

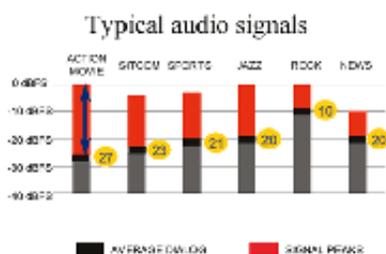
 [Print Page](#)

A Closer Look at Audio Metadata

7/24/2002

Last time we introduced the idea of Audio Metadata, or "data about the audio data," and this time we will explore it in greater detail. Properly set metadata allows a 5.1-channel program to be reproduced correctly by mono, stereo, and 5.1-channel decoders, and it controls the dynamic range of the program so that it fits the listening environment. Metadata must be set correctly for all of this to work the way it was intended.

First things first. It is worth pointing out that there are two types or levels of audio metadata - consumer and professional. Consumer metadata is present in every Dolby Digital (AC-3) bitstream and is used by Dolby Digital decoders to optimize the decoded audio. Professional metadata, on the other hand, is a superset of consumer metadata and is never actually transmitted to consumers. Professional metadata carries up to eight consumer metadata streams, or programs, and other control parameters.



(click thumbnail)

Figure 1. Multiple programs and/or channels reach the Dolby Digital decoder with different dialogue level settings

Professional metadata is used to feed multiple Dolby Digital encoders that extract the appropriate consumer metadata program and pass it on to... you guessed it, the consumer.

Although there are some 27 consumer metadata parameters, we will explore a critical few that I have found to be most important.

THE THREE D'S OF AUDIO METADATA

Dialogue Level (a.k.a. dialnorm)

The dialogue level (also known as dialogue normalization or dialnorm) setting represents the average loudness of dialogue in a presentation. This parameter controls an attenuator in the Dolby Digital decoder that normalizes the average audio output of the decoder to a preset level. This ensures that a consumer can watch a television program without having to adjust the volume control during

commercial breaks or when switching channels.

The proper dialogue level setting is determined by measuring its long-term A-weighted loudness equivalent, or $Leq(A)$, and there are products available to do this in a manner far simpler than the name of the measurement might suggest. The scale used in the dialogue level setting ranges from -31 to -1 dB in 1 dB steps, with -31 indicating no attenuation and -1 indicating 30 dB of attenuation applied in the consumer decoder. What was that?

I know, it sounds counter-intuitive, but, surprisingly, it actually makes sense. Dolby Digital decoders standardize average loudness to -31 dBFS $Leq(A)$, (31 dB below 0 dB full-scale digital, averaged over time). When a decoder receives a relatively quiet input signal, such as a feature film with a dialogue level setting of -31, it is assumed that the program already matches the target level of -31 dB $Leq(A)$ and therefore requires no further attenuation.

On the other hand, a louder program such as live music may require attenuation to bring its $Leq(A)$ to -31 dB. For example, when the dialogue level parameter setting is -21, the decoder will apply 10 dB of attenuation to the signal, when the setting is -11, it applies 20 dB of attenuation, and so on.

A simple way to figure out the attenuation that will be applied is to add 31 to the dialogue level setting. For example, $31 + (-31) = 0$ dB of attenuation, while $31 + (-20) = 11$ dB of attenuation. Due to the rather coarse 1 dB per-step resolution, the dialog level setting is meant to change only at program transitions and is not really a good method for gain-riding program audio.

Properly setting the dialogue level parameter not only ensures that program-to-program and channel-to-channel loudness variations are controlled, it also forms the foundation for the dynamic range control (DRC) system included in the Dolby Digital system. If set incorrectly, the dialogue level parameter can cause DRC to react incorrectly to the audio it is processing.

Dynamic Range Control (DRC)

Different listening environments present a wide variety of dynamic range requirements. Obviously a quiet movie does not fit well into a noisy environment, nor does a loud movie fit into a quiet environment. The classic solution has been to dramatically reduce the dynamic range of the audio prior to transmission, then the audio level can be set once by each viewer to suit his or her environment. The unfortunate side effect is that audio impact is lost. Explosions, dialogue, and background grasshoppers are all reproduced at the same loudness and the program can sound, well, flat - to say the

least.

Happily, there is another solution. Dolby Digital provides a Dynamic Range Control (DRC) system that is rather unique. Based on a pre-selected DRC profile, the Dolby Digital encoders calculate and send DRC metadata along with the original audio signal. The DRC metadata can then be applied to the signal by the decoder to reduce the signal's dynamic range. On many decoders, DRC can optionally be scaled back or even disabled so that the original dynamic range of the audio is delivered.

This unique consumer-side dynamic range processing allows the kitchen DTV set to have restricted dynamic range so that quiet audio can be heard above background noise, while simultaneously the large DTV set in the family room can have unrestricted dynamic range and can stomp on the background noise (and possibly the neighbors). DRC helps to provide the best possible presentation of program content in virtually any listening environment, regardless of the quality of the equipment, number of channels, or ambient noise level.

