

Audio Metadata: Key For Producer Creativity and Tool for Broadcaster Flexibility

HPA Technology Retreat
February 24, 2006

Kenneth Hunold
Broadcast Applications Engineer
Dolby Laboratories, Inc.



CONFIDENTIAL INFORMATION

Why Is There Audio Metadata?

- A bit of history
- What about NTSC audio can be improved?
- How do we address these perceived shortcomings?

NTSC Television Audio

- Sound modulation changed from AM to FM in 1941
 - Even though FM is capable of “High Fidelity,” TV audio is not regarded as a high-quality medium
- TV sets have small speakers, sometimes poorly positioned for clarity and intelligibility
- Prior to the 1970’s, Network TV audio distribution was typically done over “5 kc lines”
 - Outside of New York, audio was decidedly “tinny”

NTSC Television Audio

- TV stations operate their audio systems as close to 100% modulation as possible
 - “Louder is better”
 - Made it easier to hear in the home
 - *Consistent loudness across the dial*
- All audio programming was processed with AGC circuits to maintain the highest modulation
- Even though audio processing improved over the years (for TV and FM) everyone continued to operate their audio system as close to 100% modulation as possible

NTSC Television Audio

- As long as all TV signals were delivered via broadcast channels, “loudness” was fairly consistent among all sources.
- Today, most TV programs get to the viewer via CATV systems
- Programming arrives at CATV head ends from a number of sources via a number of different paths.
 - Discrete audio and video
 - Satellite (analog and digital)
 - ATSC broadcast
- To compound this situation, television programs are often mixed without a “standard” for loudness

Audio Challenges For DTV

- Now we can deliver *up to* six channels of audio (5.1)
- Program Loudness
 - Cornerstone of consistency and compatibility
- Monitoring Setup
 - Speaker placement and SPL
- Downmix Compatibility
 - Simultaneous 5.1 and Stereo delivery
 - Stereo-compatible 5.1 mixes

Audio Mixing for Television

- There is no universal monitoring level
 - Film mixing is standardized to 85 dB SPL (C-weighted, slow response) at reference level
 - Television is generally mixed at a lower SPL, often between 79 and 82 dB
- As a result, audio levels and the resultant loudness vary from station to station and program to program
 - This is not always due to carelessness or inattention to detail
 - Sometimes these levels are adjusted in Master Control, or are left to be corrected by the audio processing equipment

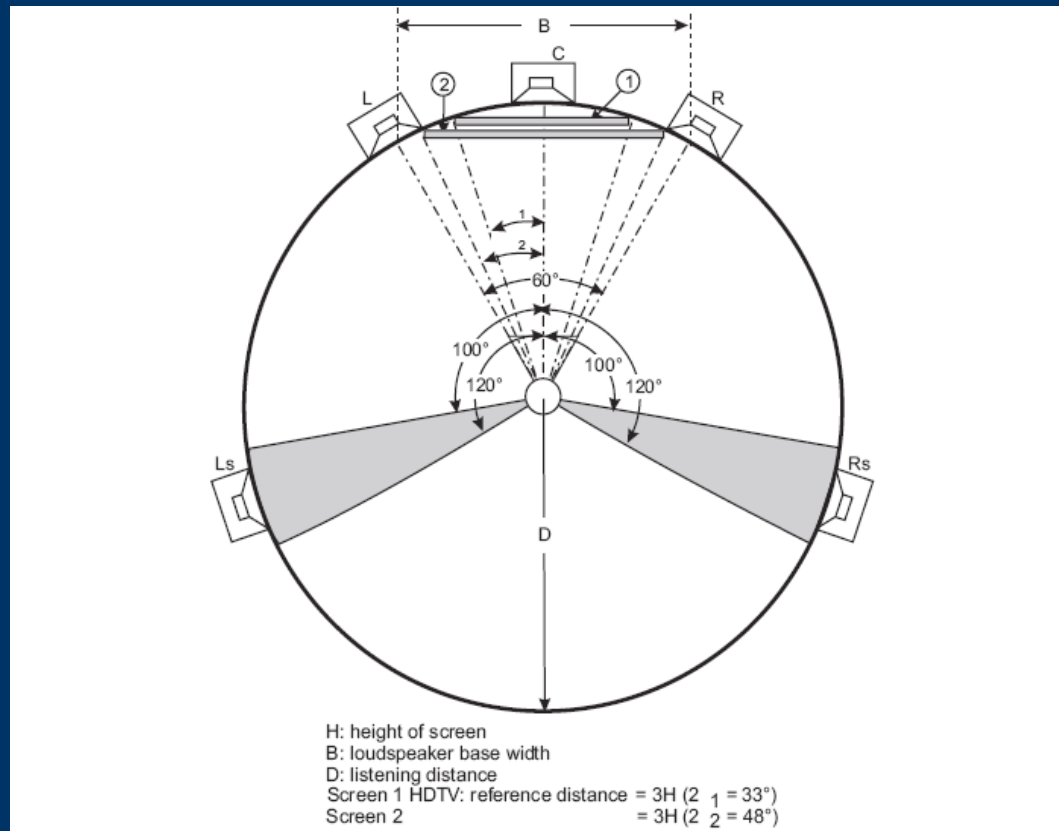
Audio Mixing for Television

- Digital audio delivery systems have greater dynamic range than earlier analog systems
 - Systems with 96 db dynamic range (16 bit systems) do not benefit from crowding everything at the highest end of the dynamic range (0 dBFS)
 - Unfortunately, that is exactly what is being done on audio CDs!
- TV Engineers have developed standards that allow up to 20 dB of headroom above the “line-up level” (SMPTE RP155)
 - There is plenty of room below the line-up level for program dynamics.
 - This is good *and* bad!

