High-Quality Audio Application of DVD

Acoustic Renaissance for Audio

 jrs@meridian-audio.com

0. Summary

This document has been prepared as a comment and proposal to all concerned in the standardisation and application of DVD for both Video and Audio.

In reviewing the current specification for the DVD Video disc [32], we conclude that remarkably few additional changes are needed to achieve the objectives originally set out in the *ARA Proposal* [3].

We further suggest that making these changes be considered urgently, since both a High Quality Audio Video Disc and a Pure Audio Disc would provide large benefits.

We review the capability of the current specification and request some minor but important changes.

The first is a proposal for a new data type in audio streams. This data type carries *losslessly compressed* linear PCM (packed) audio[1], and is quite distinct from the currently recognised streams carrying lossy-compressed (data-reduced) channels such as *MPEG Audio*, *DTS* or *AC-3*.

The second provides two ways in which hierarchical or Ambisonic multichannel audio streams can be correctly handled or identified, namely:

- That when a multichannel speaker feed is conveyed, a flag (or data type) can indicate that the original source was an Ambisonic B-Format.
- That when a multichannel speaker feed is conveyed, a flag (or data type) indicates that the original source was encoded according to the Gerzon hierarchy, see [15] and [5].

The third proposal calls for a form of pre-emphasis for use at 96kHz sampling rate.

These proposals do not add cost to any currently proposed video players, and are structured in such a way that only the more advanced applications will require extra hardware.

A glossary is included in section 13.

[1] Lossless compression and decompression is a process that returns the original input data exactly. Its advantage is that in the compressed form, the data occupies less space on the disc, reducing the maximum data rate. The degree of compression is very high at high sampling frequencies. Lossless compression is quite unlike lossy psychoacoustic-based compression, which does **not** return the original data intact. For more details, see the glossary in section 13.

For ease in distinguishing between lossless and lossy compression in this document, we have adopted the term ‘packed’ to describe losslessly compressed linear PCM audio.

Part 1: Background

1. Introduction

In April 1995, *Acoustic Renaissance for Audio* published its original document 'A Proposal for the High-Quality Audio Application of High-Density CD Carriers'. (*The ARA Proposal*). [3]

That document set out a number of points that were felt to be important when specifying a pure-audio application for high-density optical discs and set out a method by which this could be achieved within the parameters known at that time. The proposed disc was referred to as 'HQAD' – *High Quality Audio Disc*.

The *ARA Proposal* was placed before the DVD syndicates, the *Advanced Digital Audio Conference* of the *Japan Audio Society,* and has been widely circulated, reviewed, re-printed and discussed.

Since the *ARA Proposal* was circulated, there have been a number of other proposals relating to DVD application – including some very important comment from the Recording Industry[2], and a summary from the *Advanced Digital Audio Conference* of the *Japan Audio Society,* of which the author is a member. [1]

For the record, the *ARA* is fully prepared to stand behind its original document and believe that the *ARA Proposal* should remain unchanged as a statement of objectives and as a method of conveying audio in the *video* stream of DVD discs.

Whilst there has been an apparent lack of activity relating to the development of a pure-audio standard, there has been enormous progress made by the DVD Consortium of Ten Companies[3] – in evolving the DVD specification, culminating in the recent version 1.0. [32]

The DVD specification in that version, shows substantially improved audio specifications over earlier releases and the DVD Consortium should be congratulated on not only avoiding a specification that invites abuse of the audio, but actually providing very high-quality options.

In our opinion, the audio specifications of the DVD Video Disc have been improved to the extent that:

1. a High Quality Audio Disc could be achieved with very few modifications, *and*
2. the resulting performance could effectively remove the need for a separate DVD Audio Specification so far as the high-density channel is concerned.
3. A High-Quality Audio-with-Video disc would be a rich addition to the industry and would impact less severely on current carriers like CD-DA than the originally-proposed HQAD.

This document reviews the current options, highlights deficiencies and makes some specific proposals.

---

[2] Represented by *IFPI, RIAA* and *RIAJ*

[3] The DVD standard is published by, and copyright of: *Hitachi Ltd., Matsushita Electric Industrial Co. Ltd., Mitsubishi Electric Corporation, Philips Electronics N.V., Pioneer Electronic Corporation, Sony Corporation, Thompson Multimedia, Time Warner Inc., Toshiba Corporation, Victor Company of Japan Inc.*

2. Purpose of this document

The purpose of this document is to encourage minor additions to the current DVD specification and to illustrate the options this disc brings to the Music industry.

It's second purpose is to illustrate how a Pure Audio disc could be achieved within the current DVD framework.

We review the differences between the capability of DVD and the proposed HQAD. This document lays out some minor changes which, if taken up, would permit the very high-quality audio application to exist *within the Video DVD application*.

This document does not review the factors leading to the audio specification, that having been covered in the *ARA Proposal*. However, the following remain vital points in our thinking:

- There is a real possibility that DVD will become the de facto standard over a decade, effectively replacing CD-DA.
- The audio community will not be content with a carrier that uses psychoacoustically-based data-reduction schemes such as *AC-3* or *MPEG* alone.
- There is a rapidly growing interest in multichannel music systems. A carrier like DVD, makes such systems possible.
- Recently audio engineers have begun to explore methods of enhancing CD resolution because it is increasingly understood that the existing CD-DA 2-channel 16-bit/44.1kHz channel is inadequate.
- There are many programs, for which music can be enhanced by stationary or moving pictures.

