network (*ie* a high-pass filter with its corner frequency well above the audible frequency range) and an oscilloscope[3]. He observed a maximum slew rate of 0.14V/µs, equivalent to a full-scale sine wave of frequency 2.2kHz. Once again, today we are able to perform this test much more easily and accurately by direct analysis of the numbers comprising a digital recording. Moreover, there is every reason to do so with DVD-A and SACD since their extended HF bandwidths hold the possibility of increased signal rate of change.

For the measurements described here, the measured rate of change in LSB per sample period was converted, in line with Baxandall's work, to an equivalent full-scale sine wave frequency using the equation

$$f = \frac{S.f_S}{\pi.2^B}$$

where S is the difference in amplitude between adjacent samples in LSB, f_S is the sampling frequency in hertz and B is the recording's resolution in bits.

Figure 7 shows five examples of rate of change results obtained from the test discs. The blue trace again represents the left channel and overlays the red trace representing the right channel.

Fig 7a was obtained from the DVD-A recording of the *1812 Overture*, one of two recordings in the test group that were duplicated on DVD-A and SACD. Fig 7b shows the equivalent result obtained from the SACD recording, which differs in two significant ways. First, the SACD version has the higher peak equivalent frequency, corresponding to about 10.5kHz full-scale. The prominent spikes towards the end of the track are caused by cannon shots and reach the highest equivalent frequency recorded from any of the test tracks. Second, ultrasonic quantisation noise clearly dominates much of the SACD result, obliterating the fine detail seen in the DVD-A traces. Two stepchanges in the quantisation noise level are clearly evident, which are presumed to indicate some gain-riding in the transfer to disc.

There are two obvious reasons why the SACD peaks may exceed those from DVD-A. First, the DVD-A track has a 44.1kHz sampling rate (despite the original batch of discs claiming 88.2kHz), so its bandwidth is restricted compared to the SACD version. Second, the SACD's ultrasonic quantisation noise can add to the rate of change due to the signal itself.

Fig 7c shows another SACD result in which the quantisation noise almost completely dominates the traces, despite the source material containing percussive sounds. In Fig 7d the quantisation noise is even more dominant and this time appears at obviously different levels in the two channels, suggesting that the balance of the recording was adjusted to increase the amplitude of the right channel.

Fig 7d might be mistaken for an SACD result were it not for the fade to zero at the track's conclusion. In fact is measured from a modern pop music recording on DVD-A, the sustained 'hash' in this case being due to the presence of unrelenting percussion throughout.

2.5 Signal-to-Noise Ratio

No attempt was made to determine the signal-to-noise ratio of the tracks listed in Table 1. But the ten 24-bit tracks listen in Table 2 were indirectly assessed for signal-to-noise performance as part of later work relating to the word length reduction of 24-bit files to 16-bit. This began as a result of casual experiments in which 24-bit files were converted to 16-bit with and without TPDF dither, and then subject to listening comparisons. It was expected that the truncated files would generally sound worse than the correctly dithered files due to the creation of signal-correlated quantisation errors. In fact the undithered files were preferred, and that preference was confirmed by other listeners.

To examine the nature of the quantisation error in the truncated files, a software utility was written to extract and amplify it so that it could be listened to and analysed. If S is a 24-bit integer sample value from the source file then, in BASIC notation, the quantisation error that results from truncating it to 16-bit resolution is given by

S/256 – CINT(S/256)

and has a maximum amplitude of 0.5. The code used to extract the quantisation error calculates

CINT(16000 * (S/256 – CINT(S/256) + RND + RND – 1)

to amplify and then quantise the error signal with TPDF dither, so as to prevent the addition of correlated errors by the quantisation.

There are three forms of extracted quantisation error that we should expect to encounter: (1) quantisation noise (*ie* uncorrelated error); (2) quantisation distortion (*ie* correlated error); and (3) any component of the 24-bit signal having an amplitude of less than 0.5LSB 16-bit. Figure 8 shows example spectra for these three forms of error.

It quickly became apparent when listening to the extracted quantisation error that for most of the time it appeared to be white-spectrum random noise, and spectrum analysis confirmed this. In only once case among the ten 24-bit recordings listed in Table 2 was the quantisation error obviously correlated with the signal. As anticipated, spectrum analysis of brief periods of silence before the start of the music (Figure 9) showed this particular track to have a significantly lower noise floor (red trace) than one in which the extracted quantisation error sounded non-correlated (blue trace). High-pass filtering the latter to remove studio sounds and then analysing it resulted in the probability density function shown in Figure 10, which confirms the noise source to be Gaussian.

The obvious candidate, bearing in mind that the recording concerned is digital in origin, is microphone self-noise. Typical studio capacitor microphones have equivalent self-noise, unweighted, of 17dB SPL or worse, which means that the microphone has to be subjected to a peak sound pressure level of approximately 110dB if it is to achieve 16-bit equivalent noise performance. In many recordings, particularly audiophile recordings of acoustic music in natural acoustics where the microphones are quite distant from the performer(s), this will not be the case. The relatively high level of Gaussian noise in the recording can then be expected to act as dither when a 24-bit recording is truncated to 16 bits.

In an attempt to establish formally whether the extracted quantisation error is random noise, the error files for the ten 24-bit recordings listed in Table 2 were subjected to an autocorrelation analysis. The quantisation error was divided into 8192-sample blocks, each of which was subject to a lag 1 autocorrelation test. Blocks which failed this test (*ie* whose content appeared not to be random) were counted and they were each subjected to a full autocorrelation analysis for sample lags out to 4096 samples.