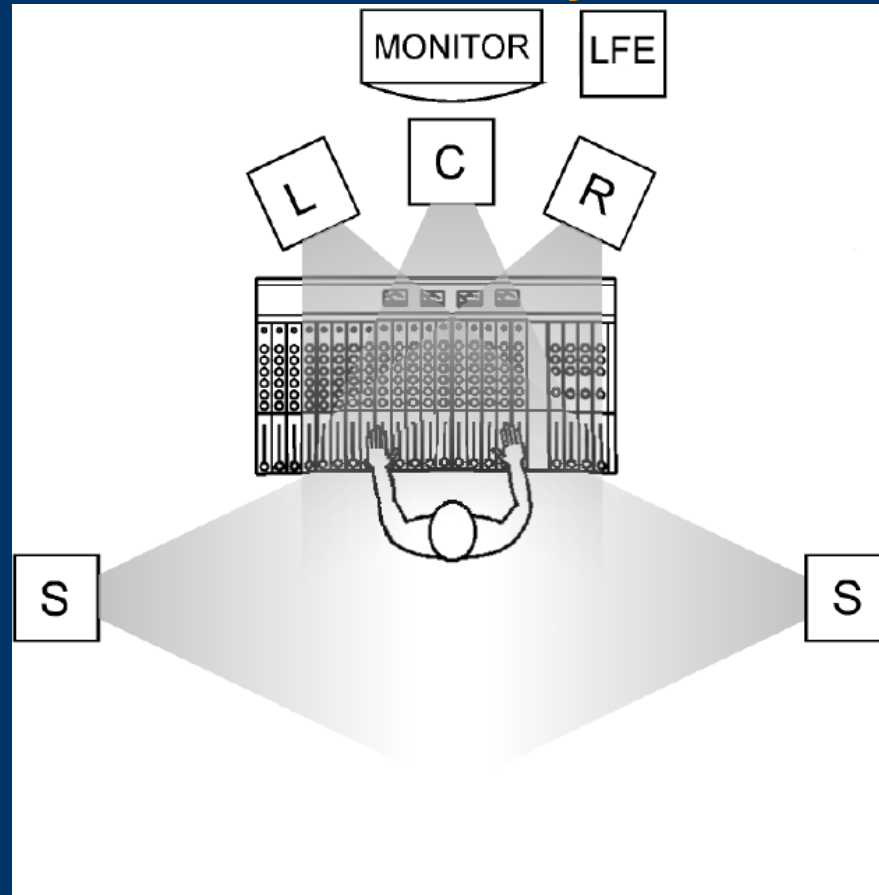
Audio for DTV

- In DTV, the feed for the DTV transmitter is often taken before any audio processing is applied
 - This is *also* good and bad!
- It allows us to step back in the audio system and avoid making a “one size fits all” program, tailored to the lowest common denominator of listener
 - But there is no gain without pain, and gain changes affecting dialogue often result in pain!

