Audio Metadata Demystified

Mike Babbitt
Dolby Laboratories Inc. - San Francisco, CA
mgb@dolby.com
Thanks for listening!!
Mike Babbitt
mgb@dolby.com
What is Dolby Digital?

• Dolby Digital is the encoding technology used to deliver audio to the home within a compliant digital television transport stream.

• Dolby Digital is a scalable technology
  • Data rate can be changed
  • Number of channels can vary from one to six (i.e. mono to 5.1 surround sound).

• Dolby Digital decoders use Metadata.
Dolby Digital nuts and bolts...

- All Dolby Digital data streams carry metadata.

- Metadata is always delivered to the home.
Why do we need Audio Metadata?

- Digital Television can easily deliver better than 100dB of dynamic range
- Cinema-style production techniques are now being used on broadcast television programs
- Consumer audio systems are better than ever
  - Maybe even better than some professional studios
- **BUT**— Not everybody has an expensive home theater system
  - Still lots of stereo and mono viewers
  - “Home Theater in a Box” units are everywhere
- Or the listening environment may be “challenged”
  - Sensitive neighbors
  - Refrigerators, dishwashers, etc. raising ambient noise levels
What concerns viewers most about television audio?

- Loudness variations
  - Different program segments are at different loudness levels
- Ambient noise
  - Late night listening at lower volume makes quiet sections within some programs difficult to hear comfortably
- Home system compatibility
  - Many consumers do not have a 5.1 home theater system, or the equipment’s capabilities are limited

*These are the same issues that viewers had with DVDs.*
Metadata to the rescue!

• Loudness variations? Dialnorm!
  • Helps to correct volume differences
    • between program segments
    • between channels, signal sources

• Ambient noise problems? Dynamic Range Control!
  • Gives a single audio program the capability to deliver both:
    • wide dynamic range for audio enthusiast
    • narrow dynamic range for general listeners

• Home system compatibility? Downmixing!
  • Serve mono, stereo and 5.1 listeners with the same audio bit stream
The “Three D’s” of Metadata

- **Dialnorm (Dialog Normalization)**
  - Describes the average program volume

- **Dynamic Range Control (DRC)**
  - “It’s too quiet…now it’s too loud!”

- **Downmixing (backwards compatibility)**
  - Multiple channels through fewer outputs
The “Three D’s” of Metadata

• Dialnorm (loudness uniformity)
  • Describes the average program volume

• DRC (Dynamic Range Control)
  • “It’s too quiet…now it’s too loud!”

• Downmixing (backwards compatibility)
  • Multiple channels through fewer outputs
Loudness Uniformity

• Level variations are undesirable
  • When switching TV channels
  • Between different programs
  • Between program segments (station breaks and commercials)
Loudness Uniformity

- People know how to set their TV volume
  - listen to the program a few seconds
  - adjust the volume until it sounds right

- People know if the next program is the same
  - listen to the new program
  - readjust volume if necessary

- This judgment is based on “dialog level”
  - people have referenced to “dialog level” since the invention of the Radio and TV
What is the “dialnorm” metadata parameter?

- It’s set by the program producer or the broadcaster
- Defined as the level of normal spoken dialogue with respect to Full Scale Digital (a.k.a. 0dBFS)
- Dialnorm values range between -31 (no level shift in the home decoder) to -1 (maximum level shift in the home decoder)
- Dialnorm also applies to other types of program material, like music videos and concerts
Dialnorm Demystified

Unity Gain

Output Level

Input Level
Dialnorm Demystified

- Output Level
- Input Level
- Program 1
  - Dialogue Level
- Program 2
  - Dialogue Level
- Comfortable Listening Level

Input Level
Dialnorm Demystified

Output Level

Program 1: Dialog Level

Program 2: Dialog Level

Input Level

Comfortable Listening Level

Ouch!
Viewer at home turns down the volume of Program 2 to match the level of Program 1.
Program 1 is now too quiet for comfortable listening, so overall volume is adjusted again.

