Information

All About Audio Metadata

Metadata, the “data about the audio data” that travels along with the multichannel audio bitstream in Dolby Digital, makes life easier for broadcasters while also increasing the creative ability of audio mixers. For broadcasters, audio metadata now means they have a set-and-forget solution, instead of monitoring, compressing, and adjusting levels all over the plant. For audio mixers, this means their vision can finally be delivered all the way into the home without compromise.

In the past, broadcasters working with multichannel audio had problems with soundtracks whose average levels fell above or below that of other programming. Also, issues arose when surround sound content was played back on television sets with stereo or mono audio output. With Dolby Digital, engineers can mix audio content differently for various playback systems and set playback levels, all at the postproduction stage, so broadcasters can deliver a more consistent audio signal and ensure that the most important audio elements come through.

Producers should have a solid grasp of audio metadata in order to take full advantage of its capabilities. Metadata is added during production or postproduction; the Dolby DP570 Multichannel Audio Tool is the latest aid for creating and monitoring the metadata. The unit performs these functions without requiring any external encoders or decoders, making the process easier and faster than ever before.

The DP570 accepts up to eight channels of PCM audio, then feeds the audio accompanied by a separate metadata bitstream to a Dolby DP571 Dolby E Encoder for further distribution, or directly to a DP569 Dolby Digital Encoder for transmission to the consumer. In this way, producers can easily assign metadata parameters that will ultimately reach consumer playback systems via programs with Dolby Digital audio.

The three Ds: dialogue level, dynamic range control, and downmixing

Metadata provides a number of key parameters that are specifically intended to change the sound of the program delivered to the consumer’s unique listening environment. These metadata parameters are known as *dialogue level* (aka *dialogue normalization* or *dialnorm*), *dynamic range control*, and *downmixing.* Although technically not an individual metadata parameter, downmixing within the consumer’s decoder is effected by specific metadata parameters, and as with the other two Ds, care must be taken in monitoring and selecting the metadata parameters that effect downmixing conditions. The engineer is ultimately responsible for optimizing the multichannel mix for best reproduction in the optimal listening environment. However, care should be taken to ensure that less optimal listening environments are accounted for, as well. On the contrary, enjoyment of a television program should not be limited to only those

Dolby Laboratories, Inc. 100 Potrero Avenue, San Francisco, CA 94103-4813 Telephone 415-558-0200 Fax 415-863-1373
Wootton Bassett, Wiltshire SN4 8QJ England Telephone (44) 1793-842100 Fax (44) 1793-842101 www.dolby.com

Dolby and the double-D symbol are trademarks of Dolby Laboratories. © 2001 Dolby Laboratories, Inc.

consumers with full-blown home theater systems. Dolby Digital and metadata combine to provide a simple and easily used method to provide for the best quality reproduction without regard to the number of speaker channels, ambient noise levels, or quality of equipment in a consumer's home.

Dialogue level

The dialogue level parameter (aka *dialogue normalization* or *dialnorm*) within the Dolby Digital stream provides a normalization value to the home decoder. This value lowers the volume of the audio to a preset level, which aids in level matching from program content to program content and media to media. The setting of this dialogue level parameter is crucial to the proper operation of home decoders and provides three main functions:

1. During the audio mastering for DTV, the mixer is endeavoring to match the audio playback level to the level of the theatrical release. Dialogue level provides the mixer a method to achieve this level match simply and with no degradation of the original audio content. In television postproduction, the level matching of dialnorm allows different content (commercials, newsbreaks, sitcoms, and so on) to be interspersed while maintaining the same comfortable listening level in the home.

2. Consumer requirements for home enjoyment of multichannel audio vary widely. For example, not everyone who watches DTV is listening in an ideal home theater environment. Some care must be taken so that each consumer can be reasonably assured that the mixer's intent will be carried through their chosen system to their ears. In addition to level matching between different programs and media, dialogue level gives the ancillary benefit of providing some measure of protection against digital clipping by lowering the overall volume a bit and providing a sort of "virtual headroom."

3. A properly set dialnorm value provides the "null band" within the dynamic range profile where the audio level is neither raised nor lowered. Without dialnorm, reduced dynamic range listening modes may not work correctly.

The term *dialogue normalization* was developed in the film industry. Films are mixed with the dialogue set to a specific level, and all other audio elements are subordinate to the dialogue. Additionally, in the film world, dialogue emanates from the center speaker, giving the film mixer a simple method of measuring this level; they simply measure the output of the center channel and make sure that the dialogue emanating from this channel is not masked by music or sound effects.

That is not to say that dialogue is required to set the dialogue level parameter. In music-only broadcasts, the dialnorm parameter can be best described as the average volume level of the program.

Dialogue level, in its simplest form, is exactly the same as turning the volume down a bit on a consumer's home stereo. However, simply adjusting the volume on a home stereo

provides none of the other dialogue level advantages of providing for reduced dynamic range listening conditions, accurate and musical dynamic range compression, and “virtual headroom” extension prior to the D/A circuitry.

