Dolby Digital Primer

Reprint of 'Dolby's Digit Squasher'

Audio Media, Issue 93, August 1998

DOLBY'S

As Dolby Digital (AC-3) gets closer to becoming standardised for the new release formats, DOUG MITCHELL reminds us of the basics.

The 'AC' of Dolby Digital (AC-3) stands for Audio Coding. As indicated by the number 3, this is the third version of the audio codec developed by Dolby Laboratories. Dolby's AC-3 coding concept is based, in part, upon previous work on the single-channel AC-2 coding algorithm which featured critical bandwidth sampling and noise shaping.

An analogy used by Dolby to describe the functionality of the AC-3 code is that of a car pool: "Suppose you needed to get 4000 people (essential information) from one place to another within one hour. The highway (available spectrum) can only carry 1000 cars in one hour. By getting the 4000 people (the essential information) to ride in only 1000 cars, the unneeded additional information (the 3000 cars left at home) is eliminated. That's high efficiency transportation, and that's what Dolby Digital is about."

Another way to think about this code might be to consider this article. Let's say that the article contains 3000 words and I want to transmit all of that information to you, the reader. You have one minute to read the article and retain all of its necessary information. However, you can only read at a rate of 150 words a minute. So, I need to encode the article in such a way that it is reduced to 150 words. To do this effectively, I might have previously informed you of a special form of shorthand. You now know this shorthand model and are able to interpret and make sense of the shortened sentences and abbreviated words. Additionally, the shorthand model is interactive so that, if it is necessary to change the method of shorthand to allow for some strange words, you are able to react to those changes as well. Following your reading of the article in one minute, you are also able to relate to your co-workers all of the information that required the original 3000 words. Best of all, you faithfully relate the information without any embellishment or errors. This analogy introduces some of the key concepts to the Dolby Digital (AC-3) code: model prediction, noise shaping, and hybrid forward/backward adaptive bit allocation.

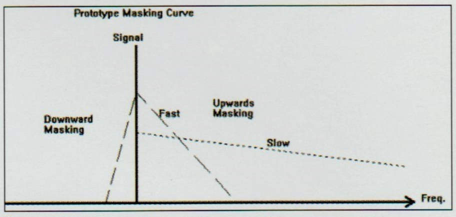

Figure 1. Upward masking envelope.

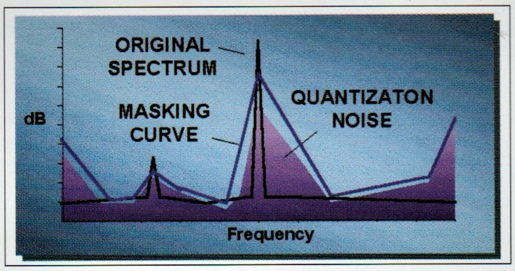

Figure 2. Noise shaping envelope.

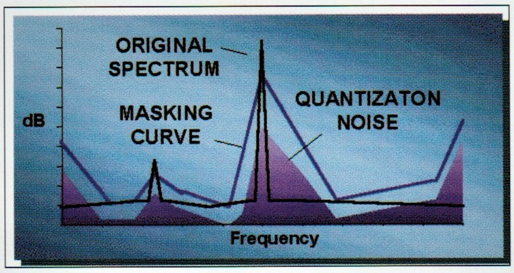

Figure 3. Noise shaping envelope at higher data rates.

I hope to explore some of the intricacies of the Dolby Digital (AC-3) code and discuss production concerns. Knowledge of the coding algorithms and the host of possibilities provided by the growing number of consumer AC-3 decoders will help audio engineers in their mixing decisions.

The Dolby Digital (AC-3) Encoder

The simplest and most basic way to describe the AC-3 encoder is to indicate that it removes information from a digital audio data flow that is determined to be unnecessary based upon models of human perception. Thus, as we'll see, even though the Dolby Digital (AC-3) codec takes advantage of several different data compression schemes, it is most appropriately described as a 'perceptual coder'. Since the removal of this information can create noise artefacts which would be perceptible, the code places the noise under selective higher dynamic areas so that it may be masked.