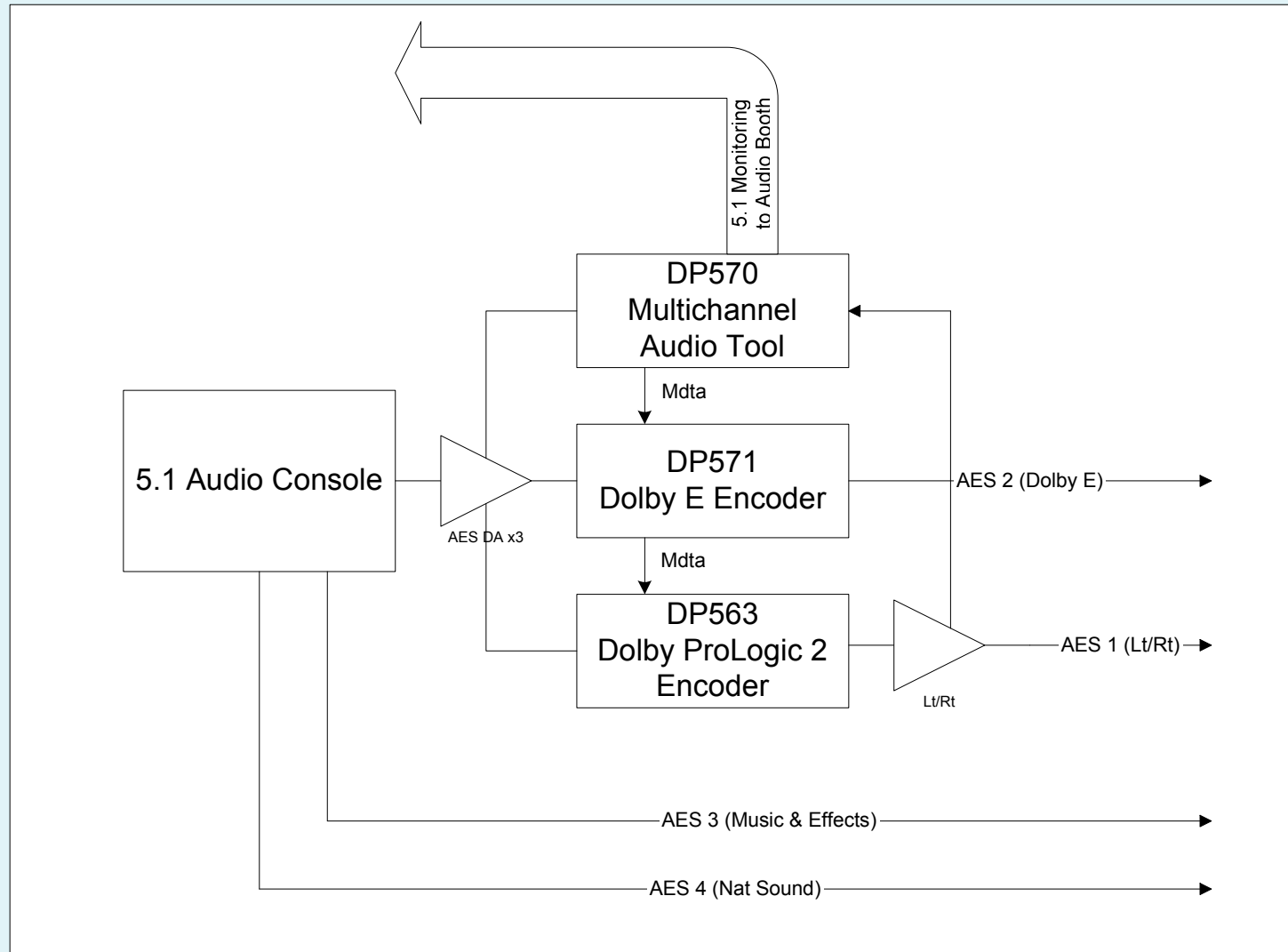
Reference Loudspeaker Placement



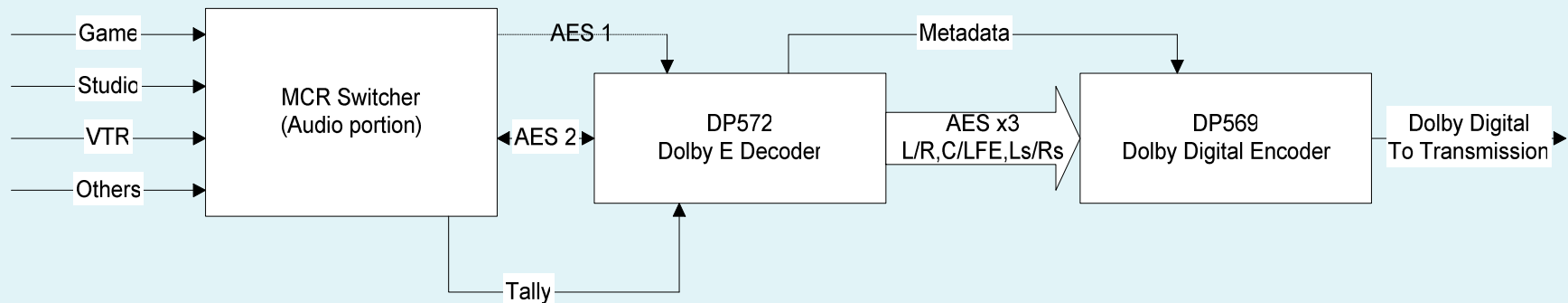
Typical 5.1-Channel Room Layout



“Generic” 5.1 Audio Room



Broadcast Center Master Control (Audio)