3. Extent of discussion and authorship

This draft document is a proposal from the Chairman of *Acoustic Renaissance for Audio*, a free body dedicated to advancing audio quality.

Part 2: Where we are now?

4. Relevant summary of DVD Version 1.0

The table below shows a comparison of the main points of the original *ARA Proposal* compared with the current DVD standard.

Parameter	ARA Proposal	DVD 1.0	Comment
Mandatory	All video players play 2-channel version.	2-channel PCM and/or either 2-channel MPEG or 5.1 AC-3.	Adequate
Channels	Up to 8	Up to 8 per stream with restrictions.	
Channel Coding	• Linear PCM • Lossless packing	• Linear PCM • Dolby AC-3 (525/60) • MPEG-1/2 (625	Lossless packing is essential and must be allowed for.
Sampling Frequency	• 48kHz *or* • 96kHz	• 48kHz (PCM, AC-3, MPEG) *or* • 96kHz PCM	Adequate
Precision (PCM)	Adjustable in the range 16 – 24 bits	16, 20 or 24 bit	Adequate
Pre and de-emphasis	• Permitted • Not with lossless packing	• Permitted as 50+15μs	Adequate
Two-channel compatibility	• By provision of Lt and Rt at high precision. • By provision of Red Book layer.	• 1 or 2 channels mandated. • All channels equivalent precision. • Mixdown specified for karaoke applications.	Adequate
Budget player compatibility.	By provision of a lossy-compressed AC-3 or MPEG mix.	By provision of *either* PCM or AC-3 (via mixdown) or MPEG to 2-channel output.	Adequate
Digital outputs	• High-speed video data stream. • Audio multichannel serial interface. • Modified SPDIF for 96kHz operation.	Not specified	Required

The table below shows the subsidiary points of the original *ARA Proposal* compared with DVD.

Parameter	ARA Proposal	DVD 1.0	Comment
Additional data			
Channel use	Header or subcode to identify channel use, including bass effects, height, other.	Extensive description in stream data, height and other use not described.	Needs extending.
Copyright information	As data packets	As data packets	Adequate
Titles, lyrics, musical score	As data packets	As sub set	Adequate
Dynamic-range control	As data packets	Only described within MPEG audio and AC-3	Desirable in all formats
Absolute sound level datum	Header or data packets	Not referenced, but implicit in AC-3.	Desirable in all formats

Part 3: Audio use of DVD

5. Audio Streams

The ARA Proposal assumed that in the pure audio application, the storage method which carries the variable-rate video bitstream in the movie disc, could be devoted to audio information. We suggested a method of lossless coding that could exploit this variable data-rate delivery.

Whilst one could envisage a version of the DVD that allowed this, it turns out that DVD has now been given the potential to carry a considerable bandwidth of audio data, but DVD has been structured for audio streams to have constant data rates – a reasonable solution for sound with pictures, and with the potential for simpler implementation of audio-only replay.

Overall, the new specifications are such that a High Quality Audio Disc can be achieved within a Video format – allowing a whole range of material to be issued with varying quality or quantity in the picture and sound sections. By not excluding moving pictures, many highly-valid music-with-picture applications are enabled.

In particular, the DVD allows the producer to achieve many different balances between the number and quality of audio streams attached to a picture that ranges from 'none' to excellent MPEG2 in a sliding scale that includes the ability to deliver 'stills' and MPEG1 pictures.

The current DVD specification places some important constraints on the audio data streams that were not envisaged in the earlier proposal, and this document attempts to show how many of the original objectives of the HQAD can be achieved – with picture capability intact.

5.1 Data Rates

The following Table 1 is reproduced (in part) from the DVD Specification[4].

Table 1	**Transfer Rates Mbps**		
Object	Total streams	One Stream	Note
Video Stream	9.80	9.80	Number of streams = 1
Audio Streams	9.80	6.144	Number of streams = 8 (max)

This table highlights some important differences from the assumptions we made in our original proposals, namely:

- The audio streams are constant rate
- There can be more than one audio stream
- The data-rate of any one stream is limited to 6.144Mbps
- The specification allows for sophisticated mix-down between channels (karaoke)

5.2 Audio Streams data rate

The limit of 6.144Mbps for one audio stream is the most significant drawback to realising the HQAD objectives. Our interim conclusions are that the majority of the original objectives we can be achieved, *however*:

- Two audio streams are necessary
- Lossless coding (as proposed) is now *important* to achieve the best result.

[4] This information is © DVD Consortium, but at the time of writing is already in the Public Domain.

6. The DVD bit budget

6.1 Pure Audio

Table 2 below, shows some bit-budget examples to illustrate high-quality audio-only use of DVD. The table assumes the following:

- DVD Version 1.0 specification, no changes assumed, *and*
- audio in 2 PCM streams, *and*
- Audio stream 0 always carries a 2-channel mix for basic player and 2-speaker compatibility, *and*
- Audio stream 1 carries a 4, 5 or 6 speaker surround mix.