Program 1
Program 2

Output Level

Input Level

Comfortable Listening Level

Dialnorm Demystified
Rx for Loudness Control
(without metadata)

Repeat as necessary until blisters appear or volume knob breaks
Dialnorm Demystified

Output Level

Program 1: Dialnorm -31
Program 2: Dialnorm -21

Input Level

Comfortable Listening Level
Dialnorm Demystified

Program 1:
Dialnorm -31

Program 2:
Dialnorm -21

The home decoder shifts the output level to match the viewer’s listening level

Comfortable Listening Level

Input Level

Output Level
What is the “dialnorm” value?

- It identifies the area of normal speech within an audio program and provides the ability to “calibrate” the listening environment.
- The decoder uses this parameter to “normalize” program volume to the preferred listening level.
- Where did “-31” come from?
  - Movie dialog ranges from -30 to -25dBFS.
  - Typical movie dialog is around -27, but sometimes even lower.
Dialnorm Demo

Typical Programming

a) without dialnorm

b) with dialnorm
Dialnorm is your friend!

It’s the most misunderstood audio metadata parameter…

- The “loudness defining parameter” in DTV
- Indicates the *average A-weighted level of speech* within a program
- Quantified using the equivalent loudness method – Leq(A) (long-term A-weighted loudness equivalent)
- The **dialog normalization** parameter is used by the decoder to “normalize” the level of different program material and allows the viewer to calibrate their listening environment
- The transmitted audio is **not modified**
- Value is a part of *audio metadata*
- It is **NOT** the Standard Operating Level of your facility..

When used correctly, dialnorm can minimize the average channel to channel dialog/loudness problems and eliminate audio problems experienced with downmixed and RF remodulated content.
The Level of Speech is Important to Listeners

- Listeners adjust their television volume controls in an effort to normalize the speech to a consistent level.

- Research has indicated that listeners agree with each other more consistently when assessing the level of programming containing speech rather than other types of signals.
Note the high correlation among listeners when assessing the level of speech (as compared to “Other” types of signals w/ heterogeneous spectra)

19 out of 21 listeners agreed with each other when judging the level of speech to within 1dB

Whereas, signals that did not contain speech (i.e. footsteps) the listeners disagreed with each other by up to ~12dB!!
Note the high correlation among listeners when assessing the level of speech (as compared to “Other” types of signals with heterogeneous spectra)
Anatomy of an audio signal

- Digital Full Scale
  - 0 dBFS
  - -10 dBFS
  - -20 dBFS
  - -30 dBFS
  - -40 dBFS
  - -50 dBFS

- Reference Level for this program
  - Original Signal: SIGNAL PEAKS
  - Dialog: DIALNORM VALUE = -20

- Average Dialog

DOLBY
Anatomy of an audio signal

Dolby Digital Decoder behavior in Line Mode

- **Digital Full Scale**: 0 dBFS
- **-10 dBFS**
- **-20 dBFS**
- **-30 dBFS**
- **-40 dBFS**
- **-50 dBFS**
- **-50 dBFS**

**DIALNORM VALUE -20**
**OVERALL PROGRAM LEVEL SHIFTED -11 dB**
**IN ORDER TO REPRODUCE DIALOG AT -31 dB**

**PROGRAM LEVEL SHIFTED -11 dB**
**DIALOG LOUDNESS AT -31 dBFS**
*(decoder operating in Line Mode)*

Decoder “normalized” dialog level based on transmitted dialnorm value.
Typical audio signals