Table 3 shows the number of failed blocks for the left and right channels of each of the recordings of Table 2. In all but one case the number of failed blocks is a very small proportion of the total, the exception being that already referred to where the extracted quantisation error was audibly correlated. Figure 11 shows the autocorrelation graphs for the 11 failed blocks of the right channel of recording number 7, of Haydn chamber music. The blocks are arranged in chronological order through the track. Note that the most obviously correlated blocks appear near the beginning and end of the file. This is often found; in fact it is typical for failed blocks to be clustered around the start and end of the track in this way. This occurs because the microphones are usually faded up and then down at these points, resulting in brief periods when there is insufficient microphone noise present to provide effective dithering. Once the microphones are at full gain, the instance of failed blocks typically falls to an extremely low frequency.

3 CONCLUSIONS

1) Some odd material has found its way on to high-resolution music carriers, *eg* 48kHz PCM on SACD and ineptly upsampled material on DVD-A despite there being an analogue source. The labelling of discs is such that it is often unclear exactly what is on offer.

2) Wide bandwidth recordings have signal content to over 40kHz, sometimes over 50kHz. With SACD some ultrasonic signal components are lost in the shaped quantisation noise. It is evident in some recordings that the use of large-diaphragm microphones is a limiting factor.

3) To prevent clipping on the most testing material the power amplifier must be able to deliver short-term power over 16dB (40×) higher than is required on a continuous basis.

4) The maximum rate of change measured was equivalent to a full-scale sine wave of approximately 10.5kHz.

5) Many (most?) 24-bit PCM recordings barely better 16-bit noise performance, if at all, because of microphone self-noise. SACD recordings must be similarly limited within the audio band.

6) Recording engineers face a dilemma: whether to maximise bandwidth (small-diaphragm microphones) or minimise noise (large-diaphragm microphones).

4 REFERENCES

1. B.-E. Edvardsen, P. W. Mitchell and P. Tribeman, ‘The Power Envelope: a better way to compare the musically useful power output of amplifiers’, NAD Electronics.

2. P. W. Mitchell, ‘A Musically Appropriate Dynamic Headroom Test for Power Amplifiers’, Preprint 2504, Audio Engineering Society 83rd Convention (October 1987).
3. P. Baxandall, ‘Audio Power Amplifier Design’ (part 1), Wireless World, January 1978.

5 FURTHER READING

K. Howard, ‘New Media Metrics’, Stereophile, April 2004 (link at www.audiosignal.co.uk/webwords).
K. Howard, ‘The Dither Habit’, Hi-Fi News, February 2005.
K. Howard, ‘Contingent Dither’, Stereophile, July 2005 (link at www.audiosignal.co.uk/webwords).

Table 1. Listing of the DVD-A and SACD tracks used for the spectral, dynamics and rate of change measurements:

DVD-As

Title	Duration	Recording	Format
Tchaikovsky 1812 Overture Telarc DVDA-70541	15:43	DSD	24/44.1
Walton Portsmouth Point Overture EMI 7243 4 92402 9 2	05:33	analogue	24/48
Fly Me To The Moon Sinatra At The Sands EMI 7243 4 92402 9 2	02:47	analogue	24/192
Hotel California Eagles: Hotel California Elektra 7559-60509-9	04:56	analogue	24/192
Easy Does It The Ray Brown Band: Soular Energy Hi-Res Music HRM 2011	03:59	analogue	24/96
The Conga Kings: Tumbao de Tamborito The Ultimate DVD Surround Sampler Chesky CHDVD221	06:24	PCM	24/96
Mujaka The Latin Jazz Trio AIX Records AIX 80011	09:39	PCM	24/96
Slang Of Ages Steely Dan: Everything Must Go Reprise 9362 48435-9	04:14	PCM	24/192