Table 2: Examples of DVD Audio Combinations
PCM on all audio input channels no Video

Example #	Stream #	Channels fs kHz	Ch. #	Precision Bits	Data Rate Mbps Audio Channel Input	Audio	Video	Total	Play mins
1	0	48	2	16	1.54	6.14	0.00	6.14	102
	1	48	6	16	4.61				
4	0	48	2	20	1.92	7.68	0.00	7.68	82
	1	48	6	20	5.76				
15	0	48	2	24	2.30	8.06	0.00	8.06	78
	1	48	5	24	5.76				
16	0	48	2	20	1.92	5.76	0.00	5.76	109
	1	48	4	20	3.84				
17	0	48	2	24	2.30	6.91	0.00	6.91	91
	1	48	4	24	4.61				
18	0	48	2	16	1.54	1.54	0.00	1.54	408
19	0	48	2	20	1.92	1.92	0.00	1.92	326
20	0	48	2	24	2.30	2.30	0.00	2.30	272
30	0	96	2	16	3.07	9.22	0.00	9.22	68
	1	96	4	16	6.14				
37	0	96	2	16	3.07	3.07	0.00	3.07	204
38	0	96	2	20	3.84	3.84	0.00	3.84	163
39	0	96	2	24	4.61	4.61	0.00	4.61	136
40	0	96	3	20	5.76	5.76	0.00	5.76	109

Attention is drawn particularly to example 15 and to example 30, where 2 and 4-channel mixes are shown at 96kHz @ 16 bit. It should be remembered that this linear PCM application that permits pre-emphasis and noise-shaping optimisation. A 96kHz 16-bit channel can have an extremely high audio performance as shown recently by Stuart and Wilson[5] [31] and as illustrated in a companion paper from the ARA. [4]

[5] Stuart and Wilson [31] recently demonstrated that with optimum noise-shaping, a 96kHz 16-bit channel can deliver a low-coloration subjective dynamic range of 22-bits.

The ARA paper [4] illustrates how this can be extended to 23-bits.

A 4-channel surround mix is suboptimal for a 5.1 speaker array, but very much better than 2-speaker feeds. In addition, the four channels could be hierarchically-encoded surround.

Consequently, we suggest that the DVD channel assignment descriptors be extended to include hierarchical formats. This is discussed more fully in an accompanying document. [5]

6.2 Audio with Pictures

Table 3 below, shows similar bit-budget examples in a high-quality use of DVD – i.e. with a video stream. Audio quality is optimised by using either constrained MPEG2 (3.5Mbps) or the MPEG1 Video stream option (at the maximum permitted rate of 1.856Mbps).

Table 3: Examples of DVD Audio Combinations
PCM on all audio input channels with Video

		Channels			Data Rate Mbps				
Example #	Stream #	fs kHz	Ch. #	Precision Bits	Audio Channel Input	Audio	Video	Total	Play mins
1	0	48	2	16	1.54	6.14	MPEG2	9.64	65
	1	48	6	16	4.61				
4	0	48	2	20	1.92	7.68	MPEG1	9.54	66
	1	48	6	20	5.76				
16	0	48	2	20	1.92	5.76	MPEG2	9.26	68
	1	48	4	20	3.84				
17	0	48	2	24	2.30	6.91	MPEG1	8.77	71
	1	48	4	24	4.61				
18	0	48	2	16	1.54	1.54	MPEG2	5.04	124
19	0	48	2	20	1.92	1.92	MPEG2	5.42	116
20	0	48	2	24	2.30	2.30	MPEG2	5.80	108
37	0	96	2	16	3.07	3.07	MPEG2	6.57	95
38	0	96	2	24	4.61	4.61	MPEG2	8.11	77
39	0	96	3	16	4.61	4.61	MPEG2	8.11	77
40	0	96	3	20	5.76	5.76	MPEG2	9.26	68

This table shows that an MPEG1 moving picture can accompany two very high-quality audio streams at 48kHz and in each case, playing time is adequate.

Particular attention is drawn to example 4, in which two 20-bit 48kHz streams are used to provide optimum 2-speaker stereo and 6-speaker surround.

The situation is not so good for 96kHz recordings – essentially full surround mixes are excluded by the data-rate restrictions – *however*, example 39 is very important for the following reasons.

- Linear PCM using 16-bits at 96kHz *and using* the proposed Pre-Emphasis and noise-shaping scheme [4] can convey a surround mix with exceptional transparency – essentially equivalent to a normal 23-bit channel at 48kHz but giving wider bandwidth, *and*
- three channels are sufficient, with hierarchical coding (see [5]) to convey a full surround experience, *and*
- it offers a high quality picture – up to 5Mbps for the picture.

7. Channel coding

7.1 Signal processing

Linear and psychoacoustically correct coding methods are known which can improve the performance of linear-PCM channels. The two principal methods are noise-shaping with or without pre- /de-emphasis and are permitted under the DVD specification. [14] [6] [2] [28]

This technology will prove to be very important when the sample rate is 96kHz, because smaller word size is dictated by the lower data rate.