In order to effectively remove this data, the encoder divides the incoming PCM digitised audio spectrum into narrow frequency bands for analysis. This filtering action of the code produces critical bandwidths, based upon psychoacoustic research, relating the width of these bands to areas of sensory perception along the basilar membrane of the ear. Some 50 bands are included in the AC-3 model.

DIGIT SQUASHER

DOLBY AC-3 PRIMER

Each of the bands created by the coder is separately analysed for spectral activity so that bits may be assigned for later reconstruction of the waveform. The removal of bits from the data stream is based upon models of human perceptibility originally developed in the 1960s and 1970s for research in speech cognition. Continued research by

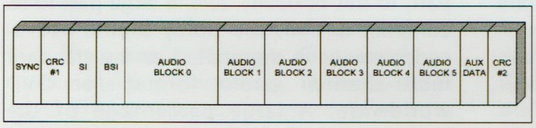

Figure 4. Dolby Digital (AC-3) serial synchronisation frame.

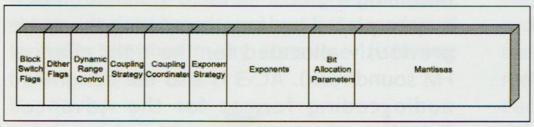

Figure 5. Dolby Digital (AC-3) audio block structure.

Dolby has improved the models upon which this bit reduction may take place. Additional emphasis has been placed upon the concepts of upward masking envelopes — the ability of a transient sound to mask adjacent spectral components of higher frequency (see Figure 1). However, the removal of these bits from the data stream also produces quantisation noise.

The AC-3 encoder works to reduce or eliminate this noise by forcing this quantisation noise under the masking envelope produced by dynamic frequency components within each critical band (see Figure 2). For this reason, the Dolby Digital (AC-3) encoder is taking advantage of data compression concepts originally developed as frequency domain audio coders, but is adding the capabilities of transform and sub-band coding as well.

Increasing the data rate from, say, 384kb/s for standard digital television, to 640kb/s (an option for DVD) increases the ability of the codec to mask noise. Higher data rates push quantisation noise levels further below the masking threshold, as indicated in Figure 3.

Dolby's long history of research in noise reduction, both in the analogue domain and in the digital domain, contributes to the effectiveness of this scheme. In addition to the noise shaping capabilities of the code, Dolby Digital (AC-3) is an effective bit management system. From the outset, the AC-3 code was designed for multi-channel applications — keeping each coded channel's information discrete. Yet, the code establishes an ingenious common bit pool so that channels with greater frequency content which might task coder operation can demand additional data. Alternatively, loud signals in one channel may provide a masking envelope for quieter activity in other channels.

The Dolby Digital (AC-3) code is formatted into a serial synchronisation frame which contains a synchronisation information (SI) header, a bitstream information (BSI) block, and six coded audio blocks (AB). Each of the coded audio blocks represent 256 audio samples. An auxiliary data field (Aux) may follow the audio blocks, and a cyclic redundancy check (CRC) field is included at the end of the frame. Assuming a sampling rate of 48kHz, each of the audio blocks represents 5.3ms. The conceptual block diagram for the encoded Dolby Digital (AC-3) synchronisation frame is displayed in Figure 4. The breakdown of audio block information is displayed in Figure 5.

The Dolby Digital (AC-3) Decoder

Not only is the code extremely efficient at removing data, it also makes use of an ingenious active decoder model so that less data needs to be continuously transmitted from the encoder. The technical term for this is 'parametric bit allocation', which takes advantage of forward and backward bit allocation techniques. What this means is that the AC-3 encoder produces its data based upon

Dolby's long history of research in noise reduction, both in the analogue domain and in the digital domain, contributes to the effectiveness of this scheme. In addition to the noise shaping capabilities of the code, Dolby Digital (AC-3) is an effective bit management system.

