AN ANALYSIS OF AUDIO FOR DIGITAL CABLE TELEVISION RECOMMENDATIONS FOR THE DIGITAL TRANSITION VIA AUDIO METADATA

Jeffrey C. Riedmiller
Dolby Laboratories Inc.
San Francisco, CA

Abstract

With an increasing number of cable systems deploying large volumes of digital set top boxes and digital services it has become increasingly apparent that large audio level discrepancies can exist between analog and digitally tiered programming. Since the local cable system usually has no control over the digital service audio levels it is up to the programmer(s), in most cases, to rectify any of these discrepancies. Programmers offering digital services are able to take advantage of some very unique tools included with their encoding systems to assist with many of these issues. These tools include the generation of audio metadata that is carried within the coded audio bitstream to the digital subscribers set top decoder. In this paper, the impacts of supplying digital programming with "valid" audio metadata vs. "default" or "invalid" audio metadata from the digital cable subscriber's perspective will be analyzed. However, the proper use of audio metadata by the programmer also assumes that the millions of deployed set top decoders have been designed with a clear understanding of the entire system as well. Recommendations for digital set top designs are needed for the cable industry and will insure that the effort to generate "valid" metadata by programmers will result in the expected behavior within the digital set tops. To this end, a reference model is also proposed in this paper.

Audio Metadata Origin and Digital Audio Delivery Requirements

The digital audio delivery method currently used by cable networks and cable system operators is the Dolby Digital (AC-3) audio coding system and has been adopted by the Society of Cable Telecommunications Engineers (SCTE) for use in digital cable systems within the United States.

When the Dolby Digital (AC-3) coding system initially developed for cinema applications was adapted into a form suitable for consumer use, the differing and practical needs of consumers were evaluated. As a result, it was decided to create a signal representation syntax that would allow a single encoded bitstream to be decoded into a form useable by nearly every potential listener. Many factors were considered when developing this syntax. The first was how a potential listener may receive the decoded content. In the case of a digital cable set top box this could be via the line level, RF or Digital Interfaces. Second, the number of playback channels in use. Thus allowing multichannel audio programs to satisfy consumers with only 2-channel decoders. Third, the amount of dynamic range desired knowing that consumer product capabilities and individual tastes vary. Fourth, provide a means for level matching between programs and channels. Finally, the allowable dynamic range on a modulated RF interface to control peak program levels similar to present day broadcast practices.

This syntax or “extra” information carried in the encoded bitstream is commonly referred to as metadata.

What is Audio Metadata?

Audio metadata is the data sent in tandem with the coded audio signal to describe this signal to the receiver/decoder. In the case of Dolby Digital (AC-3), audio metadata is used along with subscriber input to control how the audio program is presented in the home, including modification of the program dynamic range or the number of playback channels. For the purposes of this paper, some of the more important metadata parameters and their uses are shown in Table 1 (Table 1 is located on the last page of this paper). These “key” metadata parameters will have the greatest effect on how a listener or subscriber experiences a program. A complete list of transmitted metadata can be found in ATSC Document A/52. It is important to note that these parameters are normally determined during the Dolby Digital encoding process by configuring the encoder.

The Informational metadata shown in Table 1 is used by the Dolby Digital decoder to optimize the decoding process by specifying the number of channels used and the program configuration. It determines where to route the decoded audio channels, whether or not the channels should be downmixed (i.e. 5.1 channel to 2 channel), and what type of service is present (i.e. Complete Main, Music & Effects, Dialog, etc).

Level Control metadata sets the overall program playback level for loudness normalization (*dialnorm*). It also determines how various mixdowns from 5.1 channels to one, two, or three channels are achieved (cmixlev, smixlev).

Dynamic range control metadata has the ability to control the dynamic range of the decoded program at the receiver. The values are calculated in the encoder based on dynamic range profiles selected by the content creator or the programmer. There are five standard profiles: Film Standard, Film Light, Music Standard, Music Light, and Speech. As noted in Table 1, a *compr* word is generated once per Dolby Digital frame (32 msec) and a *dynrng* word is generated once per block (5.3msec). In addition to generating dynamic range control words for presentation purposes, the encoder also generates control values to prevent clipping in the decoder. This is necessary due to the higher levels produced by combining channels if downmixing is required. The details of Dynamic Range Control will also be discussed later in this paper.

Digital Set Top & Subscriber Requirements

In most cases the digital cable set top box has three primary interfaces for supplying audio to subscriber owned equipment. The two most common interfaces include the ch.3/4 remodulator (via RF output) and the analog baseband L&R line level outputs (via RCA connectors). The third output, if available, gives the subscriber access to the raw Dolby Digital (AC-3) bitstream which can be decoded externally by a home theatre system, if desired. It is also important to note that in two of these cases (i.e. baseband and RF), the decoded digital program must satisfy and provide the digital subscriber a seamless listening experience when switching between digital and analog channels.

At first glance, this level of operability may seem impossible to achieve. However, with a thorough understanding of the Dolby Digital audio delivery system and the proper application of audio metadata these requirements can indeed be met, and therefore increase subscriber satisfaction.

Since Dolby Digital (AC-3) programs need to coexist with analog (NTSC) program sources, the gain structure of the set top box should satisfy the following criteria: 1. All sources whether analog or digital from the set top should provide similar dialog levels at the line level outputs and in any operating mode. 2. In most cases the set top includes a RF modulator and the dialog level from digital sources should match that of typical analog (NTSC) broadcast practices. Note that this can only be accomplished with the Dolby Digital decoder operating in RF Mode. 3. The line level outputs should provide, when connected to the line inputs of a television, similar dialog levels to that of the television's tuner.

The last paragraph referred to the "dialog level" of a particular source, analog or digital. This level is used as a metric to establish a subjective level match for the subscriber from any type of source. This unique approach to level matching may seem quite different at first glance but experience has shown that a very large percentage of television viewers base their listening enjoyment of a particular program on how intelligible the dialog is. Thus, the reproduced dialog level now becomes a reference for the programmer and sound mixer to use to ensure the consistent reproduction of their programming. In the case of motion picture mixing, the dialog loudness is largely adjusted, and thus normalized, by ear in a room similar to a typical theatre with the acoustic "gain" carefully set to a standard value. The motion picture industry has been performing this "dialog normalizing" practice for years to compensate for the inability of theatres to adjust levels on a film-by-film basis.