The *ARA* have proposed a form of pre-emphasis to be used when the sampling rate is 96kHz. [4] This method permits reduced data rate by adding 2-bits to the audio performance of the 96kHz channel. The document also illustrates how adding noise-shaping to that pre-emphasis can give an effective resolution of 23-bits with a 16-bit 96kHz channel – in other words, 16-bits at 96kHz is all that we require.

7.2 Lossless coding

In the *ARA Proposal* we strongly recommended that the high-quality audio channels be losslessly coded (packed). Signal processing has advanced to the state where the data-reduction benefits of such coding are too good to pass by. Unlike perceptual or lossy data reduction, lossless coding does not alter the final decoded transmitted signal in any way, but merely 'packs' the audio data more efficiently into a smaller data rate.

7.3 Lossless coding for flexible wordlength and bandwidth

Existing lossless audio data compression systems are optimised for reducing *average data rate*, but not for reduction of *peak data rate* or for optimum results at high sampling rates such as 96kHz. We have determined simple-to-decode methods optimised for these latter requirements and have been studying the special requirements of DVD audio.

The process of packing PCM becomes more efficient as sampling rate is increased. For example, packed 96kHz audio does not double the data rate of packed 48kHz as you would expect; the increase is more like 30%.

It is possible to design the lossless-coding specification in such a way that at the mastering stage the record producer can make a personal trade-off between playing time, frequency range, number of active channels and precision. The packed channel can convey this choice implicitly in its control data, and the system operation will be transparent to the user.

This scenario has the following benefits:

- A producer mastering at 48kHz can control the incoming precision of each channel – and trade playing time or channels for noise-floor.
- A producer mastering at 96kHz can also trade bandwidth for playing time, active channels and precision.

By way of examples:

a) playing time or precision may be extended by pre-filtering information above 30kHz

b) playing time or precision may be extended by only supplying a 2, 3 or 4-channel mix.

The technical standard for lossless coding can specify the maximum input and output wordlength, possibly as 24 bit. In addition, the standard can be arranged so that choices regarding input wordlength, number of active channels and bandwidth are automatically handled by the coding, without manual intervention by producer or end-user.

7.4 Lossless coding in the DVD

We strongly advocate that the DVD specification be altered to permit a use of the private stream for losslessly-coded audio.

Our concept is that a High-Quality Audio-Video Disc be considered, in which there are two audio streams as follows:

- Audio stream 0 always carries a 2-channel PCM, MPEG or AC-3 mix for basic player and 2-speaker compatibility, *and*
- Audio stream 1 carries a 4, 5 or 6 speaker surround mix, losslessly encoded.

Lossless packing offers the opportunity to make a much better product. It allows us to convey more precision and more channels. It also gives a lot of open-ended flexibility to the user – as can be seen in some of the examples quoted in Tables 5 and 6 on pages 11 and 12.

7.5 Lossless coding specifically designed for this DVD application

We propose that for this application, a lossless coding scheme is allowed for that has the following properties:

- Output data filled out to a *constant data-rate* to meet the disc constraints.
- Output data-rate assumed to be lower than PCM input at 48kHz
- Output data-rate assumed to be significantly lower than PCM input at 96kHz.
- Input wordlength *continuously adjustable* between 16 and 24 bit
- Bandwidth *continuously adjustable* between 22kHz and 48kHz

We have prepared a study and are now aware of relatively simple-to-decode packing and unpacking techniques that allow the lossless data compression shown in the table below for five associated channels and a player latency-model of 30ms. (Compression is shown as the saving in bits per sample per channel).

Table 4	**Data-rate reduction: bits/sample/channel**	
Sampling kHz	**Peak**	**Average**
48	2	7
96	6	9

Table 9 shows the relative coding efficiency for linear PCM with and without noise-shaping, pre-emphasis and lossless coding.

8. The HQAVD bit budget

Tables 5 and 6 below, show some bit-budget examples to illustrate high-quality audio + video use of DVD Video. Table 5 deals with 48kHz sampling rates, Table 6 shows similar cases where the rate is 96kHz. These assume the following:

- DVD Version 1.0 specification extended to permit losslessly-coded audio streams, *and*
- MPEG2 Video at 3.6Mbps maximum when possible, *or*
- MPEG1 Video stream at maximum rate when possible, *or*
- *MPEG2* still pictures, *and*
- Audio stream 0 (PCM) always carries a 2-channel mix for basic player and 2-speaker compatibility, *and*
- Audio stream 1 (Lossless coding) carries a 4, 5 or 6 speaker surround mix.

Tables 5 and 6 each have five columns describing data rate in Mbps. The first, labelled 'Input', is the worst-case rate of data in the uncompressed recording being fed to the mastering process. The second column, labelled 'Stream, gives the expected maximum data rate in the packed channel – i.e. on the disc. The third column gives the total data-rate for the two audio streams. The remaining columns show if video rates of either 3.5Mbps or 1.845Mbps are possible and therefore the total disc data rate – which is used to compute playing time.