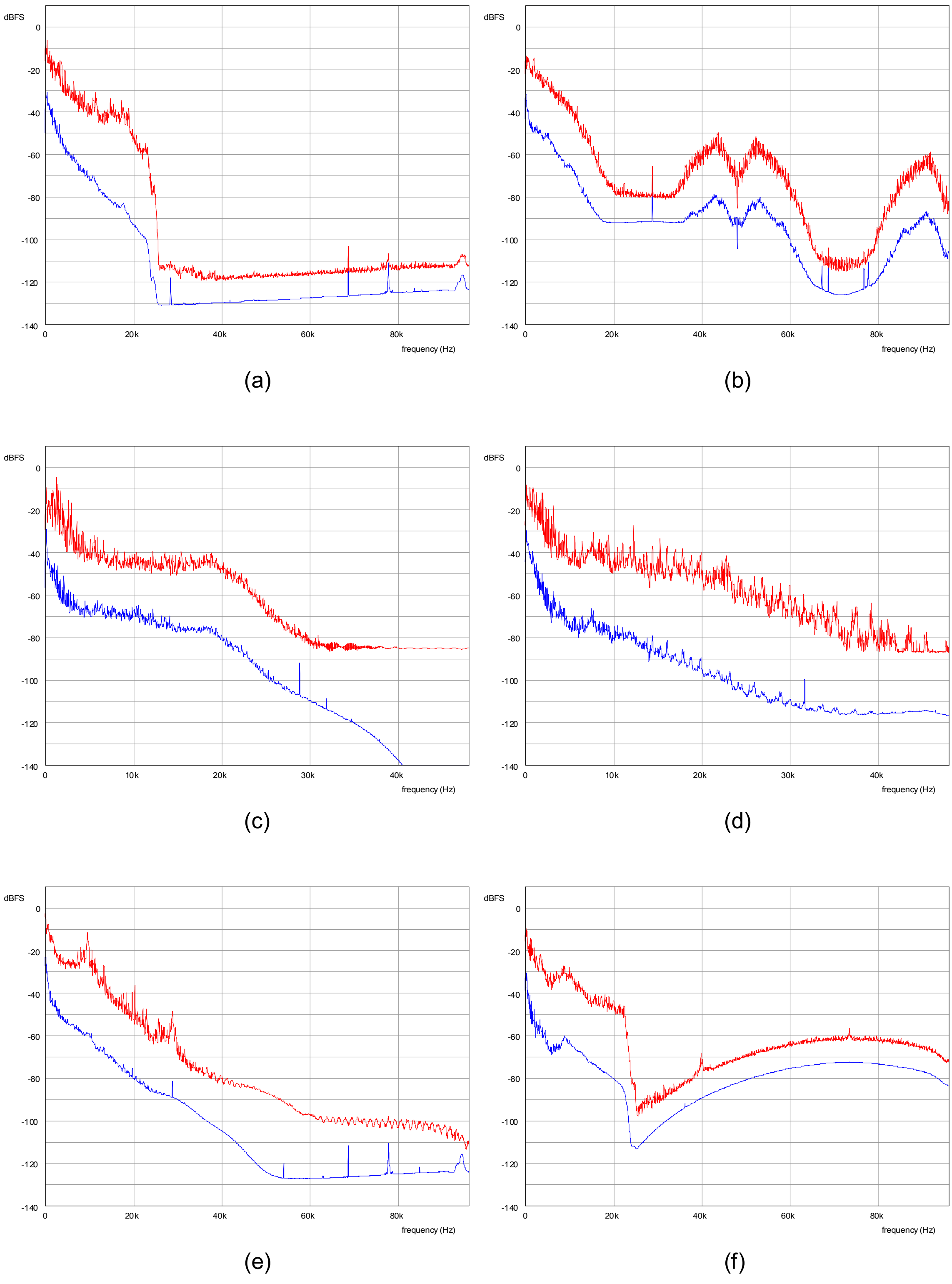
SACDs

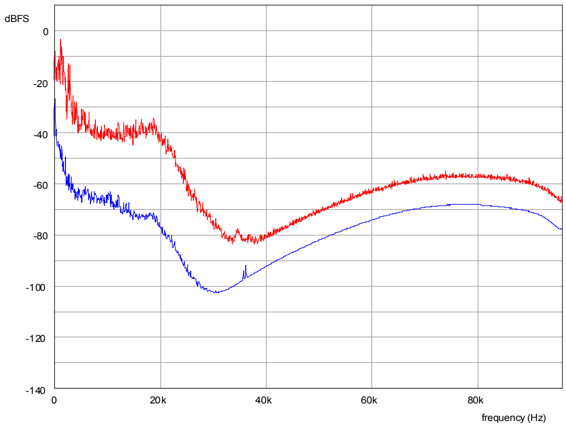
Title	Duration	Recording
Tchaikovsky 1812 Overture Telarc SACD-60541	15:44	DSD
Easy Does It The Ray Brown Trio: Soular Energy Groove Note GRV1015-3	03:59	analogue
Blues Opus 3 Opus 3 Showcase Opus 3 CD-21000	04:34	DSD
One Two Free The Steve Davis Project: Quality of Silence DMP SACD-04	03:46	DSD
Forget About It Alison Krauss: Forget About It Rounder SACD 11661-0465-6	03:28	PCM
Dancing In The Dark Diana Krall: The Look Of Love Verve 589 597-2	05:48	DSD
Simple Waltz Clark Terry & Max Roach: Friendship Eighty-Eight's VRGL 8805	03:08	DSD

Table 2. Listing of the 24-bit LPCM recordings used for the quantisation error analysis:

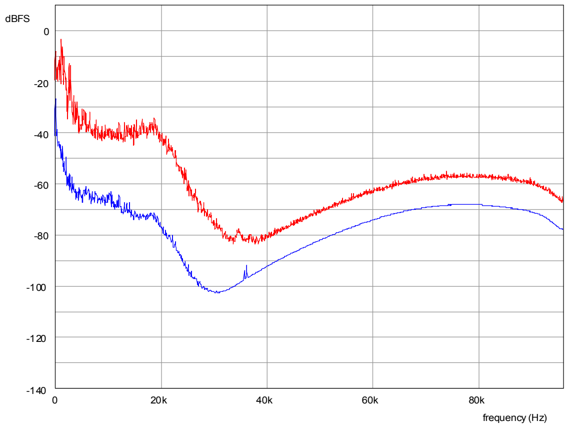
Track ref	Artist(s)	Track	Album	Label	Number
1	Dave's True Story	Daddy-O	Sex Without Bodies	Chesky	CHDVD174
2	Sara K	Brick House	Hobo	Chesky	CHDVD177
3	John Basile Quartet	Desmond Blue	The Desmond Project	Chesky	CHDVD178
4	Latin Jazz Trio	Doña Olga	Latin Jazz Trio	Aix Records	AIX 80011
5	Zephyr	Now Is the Month of Maying	Voices Unbound	Aix Records	AIX 80012
6	Peppino D'Agostino	Desert Flower	Acoustic Guitar	Aix Records	AIX 80013
7	Pro Arte Trio	Trio No1, Menuetto	Haydn Piano Trios	Aix Records	1340 AX
8	George Enescu Quintet	Scarlatti Six Sonatas No2	Scarlatti/Beethoven	Aix Records	1341 AX
9	The Ray Brown Trio	Easy Does It	Soular Energy	Hi-Res Music	HRM 2011
10	Robert Silverman	excerpt	Diabelli Variations	Stereophile	STPH017-2