a known psychoacoustic masking model as described above. The parameters of this psychoacoustic model are also included in the AC-3 decoder. Therefore, the encoder does not waste bits defining the model. In this way, more of its bit allocation may be utilised to describe the audio itself. However, if alterations in the model are required due to tasking spectral material presented to the system, the model may be altered at the encoder. In this case, the encoder sends delta bits to the decoder so that alterations in the internal decoder model may be achieved. A conceptual model of a hybrid

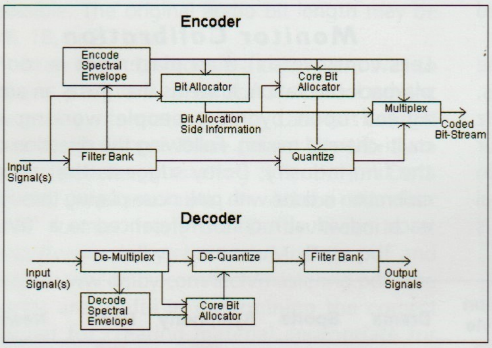

Figure 6. Hybrid forward/backward bit allocation model.

forward/backward bit allocation system is shown in Figure 6.

Other factors regarding the efficiency of the AC-3 code at its initial development were compatibility with existing playback systems already in the consumer market and the ability to control, if necessary, the dynamic range of playback. Both of these issues are addressed at

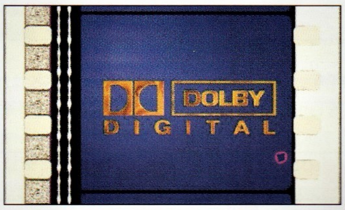

Figure 7. 35mm film showing the Dolby Digital (AC-3) code between the sprocket holes.

the original encoding stage during production of the AC-3 data stream. Additional language in the code, known as 'metadata', may be generated and is included in the AC-3 Bit Stream Information (BSI) data block. This metadata may be utilised to direct the downmixing compatibility

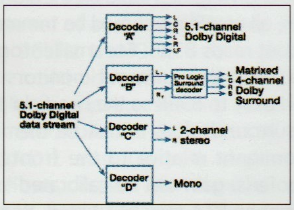

Figure 8. Single AC-3 bitstream appropriate for different listening environments.

of the system. Information for the control of the dynamic range is included within the six blocks of coded audio data following the bit stream information block.

Early in the development of standards for high definition television, a matrixed form of audio delivery (similar to Dolby Pro Logic) was envisioned. However, following the research by Dolby, culminating with the 1992 release of films including AC-3 coded as pixel information between the sprocket holes (Figure 7), it was determined that a discrete multi-channel format should be utilised instead. As the acceptance and use of Dolby Digital films increased, and as talks continued regarding new high definition television standards and DVD made its way to the consumer market, it became apparent that a compatible audio bit stream would be of extreme importance. The features of the Dolby Digital (AC-3) decoder ensure this compatibility.

Downmixing

The AC-3 decoder implemented in consumer units is designed to be able to receive the AC-3 data code and flexibly downmix the discrete channel data to appropriate speakers/configurations. As mentioned above, under the operation of the AC-3 encoder, the Dolby AC-3 bitstream is composed of all relevant data for each channel of information to be sent. It is the job of each consumer decoder to translate the information as appropriate to the given listening environment (Figure 8).

Even though the original AC-3 code may include the information for a 5.1 channel presentation, the decoder can effectively downmix the bitstream to stereo, mono, and even Dolby Surround L/R two-channel matrix for decoding to LCRS (Left, Centre, Right, Surround).

The consumer AC-3 decoder (most familiar now in DVD applications) allows the consumer to determine their respective speaker set-ups. Devices designed to receive the digital AC-3 source via either line (SPDIF) or RF methods are equipped to allow the consumer to alter the software built into the decoder. Generally, this may be referred to as 'Speaker Mode'. The consumer is able to define front (stereo L/R) and rear speakers as being Large or Small; rear speakers as Large, Small, or None; centre speaker as Large, Small, or None; and sub-woofer (LFE) speaker as Discrete LFE (Mode 1), Redirected Bass from L/C/R (Mode 2), or None. Once entered, this information will determine the type of downmixing or bass management action performed by the decoder.