MOVIE   TV DRAMA   SPORTS   SYMPHONY   ROCK   NEWS

-40 dBFS
-30 dBFS
-20 dBFS
-10 dBFS
  0 dBFS

AVERAGE DIALOG
SIGNAL PEAKS

27  24  21  20  10  20
Normalized audio signals

At the output of the Dolby Digital decoder

All dialog now reproduced at -31 dBFS
Dynamic Range is Preserved

MOViE  TV DRAMA  SPORTS  SYMPHONY  ROCK  NEWS

-40 dBFS
-30 dBFS
-20 dBFS
-10 dBFS
0 dBFS

AVERAGE DIALOG
SIGNAL PEAKS
Dynamic Range is Preserved

Movie TV Drama Sports Symphony Rock News

0 dBFS

-10 dBFS

-20 dBFS

-30 dBFS

-40 dBFS

Average Dialog

Signal Peaks
Separating Commercials from Content

- 5.1 Program Start
- Stereo and 5.1 Commercials
- 5.1 Program Continues
- Stereo and 5.1 Commercials
- 5.1 Program Continues
- Stereo and 5.1 Commercials
- 5.1 Program Continues
- Stereo and 5.1 Commercials
- 5.1 Program Continues

Stereo and 5.1 Commercials
Commercials and Interstitials during Primetime 5.1 Programming

Speech-based measurements using Leq(A), includes both stereo and 5.1 content

- -10
- -15
- -20
- -25
- -30
- -35
- -40
- -45
- -50

29-Jun
Primetime 5.1 Programming w/o Commercials and Interstitials

Speech-based measurements using Leq(A)

19-Jun, 21-Jun, 26-Jun, 28-Jun, 29-Jun
Commercials and Interstitials during Primetime 5.1 Programming

Speech-based measurements using Leq(A), includes both stereo and 5.1 content

Primetime 5.1 Programming w/o Commercials and Interstitials

Speech-based measurements using Leq(A)
Commercials and Interstitials during Primetime 5.1 Programming

Speech-based measurements using Leq(A), includes both stereo and 5.1 content
The “Three D’s” of Metadata

• Dialnorm (loudness uniformity)
  • Describes the average program volume

• DRC (Dynamic Range Control)
  • “It’s too quiet…now it’s too loud!”

• Downmixing (backwards compatibility)
  • Multiple channels through fewer outputs
Dynamic Range Control (DRC)

• Some people want full dynamic range
• Some do not!
  • listening conditions vary
    • ambient noise problems
    • late night listening
      • the kids are asleep
      • the neighbors are complaining
  • product capabilities vary
    • large home theaters
    • small stereo or mono TVs
The level of audio that falls above the dialog area (as defined by the dialnorm value) is cut. Audio that falls within the dialog area (a.k.a. the “null band”) is unaffected.

The level of audio that falls below the dialog area (as defined by the dialnorm value) is boosted.
Dynamic Range Control (DRC)

- The audio is delivered unchanged
- The viewer’s decoder implements DRC as instructed by the consumer or as determined by the equipment’s capabilities
- The viewer at home determines the appropriate dynamic range for playback
- Dialnorm defines the “null band” where normal speech resides
- DRC metadata defines how much control is required to prevent audio peaks from exceeding the consumer’s threshold
Dynamic Range Control Demo

with Dialnorm set properly
Dynamic Range Control

• Digital coding systems can easily deliver extreme amounts of dynamic range > 100dB (AC-3, aka Dolby Digital)

• The dynamic range needed by different listeners varies
  • Full dynamic range for home theatre enthusiast
  • Narrow dynamic range for portable TV
  • Audio should be sent without processing, but in a form usable for each type of application (headend IRD, STB, etc)

• The AC-3 bit stream includes a dynamic range control element
  • Used by decoders to implement dynamic range limiting
  • Protects 2 ch downmix of 5.1 ch service against overload
  • Use may be altered or defeated under control of the listener
  • The user may enjoy either wide or narrow dynamic range from the same bit stream
Dynamic range control allows...