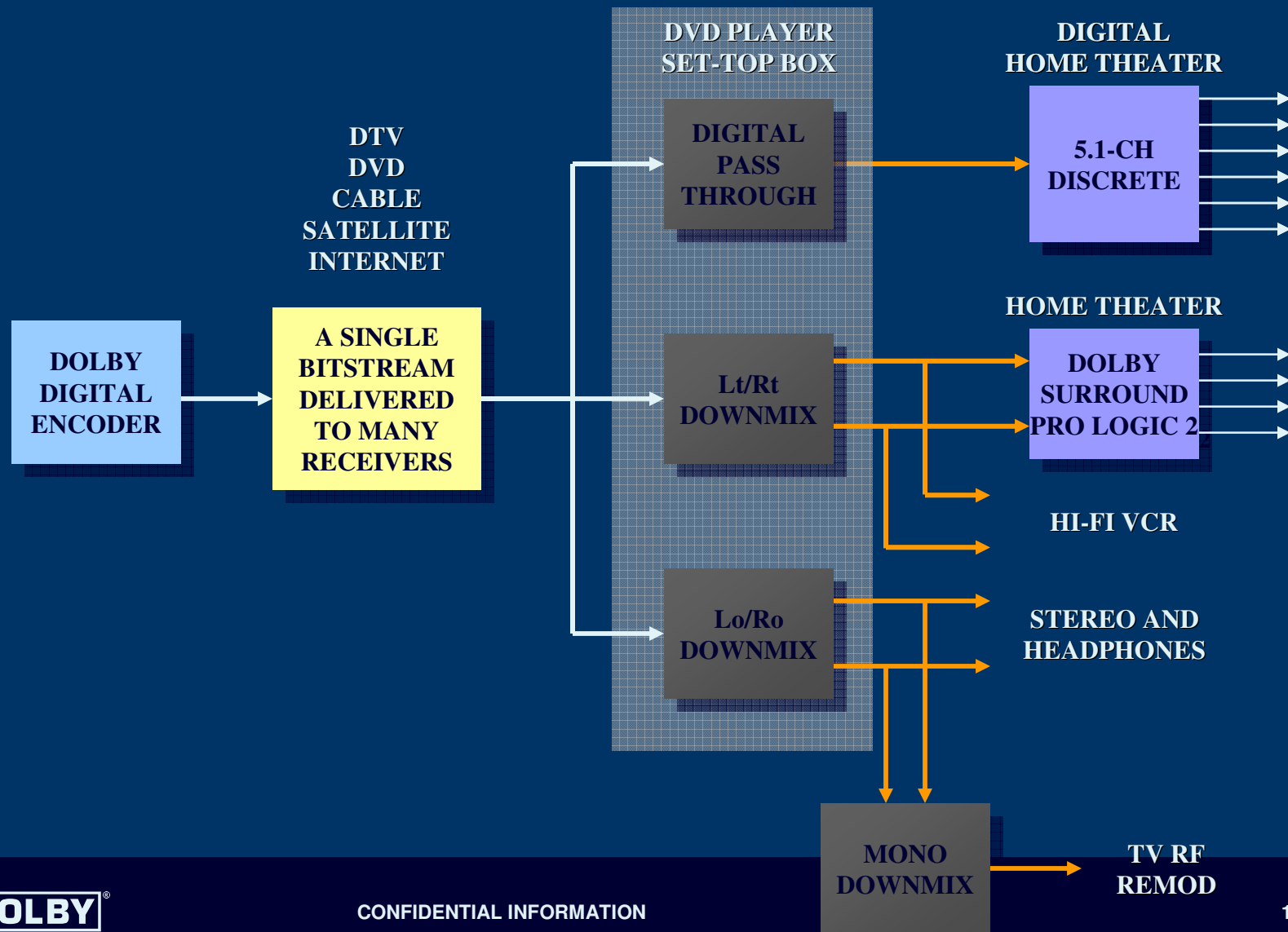
Mezzanine Audio Compression

- Live HD program backhaul of six uncompressed audio channels (data compression) is very rare
- Bit rate reduction of audio signals is used to make room for more video data
- All North American broadcast networks use Dolby E for live program audio backhaul.
- Dolby E provides 8 channels of audio *and audio metadata* in a single AES “pair”
- In the Broadcast Center, audio and metadata can be decoded and routed discretely, if desired (but use caution...)

Downmixing Compatibility

- Not everyone will hear a 5.1-channel production “in 5.1”
 - Viewer listening conditions vary
 - The quality and capabilities of consumer’s equipment varies
- Examples include:
 - Built-in TV speakers
 - Late Night listening (spouse, children, and/or neighbors are asleep)
 - “Channel 3” RF output to TV
 - Oh yes, high quality Home Theatre
- Audio control room monitoring is very important!
 - “What you hear is what you get”
 - Mixers need to hear what the viewer will hear (in as many different situations as they choose to check)