If the original program content is created and transmitted in a stereo, mono, or L/R matrix format, information concerning this may be contained in the BSI block of the AC-3 data code. In this manner, proper decoding of the program may take place without any readjustment of the system on the part of the user. This is especially attractive for digital television service providers, since much of the content for initial transmission will not have been produced for 5.1 channel reproduction.

As implied above, in the Speaker Mode settings, the AC-3 decoder can also accommodate bass management. For set-ups which do not include a sub-woofer, low-frequency information may be routed to wide range front speakers of the system which can handle the information (Large).

Mono compatibility represents special concerns — especially with regard to the reception of matrixed Dolby Surround L/R material. Producers of DVDs or broadcast digital television program material which is coded in L/R will have to be extremely careful in monitoring their mixes prior to encoding due to the fact that mono decoding systems will cancel the surround information.

The AC-3 Decoder, Loudness And Dynamic Range

An ingenious use of metadata in the AC-3 code is the capability to control loudness and dynamic range of program material at the decoder. This is accomplished with separate metadata words which define the Dialogue

Not only is the code extremely efficient at removing data, it also makes use of an ingenious active decoder model so that less data needs to be continuously transmitted from the encoder. The technical term for this is 'parametric bit allocation', which takes advantage of forward and backward bit allocation techniques.

Normalisation and the Dynamic Range Control. These were implemented into the code for a variety of reasons — both aesthetic and technical. We can currently observe radical differences in subjective loudness between various broadcast stations or between broadcast program material which allows for natural dynamics and commercials which are more heavily compressed. Dialogue normalisation may be thought of as a 'program balancer'. It functions just like a volume control to equalise the subjective loudness of programs (regardless of what the original dynamic range was). Dynamic Range Control allows the user to reduce the dynamic range of a program, if necessary. For example, watching an action/adventure film while others in the house are asleep. A separate compression metadata word is used to instruct the decoder to compress the audio signal further to avoid clipping signals when the decoder is driving an RF remodulator in a mono downmix mode — ie. over a conventional television tuner.

Dialogue Normalisation

Dialogue normalisation metadata does not cause compression to occur. Rather, it adjusts the volume of playback through the decoder by interacting with the system volume control. The information for program dialogue normalisation, referred to as 'dialnorm', is entered at the encoding stage. The reference used for setting the dialnorm value is spoken dialogue at a normal level. The chart in ➤

Figure 9 shows relative levels of various programs with respect to a normal dialogue level.

Dolby has settled upon the use of 31dB below digital full scale to represent the optimum dialogue normalisation. The determination of this value is based, in part, upon the standardisation within the film industry to mix dialogue at this level. Using -31 as a dialnorm figure produces the results shown in Figure 10 for the various program levels.

For program material other than speech, the dialnorm settings will vary and are obviously somewhat subjective. To make the proper dialnorm for music productions (other than by use of Dolby's metering system provided on the Dolby 561B encoder), Dolby recommends that the producer have a number of people listen to specially prepared reference level material from Dolby and compare it to the music level. Adjustments can be made to raise or lower the dialnorm setting until the two seem the same with respect to volume. Obviously, the correct dialnorm setting is the point where no change in system volume would be required by the end user. Dialogue normalisation values can be selected from 1 to 31.

Dynamic Range Control

Unlike dialnorm metadata, dynamic range (dynrng) information does cause compression to occur at the decoder. Dolby has implemented the dynrng metadata to account for widely varying listening conditions and environments which may be encountered by AC-3 decoders. It is expected that many consumer listening environments may be noisier than the original production setting, thus requiring compression of the overall dynamic range capable in a digital system. Also, compression may be utilised to prevent dynamic material from being presented at an obnoxiously loud level.

The metadata used to control dynrng may be included in each of the audio data blocks. Therefore, in a program encoded for 48kHz audio, control words for compression may occur every 5.3ms. No noticeable 'zipper' effects in gain change will be noticed and an additional smoothing effect between data blocks is employed as well.