Figure 1. Peak (red trace) and average (blue trace) spectra for various tracks from Table 1:

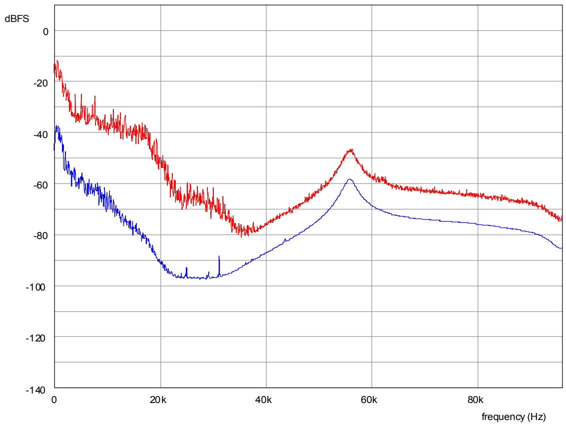




(g)



(h)



(i)

Figure 2. The block stepping technique used to determine peak levels across different time frames:

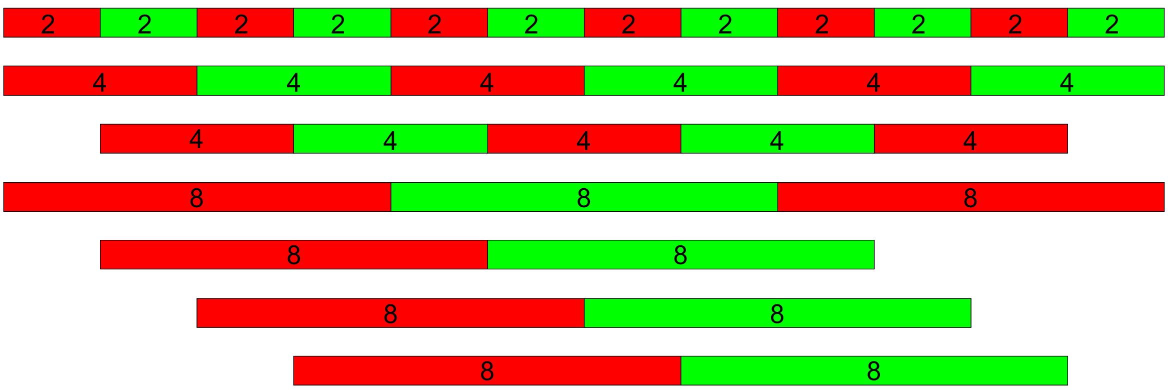


Figure 3. NAD's Power Envelope curve compared with a numerical analysis of CD data:

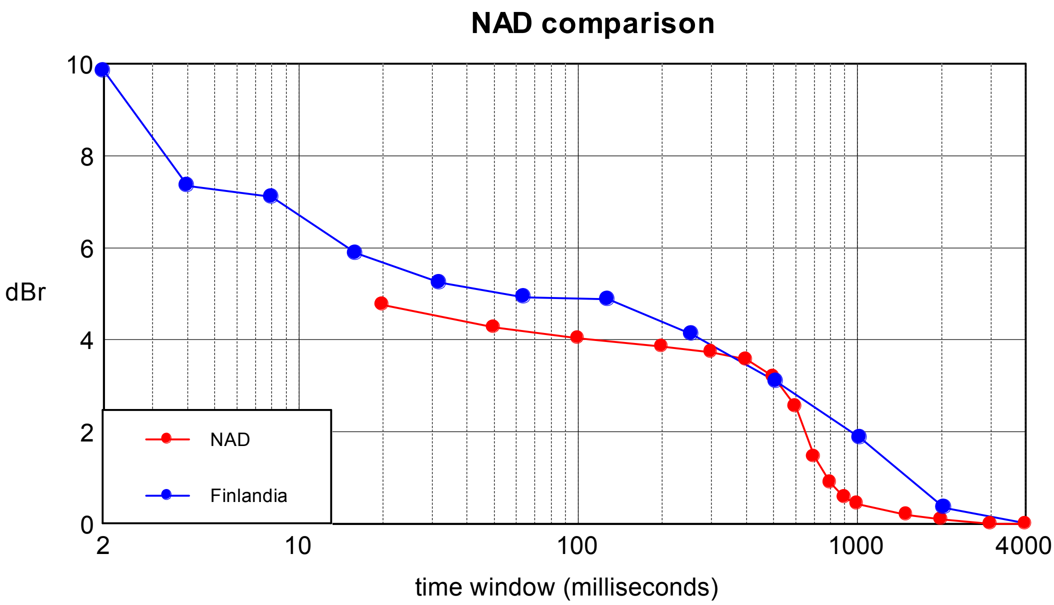


Figure 4. Example histogram of peak level occurrences:

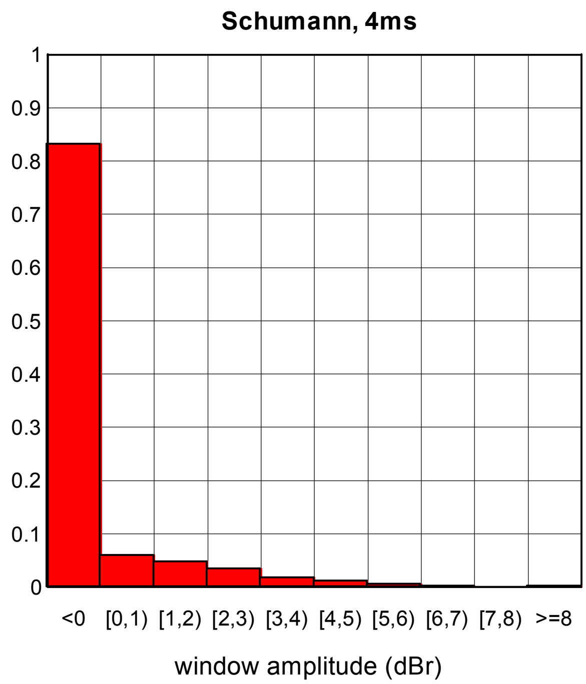


Figure 5. Comparison of results obtained with 2ms block stepping alone (blue trace) and further single-sample stepping (red trace):

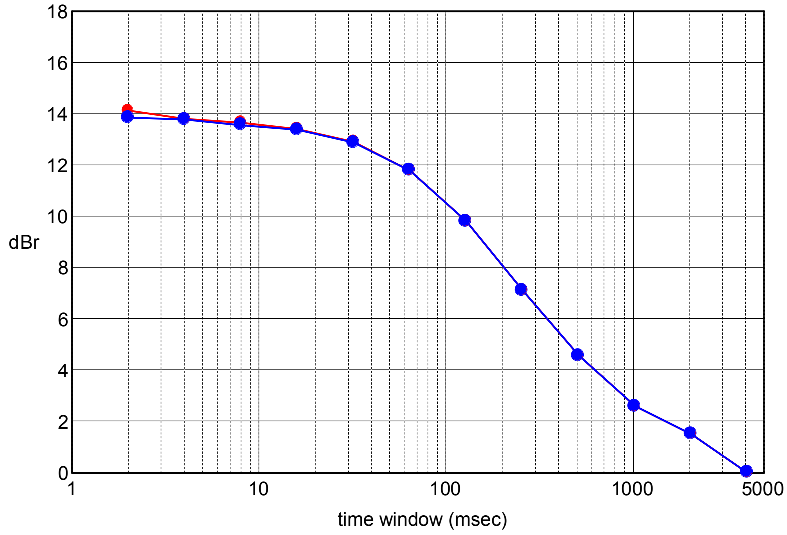


Figure 6. Peak level analysis of four example tracks from Table 1:

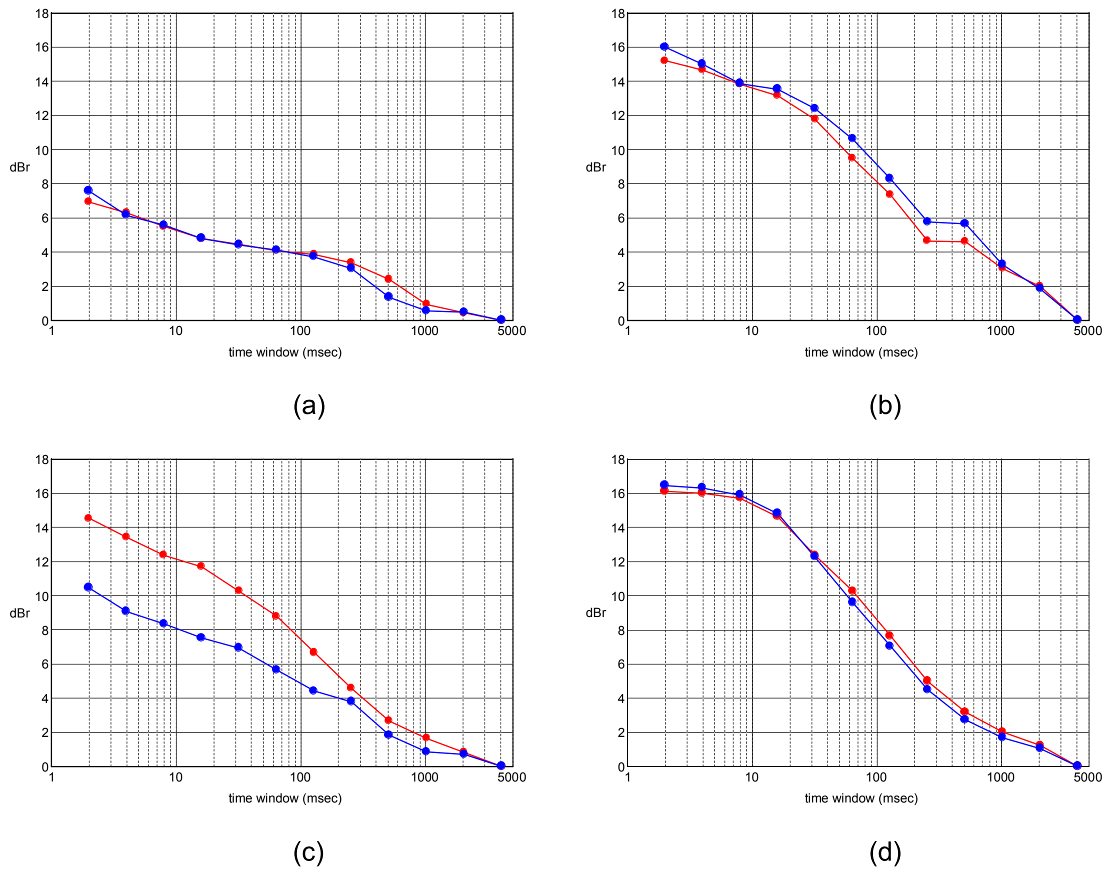
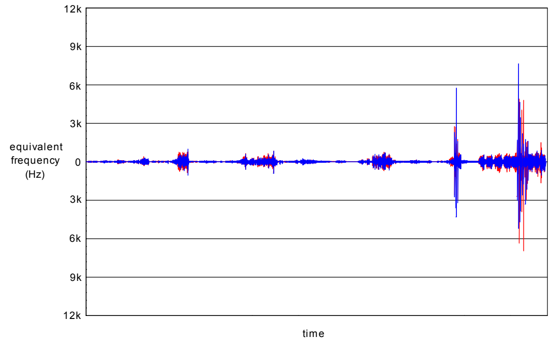
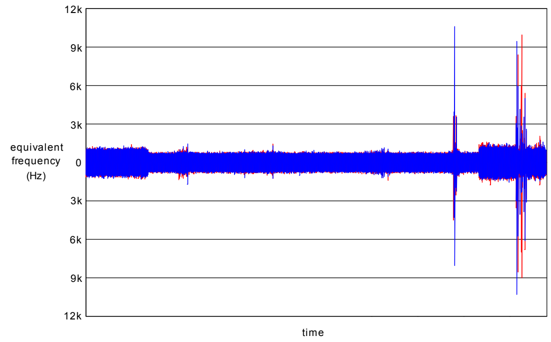


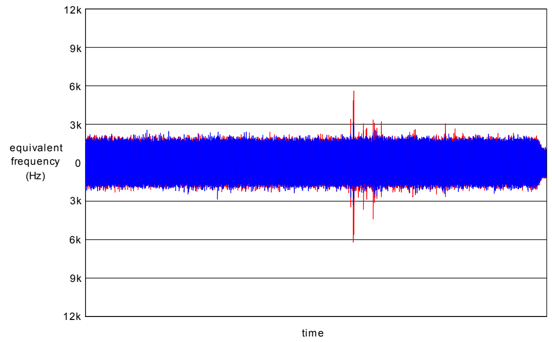
Figure 7. Rate of change graphs for five example tracks from Table 1:



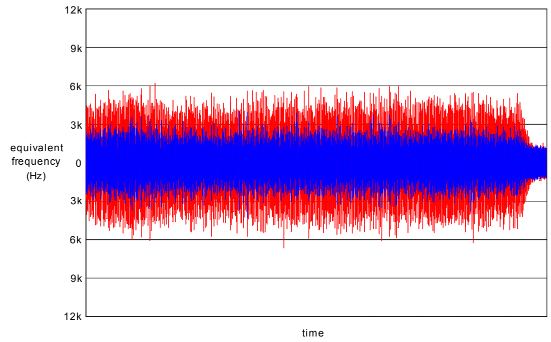
(a)



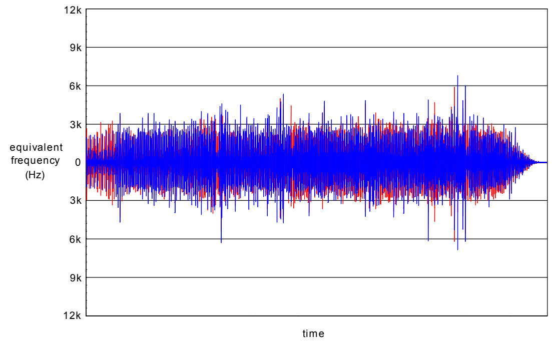
(b)



(c)

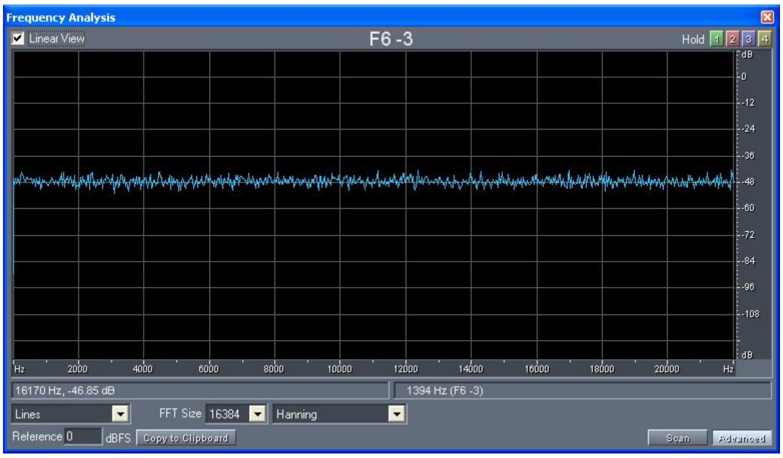


(d)

