TVTechnology

Print Page

Audio Metadata: You Can Get There From Here

8/21/2002

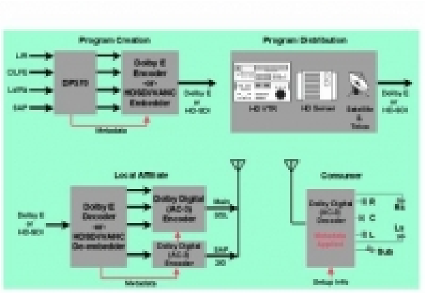
Last time we explored the three D's of audio metadata: Dialogue Normalization, Dynamic Range, and Downmixing, as well as discussing the differences between consumer and professional metadata. This time we will get down to brass tacks and try to answer three questions: How and where do these and other metadata parameters get created? How does metadata make it from the point of creation all the way to the Dolby Digital (AC-3) encoder to be sent on to consumers? and finally what happens if there is no metadata or if it is set incorrectly? The system looks great on paper and works well in many situations, but these three questions raise significant practical issues that must be overcome.

Before we delve into this, I did want to bring up two additional metadata parameters that were not discussed last time but are also important. The first is called Audio Coding Mode (or acmod), and it describes the number and type of audio channels present. The notation is Front Channels/Back Channels followed by an "L" if an LFE signal is present. For example, a 2/0 program means that there are two front channels, namely left and right, and no others. A 1/0 program means that it is mono and the audio comes from the center channel. A 3/2 program is then Left, Center and Right, and two surround channels; 3/2L is the same program with an LFE channel.

The second additional metadata parameter is called Dolby Surround Mode (or dsurmod), and it indicates whether a 2/0 (two-channel) program is Dolby Surround encoded. Although this value is informational, some newer decoders may use it to automatically trigger the Dolby Surround decoder when a program switches from 3/2L to 2/0 to keep all 5.1 speakers reproducing audio, and more importantly keep dialogue centered as we have discussed previously. My decoder does not pay attention to this value, but because I told it to always Pro Logic decode two-channel audio, it switches all by itself.

METADATA AUTHORING

Metadata can be created or authored using several different pieces of equipment. From the time Dolby Digital (AC-3) was first released until about three years ago, metadata was set in the encoder itself via some sort of user interface. Twenty-seven parameters lead to at least twenty-seven layers of menus to dig through - definitely not for the faint of heart. Luckily for users, the encoders also have a set of default values. However, this is sometimes unlucky for the viewers as the values were likely set at Dolby's factory in Brisbane, Calif., and haven't been touched since.



(click thumbnail)
Fig. 1 - Simplified signal flow from program production to consumer. HD-SDI refers to embedded audio within the VANC space.

Dolby then developed a product called the DP570 Multichannel Authoring Tool for use during the program creation stage. The unit allows metadata values to be authored at the point where the final audio mix is created. Most importantly, it allows the mixer to hear, in realtime, the effects of the changes being made. The unit can emulate playback systems that range from full-blown 5.1 all the way down to mono (through a two-inch speaker if the user has one) and allows up to eight programs to be authored simultaneously. Thankfully, the DP570 has a PC-remote control program that provides metering of all audio channels and allows relatively straightforward control of metadata and monitoring functions. Right now, these units are being used by post-production studios from coast to coast for the creation of 5.1 channel programs that can be seen on HBO, Showtime, Starz! Encore and others.

DISTRIBUTION

How do we get that metadata from the post-production studio through the network and to the affiliate's Dolby Digital (AC-3) encoder? Good question, and here are a few general suggestions.

The first - and most labor-intensive - is called sneakernet. It involves using the DP570's remote control to print a copy of all the metadata values, then "sneakernetting" those sheets of paper to each Dolby Digital (AC-3) encoder (remember, it is 27 values for one program, so hopefully there are only a few programs) and manually entering the values. What's that? Your staff has better things to be doing with their time (and sanity)? I agree.

The second solution involves using another technology called Dolby E, so named because it came after Dolby D, a.k.a. AC-3. Developed in 1999, Dolby E accepts up to eight PCM audio channels via four AES pairs and metadata and encodes them into a 1.92 Mbps data package that fits nicely into the space of a single 20-bit 48 kHz AES pair. As long as the digital path contains no sample rate conversion or level shifts, the Dolby E signal can be routed, switched and stored just like a standard AES signal. Importantly, all the audio and metadata are kept in-sync as one signal then decoded back to baseband PCM audio and metadata just prior to the Dolby Digital (AC-3) encoder.