It is important to realise that the settings of the dynrng control are based upon the dialnorm setting. One way of thinking about this might be to consider the dynrng function to be a compander circuit with the dialnorm setting functioning as the threshold level over which gain reduction occurs and under which increase in level is initiated. Again, depending upon the functionality of the decoder, the end user may choose not to engage the compression action at all or may choose to only engage program gain increase or reduction. Some decoders may allow users access to each of these functions. Others may only allow the full companding circuit to be engaged or disengaged.

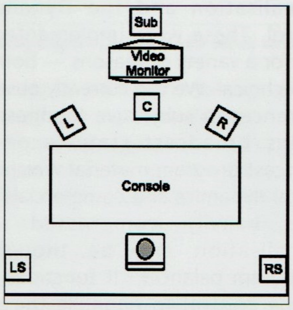

Figure 11. Room layout with two surround speakers standardised by Dolby.

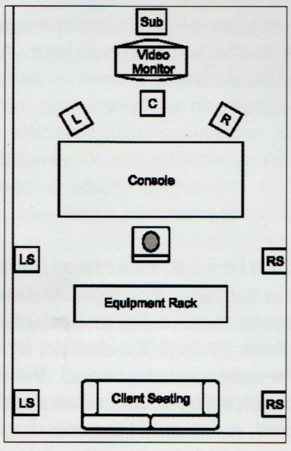

Figure 12. Room layout with four surround speakers standardised by Dolby.

Production

Perhaps the most controversial aspect of mixing for multi-channel production lies in the placement and functionality of control room monitors. On this subject there are a fair number of opinions, but some of the debate may be alleviated by following Dolby's recommendations for monitor set-up and calibration. Dolby has developed some standardised room layouts, displayed in Figures 11 and 12.

Additional suggestions on monitor placement have been issued by the International Telecommunications Union in the guidelines released under ITU-R [BS.775-1] Multi-channel Stereophonic Sound Systems With and Without Accompanying Picture. The ITU specifies that all speakers lie on the circumference of a circle with the front left and right speakers forming a 60-degree angle in front of the engineer and the rear left and right speakers positioned between 100 and 120 degrees behind the engineer's position.

Both Dolby and the ITU suggest that the centre speaker be identical in radius distance from the engineering position as the front left and right. If room construction prevents the positioning of the centre speaker the same distance away from the engineering position as the front left and right, it is suggested that a time delay be utilised for compensation. Additionally, the centre speaker should lie in the same horizontal plane as the front left and right.

Monitor Calibration

Less controversial than placement is room playback calibration. It seems that this is an area agreed upon by most people working in multi-channel mixing. Following the direction of the film industry, Dolby suggests that room calibration is done with pink noise playing through each individual monitor referenced to a '0VU' setting (-20dBfs). Using an SPL meter set to measure a C-weighted scale on the slow response ballistics, all amplifiers should be trimmed so that the output reads 85dB. For smaller rooms Dolby suggests that the surround monitors might be trimmed back to 83dB so the nearfield placement of the surrounds does not cause them to sound too prominent relative to the fronts. The LFE sub-woofer(s) gain can be calibrated in a similar manner or an RTA may be utilised. Regardless of methodology, the sub-woofer setting is altered due to the fact that it is receiving only 1/10 of the full audio spectrum. If using an SPL meter, make sure that the amplifier for the sub-woofer is receiving the pink noise followed by a 120Hz low-pass filter. The reading for the sub-woofer should be 10dB higher than the rest of the system (95dB). Speaker response curves are displayed in Figure 13.

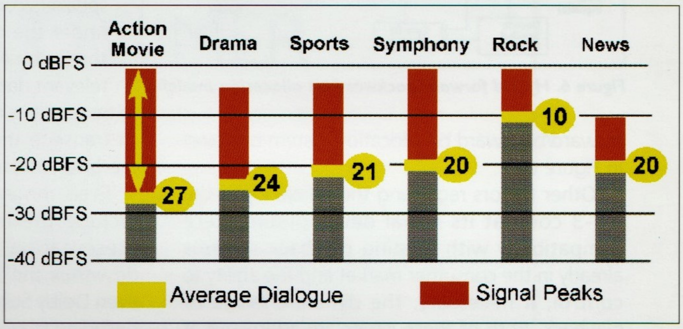

Figure 9. Relationships of various programs with respect to dialogue level.