On the other hand, many television broadcasters still employ the same practices and audio processing techniques that have been around for some time. First and foremost, the dynamic range of the program is often limited and the level of speech is such that waveform peaks may frequently reach 100% modulation, which, in turn, leaves a very limited amount of headroom for music and sound effects to make a dramatic impact on the listener. However, Hollywood theatrical films offered by premium cable programmers' use the same soundtrack mixed for theatres that have reproduction systems with virtually unrestricted dynamics capabilities. These films often have far greater dynamic range and headroom requirements than typical television programming. Also note that because of this increased headroom, dialog will be placed lower in level within the mix, and will not generally match normal television program dialog levels. This is due to their different peak-to-average ratios. For reference, movies typically have a dialog level of ~ 27dB LAeq below maximum level and typical NTSC broadcasts having a typical dialog level of ~ 17dB LAeq below 100% modulation. (LAeq will covered in the next section)

By now it should be obvious that presenting digital subscribers with these two very different types of programming while simultaneously keeping their equipment from being driven into overload could certainly cause complaints. This could result from the differences in the high peak-to-average ratio (i.e. wide dynamic range) sources such as film and a low peak-to-average ratio (i.e. narrow dynamic range) source such as a news broadcast. Often, a premium cable programmer solves this by processing their audio programming to more closely match current television broadcast practices. By restricting the dynamic range and therefore increasing the average dialog level of the program, sources can be made to match. While this solution is quite acceptable for programming that does not require the dynamic range of typical theatrical releases, it forces this restriction on all programming and therefore ignores the capabilities of the Dolby Digital system as well as the intentions of the original sound mixer.

With the Dolby Digital system, there are no assumptions about the program levels, so the source can be presented to the encoder with almost any operating practice or philosophy. The system is quite capable of delivering programming with differing average dialog levels and dynamic range without potentially annoying subscribers as they switch through various types of programming that exist on a typical cable system. However, this can only be accomplished when the system is utilized and implemented properly by the programmer, set top manufacturer and in some cases the local cable system. For the programmer itself, it is extremely important to understand the use of Dolby Digital metadata and the parameters that generate and control it in their encoders. One of the most important encoder parameters is the Dialog Normalization value – *dialnorm*. Once this value is entered in the encoder, it is used by the encoder and carried in the bitstream to the decoder as a part of metadata. The *dialnorm* value provides, among other things, the digital subscriber with normalized dialog levels between differing types of digital sources and programs. The next two sections will cover this key metadata parameter in detail.

Level Measurement Practices

Broadcasters usually control their program levels by a volume unit (VU) meter or a peak program meter (PPM) and it is important to note that both are used to read signal voltages, and therefore make no attempt to measure subjective loudness. Thus several different voices, adjusted in level so that they all deflect meters to the same mark, may sound somewhat different in level to the listener. Over the years there have been several attempts to design measurement devices (i.e. meters) which give the broadcaster results that more closely match subjective loudness. One measure that gives us results that correlate closer with subjective loudness is called the equivalent loudness method (LAeq) which is the long-term average A-weighted level. Leq itself can be defined as the level of a constant sound, which in a given time period has the same energy as a time-varying sound. Matching the levels of different voices so that they give the same LAeq measurement will deliver closer subjective loudness than either PPMs or VU meters. PPM and VU meters are also frequently used to measure and/or align to a pre-determined “house” reference level, and thus only have an arbitrary relationship to the dialog or speech level within a given program. For example, if a VU meter and a PPM meter are calibrated to display a reference tone equally, and speech that averages 0VU is applied to both, the PPM meter will indicate levels considerably above its reference level and possibly above the maximum permitted level. On the other hand, speech that averages at the PPM reference, will most likely indicate many dB below 0VU (our original reference). This confirms the very important idea that the reference level is not the same as the speech or dialog level of a program. Instead, perhaps we should define the relationship between reference levels and the actual dialog (or speech readings) of a given program. With this relationship known, we can then easily describe how to make programming controlled with VUs or PPMs approximately deliver standardized dialog level in Dolby Digital (AC-3).

The standardized dialog level for Dolby Digital (AC-3) is defined as –31dBFS and is quantified using the equivalent loudness method that is A-weighted (LAeq) as stated above. This is the level of spoken dialog present in a single channel, such as the center channel of a 5.1-channel program. For stereo and downmixed (i.e. 5.1 channel downmixed to 2-channel as in the case of digital cable set top boxes) programs, the level is equivalent to –34dBFS LAeq in both channels simultaneously. This is because dialog will be present in both the left and right channels and will combine acoustically for a ~3dB increase in level.

As stated earlier, dialog levels in conventional analog television audio are roughly normalized, in many cases, by the use of automatic dynamics processing. Recent measurements indicate that they are typically ~ -17dBFM, where 0dBFM would equal the maximum permitted FM carrier deviation (25kHz in the U.S.) More detail on the importance of this will be presented in the following sections.

Dialog Normalization – *dialnorm*

Perhaps the most often misunderstood Dolby Digital (AC-3) metadata parameter is the Dialog Normalization value or *dialnorm* value. The Dialog Normalization parameter describes the long-term average dialog level of the associated program and is specified on an absolute scale that ranges from –1 dBFS (dB Full Scale Digital) to –31 dBFS LAeq.

Encoded Dolby Digital elementary streams carry the audio signals that are fed into the encoder with no changes to the level or the program's dynamics. It is the metadata, and in this particular case the *dialnorm* value, that is used by the decoder to adjust the reproduced level of audio programs which will then reproduce the dialog at a consistent or uniform level. It is also important to note that all decoders are required to make use of this metadata parameter and apply the proper level normalization/attenuation (based on the transmitted *dialnorm* value) to the decoded audio program.

As an example of what the dialog normalization parameter is capable of, imagine switching between a news program and a wide dynamic range movie. In order of magnitude, these items may have average dialog values of about –14 and -28 dBFS LAeq respectively. Thus, if a listener sets the playback level (using a volume control) to comfortably reproduce the news program and then switches to the movie, its dialog will be about 14 dB (28–14) quieter than the voice of the newsreader, and probably unintelligible. This will likely force the listener to adjust the volume control every time they switch between the two. In this example, note that the quieter source is the movie dialog. The reason for this as stated earlier is movie mixers use a standardized acoustic level for dialog while mixing. With digital formats this is equivalent to between -25 and -31 dBFS LAeq. Since movies constitute a large portion of the material to be conveyed by Dolby Digital, this standardization is retained. This means that *all* dialog (movies, sports, news, etc) should emerge from a Dolby Digital decoder at about -31 dBFS LAeq. However, for this to be the case the programmer **must** properly set the *dialnorm* value in the Dolby Digital encoder. In the example above, *dialnorm* needs to command 17 dB of attenuation on the news program, and only 3 dB on the movie to create an acceptable level match between the two upon decoding. In other words, *dialnorm* in a sense acts as an automatic volume control. In the case of programming that does not contain dialog, such as music, *dialnorm* should be adjusted appropriately to match programming that includes spoken dialog.