Table 5: Examples of DVD Audio Combinations @ 48kHz
Stream 0 PCM; Stream 1 Lossless Multichannel: with Video

		Channels			Data Rate Mbps					
Example #	Stream #	fs kHz	Ch. #	Precision Bits	Input	Audio Channel Stream	Audio	Video	Total	Play mins
1	0	48	2	16	1.54	1.54	5.57	MPEG2	9.07	69
	1	48	6	16	4.61	4.03				
2	0	48	2	16	1.54	1.54	6.14	MPEG2	9.64	65
	1	48	6	18	5.18	4.61				
3	0	48	2	16	1.54	1.54	6.43	MPEG1	8.29	76
	1	48	6	19	5.47	4.90				
4	0	48	2	20	1.92	1.92	7.10	MPEG1	8.96	70
	1	48	6	20	5.76	5.18				
5	0	48	2	20	1.92	1.92	6.82	MPEG1	8.67	72
	1	48	6	19	5.47	4.90				
6	0	48	2	20	1.92	1.92	7.54	MPEG1	9.39	67
	1	48	6	21.5	6.19	5.62				
7	0	48	2	24	2.30	2.30	7.49	MPEG1	9.34	67
	1	48	6	20	5.76	5.18				
8	0	48	2	16	1.54	1.54	4.96	MPEG2	8.46	74
	1	48	5.1	16	3.92	3.43				
9	0	48	2	16	1.54	1.54	5.45	MPEG2	8.95	70
	1	48	5.1	18	4.41	3.92				
10	0	48	2	16	1.54	1.54	5.70	MPEG2	9.20	68
	1	48	5.1	19	4.65	4.16				
11	0	48	2	20	1.92	1.92	6.33	MPEG1	8.18	77
	1	48	5.1	20	4.90	4.41				
12	0	48	2	20	1.92	1.92	6.08	MPEG2	9.58	65
	1	48	5.1	19	4.65	4.16				
13	0	48	2	20	1.92	1.92	6.69	MPEG1	8.55	73
	1	48	5.1	21.5	5.26	4.77				
14	0	48	2	24	2.30	2.30	7.57	MPEG1	9.42	67
	1	48	5.1	23.5	5.75	5.26				
15	0	48	2	24	2.30	2.30	7.58	MPEG1	9.44	66
	1	48	5	24	5.76	5.28				
16	0	48	2	20	1.92	1.92	5.38	MPEG2	8.88	71
	1	48	4	20	3.84	3.46				
17	0	48	2	24	2.30	2.30	6.53	MPEG1	8.38	75
	1	48	4	24	4.61	4.22				
18	0	48	2	16	1.54	1.54	1.54	MPEG2	5.04	124
19	0	48	2	20	1.92	1.92	1.92	MPEG2	5.42	116
20	0	48	2	24	2.30	2.30	2.30	MPEG2	5.80	108

Table 6: Examples of DVD Audio Combinations @ 96kHz PCM on all audio input channels with Video

Example #	Stream #	Channels fs kHz	Ch. #	Precision Bits	Data Rate Mbps Audio Channel Input	Stream	Audio	Video	Total	Play mins
30	0	96	2	16	3.07	3.07	6.91	MPEG1	8.77	71
	1	96	4	16	6.14	3.84				
31	0	96	2	16	3.07	3.07	7.77	MPEG1	9.63	65
	1	96	5.1	15.6	7.64	4.70				
32	0	96	2	16	3.07	3.07	8.83	Still	8.83	71
	1	96	6	16	9.22	5.76				
33	0	96	2	16	3.07	3.07	8.95	Still	8.95	70
	1	96	5.1	18	8.81	5.88				
34	0	96	2	20	3.84	3.84	9.22	Still	9.22	68
	1	96	4	20	7.68	5.38				
35	0	96	2	20	3.84	3.84	9.23	Still	9.23	68
	1	96	5.1	17	8.32	5.39				
36	0	96	2	20	3.84	3.84	7.58	MPEG1	9.44	66
	1	96	3	19	5.47	3.74				
37	0	96	2	16	3.07	3.07	3.07	MPEG2	6.67	94
38	0	96	2	20	3.84	3.84	3.84	MPEG2	5.70	110
39	0	96	2	24	4.61	4.61	4.61	MPEG2	8.21	76
40	0	96	3	16	4.61	4.61	4.61	MPEG2	6.46	97
		AC-3 with 96kHz								
50	0	48	AC-3	AC-3	0.45	0.45	6.32	MPEG1	8.18	77
	1	96	6	16.2	9.33	5.88				
51	0	48	AC-3	AC-3	0.45	0.45	5.34	MPEG2	8.94	70
	1	96	5.1	16	7.83	4.90				
		96kHz with 30kHz bandwidth								
44	0	48	2	20	1.92	1.92	7.68	MPEG1	9.54	66
	1	96	6	18	10.37	5.76				
45	0	48	2	20	1.92	1.92	7.80	MPEG1	9.65	65
	1	96	5.1	20	9.79	5.88				
		AC-3 with 96kHz @ 30kHz bandwidth								
52	0	48	AC-3	AC-3	0.45	0.45	6.21	MPEG1	8.06	78
	1	96	6	18	10.37	5.76				
53	0	48	AC-3	AC-3	0.45	0.45	4.36	MPEG2	7.96	79
	1	96	5.1	16	7.83	3.92				

Important Note: These figures assume a lossless packing scheme optimised for peak rate reduction, but filled to maintain that constant rate. They also assume signals having relatively moderate compressibility and the compression ratios given in Table 3 on page 10.