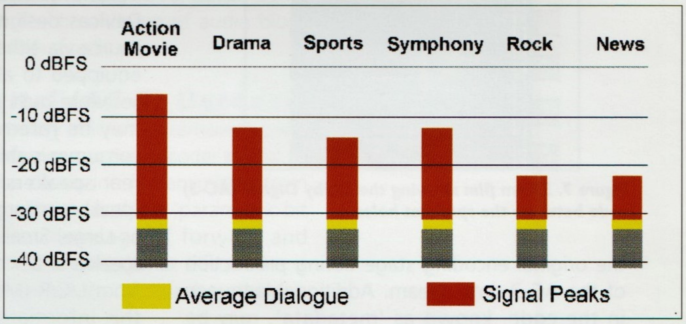

Figure 10. Output levels of various programs using dialnorm -31 as a reference.

Regardless of multitrack recording options being employed, there are a few other considerations to keep in mind when producing audio for ultimate AC-3 coding. Not the least of these is channel assignments. Dolby and ITU have proposed the following Channel-to-Track layout.

The data presented to the AC-3 encoder must be in linear PCM format, but there are options on sample rate and bit depth. Sampling frequency is normally 48kHz, but 44.1kHz and 32kHz are also possible. The original audio bit length may be 16, 18, or 20 bits.

In an effort to provide a convenient way to allow program producers to document necessary information, Dolby Labs has created two handy forms. These forms, the Mix Data Sheet and the Mastering Information Data Sheet, are available in Adobe Acrobat .pdf format from the Dolby website at http://www.dolby.com/tech/MixData.pdf and http://www.dolby.com/tech/mastering.pdf. The forms are useful in determining the correct format for creating material appropriate for coding in AC-3 format.

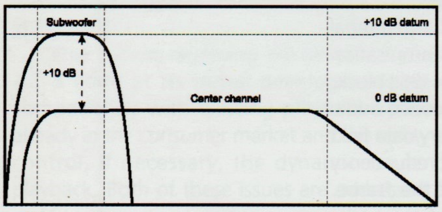

Figure 13. RTA curve for 5.1 multichannel monitoring.

For those who wish to encode their own AC-3 information rather than relying on a DVD mastering facility or digital broadcast head end, there are an increasing number of options. Recently Dolby announced the introduction of the model DP569, a professional real-time 5.1 channel encoder with a list price of $5000 — a significant reduction from its $19,600 predecessor, the DP561B. Several digital audio workstations and software vendors have entered the AC-3 coding field as well. Sonic Solutions markets an entire DVD Audio Workstation which includes SonicStudio High-Density, DVD Producer authoring, DVD PrePlay proofing, Sonic Dolby Digital 5.1 real-time surround encoding, and a complete set of audio and CD prep software including 96kHz NoNOISE and High-Density Suite for multi-channel 24-bit 88.2/96kHz audio. The complete system is priced at $59,999. Studio Audio & Video Ltd, makers of the SADiE and Octavia digital audio workstations, have announced a new partnership with the Daikin Scenarist DVD authoring system. Sonic Foundry, makers of Sound Forge software, have entered the field as well with Soft Encode which carries a suggested list price of $1995 for the 5.1 encoding system (Figure 14).

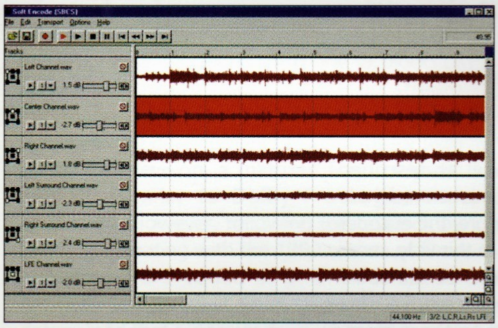

Figure 14. SoftEncode from Sonic Foundry.