Remember that the *dialnorm* value itself is **not** the standard operating level of a facility (i.e. -20dBFS = + 4dBu or 0VU) but the equivalent of the spoken (not shouting or whispering) dialog level with respect to digital full-scale. The attenuation introduced in the decoder is (31 + *dialnorm* value) with the *dialnorm* value being negative. Speech in a single channel (i.e. center channel of a 5.1 channel program), with an LAeq value of -31 dBFS should have -31 entered in the encoder, which commands 0 dB of attenuation in the decoder. If speech is present in two channels simultaneously (a stereo program) with an LAeq value of –34dBFS in each channel, –31 should be entered in the encoder. Similarly, the news bulletin (mono) with speech at -14 dBFS LAeq requires a *dialnorm* setting of -14, giving 31 + (-14) = 17 dB of attenuation so that speech from the newsreader comes out of the decoder at -31 dBFS. If the source material is recorded at a lower level resulting in peaks that do not approach digital full-scale, less attenuation is needed, and the *dialnorm* value moves closer to its minimum value of -31. Thus, if the news bulletin used in the example above had a dialog level of -20 dBFS LAeq rather than -14 at its source, a *dialnorm* setting of -20 would yield the standard level (-31 dBFS) at the decoder outputs while operating in Line Mode.

This unique feature alone has the potential to minimize the average channel to channel dialog level problems that have plagued cable television for years and subsequently has forced many premium service programmers to process their audio to match the typical dialog levels and dynamic range of other non-premium programming. The Dialog Normalization parameter enables the premium service programmer to achieve an acceptable program to program level match while minimizing the need for the traditional and irreversible heavy processing that is commonly applied to most non-premium type programming. This leaves ample headroom for feature films regularly offered by premium network providers while simultaneously maintaining an acceptable level match with typical programming that has limited dynamic range. If the entire delivery system is to function as intended, an understanding of set top decoder operating modes is required. It is also critical to ensure that the decoder is operating in the correct mode, which can differ in certain applications. Unbeknownst to many cable programmers and local cable system technical personnel, the digital set top that their particular system may deploy can potentially offer and/or default to different operating modes which impact the subscriber in different ways. These modes may disable the ability to offer the subscriber an acceptable level match between analog and digital programming.

Digital Set Top Decoder Operation

This section defines the importance of digital set top decoder operating modes and how each of them are applied to allow us to satisfy our three original requirements which are restated here:

Digital Set Top Output Level Requirements:

1. All sources whether analog or digital from the set top should provide similar dialog levels at the line level outputs and in any operating mode.
2. In most cases the set top includes an RF modulator and the dialog level from digital sources should match that of typical analog (NTSC) broadcast practices. Note this can only be accomplished with the Dolby Digital decoder operating in RF Mode.
3. The line level outputs should provide, when connected to the line inputs of a television, similar, dialog levels to that of the television's tuner.

Dolby Digital decoders found in consumer products, in general, can operate in two modes. Each of these modes has a specific application, and care **must** be taken when the set top is designed and deployed to insure that the intended mode is used by default.

Decoders can be found in many places including Digital Cable set top terminals, Home Theatre systems, consumer satellite Integrated Receiver Decoders (IRD), DBS receivers and Commercial IRDs that are used in cable headends, and Cable turnaround uplink facilities. It is important to note that the default operating modes in each of these cases may vary and are based on that particular device's function within a given system.

Line Mode Operation:
Line Mode operation generally applies to the baseband line level outputs from two-channel set-top decoders, two-channel digital televisions and multichannel Home Theatre decoders. It is important to note that Line Mode operation is a requirement for all digital cable set top boxes that have analog baseband (line level) outputs. With respect to consumer type applications, a decoder's outputs operating in this mode will typically be connected to a much higher quality sound reproduction system than that found in a typical television set. In this mode, dialog normalization is enabled and applied in the decoder <u>at all times</u>. Further, in this mode the normalized level of dialog is reproduced at a level of -31 dBFS LAeq, but **ONLY** when the transmitted *dialnorm* value has been correctly adjusted for a particular program. In general, with the reproduced dialog level at –31dBFS, this mode allows wide dynamic range programming to be reproduced without any peak limiting and/or compression applied as may be intended by the original program producers. Further, since the Dolby Digital (AC-3) digital audio coding system can provide more than 100dB of dynamic range, there is no technical reason to encode the dialog at or near 100% as is commonly practiced in analog television systems. This allows the delivery system to meet one of its goals of being able to deliver high impact cinema type sound to the digital subscriber's living room.

RF Mode Operation:
RF Mode is intended for products such as cable and satellite set-top terminals that generate a monophonic and/or downmixed signal for transmission via the channel 3/4 remodulator that feeds the RF (antenna) input of a television set. This mode was specifically designed to match the average reproduced dialog level and dynamic range of digital sources to those of existing analog sources such as NTSC and analog cable TV broadcasts. In this operating mode dialog normalization is enabled and applied in the decoder <u>at all times</u>. However, the dialog level in this mode is reproduced at a level of -20 dBFS LAeq **ONLY** when the transmitted *dialnorm* value is valid for a particular program. The Dolby Digital decoder introduces an +11 dB gain shift and thus the maximum possible peak to dialog level ratio is reduced by 11dB. This is achieved by compression and limiting internal to the decoder. It is important to note that digital set top boxes which include an RF modulator are required to provide RF Mode in addition to Line Mode operation and also **must** include some way of changing operating modes.

Figure 1 compares the signal relationships in the decoder for both Line and RF operating modes. Notice the reproduced dialog level and dynamic range of each mode.