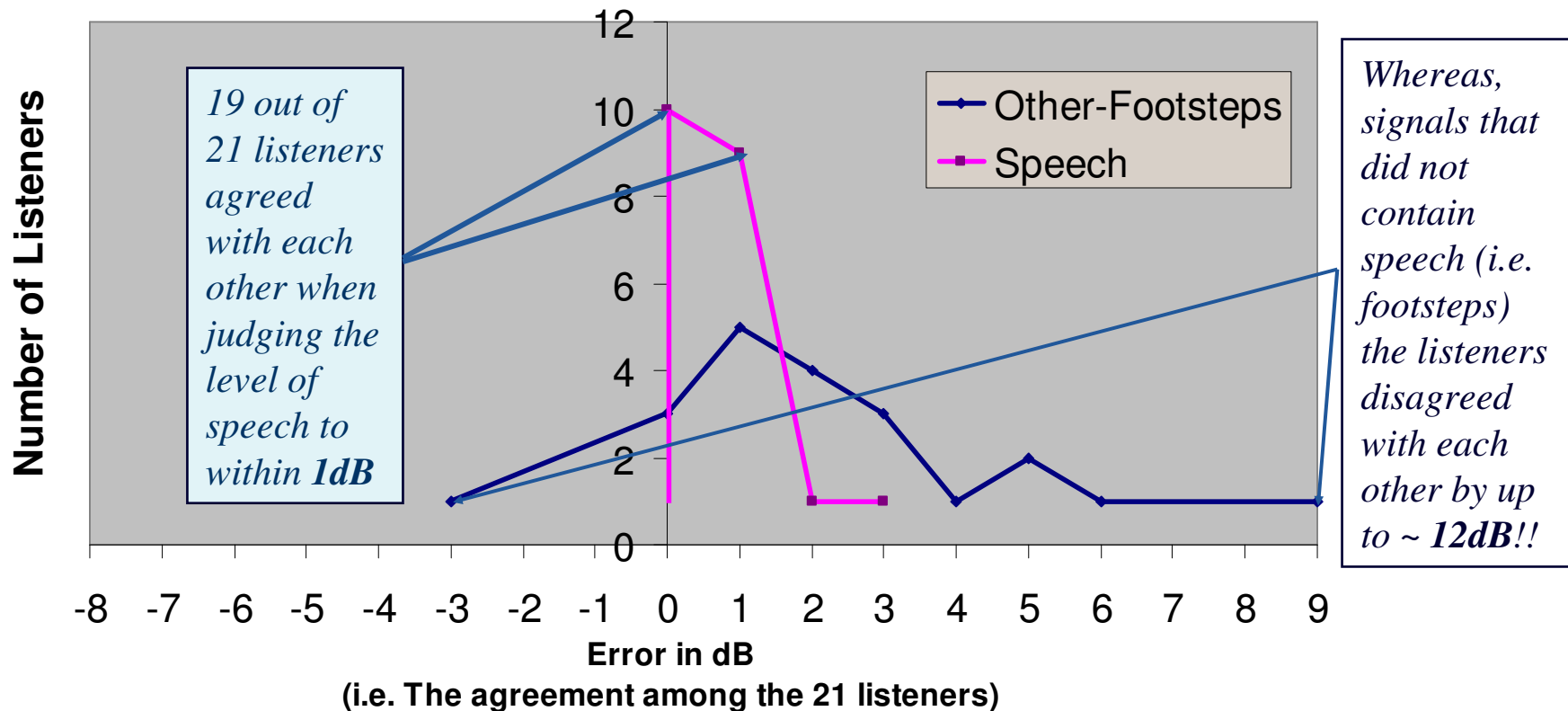
Downmixing possibilities



Loudness Measurement

- The loudness of dialogue is used as the metric by which a viewer judges the program's overall loudness
 - Research has shown that listeners are very consistent when judging the level of speech
 - Listeners are not as consistent when judging the loudness of other sounds (music, sound effects, etc.)
- dialogue loudness forms the “center of gravity” of a mix
 - Listeners will adjust their volume control so that the announcer or actor is clearly and comfortably heard
- The loudness of dialogue is used match the perceived loudness of programs *without utilizing AGC*