The Dolby Digital stream carries metadata for the two possible operating modes of the decoder. The operating modes are known as Line Mode and RF Mode due to the type of output with which they are typically associated. Line Mode is relatively light dynamic range compression and is typically used on decoders with six- or two-channel line-level outputs. RF Mode is designed for products such as set-top boxes that have RF re-modulated (i.e. Channel 3 or 4) outputs. RF Mode is heavier dynamic range compression and the peaks are limited to prevent severe overmodulation of television receivers. Full-featured decoders allow the consumer to select whether to use DRC and if so, how much. Thankfully, the consumer sees simple options such as Off, Light Compression, and Heavy Compression instead of None, Line Mode, and RF Mode.

Six preset DRC profiles are available in the Dolby Digital system: Film Light, Film Standard, Music Light, Music Standard, Speech, and None, and each can be chosen separately for Line and RF Modes. The station - or ideally the content producer - chooses which of these profiles to assign to each mode. When the consumer or decoder selects a DRC mode (i.e. apply DRC fully, not at all, or somewhere in-between), the chosen profile is applied to the decoded audio.

In addition, signal peaks can be limited to prevent clipping during downmixing through the use of overload protection metadata. For example, consider a 5.1-channel program with signals near digital full scale on all channels being played through a stereo set-top box. Without some form of attenuation or limiting, the output signal would obviously clip as the 5.1 channels are being downmixed to stereo. Proper setting of dialogue level and DRC parameters can prevent clipping, but just in case, protection DRC can kick in and maintain control, although it is best to avoid this. It is important to note that even if the "None" profile is chosen, protection DRC is still active.

Downmixing

Downmixing allows a multichannel program to be reproduced over fewer speaker channels than for which the program is optimally intended. Simply put, downmixing allows consumers to enjoy a 5.1 channel broadcast regardless of how many speakers they have.

Set-top boxes, used for the reception of terrestrial, cable, or satellite, typically offer an analog mono signal modulated on the Channel 3/4 output, a line-level analog stereo signal and an optical or coaxial digital output. The analog stereo output is a downmixed version of the decoded Dolby Digital bitstream while the digital output delivers the Dolby Digital bitstream to a downstream decoder.

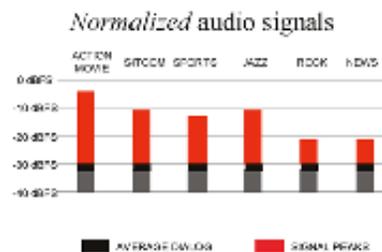
In each of these devices, the analog stereo output is one of two different stereo downmixes. One type is a surround-compatible downmix (left-total/right-total, or Lt/Rt) of the multichannel source program suitable for Dolby Surround Pro Logic or other matrix decoding. The other type is a simple stereo downmix (called a left-only/right-only, or Lo/Ro) suitable for playback on a two-channel stereo system or on headphones, and from which a mono signal is derived for use by an RF re-modulator.

The only difference between the downmixes is how the surround channels are handled. The Lt/Rt downmix sums the surround channels, attenuates them 3 dB (i.e. multiply by 0.707) and adds them out-of-phase to the left channel and in-phase to the right channel. This allows a Pro Logic home theater decoder to produce L/C/R/S channels when connected to a stereo set-top box or DTV receiver. Conversely, the Lo/Ro downmix adds the right and left surround channels discretely to the left and right speaker channels. This preserves the stereo separation for stereo-only monitoring and produces a mono-compatible signal. Lt/Rt is the default selection in all consumer decoders with the exception that the mono signal feeding the RF re-modulator output is derived from a Lo/Ro downmix.

The formula for Lt/Rt compatible and Lo/Ro stereo downmixes are:

$$Lt = L + (0.707 * C) - (0.707 * Ls) - (0.707 * Rs)$$

$$Rt = R + (0.707 * C) + (0.707 * Ls) + (0.707 * Rs)$$



(click thumbnail)

Figure 2. Once the dialogue level setting is applied to the decoded audio, all programs are "lined up" or normalized. Note the signal peaks are not affected by this action.

$$L_o = L + (\text{cle}v * C) + (\text{slev} * L_s)$$

$$R_o = R + (\text{cle}v * C) + (\text{slev} * R_s)$$

With L_o/R_o you can see that there are separate metadata parameters included in the downmix formula. The metadata parameters $\text{cle}v$ (center level) and slev (surround level) default to 0.707 (i.e. -3dB just like in L_t/R_t) but can be adjusted to fine-tune the stereo downmix.

What about the LFE channel? Well, it is simply discarded. Be very careful about audio that is only in the LFE channel for the logical reason that not all consumers will hear it.

All of this stuff looks good on paper, but does it really work? Sometimes yes, sometimes not quite. Next time we will wrap up metadata with a discussion of some real world issues. Thanks to everyone for your continued support and suggestions; it's great to hear from old friends and new ones. I appreciate all of your comments and ideas and will  continue to do my best to address them. [Print Page](#)