- Note the dialog level remains consistent!!
Dynamic Range Control (DRC)

Dialnorm value is used as threshold for DRC
Dynamic Range Control (DRC) Benefits

- DRC can increase viewer satisfaction
  - High-level program dynamics are controlled
  - Low-level program dynamics are controlled
  - Narrow dynamic range is sometimes preferred
    - ~15dB dynamic range essential to intelligibility in a typical domestic environment
- Need to fit with existing programming & practices
- Some of the people some of the time want the full dynamic range - non-destructive!
- Pre-processed material can be passed through without additional processing via the “Null Band” in DRC algorithm
A properly normalized audio signal
The output of a Dolby Digital decoder in the home

Decoder “normalized” dialog level based on transmitted dialnorm value
Dialnorm Operation
Mandatory behavior within a consumer Dolby Digital decoder

Program 1
Speech level = -31
Dialnorm = -31
Output = -31

Program 2
Speech level = -20
Dialnorm = -20
Output = -31

Program 3
Speech level = -27
Dialnorm = -27
Output = -31

Program 4
Speech level = -15
Dialnorm = -15
Output = -31
A properly normalized audio signal

The output of a Dolby Digital decoder in the home

Decoder “normalized” dialog level based on transmitted dialnorm value
What happens when Dialnorm is set wrong?

Decoder “normalized” dialog level based on transmitted dialnorm value
Dialnorm Operation
Mandatory behavior within a consumer Dolby Digital decoder

Program 1
Speech level = -31
Dialnorm = -31
Output = -31

Program 2
Speech level = -20
Dialnorm = -31
Output = -20

Program 3
Speech level = -27
Dialnorm = -31
Output = -27

Program 4
Speech level = -15
Dialnorm = -31
Output = -15

Dialnorm does not match measured level of speech
Output Level of Speech
What happens when Dialnorm is set wrong?

Decoder “normalized” dialog level based on transmitted dialnorm value
DRC Demystified

Input Level

-31 dialnorm

Output Level

Dialog

Input Level
Dynamic Range Control

Demo

with Dialnorm set improperly
The “Three D’s” of Metadata

- Dialnorm (loudness uniformity)
  - Describes the average program volume

- DRC (Dynamic Range Control)
  - “It’s too quiet…now it’s too loud!”

- Downmixing (backwards compatibility)
  - Multiple channels through fewer outputs
Downmixing

From Full, 5.1 Channel Home Theaters

To small Stereo or Mono TVs
3 Types of Downmixing

- **Surround downmix**
  - Lt / Rt
    - Left total / Right total, for Pro Logic or Pro Logic II decoding

- **Stereo downmix**
  - Lo / Ro
    - Left only / Right only, for headphones or stereo televisions

- **Mono downmix**
  - From Lo / Ro, for mono televisions
Downmixing process

• Dialnorm value defines “null band” where normal speech resides
  • The “null band” is not affected by DRC

• DRC controls peaks and prevents clipping when channels are summed
  • The level of signals above speech are lowered in level
  • The level of signals below speech are raised in level
  • Predicted overload conditions are prevented

• Downmix metadata defines how the channels are summed
Receiver downmixing options

• All channels in the original audio program are delivered to the viewer untouched

• Product or user determines downmix style
  • The viewer may listen to the program in any realistically conceivable audio or home theater environment from mono to full-range 5.1

• Downmix metadata defines how the channels are summed
The “Three D’s” of Metadata

• They all work together:
  • Dialnorm identifies where dialog is placed and allows the decoder to normalize program volume to a level selected by the viewer
The “Three D’s” of Metadata

• They all work together:
  • Optional consumer DRC modes raise the audio level below, and lower the audio level above the dialnorm range
The “Three D’s” of Metadata

- They all work together:
  - **Downmixing** uses DRC to prevent clipping when fewer speakers are used

From full-blown 5.1 home theaters to small mono televisions, everyone receives the same signal!
The “Three D’s” of Metadata

- They all work together
Conclusion

• Speech levels are important to:
  • Your viewers
  • And STB decoders, too!

• Measure loudness not level (i.e. Leq, not VU or peak)

• Check dialnorm on Dolby services

• Measure your channels
Loudness and Audio Metadata Demystified

• Documents and guidelines available at www.dolby.com

• Technical information and software downloads for your Dolby broadcast products are available at: www.dolbysupport.com
Thanks for listening!!

Mike Babbitt

mgb@dolby.com