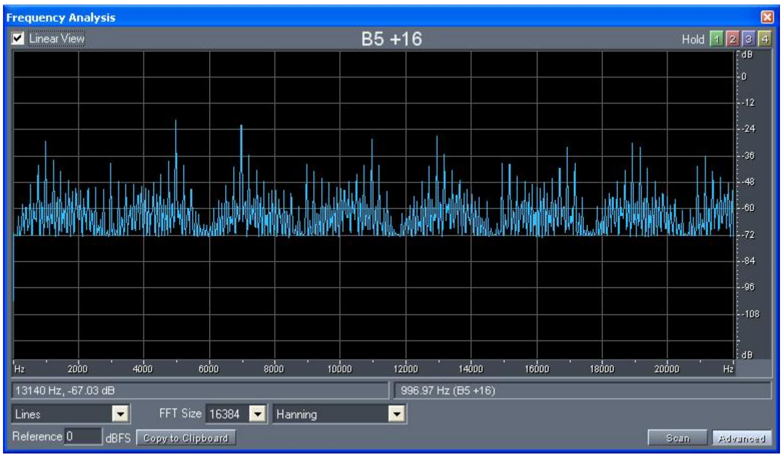


(d)

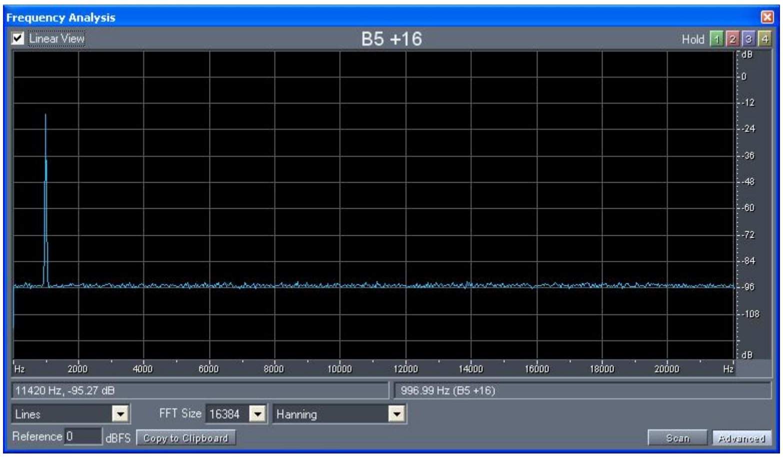
Figure 8. Example spectra of the three types of extracted quantisation error:



997Hz 0dBFS, 24-bit file truncated to 16 bits



997Hz -90dBFS, 24-bit file truncated to 16 bits



997Hz -100dBFS, 24-bit file truncated to 16 bits

Figure 9. Noise floors of tracks 2 (blue trace) and 4 (red trace) from Table 2:

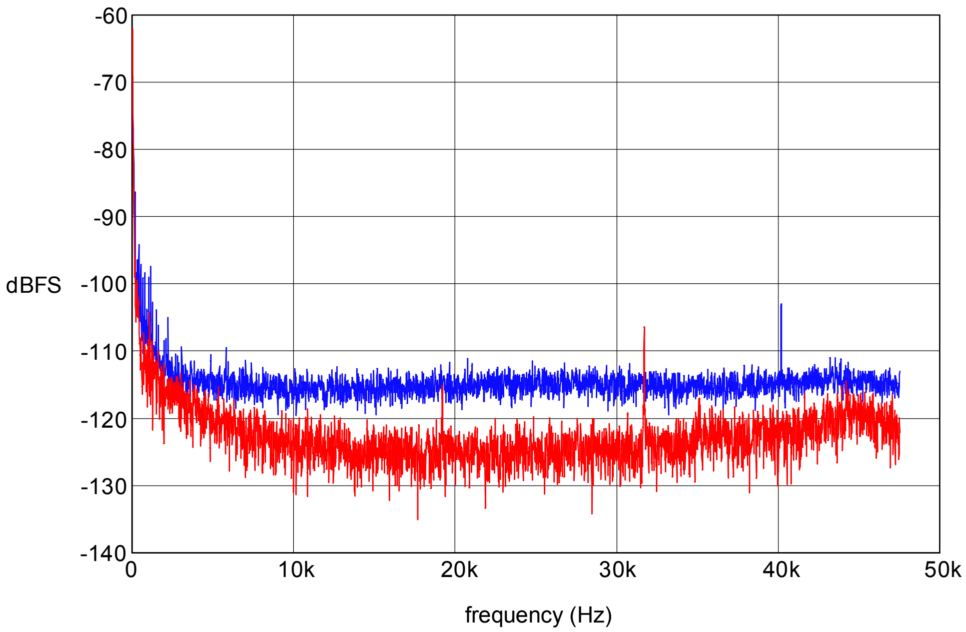


Figure 10. PDF of the noise from track 2, Table 2, high-pass filtered to remove studio sounds:

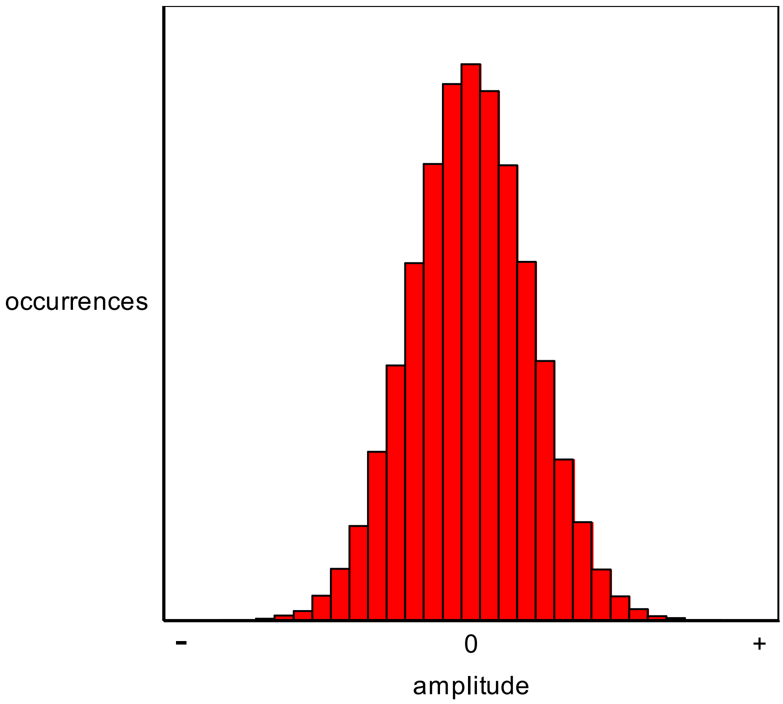


Figure 11. Autocorrelation graphs for the 11 failed frames of the right channel of track 7, Table 2:

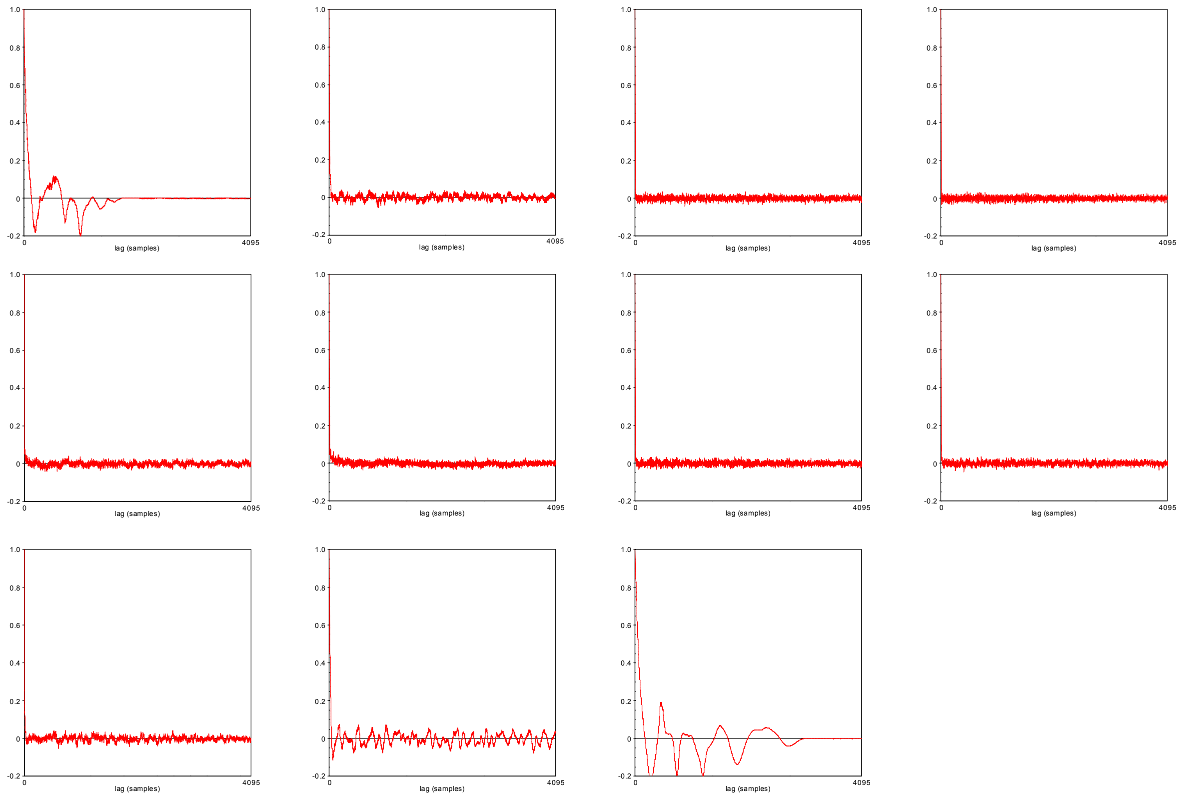


Table 3. Numbers of frames failing the autocorrelation test for each of the tracks of Table 2:

Track ref	Channel	Failed frames
1	left	1/2355
	right	1/2355
2	left	4/4241
	right	3/4241
3	left	1/2956
	right	1/2956
4	left	1182/5339
	right	797/5339
5	left	5/984
	right	6/984
6	left	9/2762
	right	13/2762
7	left	15/3383
	right	11/3383
8	left	0/3637
	right	0/3637
9	left	1/2800
	right	2/2800
10	left	0/315
	right	0/315