Note the significance of examples 51 and 53; MPEG2 pictures are accompanied by full surround using 96kHz audio (probably with the new pre-emphasis).

9. Pure Audio Disc Proposal

This document has focused on ways in which high-quality audio can be conveyed on the current DVD and illustrated how the result could be radically improved for wide-band (96kHz sampling) and multichannel use by permitting lossless coding in the private stream – thus better attaining the objectives of both *ADA* and *ARA*.

In this application many more useful options are enabled at both 48kHz and 96kHz sampling rates.

In the original *ARA Proposal* we pointed out that Red Book compatibility would be an important feature of a pure audio application.

We would like to make a proposal based on the following key points.

1. That the DVD Audio disc be based on the spirit of the current DVD specification.
2. The DVD specification be adjusted to permit losslessly-coded packed audio which is filled out to a constant data rate to make more efficient use of the data rate and capacity constraints.
3. The DVD specification be adjusted to flag hierarchically-encoded multichannel material to allow useful results to be obtained when fewer 3 or 4 channels are available to convey the audio.
4. The DVD specification be adjusted to permit a new form of pre-emphasis at 96kHz. This pre-emphasis gives a 2-bit addition to each channel's effective dynamic range – and it can be combined with noise-shaping techniques to increase the effective performance by 7 bits.
5. That if a pure audio disc is envisaged, it uses constant data-rate in all streams.
6. That the disc is constructed with a Red Book layer as well as a High-Density layer.
7. That the audio disc should have the ability to carry 8 streams, any one of which can be Linear PCM, Losslessly Packed PCM, DSD, MPEG1/2 audio or AC-3. The relative coding efficiencies of PCM with and without noise-shaping, pre-emphasis or lossless packing are shown along with DSD in Table 9.
8. A player still needs to handle only the current mandatory streams – i.e. PCM and AC-3 or MPEG audio. Higher quality audio comes from reading one of the other options.

Tables 7 and 8 extend the examples already given in Tables 1 and 2 (for audio only) and illustrates the bit budget for examples which include DSD in Stream 0.

10. Conclusion

This document proposes that the DVD standard be modified to allow more flexible audio use. In particular, we advocate the introduction of two additional data types in audio streams, namely:

- Losslessly packed linear PCM packed out to a constant data rate
- Delta-coded like *Sony*'s DSD which uses 64x 1 bit coding

We also advocate some minor changes so that:

- Pre-emphasis at 96kHz or hierarchical coding can be used to make the storage more efficient.

We also advocate the addition of a Red Book layer when the disc has a pure audio application.

We do not necessarily advocate the removal of the picture stream. The DVD specification is sufficiently flexible that that can be included at the user's discretion.

Table 7: Examples of DVD Audio Combinations
Stream 0 DSD; Stream 1 Lossless PCM @ 48kHz

Example #	Stream #	Channels fs kHz	Ch. #	Precision Bits	Data Rate Mbps Input	Audio Channel Stream	Audio	Play mins
61	0	3072	2	1	6.14	6.14	9.57	65
	1	48	5.1	16	3.92	3.43		
62	0	3072	2	1	6.14	6.14	9.22	68
	1	48	4	18	3.46	3.07		
63	0	3072	2	1	6.14	6.14	9.60	65
	1	48	4	20	3.84	3.46		
64	0	3072	2	1	6.14	6.14	8.16	77
	1	48	3	16	2.30	2.02		
65	0	3072	2	1	6.14	6.14	8.74	72
	1	48	3	20	2.88	2.59		
66	0	3072	2	1	6.14	6.14	9.31	67
	1	48	3	24	3.46	3.17		
67	0	3072	2	1	6.14	6.14	7.87	80
	1	48	2	20	1.92	1.73		
68	0	48	2	24	2.30	2.11	2.11	297
69	0	3072	2	1	6.14	6.14	6.14	102
70	0	48	6	24	6.91	6.34	6.34	99

Table 8: Examples of DVD Audio Combinations
Stream 0 DSD; Stream 1 Lossless PCM @ 96kHz

Example #	Stream #	Channels fs kHz	Ch. #	Precision Bits	Data Rate Mbps Input	Audio Channel Stream	Audio	Play mins
80	0	3072	2	1	6.14	6.14	9.98	63
	1	96	4	16	6.14	3.84		
81	0	3072	2	1	6.14	6.14	10.84	58
	1	96	5.1	15.6	7.64	4.70		
82	0	3072	2	1	6.14	6.14	11.90	53
	1	96	6	16	9.22	5.76		
83	0	3072	2	1	6.14	6.14	12.02	52
	1	96	5.1	18	8.81	5.88		
84	0	3072	2	1	6.14	6.14	11.52	54
	1	96	4	20	7.68	5.38		
85	0	3072	2	1	6.14	6.14	11.53	54
	1	96	5.1	17	8.32	5.39		
86	0	3072	2	1	6.14	6.14	9.89	63
	1	96	3	19	5.47	3.74		
87	0	96	2	16	3.07	3.07	3.07	204
88	0	96	2	20	3.84	3.84	3.84	163
89	0	3072	2	1	6.14	6.14	6.14	102
90	0	96	3	16	4.61	4.61	4.61	136