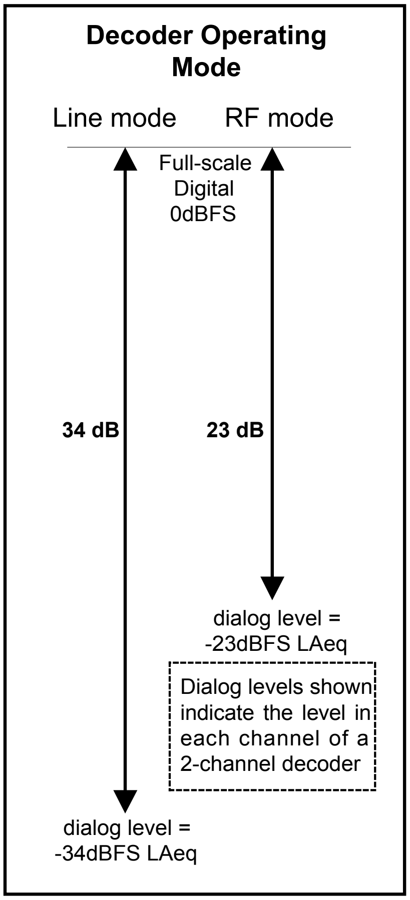


Figure 1

Decoder Operating Mode Levels - Line vs. RF

Digital Set Top Gain Structure

The digital cable set top is required to provide the subscriber with both analog and digital programming. It is imperative that digital set top manufacturers fully understand both decoder operating modes as well their applications within a given set top design.

The audio subsystems within the set top include:

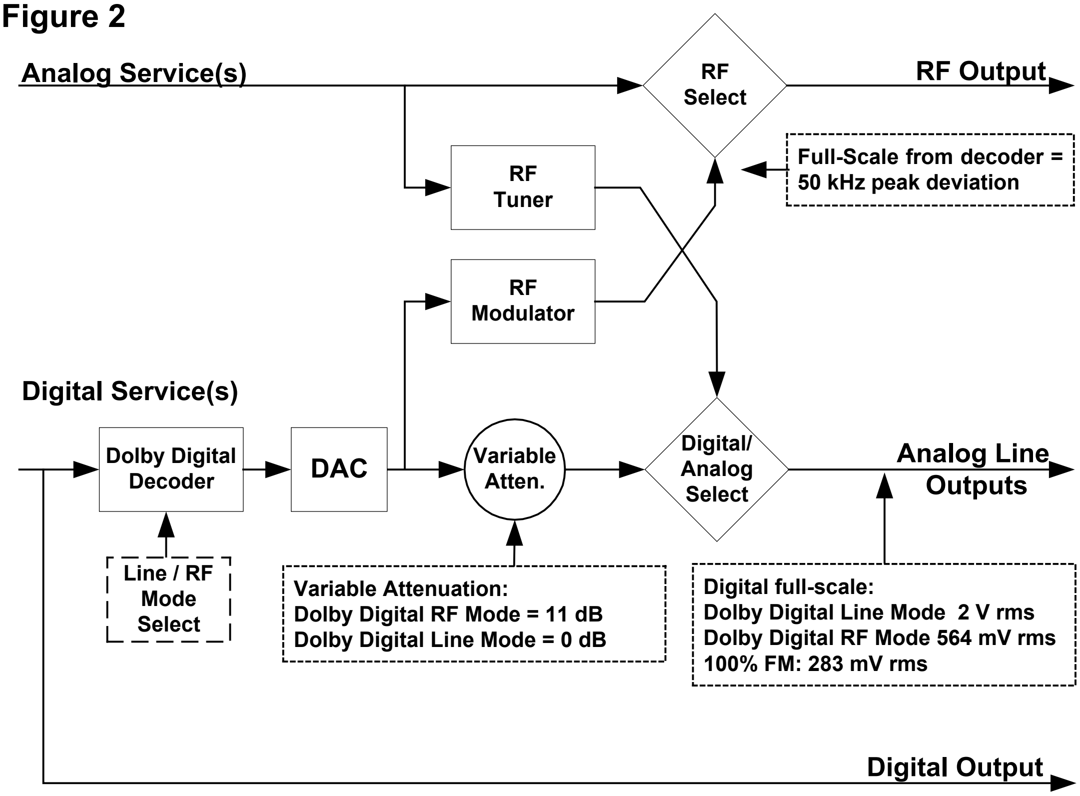
1. NTSC Tuner
2. 2-channel Dolby Digital decoder IC and associated digital to analog converter
3. Channel 3/4 FM modulator

Beginning with the FM modulator, we shall define 0dBFM as the maximum FM deviation (25kHz) for NTSC broadcasts. Both theory and recent measurements indicate that NTSC tuners have ~ 6dB of headroom above 25kHz deviation while receiving a mono signal. As will be seen, this works to our advantage.

With a Dolby Digital decoder operating in RF Mode, the normalized dialog level is a –20dBFS LAeq (–23dBFS LAeq in each channel), since most digital cable set tops utilize a 2-channel decoder. In some set top designs the output of the decoder IC may feed digital multipliers, used for digital control of the volume before being presented to the digital-to-analog converters (DACs). With this in mind, the gain structure should be provisioned initially with volume controls set to unity gain.

To set the proper amount of gain into the modulator for digital sources, the two analog signals from the DACs should be combined and fed to the mono RF modulator with gain such that a correlated tone (i.e. 400Hz) in both channels (L&R) at 0dBFS yields +6dBFM (200% modulation). When the set top is tuned to a digital source AND this particular source has the correct dialog normalization value associated with it, the dialog should emerge from the decoder operating in RF Mode at –23dBFS LAeq in each channel. Referring to Figure 3, it can be seen that with the RF modulator gain structure set as indicated above, the –23dBFS dialog becomes –17dBFM and matches the typical dialog level of NTSC broadcasts. On the contrary if the Dolby Digital decoder is operating in Line Mode the decoded dialog level for the same program must be 11dB lower.

Figure 2



Digital Cable Set Top Terminal - NTSC Version

Recall that in Line Mode peaks can be 11dB higher with respect to the dialog level and in order to avoid overload the average level must then be 11dB lower

Therefore, the normalized dialog level in Line Mode being –34dBFS LAeq in each channel, then becomes –28dBFM. Obviously, under these conditions the digital subscriber will experience a severe level mismatch between analog and digital sources. Hence the requirement, as stated earlier: digital set tops which include an RF output **must** also include the capability of placing the internal Dolby Digital decoder into RF Mode. Note the dialog levels for both operating modes in Figure 3.

The line level outputs on digital cable set top boxes will, in most cases, feed the line inputs of a stereo television, VCR or a home theatre system. And since it may be desirable for set top output levels to match the standard output levels of other consumer audio equipment such as CD and DVD players we must consider the following. Many consumer audio products (with line level inputs) are usually designed to accept a level of 2Vrms maximum and not much more. Therefore the digital set top box must be designed such that digital full-scale signals are able to deliver 2Vrms at the line level outputs.