Another distribution technique gaining quite a bit of popularity involves embedding the eight channels of audio along with the video in the High Definition Serial Digital Interface, or HD-SDI (see SMPTE 292M and 299M). What about metadata, you might ask? Well, it turns out that there is some extra space in the HD-SDI signal and it is called Vertical Ancillary data, or VANC (see SMPTE 334M). Along with closed-captioning data, the audio metadata is embedded into the VANC space of the HD-SDI stream. Again, all the audio, video and metadata for a program are kept in-sync as one signal that can be routed, switched and stored.

If a post-production facility used a DP570 to author the metadata, then used either Dolby E or HD-SDI and VANC to store the audio and metadata on the final program tape, it could be delivered to the network for distribution to affiliates. We will save the gory details of distribution for a future article, but for now let's assume that it works. The affiliates will take the signal and de-embed and/or decode to baseband PCM and metadata, feed these signals to their Dolby Digital (AC-3) encoders, then transmit the signal to consumers whose receivers will decode the audio and apply the metadata and reproduce the audio. Fig. 1 shows a typical signal flow for the process.

METADATA PROBLEMS

I have some alarming news for readers. Not one terrestrial broadcast network is currently supplying its affiliates with an audio metadata stream. I know, I know, you are saying to yourself "but I tune in DTV stations all the time and the audio sounds normal." True enough, and the signals do in fact have metadata - remember that the Dolby Digital (AC-3) encoder has a default set of values that get plugged into the bitstream during the encoding process.

Here is what happens: If the Dolby Digital (AC-3) encoder receives a valid metadata stream, it uses some of the information to configure the encoder and combines the rest with the encoded audio to create a complete Dolby Digital AC-3 stream that will be sent on to consumers. If the encoder does not receive a valid metadata stream, either because it is not being sent or due to data errors, the encoder can continue to use the last good metadata values it received or revert to internal preset values. Although this will keep audio of some sort on the air, at best it will not be optimized and at worst there will be channels missing.

As this behavior is user-configurable, I would strongly recommend giving some serious thought to the choices you make before you start passing on external metadata from the network or other sources. If the metadata disappears for some reason, you certainly do not want the dialogue to go with it because the encoder dropped back to 2/0 during a 3/2 program (the center channel would be ignored!). For most situations, I would recommend that the encoder be set to revert to values that were as "large" as the largest program. This means that at a minimum, Audio Coding Mode would be set to 3/2 or 3/2L and Dialogue Level set to -27. The rest of the values are not as critical as these two and can remain at default values.

One of the questions I am asked most often is "what happens if someone tries to trick the system and set the dialogue level setting incorrectly to purposefully make a program or a commercial louder than it should be?"

To recap, the dialogue level metadata parameter represents the long-term A-weighted loudness of the dialogue of a given program and controls a 1-dB-per-step attenuator present in every Dolby Digital (AC-3) decoder, which is used to normalize the decoded audio to a target value of -31 dBFS. A program that has relatively quiet audio might have a dialogue level of -31, and the system will therefore apply no attenuation after decoding. A loud program might have a dialogue level of -11 dBFS and the system will apply 20 dB of attenuation to make the program match the target of -31. So, what happens if a devious commercial producer makes a loud commercial (is there any other kind?) and sets the dialogue level metadata parameter to -31? The loud commercial becomes even louder. The dynamic range system built into Dolby Digital (AC-3) will help somewhat but not completely, especially if you have decided to watch a movie with full dynamics and have told your decoder to ignore the dynamic range control metadata.

This leads us nicely into our topic for next time. We will investigate the first of two of arguably largest complaints with television audio: Loudness. Using some of what we have learned so far we will uncover some of the root causes of loudness problems in both analog and digital television audio and discuss some potential solutions. For anyone that might be interested, I have made two metadata drawings available on my Web site at www.linearacoustic.com/downloads.htm. They depict how and where metadata is used by both the Dolby Digital (AC-3) encoder and decoder. As always, thanks for your continued support and feedback! **Print Page**