Table 9 Showing Coding Efficiencies				
Type bits @ rate	**Noise-shaping**	**Pre-emphasis**	**Lossless coding**	**Rate Mb/s**
21 @ 48k	N	N	N	2.016
18 @ 48k	Y	Y	N	1.728
20 @ 96k	N	N	N	3.840
16 @ 96k	Y	N	N	3.072
14 @ 96k	Y	Y	N	2.688
18 @ 96k	N	Y	Y	1.536
1 @ 64x	Y	N	n/a	6.144
8 @ 8x	Y	N	N	6.144

Part 4: Supporting information

11. Contact Addresses

Acoustic Renaissance for Audio, Meridian Audio, Stonehill, Stukeley Meadows, HUNTINGDON PE18 6ED, England. **Phone**: +44 1480 52144 **Fax**: +44 1480 451587

Email: ara@meridian-audio.com **Web:** http://www.meridian-audio.com/ara

12. Bibliography and references

1 ADA 'Proposal of Desirable Requirements for the Next Generation's Digital Audio', *Advanced Digital Audio Conference* of the *Japan Audio Society*, Tokyo (1996)

2 Akune, M., Heddle, R.M. and Akagiri, K. 'Super Bit Mapping: Psychoacoustically Optimized Digital Recording', *AES 93rd Convention* San Francisco, preprint 3371 (1992)

3 *Acoustic Renaissance for Audio,* 'A Proposal for High-Quality Application of High-Density CD Carriers', private publication, http://www.meridian-audio.com/ara (April 1995)

4 *Acoustic Renaissance for Audio,* 'DVD: Pre-emphasis for use at 96kHz or 88.2kHz', private publication, http://www.meridian-audio.com/ara (October 1996)

5 *Acoustic Renaissance for Audio,* 'DVD: Application of Hierarchically Encoded Surround Sound – including Ambisonics', private publication, http://www.meridian-audio.com/ara (October 1996)

6 Craven, P.G. and Gerzon, M.A. 'Compatible Improvement of 16-Bit Systems Using Subtractive Dither' *AES 93rd Convention* San Francisco, preprint 3356 (1992)

7 Dunn, J. 'High Dynamic Range Audio Applications for Digital Signal Processing', *93rd AES Convention*, San Francisco, preprint 3434 (Oct. 1992)

8 Fielder, L. 'Dynamic Range Issues in the Modern Digital Audio Environment' *Proceedings AES UK Conference* '*Managing the Bit Budget*', 3–19 (May 1994)

9 Hawksford, M.O. 'Digital Frontiers', *HiFi News*, **40**, no.2, 58–59 and 106 (Feb. 1995)

10 Gerzon, M.A. 'Periphony: With-Height Sound Reproduction' *J. Audio Eng. Soc*., **21**, 2–10 (Jan/Feb 1973)

11 Gerzon, M.A. ‘Ambisonics in Multichannel Broadcasting and Video’ *J. Audio Eng. Soc*., **33**, 859–871 (Nov 1985)

12 Gerzon, M.A. ‘Optimal Reproduction Matrices for Multispeaker Stereo’ *J. Audio Eng. Soc*., **40**, 571–589 (July/Aug. 1992)

13 Gerzon, M.A. ‘Problems of error-masking in audio data compression systems’ *AES 90th Convention*, preprint 3013 (Feb. 1991)

14 Gerzon, M.A., Craven, P.G., Stuart, J.R. and Wilson, R.J. ‘Psychoacoustic Noise Shaped Improvements in CD and Other Linear Digital Media’ *94th AES Convention*, Berlin preprint 3501 (March 1993)

15 Gerzon, M.A. ‘Hierarchical System of Surround Sound Transmission for HDTV’, *92nd AES Convention*, Vienna, preprint 3339 (March 1992)

16 Gerzon, M.A. and Barton, G.J. ‘Ambisonic Decoders for HDTV’, *92nd AES Convention*, Vienna, preprint 3345 (March 1992)

17 Gilchrist, N.H.C. ‘DRACULA: dynamic range control for broadcasting and other applications’ **18**, 36–47, Tonmeistertagung, Karlsruhe (article in English) (1994).

18 Hoeg, W., Gilchrist, N., Twietmeyer, H. and Juenger, H. ‘Dynamic Range Control (DRC) and music/speech control (MSC)’ *EBU Technical Review*, 56– 70, (Autumn 1994)

19 Komamura, M. ‘Wideband and wide dynamic-range recording and reproduction of digital audio’ *AES 96th Convention,* Amsterdam, preprint 3844 (1994)

20 Meares, D.J. ‘Perceptual Attributes of Multichannel Sound’ *Proceedings of AES 12th International Conference ‘The Perception of Reproduced Sound’*, 171–179 (June 1993)

21 Meares, D.J. ‘High definition sound for high definition television’, *Proceedings of the AES 9th International Conference ‘Television Sound, Today and Tomorrow’*, Detroit, 187– 215 (Feb. 1991)

22 Meares, D.J. and Stoll, G. ‘Sound systems in Digital Television’, Technical Module of DVB, Document 1172, *EBU* (1993)

