



IEEE P1722a – Three Discussion Topics about Transporting Audio + Metadata via AVB

Ethan Grossman
Staff Engineer, Dolby Laboratories, Inc.
December 9, 2013



Three Discussion Topics – Overall Summary

1. Transmitting Dolby E (and SMPTE 337M) over AVB

- Dolby E is a coded audio (i.e. non-PCM) format used by the Broadcast industry
- SMPTE 337M is a format for sending non-PCM data over AES3
- Question: How to send Dolby E (and any other coded audio) over AVB?
- Discussion: It turns out that IEC 60958 in 1722-2011 does work but is not well documented. Propose to create annex to explain – need exact wording.

2. Specifying Channel Configuration for multi-channel audio in AVTP Audio

- Question: Is the current spec correct, necessary, sufficient?
- Discussion: A single “correct” 5.1 mapping standard does exist. Propose to modify current spec accordingly, further attempts at channel config fall under topic of general metadata (topic 3)

3. Transmitting Metadata with PCM Audio over AVB

- Question: How to ensure that paths exist to send and associate MD with audio over AVB
- Discussion: We identify 3 ways to transmit metadata over AVB – propose to draft an annex explaining strategy for each, and add (apparently few, minor) modifications to P1722a

Transmitting Dolby E and SMPTE 337M over AVB



- Dolby E is a coded audio format used by the Broadcast industry
 - Coded audio: 8 channels PCM plus Metadata, compressed to fit in 2ch PCM BW
 - Metadata: Data fields that tell decoder how to play back that PCM frame
 - Examples: channel config, sample rate, surround mix level, copyright, dynamic range...
 - Frame: Dolby E frames are aligned on video/film frame boundaries
 - Dolby E is “always” transmitted via SMPTE 337M (non-PCM “bursts” on AES3)
- IEC 60958 as cited in 1722-2011 does work, but not well documented
- Proposal: **Create an annex to P1722a to explain:** IEC is for consumer, SMPTE is for professional use
 - IEC 60958-1 (2008) Sec 5.3: first “Channel Status” bit set indicates “Professional use”
 - “Consumer use” setting (cs bit not set) explicitly refers to IEC 60958-3, IEC 61937
 - Professional “implies” SMPTE 337M (family) but not “formally” since SMPTE not an international standard
 - To support current and future formats, implementations of SMPTE 337M should support 24-bit variant of SMPTE 337M (not used yet so sometimes omitted by implementations, soon to be used for metadata)
 - Data Type values are different between IEC and SMPTE
 - Format can differ: Example: DD+ formatted in 16-bit samples for 61937 but 20-bit for 337M

Specifying Channel Configuration in AVTP Audio



- Question: How should channel configuration be spec'd for AVTP Audio?
- Discussion: There is a single "correct" channel order for 5.1
 - Amendment 1:2012 to SMPTE 2035-2009, Table 2 – 12-Track Audio Channel to Audio Track assignments by Audio Program type, e.g. entry 13d (L,R,C,LFE,Ls,Rs)
 - Beyond that, there is no single spec we can reference.
 - Solving this for the general case falls under the category of "audio metadata" (topic 3)
 - Propose to **change P1722a draft 6 Table 8.4. 5.1 Channel Layouts (from L,R,LFE,C,Ls,Rs) to match SMPTE layout above**, and not attempt to do any more than that.
 - Additional related specs and options are reviewed in Appendix A.

Transmitting Metadata with PCM Audio over AVB



- Background
 - Analogous to Axon proposal about separating MD from Video
- MD increasingly used for many purposes in audio delivery
 - Channel layout beyond standard 2 and 5.1 standards (topic 2)
 - x,y,z position info per track for live rendering e.g. Dolby Atmos
 - Interactive applications e.g. internet broadcast to home (user responses, etc.)
 - “Perceived loudness” management in broadcast chain
 - Etc.

Transmitting Metadata with PCM Audio over AVB



- Issues with current MD implementation
 - MD is bound to coded audio formats e.g. Dolby E, DD, DD+, AAC, etc.
 - Limits use generally
 - Many encode-decode steps e.g. in broadcast workflow
 - As soon as MD is extracted from coded audio stream, loses association with PCM
 - No system for keeping it sample-accurate aligned to PCM stream, e.g. in complex broadcast path
 - MD in current formats is not “secure” (signed) – has checksum but audio or MD can be modified and checksum regenerated – can’t differentiate from original
 - Example: Enforcement of “CALM” act (loudness) - can’t verify compliance of a given stream – thus to be dead sure it complies, must reprocess – extra step, and degrades audio if it was compliant
 - MD in current formats is not “extensible” – fixed set of values
 - MD is not consistent in value or format – different for various codecs