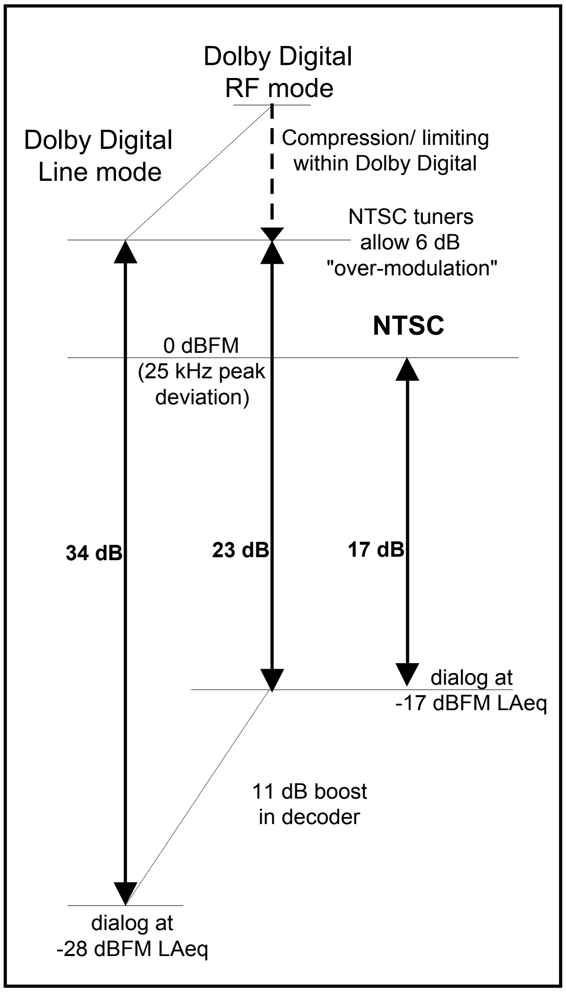


Figure 3

Set Top Gain Structure

To satisfy one of our earlier requirements of having matched dialog levels in both decoder operating modes at the line level outputs, we shall re-state that full-scale (0dBFS) signals from a Dolby Digital decoder operating in Line Mode should give 2Vrms at the line output connectors. Hence, the standard normalized dialog level (quantified using the LAeq method) in Line Mode will be 34dB below 2Vrms and equates to 40mV rms in each channel. If the decoder is deliberately set or defaults to operating in RF Mode, it becomes necessary to remove the 11dB of boost applied in RF Mode before applying the signal to the line level outputs **ONLY**. See figure 2 & 4. This will then produce the identical LAeq dialog level at the line level outputs of the digital set top box as when the decoder is in Line Mode (40mV rms).

For analog services (NTSC) delivered to the digital set top, our gain structure assumes that the normal dialog level is about –17dBFM. If the NTSC dialog level is to match digital sources, the maximum deviation of 0dBFM should be 17dB above the normalized dialog level of the Dolby Digital decoder (40mVrms). Therefore, 17dB above 40mVrms results in 0dBFM being equal to 283mVrms. See Figures 2 and 4 for more information.

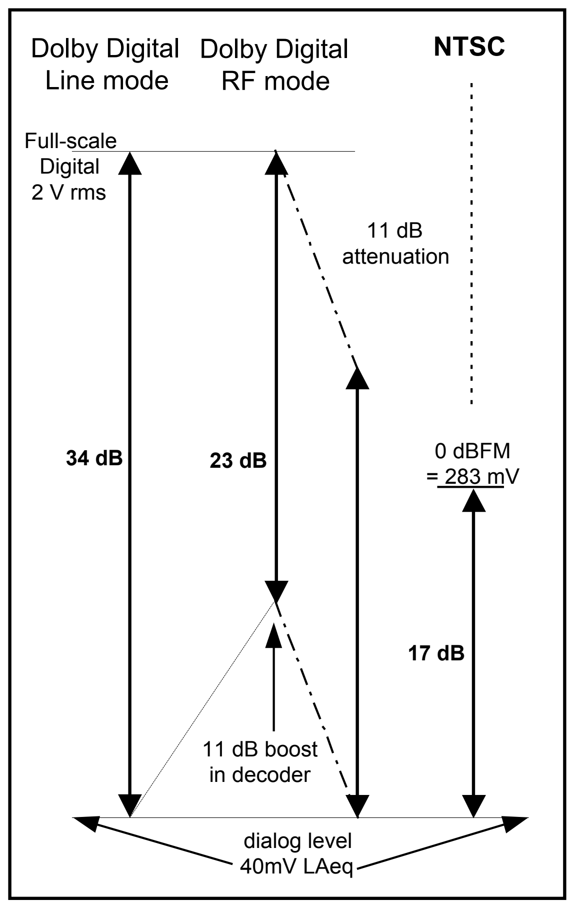


Figure 4

Set Top Box Line Output Levels

Dynamic Range Control

The Dolby Digital system conveys audio without altering its dynamics. Unlike any previous broadcast delivery mechanism, it gives the listener the option to hear the program as the original program producer intended, even if that means that it ranges from barely audible to extremely loud (i.e., wide dynamic range).

Typical analog broadcast processing usually forces the audio program level towards full modulation for a substantial portion of the time, lowering the peak-to-average ratio thereby eliminating most of the dynamic range. A benefit that comes with this is an approximate normalization of the average listening level. In other words, the same device that reduces the dynamic range determines the average volume. With Dolby Digital there are no technical pressures to reduce dynamic range, and the average volume is addressed by the *dialnorm* parameter discussed earlier. Further, the need for dynamic range compression can be considered to be independent of average listening levels.

The Dolby Digital system utilizes a unique approach to applying dynamic range compression to audio program material. The purpose of the algorithm itself is to make the decoded audio levels closer to the dialog normalization level, amplifying material that is lower in level and attenuating material that is above the dialog level. Rather than reducing the dynamic range in a non-reversible way, the Dolby Digital encoder generates compression gain words that are carried along in the Dolby Digital bitstream as part of metadata. These gain words are calculated based on a number of separate input parameters, including the selected dynamic range compression preset (see the compression profile section), the level of the program material itself, and the program dialog normalization value. Hence, setting the *dialnorm* value properly is a critical first step in calibrating the dynamic range compression system.