23 Ohashi, T., Nishina, E., Kawai, N., Fuwamoto, Y. and Imai, H. ‘High Frequency Sound Above the Audible Range Affects Brain Electrical Activity and Sound Perception’, *91st AES Convention*, New York, preprint 3207 (Oct. 1991)

24 Ohashi, T., Nishina, E., Fuwamoto, Y. and Kawai, N. ‘On the Mechanism of 'Hypersonic Effect’, *Proceedings Int'l Computer Music Conference*, Tokyo, 432–434 (1993)

25 Oomen, A.W.J., Groenwegen, R.G., van der Waal, R.G. and Veldhuis, R.N.J. ‘A Variable-Bit-Rate Buried-data Channel for Compact Disc’ *J. Audio Eng. Soc*., **43**, 23–28 (Jan./Feb. 1995)

26 Stuart, J.R. and Wilson, R.J. ‘A search for efficient dither for DSP applications’ *AES 92nd Convention*, Vienna, preprint 3334 (1992)

27 Stuart, J.R. ‘Noise: Methods for Estimating Detectability and Threshold’ *J. Audio Eng. Soc*., **42**, 124–140 (March 1994)

28 Stuart, J.R. and Wilson, R.J. ‘Dynamic Range Enhancement Using Noise-shaped Dither Applied to Signals with and without Pre-emphasis’ *AES 96th Convention*, Amsterdam, preprint 3871 (1994)

29 Stuart, J.R. ‘Auditory modelling related to the bit budget’ *Proceedings of AES UK Conference ‘Managing the Bit Budget’*, 167–178 (1994)

30 Stuart, J.R. ‘Perceptual issues in Multichannel environments’ *97th AES Convention*, San Francisco (1994)

31 Stuart, J.R. and Wilson, R.J. ‘Dynamic Range Enhancement using Noise-Shaped Dither at 44.1, 48 and 96 kHz’ *AES 100th Convention*, Copenhagen, (1996)

32 Toshiba et al., ‘DVD Specifications for Read-Only Disc’, Version 1.0 (August 1996)

33 Vanderkooy, J. and Lipshitz, S.P. ‘Digital Dither: Signal Processing with Resolution Far Below the Least Significant Bit’ *AES 7th International Conference – Audio in Digital Times*, Toronto, 87–96 (1989)

13. Glossary

AC-3 A system for perceptually encoding at a reduced data rate both two-channel stereo and 5.1-channel surround sound. *AC-3* has been developed by *Dolby Laboratories*; it is used in many motion-picture films and *LaserDisc* releases, and has been selected for television broadcast in the USA.

Ambisonics A method of recording and playing back directional sound over all horizontal directions, or the full sphere of directions including height, based on transmitting directional components of the sound field rather than loudspeaker feeds, and of reproducing the sound field by deriving signals psychoacoustically optimised for the user’s specific loudspeaker layout.

CD-DA Red Book CD for Digital Audio.

DVD Digital Video Disc – a high-density disc carrying *MPEG-2* encoded variable-rate video with lossy-compressed audio. More recently redefined as Digital Versatile Disc. A collective term for the new generation high-density CDs.

DSP Digital signal processing.

HQAD High-Quality Audio Disc. New format high-density CD applied to audio, as proposed in this document.

Lossless compression A process by which the data of a PCM audio signal can be more efficiently packed into a channel. Although lossless compression of audio does not work in the same way, users of computers will be aware of algorithms such as ZIP and LZW that allow more efficient use of disc storage. Lossless compression of audio has the same effect: less space is used on the disc, which has the important effect of reducing the data rate. Unlike lossy compression, lossless compression systems return the *input data exactly* from a decoder. For clarity, in this document losslessly compressed PCM is referred to as ‘packed audio’.

Lossy compression A process by which an audio signal is examined from a human-psychoacoustic viewpoint. An algorithm attempts to estimate and remove the inaudible components of the signal. The remaining ‘audible’ component is efficiently coded in the output channel. Lossy compression schemes include *MPEG* audio, *PASC, AC-3* and *MUSICAM*. The data recovered from a matched decoder is *not* identical to the original input, although it may sound very similar.

MPEG ‘Motion Picture Experts Group’ refers to standards for perceptual coding at a reduced data rate of video and sound signals. *MPEG-1* and *MPEG-2* are respectively video-coding standards for medium and high-quality use, and *MPEG-1* layers 1, 2 and 3 are systems for perceptually encoding two-channel stereo sound.

Packed audio The data resulting when a linear PCM audio stream is losslessly compressed.

Packing The process of losslessly compressing linear PCM audio.

PCM Pulse code modulation. A method of coding whereby a signal is represented by a discrete-sampled series.

Unpacking The process of decoding losslessly compressed (packed) audio back into the original linear PCM full-rate data.

14. Acknowledgements

Dolby, Dolby Surround, Pro Logic and *AC-3* are trademarks of Dolby Licensing Inc.
MUSICAM, PASC, MPEG and *DVD* are registered trademarks.
Ambisonics is a registered trademark of Nimbus Records Ltd.
DVD is a registered trademark *Toshiba* and the ten companies.