For those who wish to encode their own AC-3 information rather than relying on a DVD mastering facility or digital broadcast head end, there are an increasing number of options. Recently Dolby announced the introduction of the model DP569. Several digital audio workstations and software vendors have entered the AC-3 coding field as well.

Conclusion

It should be evident from this discussion that the Dolby Digital (AC-3) codec is much more than the coding process itself. The bitstream produced by the codec also describes the channel delivery capabilities and the method with which these will be reproduced. It is a complete code, yet it has been defined to allow for future expandability.

Because of the backward/forward bit allocation process described earlier, modifications or enhancements to the code may be implemented without rendering the decoders already in the field obsolete. Due, in part, to this flexibility, and in large part to its compact design, the Dolby Digital (AC-3) codec has been designated as the standard multi-channel audio format for DVD worldwide. A large percentage of the DVD players being sold worldwide are equipped with AC-3 decoding, thus AC-3 is becoming a global de facto standard. It also may be placed on laser discs (using the space previously allocated for the right channel FM soundtrack). AC-3 is also the designated audio coding format for the Advanced Television System Committee format for digital television.

Some items to be on the look out for will be automatic, quantifiable methods for determining dialnorm and dynrng values during production, and editing systems which will allow the user to scrub edit AC-3 coded files. According to Dolby, it will be possible to 'rock the reels' — playing the data forward and backward — to allow for fine editing precision.

As the issue of the forthcoming DVD-Audio format continues to be discussed, one certain consideration for this format should be the inclusion of an AC-3 coded bitstream on DVD-Audio discs. The payload represented by the data is minimal, yet the inclusion of this would ensure compatibility with the DVD-Video and DVD-ROM units already on the market. Doing this would also prevent the necessity of separate inventory and resulting consumer confusion. ❑

FOR FURTHER INFORMATION ON DOLBY AC-3:

http://www.dolby.com/digital/encodvw.pdf — Dolby Digital Professional Encoding Manual in .pdf format, Dolby Laboratories, Inc.
http://www.dolby.com/tech/parametr.html — Parametric Bit Allocation in a Perceptual AudioCoder, Grant A. Davidson, Louis D. Fielder, and Brian D. Link, Dolby Laboratories, Inc.
http://www.dolby.com/tech/ac3flex.html — AC-3: Flexible Perceptual Coding for Audio Transmission and Storage, Craig C. Todd, Grant A. Davidson, Mark F. Davis, Louis D. Fielder, Brian D. Link, Steve Vernon, Dolby Laboratories Inc.
http://www.dolby.com/tech/desgnac3.pdf — Design and Implementation of AC-3 Coders in .pdf format, Steve Vernon, Dolby Laboratories, Inc.
http://www.dolby.com/tech/ac-3mult.html — The AC-3 Multichannel Coder, Mark F. Davis, Dolby Laboratories Inc.
http://www.dolby.com/tech/mastering.pdf — Mastering Data Sheet in .pdf format, Dolby Laboratories.
http://www.dolby.com/tech/MixData.pdf — Mix Data sheet in .pdf format, Dolby Laboratories.
ftp://ftp.atsc.org/pub/Standards/A52/ — Advanced Television Systems Committee, "Digital Audio Compression (AC-3) Standard".
http://www.zoran.com/audio.htm — Zoran Microchip Audio Information.
http://www.dvdcreator.com/ — Sonic Solutions DVD Creator Information.
http://www.sadie.com/ —Studio Audio announce further collaboration with Daikin's Scenarist DVD.
http://www.soundforge.com/SoftEncode/default.html — Sonic Foundry SoftEncode Information.

INFORMATION

Ⓐ **Dolby Laboratories Inc., Wootton Bassett, Wiltshire SN4 8QJ, England.**
Ⓣ **+44 1793 842100.**
Ⓕ **+44 1793 842101.**
Ⓦ **www.dolby.com**