Transmitting Metadata with PCM Audio over AVB



- MD subject of standardization efforts, but not done yet
 - ITU Annex 6 to Document 6B/128 - Specification of an audio definition model - a proposal for flexible multichannel audio metadata (WAV file chunk based –intended for distribution format, not necessarily for streaming)
 - ETSI TS 102366 – Annex H - Extensible structure for delivery of metadata
 - Currently specified in the context of AC-3, Enhanced AC-3, but not stand-alone

Transmitting Metadata with PCM Audio over AVB



- Goals for P1722a support of MD for audio
 - Don't assume we can fully specify MD packaging (not as easy as Axon case)
 - Understand at least the three use cases presented below
 - Ensure we can support those use cases

Transmitting Metadata with PCM Audio over AVB



Use Cases (1/3)

1. MD is streamed as 337M (either as part of coded stream or separately)
 - As noted in topic 1, this is already supported in 1722-2011
 - Do devices generally know to turn off Sample Rate Conversion for 337M?

Transmitting Metadata with PCM Audio over AVB



Use Cases (2/3)

- MD is streamed as AES3 or AVTP Audio samples (i.e. not 337M)
- Affects both 1722-2011 AES3 path and AVTP Audio
 - Data path must be “bit accurate” – what goes in goes out (**any holes in this?**)
 - Addition to P1722a: **Need to designate a bit to flag PCM stream as “not suitable for decoding with DAC” (or some other wording).**
 - Processors should (shall?) automatically turn off SRC when this bit is set
 - Incidental Recommendation for AVTP Audio
 - **Include 20-bit format** (“optimal”)
 - Why is 32-bit included?



Use Cases (3/3)

- MD is sent as an opaque block based on an OUI
 - Use TSCS for this like Axon proposal?
 - Don't define a 1722 OUI yet because no standard to point to as container yet
- Question: Does it make sense to define a stream that includes both Audio and MD, in order to keep them associated with each other? Or do separate streams make more sense? (See next page)
- Propose to **create annex to P1722a to explain how to implement each of these 3 use cases along the lines described above.**

Transmitting Metadata with PCM Audio over AVB



- Possible Audio + MD Stream Structure



- Number of samples per sample block can be different e.g. Dolby Digital frames have 1536 samples (i.e. not related to picture frame as with Dolby E)
- In Dolby E, the MD block applies to approx 1 (video/film) frame period
- A MD block can contain one or more MD fields that apply to that same frame
 - Each field has start offset and duration (in samples) (field is 11 bits, i.e. 2k samples)
- Number of samples per frame changes with picture format, including drop-frame cases in which frame contains non-integral number of samples (e.g. 1601 to 1602 to match 29.97fps)

Transmitting Metadata with PCM Audio over AVB



- Structure of Audio Sample Block
 - Use AVTP Audio syntax in header?
 - How to delineate (Video/Film) frames?
 - Should audio and MD be separate streams?
 - Desire to keep audio + MD tightly associated
 - Could be separate streams since timing is good, have IDs for each – need “signable” association
- Structure of Metadata Block
 - See Appendix B for an example of a container structure

Appendix A: Other Channel Configuration Options



- EBU R 123 (2009) “Track allocation for File Exchange”
 - has a big table of possible mappings (key is like “8a”)
 - Issues: Fixed channel maps, dated, European broadcast specific.
- SMPTE ST 2035 (2009 , 2012) – “Audio Channel Assignments for Digital Television Recorders”
 - similar table with similar key
 - Issues: Fixed channel maps, dated, American broadcast specific.
- SMPTE ST 377-4-2012 – “MXF Multichannel Audio Labeling Framework”
 - Issues: Complex, does more than we need.
- SMPTE 32NF – “SDI Audio Channel Labeling” (work in progress)
 - Issues: Don’t have full details, may not be done yet.
- Do our own using an array
 - Define a large table of individual channel “names” (“L”, “R” etc.) (Dolby has a table, not standard)
 - Have the header specify an array of indexes into that table to describe mapping of each channel.
 - Issues: Uses more space. How often to update?
- Do our own using a set of fixed layouts and a simple index into that set
 - Issues: the possible layouts we provide may not work for everyone
- Remove what we’ve done
 - Maybe replace with informative (non-normative) text?

Appendix B: Example of MD formatting

ETSI TS 102366 – Annex H - Extensible structure for delivery of metadata



- The data structure consists of a container carrying one or more data payloads
 - Each payload has a unique payload ID, and the payloads can be sequenced in any order
 - Payload Config contains start offset, duration, etc.
 - Protection data follows the final payload in the container
 - Allows verification that the container and the payloads within the container are error-free.
 - Allows “signing” the MD + Associated Audio as being “untampered”