The DRC (Dynamic Range Control) algorithm supplies two types of gain words to the decoder, *dynrng* and *compr*. The *dynrng* variable is applied to the decoded audio when the decoder is operating in Line Mode. This control signal is generated based on the original program producer's artistic choice of a compression profile preset, which is included in the Dolby Digital encoder. Since many of consumer decoders default to Line Mode, the programmer may choose, via the compression profile preset, the one that suits the needs of most of their audience. Individual listeners may have the option to choose to decode the program with all of its original dynamic range depending on the type of decoder they have. However, studies and experience have shown that in most cases the majority of television viewers would never want or need the full dynamic range of the audio that this system is capable of delivering.

The *compr* variable is applied to the decoded audio when the decoder is operating in RF Mode. This control signal is generated to insure that the peak modulation of the RF remodulator is controlled to an acceptable value when the dialog is modulated to a value that is similar to current analog NTSC practices. This provides the listener who chooses to use the RF output of a set top terminal with acceptable results without affecting listeners that may be using the line level outputs and/or an external Dolby Digital decoder in a home theatre system.

Downmixing, Overload Protection and Dynamic Range Control

When a multichannel program is downmixed within a Dolby Digital decoder, the downmix coefficients are generally fixed values. Downmixing is performed in the digital domain (except in the case of a 2-channel to mono downmix), and obviously there is the possibility that the downmix will overload/clip the digital-to-analog converters (DACs) in the decoder. If the fixed downmix coefficients in the decoder were chosen so that downmixing a multichannel program would never overload the DACs, many of these downmixed programs would sound quieter than the same program reproduced in a multichannel mode or a mono or stereo program that did not require downmixing. With this potential problem in mind, the actual downmix coefficients were chosen to give a more satisfactory match in output level between downmixed and non-downmixed programs. However, the caveat with this choice is that it might result in output overload while downmixing on the rare occasions that a multichannel program approached digital full-scale on all channels simultaneously. Remember that most programming typically demands level attenuation within the decoder via the indicated *dialnorm* value. This attenuation is applied in the digital domain prior to downmixing, and in turn reduces the probability of overload the DACs. However, the 11 dB of boost provided by the decoder in RF Mode does once again increase the probability of overloading/clipping the DACs while downmixing. These different situations were taken into account and are addressed within the dynamic range compression algorithm.

To assist the system in predicting a possible overload condition the encoder generates several possible downmixes in parallel with the computations taking place that generate *dynrng* and *compr* gain words. These downmixes are used to estimate the worst-case peak level at a decoders output while taking into account the value of *dialnorm* that is currently being used. This separate process then calculates the gain reduction needed to prevent an overload condition for decoders operating in both Line & RF Modes. The two values of gain reduction derived from this process, one for Line Mode and one for RF Mode, are compared to the gain reduction values demanded by the selected dynamic range compression profile. The larger value is then substituted for the compression figure in the *dynrng* and *compr* control word. With typical multichannel programming, protection limiting is rarely needed or necessary when the accompanying *dialnorm* value is correct. As an example, imagine the highly unlikely case where a five-channel program reached full-scale simultaneously in all channels. This represents a sound roughly 30 dB higher than standard dialog level. In this case, downmixing with the decoder operating in Line Mode and preventing an overload would require that a total of 11 dB of gain reduction be applied from *dialnorm* and dynrng. In other words, the attenuation requested by both the *dialnorm* value and the calculated dynrng gain word (actual value based on compression profile selected in the encoder) must be 11 dB or more to keep heavy protection limiting from taking place. In RF Mode, this condition could require 22 dB of gain reduction (*dialnorm* + *compr*). This may be a good reason to always use an artistic compression profile in the encoder when given the choice.

When the bitstream is decoded, the decoder itself is responsible for applying these gain words to the reproduced audio. The particular gain words used are determined by the operating mode that the decoder is in. The decoder can be instructed to provide the full amount of compression indicated by the gain words, reduce or scale the amount of compression applied, or even apply no compression at all (as long as the decoder is operating in Line Mode and not performing a downmix). This allows the end user to adjust the amount of dynamic range compression applied based on their individual taste.

Compression Profiles & Presets

As stated earlier, in many applications Dolby Digital (AC-3) will be compared and heard along side analog terrestrial and cable programming sources. There are several Dynamic Range Compression profile presets included in the current Dolby Digital encoder that permit degrees of processing that are similar to present day broadcast practices. They all include a null band, which is a region in the middle of the dynamic range where the gain is fixed at unity. (i.e. no boost or cut) The ends of this null band are referenced to the value of *dialnorm* so that dialog or average program lies within the null band and is not subject to gain variation. Applying dynamic range compression in this way provides two important advantages. First, program material with an already restricted dynamic range will generally remain within the null band and in turn, is not subjected to any further dynamic range compression, and second, dialog does not modulate background noises. The result is that this type of processing reduces the dynamic range without producing audible side effects such as transient distortion or gain pumping often associated with broadcast processors. Figure 5 is a graphical representation of unity gain within the compression profile null band.

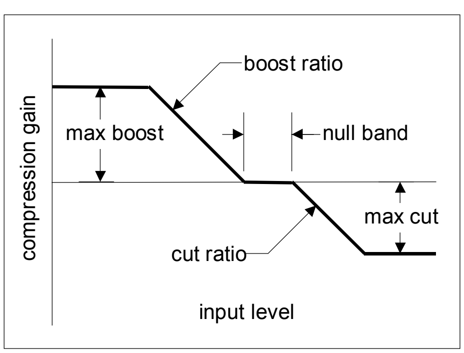


Figure 5 – Example Dynamic Range Profile

It is assumed that a typical wide dynamic range program, such as a film, has peak levels at least 10 to 15dB above the dialog peak levels and will probably be unacceptably loud for many subscribers. Thus, if the dialog level lies within the null zone, typically 10dB wide, the only way to restrict these changes in level is to use a high compression ratio above the top end of the null zone. Below the null zone, the maximum amount of gain to be applied is cause for concern. In film, raising low-level sounds excessively would sometimes reveal unwanted background disturbances, such as camera and traffic noise, which the original program producer did not intend to be audible. Since the sounds intended to be heard are rarely much quieter than dialog, film soundtracks would only require small degrees of low-level boost. Thus the low level compression ratio is not critical and limiting the maximum gain for film soundtracks to only about 6dB will provide acceptable results.

Currently, there are five Dynamic Range Compression profiles in addition to None included in the Dolby Digital Encoder. These profiles are: Film Standard, Film Light, Music Standard, Music Light and Speech. For more information on the specifics of each compression profile please consult the Dolby Digital Professional Encoding Guidelines Manual.