Speech vs. Other (Footsteps) - Correlation Histogram

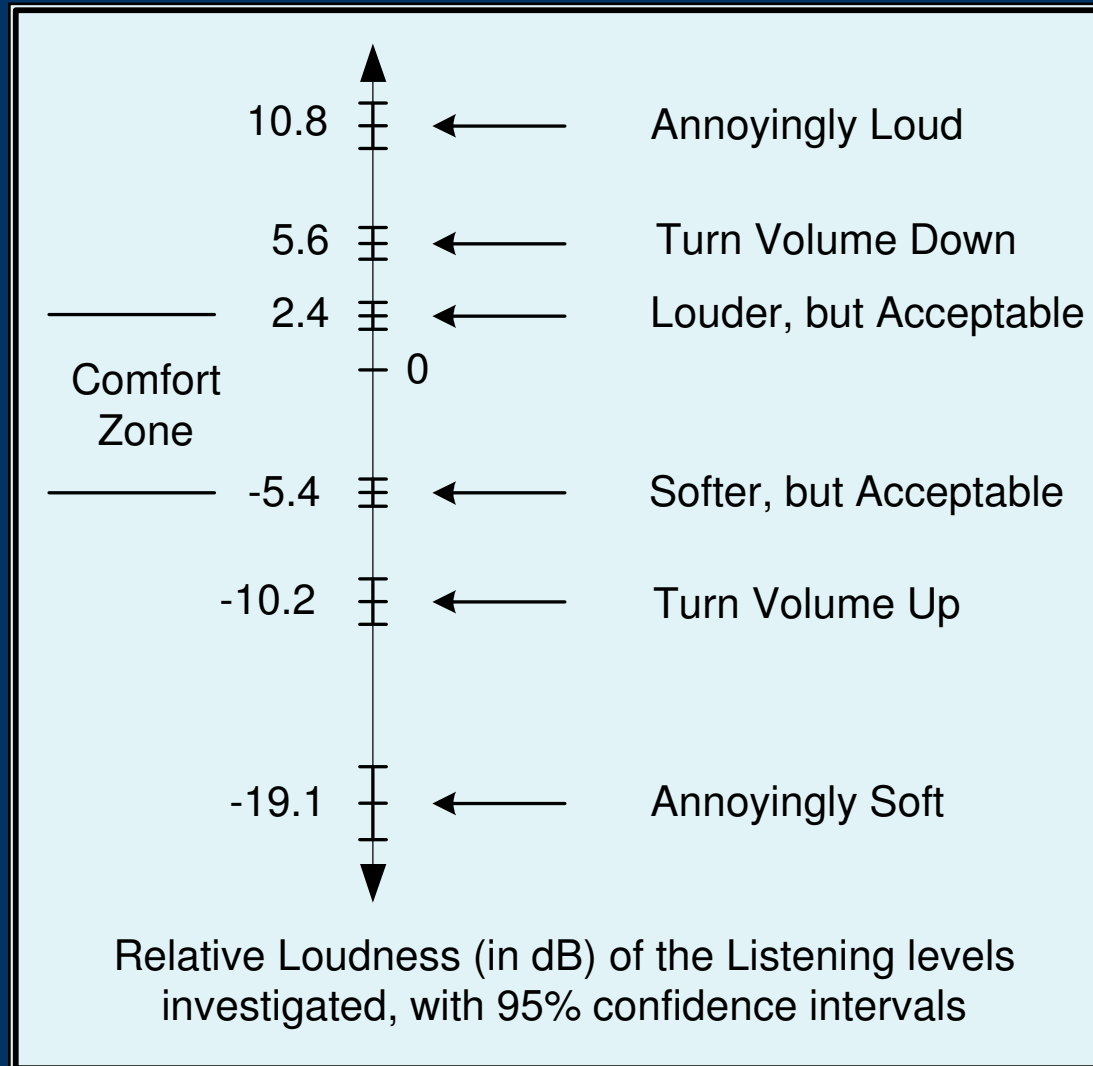


Note: The high correlation among listeners when assessing the level of speech (as compared to “Other” types of signals w/ heterogeneous spectra)