Dialogue level in a properly implemented Dolby Digital decoder can not be defeated. Dialogue level does not assert any compression or expansion on the program material, but simply lowers the volume of the audio to a standardized level.

Dialogue level works in partnership with the dynamic range control profiles. The Dialogue level parameter determines the area within the audio level of the program where the dynamic range profiles are inactive and sets a sort of “null band” between the soft and loud portions of the program. Here, no audio processing occurs.

Dynamic range control

Dynamic range control (sometimes referred to as *dynamic range compression* or *midnight mode*) gives the consumer the flexibility to listen to program audio with a reduced dynamic range. Compression of the dynamic range lets viewers watch television without disturbing the neighbors. This control is optional and can be turned off in most Dolby Digital decoders.

Dynamic range control within the Dolby Digital data stream consists of two parameters or “profiles”: RF Mode and Line Mode. These two parameters do not change the content of the encoded audio within the bitstream. They are used to adjust the extremes of the program material within the listening environment to account for those instances where it is preferable or necessary to listen to the program at a reduced dynamic range.

RF Mode is designed for peak limiting situations where the decoded program is intended for delivery through an RF input on a television, such as through the antenna output of a set-top box. The RF Mode Profile is also used for the common “midnight mode” feature on consumer decoders, which provides dynamic range compression to ensure that an action movie won’t wake up the neighbors.

Line Mode provides a lighter type of compression, and also allows user adjustment of the low-level boost and high-level cut parameters within a home decoder. This adjustment or “scaling” of the boost and cut areas allows the consumer to customize the audio reproduction for their specific listening environment.

At lower volumes, the softer portions of a program (whispers and soft-spoken dialogue) are more difficult to hear. If the viewer increases the volume, however, the louder portions (explosions, onscreen arguments, gunshots, etc.) become too loud for comfortable listening. Alternatively, in an environment with a high level of background noise, quieter portions of the program will be drowned out by the ambient noise.

When dynamic range profiles are asserted within the decoder, the decoder raises the level of the softer portions of the program while lowering the level of the louder portions, allowing the user to enjoy the movie without having to continually reach for the volume control. Once again, this ability to scale the amount of compression only applies to the Line Mode Profile and is dependent on the feature set available in the consumer's home decoder.

Because of the relationship between dialogue level and dynamic range control, it is necessary to select the appropriate dialnorm value prior to previewing dynamic range profiles. As the amount of dynamic range compression will be ultimately selected by the consumer for their own specific listening needs, it is important to preview the source mix through each preset before selecting one to include in the metadata stream.

Downmixing

Downmixing is a feature within Dolby Digital that allows a multichannel program to be reproduced by fewer speaker channels. Simply put, downmixing allows consumers to enjoy a digital television broadcast without requiring a complete 5.1-channel home theater setup.

As with stereo mixing, where the mix is monitored in mono on occasion to maintain compatibility, multichannel audio mixing requires the engineer to reference the mix with fewer speaker channels to assure compatibility for downmixing situations. In this way, metadata allows Dolby Digital to be an "equal opportunity technology" so that every consumer who receives a Dolby Digital data stream will be able to enjoy the best audio reproduction possible irrespective of the number of channels in their playback system.

Set-top boxes for terrestrial, cable, or satellite digital television will typically offer a mono signal on the RF/Antenna output, a line-level analog stereo signal, and a optical or coaxial digital output to convey the Dolby Digital stream to a decoder.

The analog stereo signals that are output from these units are most often internally downmixed from the original Dolby Digital multichannel audio program to ensure compatibility with nondigital home theater systems. This analog output can be one of two different types of stereo signals. One type is a stereo-compatible Lt/Rt downmix suitable for Dolby Surround Pro Logic decoding. The other type is a simple stereo Lo/Ro downmix suitable for playback on a stereo hi-fi or via headphones. That same signal is used for deriving the mono signal for use on the RF/Antenna output. Certain metadata parameters allow the engineer to select which stereo analog signal is preferred as the default signal at these output connections and how that stereo downmix is constructed. In addition, within the metadata parameter library there are separate parameters for the adjustment of the Lo/Ro and Lt/Rt downmix conditions.

During downmixing, the adjustment of dynamic range control parameters is limited, and this is no different when using the signals present at the stereo or RF/Antenna outputs.

The mono signal present at the RF/Antenna output of a set-top box will be applied with maximum DRC (RF Mode Profile) at all times to help prevent RF overmodulation.

Certain metadata parameters assist in achieving an appropriate downmix, helping to ensure that the intention of the engineer/content producer will translate across these environments. Specifically, metadata provides control over how certain speaker channels are “folded” into the resulting downmix.

While the engineer must optimize the multichannel mix for reproduction in an ideal listening environment, it is important to preview the mix in downmixing conditions to ensure compatibility with different playback systems when selecting the downmixing metadata parameters. These previews can be achieved in real time using the DP570 Multichannel Audio Tool.