Also note, if the “None” compression profile is selected, the dynamic range compression algorithm generates the desired compression gain words set to 0 dB gain (i.e., no cut or boost). However, the “None” profile does not disable overload protection and it is possible that the actual compression gain words in the Dolby Digital bitstream will be less than 0 dB.

Current Observations and Practices

A recent analysis has indicated that much of the digitally distributed cable programming is encoded with incorrect *dialnorm* values and in many cases is utilizing the preset default values that many Dolby and Dolby licensed encoder manufacturers have implemented. This most likely stems from confusion over the purpose of *dialnorm* and how to properly set it for a given program. Note that default values may vary from manufacturer to manufacturer, further aggravating the problem. The documentation that usually accompanies these very complex encoding systems typically does not include clear explanations on how *dialnorm* and other metadata parameters inter-relate, how to set them, and in the end, how they affect the listeners and subscribers. At least one manual simply states that setting the *dialnorm* value to –31 will cause no attenuation to take place in the decoder. While this statement is true, it should be realized by this point in the paper some of the serious drawbacks of doing so.

Further complicating things, some of these potential drawbacks become increasingly worse if the correct *dialnorm* value for a given program is a greater distance (with respect to dialog level) from a typical default *dialnorm* value of –31. Take the case of a program that regularly reaches 0 dBFS and has a very limited dynamic range such as rock music. Encoding this type of program with a *dialnorm* value of -31 appears logical, due to the fact that no deliberate attenuation will take place in a decoder operating in Line Mode. Unfortunately it will, more than likely, cause audible side effects with decoders operating in RF Mode due to protection limiting, and may lead to subscriber complaints. Since a large percentage of digital cable subscribers are utilizing the channel 3/4 RF output of their digital set top terminal to feed their televisions, properly setting the Dolby Digital metadata is extremely important.

Recall that a Dolby Digital decoder operating in RF Mode reproduces the dialog level (based on a valid transmitted *dialnorm* value) at –20 dBFS LAeq which is an +11dB increase in dialog level when compared to a decoder operating in Line Mode of -31dBFS LAeq. For this example, when the encoder calculates and creates the *compr* gain words (see the earlier section on Overload Protection) used by a decoder operating in RF Mode, they are given the responsibility for at least 11dB of overload protection limiting to keep the set top decoder DACs from clipping. This type of protection limiting can be aggressive at times, especially in the case where an artistic dynamic range compression profile was not chosen in the encoder. Had the *dialnorm* value been set correctly in this case, there would be minimal or no overload/heavy protection limiting applied in a decoder operating in RF Mode.

While the correct setting of *dialnorm* for this program will minimize any side effects from heavy protection limiting with decoders operating in RF Mode, it will also lower the reproduced level for decoders operating in Line Mode. This behavior is expected. In the case of popular music productions, with recorded levels frequently approaching 0 dBFS and average levels not too far below that, some attenuation in the decoder is required to better match the average levels found in typical analog and digital broadcast programming.

Recently, many premium service cable networks have installed systems that enable them to offer multichannel audio to their subscribers. With some of these systems it is now possible to easily carry all of the proper audio metadata generated in production, through the entire distribution system, and eventually to the subscriber's digital set top terminal and/or decoders placed in a cable headend or turnaround uplink. When a service such as this is directly compared to another programmer's service, which is using default and most likely an incorrect *dialnorm* value, it can potentially further aggravate the level discrepancies. In the worst case, this scenario applies to cable systems that are supplying digital subscribers with programming with valid metadata (in this case a correct *dialnorm* value) and simultaneously offering other programming without valid metadata, all in the same channel line up delivered to the subscriber's set top terminal.

For example, a movie typically has a *dialnorm* value of –27 dBFS LAeq. Assuming that the Dolby Digital decoder is operating in Line Mode (dialog reproduced at –31 dBFS LAeq) a program having a *dialnorm* value of –27 (movie) would require 4 dB of attenuation in the decoder to "normalize" the dialog level. In contrast, programming with an actual *dialog* level of -21 and a transmitted *dialnorm* value incorrectly set at –31 would request no attenuation or normalization take place in the decoder. The net result of this is that the two programs are now 10 dB apart upon decoding. This is a good example of how incorrectly setting *dialnorm* on only a few services can affect the perception of those services that have it set correctly. Since some programmers, especially those dealing with legacy content, may never have the need to offer their subscribers multi-channel audio and/or have a system in place to correctly set the audio metadata (in this case *dialnorm*) on a program by program basis configuring *dialnorm* appropriately becomes a challenge.

For this particular broadcaster, if most of their source material is already uniform in dialog level, possibly through the use of processing, but not at the normalized level of Dolby Digital then a "static" or fixed *dialnorm* value may be appropriate. In other words, utilizing *dialnorm* to introduce a fixed offset in the decoder to make up for the offset between the plants operating level and the normalized Dolby Digital level. Implementing a fixed *dialnorm* value, in this case, still allows the programmer the ability to easily make minor adjustments if needed. This practice of setting *dialnorm* to a valid "static" value has been implemented on several services and has been quite successful in providing much closer channel-to-channel level matches while avoiding overload/protection limiting distortions. A critical benefit is that all of this could take place without the need for the programmer to significantly modify their internal operating procedures and practices.

Analog Tiered Subscribers

As stated earlier, many cable programmers are beginning to use digital compression methods to transport their signals to both analog and digital cable systems. The main difference is where the digital content is decoded. For an analog system or tier, the service is decoded to baseband analog in the cable system's headend facility using a commercial Integrated Receiver Decoder – IRD. The decoded signal in analog form is then modulated and combined with other services/channels onto the cable plant for delivery to the subscriber. Since all of the subscribers on this type of cable system (analog) will receive the signals via the RF Tuner input on their television or set top terminal, typical wide dynamic range movies from a premium service with the headend IRD operating in Line Mode may cause subscriber annoyance when comparing this level to non-premium programming. In this application, Line Mode may offer greater dynamic range than what the analog subscriber would need or want. This is due to the differences in operating practices with various programmers with respect to the average dialog levels below full-scale digital.

A more appropriate IRD operating mode for this application (if available) may be RF Mode compression but without the 11dB boost in overall level reflected at the commercial IRDs outputs. This can be accomplished with –11dB of attenuation placed in series with the line level outputs when the IRD is operating in this condition. This mode limits the dynamic range to that which is typically found in traditional analog NTSC programming. This will give the headend technician a better chance at fine-tuning the audio modulator for this service to match that of other processed or restricted dynamic range programming without gross over-modulation. If RF Mode is not available in a particular IRD, the use of Line Mode and applying *dynrng* scale factors of 1.0 for both cut and boost may also provide satisfactory results in many cases.