LM100 Broadcast Loudness Meter Display



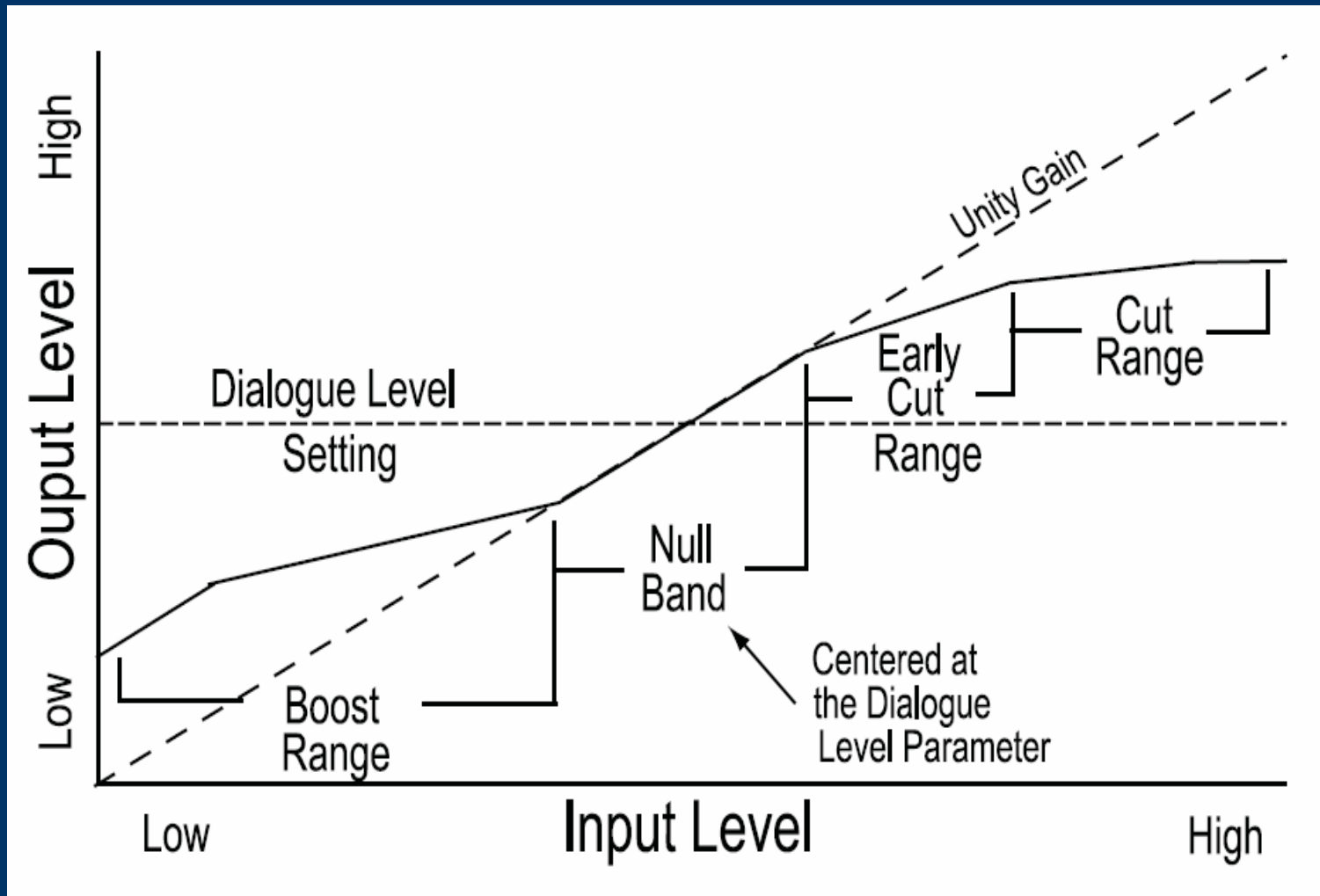
Comfort Zone Test Results



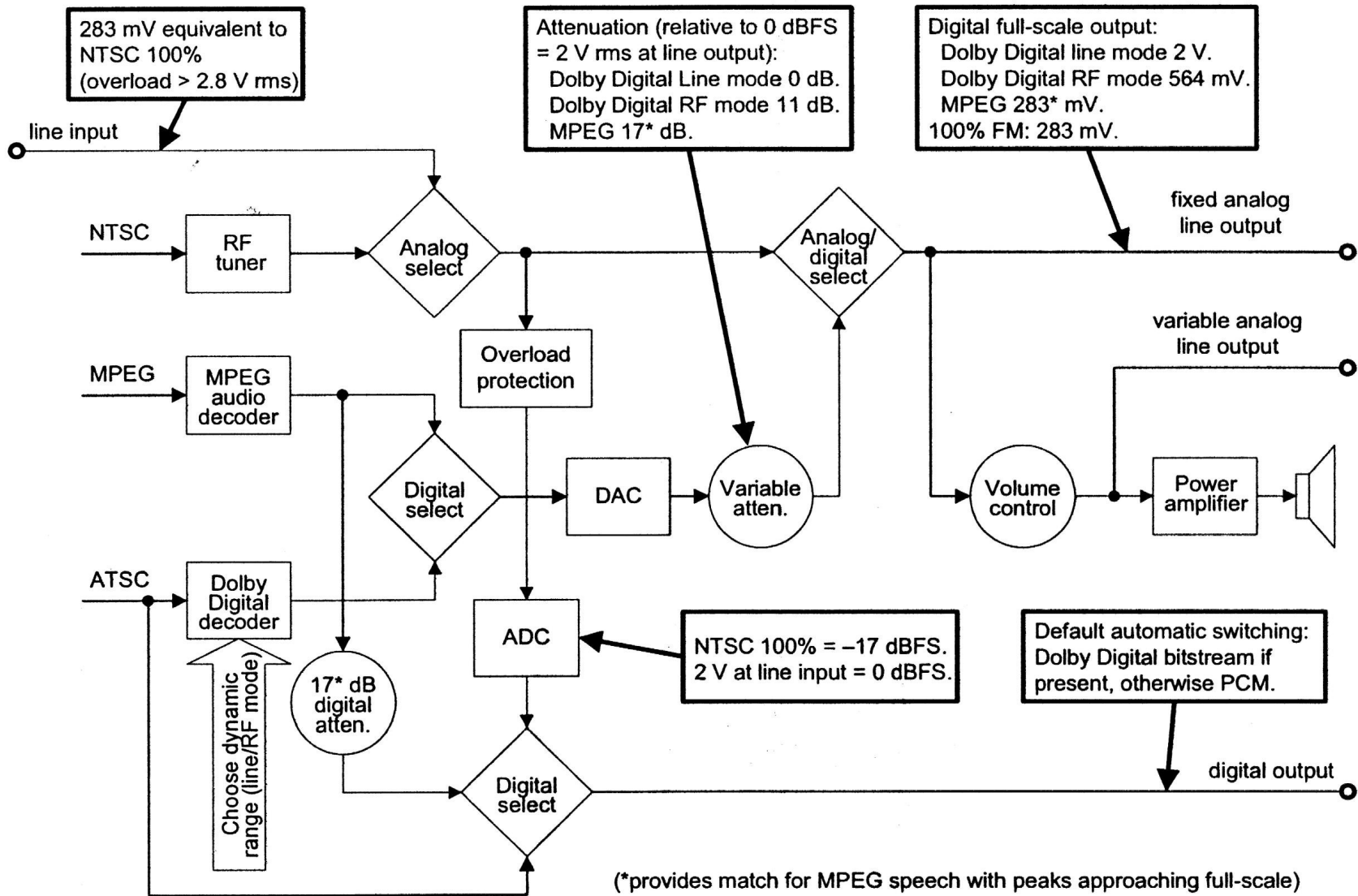
Dynamic Range Profiles

- The dynamic range of a program can be restricted by the listener via the use of Dynamic Range Profiles.
 - Reduce very loud portions of the program to protect equipment or sanity
 - Raise very quiet portions above the noise floor in a typical home
 - Air conditioning noise, refrigerator noise, outside traffic, etc.
- If the consumer has a nice (and quiet) home theatre, dynamic range control can be turned off, allowing the listener to hear the program exactly as mixed.

Dynamic Range Profiles



Set-top Box Block Diagram



Downmix Settings

- Artistically, producers often strive for the most exciting 5.1 mix that they can get
- “Effects” can sometimes be overpowering when this 5.1 mix is heard in stereo
- The amount of surround and center channel contribution to the stereo mix can be controlled *independently from the 5.1 mix*
- The stereo mix is adjusted via metadata, instead of by compromising the 5.1 mix that was originally created

Creating/Deciphering the Audio Delivery Spec

- Producers and Programmers live and die by the Program Delivery Specification Document
- Delivery Specifications currently lack consistency in terminology
 - Sometimes, terms in the document appear to be contradictory
- Some audio metadata parameters are dictated by the broadcaster
- Other audio metadata parameters should be left to the artistic community and passed to the viewer by the broadcaster
- “Trust, but verify”
 - Quality control includes many of the parameters listed in the delivery spec, and this should be expanded to include audio metadata

Broadcaster-specified Audio Metadata

- Input channel configuration
 - L/R, C/LFE, Ls/Rs
 - Conforms to SMPTE 320M
 - Required layout for Dolby E
 - “Channels 7/8”
 - Second-language SAP
 - “International” or Natural Sound
- Channel Mode
 - 3/2L, 2/0, 1/0, etc.
 - 5.1, 5.1+2, 5.1+2x1, etc.
- Dolby Surround Mode
 - Can turn on Pro Logic decoders during 2-channel programs

Broadcaster-specified Audio Metadata

- Dialogue Level (Dialnorm)
 - Strictly speaking, this does not need to be set to a particular value, as long as the dialogue level matches the Dialnorm parameter setting
 - If it *is* specified, the dialogue loudness must be measured and the level adjusted so that the measured loudness matches the Dialnorm parameter
- Dialnorm is a metadata parameters that **must** be accurately set
- From ATSC A/53D, Annex B, Section 5.5, Dialogue Level:
 - “The value of the dialnorm parameter in the AC-3 elementary bit stream shall indicate the level of average spoken dialogue within the encoded audio program. Dialogue level may be measured by means of an “A” weighted integrated measurement (LAeq). (Receivers use the value of dialnorm to adjust the reproduced audio level so as to normalize the dialogue level.)”

“Artistic” Metadata Parameters

- Dialogue Level – two ways to work:
 - Mix the way you wish, and tell the system what you did by setting Dialnorm to match the resulting dialogue level
 - Start with a specified Dialnorm value, and mix the program so that the measured value matches the specified value
- DRC Profile
 - Choose the profile that “sounds best” for your program
 - Check the mix with both DRC profiles (“Line” and “RF”)
 - Avoid using the “None” profile
 - Viewers will not be able to enable dynamic range compression
 - If Dialnorm is not set correctly, it will be impossible for your program to match the loudness of other programs

“Artistic” Metadata Parameters

- Downmixing preferences
 - The amount of surround channel information heard by stereo or mono listeners (-3 dB, -6 dB, or none)
 - The amount of center channel information heard by stereo or mono listeners (+3 dB, -3 dB, -6 dB)
- Copyright bit
 - Indicates whether the program is protected by a copyright
- Original Bitstream
 - Indicates whether the bitstream is an original or a copy

Mix Facility-specific Metadata (Pre-processing)

- Surround 3 dB Attenuation
 - Mixing rooms are calibrated differently for film mixing as compared to TV/DVD mixing
 - In Film, the surround channels are calibrated 3 dB lower in SPL than the front channels
 - In TV/DVD, the surround channels are calibrated at the same SPL as the front channels
 - If TV/DVD programs are mixed in a room calibrated for Film mixing, the Surround 3 dB Attenuation must be ON to create a proper TV/DVD mix
- LFE Filter
 - The LFE channel must be low-pass filtered at 150 Hz
 - This filter ensures that the LFE channel does not contain frequencies above 150 Hz

Conclusions

- Audio Metadata has been shown to be almost as important as the audio itself
- Proper use of metadata by the program producer results in the faithful reproduction of the program (as the Producer intended) over the widest possible range of listening conditions
- Proper use of metadata by the broadcaster results in a consistent presentation of the program and serves the needs of the broadcaster by automatically controlling certain audio functions and reducing the number of viewer complaints
- Programs can be monitored during the production and broadcast for compatibility, and with consistent results at each stage