Set up and calibration of the downstream analog modulator in the cable headend is similar to the set top box operating in RF Mode example covered earlier. For a monophonic analog modulator, two analog signals from the IRD DACs should be combined and fed to the modulator with gain such that identical (i.e. correlated) tones (400Hz) present in both channels of the decoder at 0dBFS yields +6dBFM (200% modulation) as in our set top example used earlier. With the gain set in this manner, programming from the IRD will more closely match other analog services within the cable plant in both dialog level below maximum and program dynamics.

As a side note, many programmers who have begun to include “valid” metadata have witnessed, depending on the material, an attenuation in the decoded level when compared to the level that was presented to the encoder(s). This behavior is normal and is due to the validity of the *dialnorm* value carried in the audio program and Dialog Normalization always being enabled in the decoder. To explain this, keep in mind that previously many programmers may have been encoding all of their programs with a “default” and perhaps an “invalid” *dialnorm* value of –31dBFS LAeq, which required no attenuation take place upon decoding.

Turnaround Uplink Facilities

Another specific decoder operating mode sometimes implemented in professional and commercial IRDs that is useful in some cable headends or a turnaround uplink is Custom Mode. This mode, among other things, allows the professional user to disable dialog normalization in the decoder. Disabling dialog normalization in a turnaround uplink may not offer any advantages since this type of facility will most likely re-encode the audio programming with locally generated Dolby Digital metadata that may or may not be correct. Consider a scenario in which a programmer is correctly setting all of the Dolby Digital encoding parameters, including *dialnorm* at their uplink. This programming is then distributed to both cable headends and turnaround uplinks. Enabling dialog normalization in the IRD at the turnaround uplink will normalize the incoming average dialog levels to –31dB LAeq. Since the incoming programming is now normalized with respect to dialog level, the downstream uplink encoder can employ a *dialnorm* value of –31 (which in many cases is default) and the control over decoded audio levels is determined at the original programmers uplink encoder. All of this assumes that the turnaround uplink facility does not process the incoming audio before re-encoding and that the programmer has set *dialnorm* properly. Consider the same situation described above but with dialog normalization disabled at the turnaround IRD. In this mode (Custom Mode), the expected dialog level normalization will not take place in the IRD because the *dialnorm* value carried in the bitstream is ignored. In turn, any differences in dialog level will be distributed to digital subscribers that exist on the downstream side of the turnaround facility, potentially annoying them. Another potential issue arises when the *dialnorm* value used in the turnaround uplink encoder is set to –31: in some cases presenting un-corrected audio levels (i.e. not at –31dB LAeq) may result in unnecessary limiting when the digital set top is operating in RF Mode.

Another very important point with regard to Custom Mode is that when a decoder is operating in this mode and is performing a downmix (i.e. 5.1 ch. > 2 ch.), 11dB of attenuation is automatically applied. This is due to the possibility of disabling *dynrng* compression and dialog normalization, which would increase the potential for overload during downmix condition. In most cases, the attenuation is compensated for externally in the analog domain, so this behavior is transparent to the user. However, it is wise to verify that this is in fact the case.

As a final note remember that most, if not all broadcasters airing a multichannel audio program will most likely have to replace their existing 2-channel or mono audio program rather than sending both simultaneously. This single multichannel audio program then has to satisfy everyone's listening needs. In most cases, a multichannel audio program will need to be downmixed to stereo or mono not only at the subscriber's digital set top but also in a headend or turnaround facility within the commercial IRD as well. Due to the vast deployment of commercial IRDs that cable networks may have, it becomes imperative that they are all operating in the intended modes as defined by the programmer. Many cable networks may not realize that their decoders are operating in inappropriate modes until the multichannel programming actually starts. It is important to note that this change in audio can affect affiliate cable systems and subscribers differently in each of these cases.

Recommended practices and procedures for commercial decoders are needed throughout this industry and will give programmers and subscribers predictable and consistent results for every type of application.

Conclusion

This paper has begun to outline the issues and opportunities surrounding the deployment of digital programming with an emphasis on the uses of Dolby Digital audio metadata to assist the cable industry in maintaining consistent levels. We have specifically focused on the Dialog Normalization parameter since it has the largest impact on the subscriber. Also provided was an analysis of a reference digital set top box design and how a proper implementation is extremely important. With a better understanding of the overall system, programmers can offer better service to subscribers with the simultaneous benefits of fewer complaints, increased satisfaction, increased customer retention, and best of all, an expanding customer base due to the increased quality of the delivered products.

References

1. Advanced Television Standards Committee A/52, *Digital Audio Compression Standard (AC-3)*, www.atsc.org
2. Society of Cable Television Engineers DVS018, *ATSC Digital Television Standard, Annex B*.
3. Cossette, Stan G & Guillen, Newton G – *New Techniques for Audio Metadata Use and Distribution*, AES Preprint 5044
4. Dolby Digital Professional Encoding Guidelines, Dolby Laboratories Inc.
5. Robinson, Charles Q & Gundry, Kenneth – *Dynamic Range Control via Metadata*, AES Preprint 5028
6. Riedmiller, Jeffrey & du Breuil, Tom

- Audio in Cable, Broadcast, and Satellite Distribution. Issues & Solutions for the Digital Transition via Audio Metadata

Acknowledgements
The author would like to acknowledge and thank Ken Gundry, Tim Carroll, Charles Robinson, Stan Cossette, Oren Williams, Craig Todd, Steve Lyman, Tom du Breuil, Dan Groustra, Ned Mountain, Mike Aloisi, and David Sedacca for their support.

Author contact information
Jeffrey C. Riedmiller
Dolby Laboratories Inc.
100 Potrero Ave.
San Francisco, CA. 94103
415.645.5000

Parameter Name	Description	Change Rate	Purpose
acmod	audio coding mode	Program	Informational
bsmod	bit stream mode	Program	Informational
dsurmod	Dolby surround mode	Program	Informational
lfeon	low-frequency effects channel indication	Program	Informational
dialnorm	dialog normalization	Program	Level Control
compr (“RF” profile)	compression gain word	Frame	Dynamic Range Control
dynrng (“Line” profile)	high rate compression gain word	Block	Dynamic Range Control
cmixlev	center mix level	Program	Level Control
surmixlev	surround mix level	Program	Level Control

Table 1 - Dolby Digital Metadata Parameters