Welcome!
2018-08-16

SMPTE Standards Update Webcasts

• Series of quarterly 1-hour online (this is 90 minutes), interactive webcasts covering select SMPTE standards
• Free to everyone
• Sessions are recorded for on-demand viewing convenience SMPTE.ORG and YouTube
• Please tell others about SMPTE’s webcasts on social media:
  @smpteconnect
  #SMPTEWebcast
Your Host

Joel E. Welch

Director of Education
SMPTE

Today’s Guest Speaker

Stan Cossette

Senior Staff Engineer
Dolby Laboratories
SMPTE ST 2098-2 Immersive Audio Bitstream – An overview

Presented By Stan Cossette

Presentation Overview

• Background and Terms
• High-level bitstream frame and element structure
• Hierarchy of the bitstream frame elements
• Audio Object coordinate system
• Extensibility, backward compatibility, and efficiency mechanisms designed into the standard
• Examples of bitstream usage
• Bitstream packaging in Cinema and IMF
Background and terms

SMPTE 25CSS Immersive Audio Suite

ST 2098-1 Immersive Audio Metadata
ST 2098-2 Immersive Audio Bitstream
EG 2098-3 Immersive Audio Renderer Expectations (in WD development)
RP 2098-4 Immersive Audio Renderer Interoperability Testing Procedure (in WD development)
ST 2098-5 D-Cinema Immersive Audio Channels and Soundfield Groups

Scope of ST 2098-2 is generic ‘...defines a coded representation (bitstream) of an audio program’, but the intended application is Digital Cinema. This affected:
- Choice of supported channels
- Choice of supported speaker configurations
- Assumptions regarding playback environment

Immersive Sound vs Surround Sound

From ST 2098-5:

**Immersive Sound**
Includes sound that emanates from sources at and beyond the Base Layer.

**Base Layer**
Refer to the nominally horizontal layer of Loudspeakers used in 5.1 and 7.1 Soundfield Configurations ( arrangement of speakers intended to reproduce a specific soundfield).

So **Surround Sound** uses the Base Layer only, while **Immersive Sound** uses the Base Layer and additional (typically height) layers or speakers. Also, note that **Immersive Audio** does not have to use **Audio Objects**
Background and terms

**Audio Channel**
Distinct collection of sequenced audio samples that are intended for delivery to a single loudspeaker, loudspeaker array or other reproduction device

**Audio Object**
Segment of audio essence with associated metadata describing positional and other properties which may vary with time

**Bed**
Soundfield Group (*group of channels associated with a Soundfield*), such as a 5.1, 7.1 or 9.1, that serves as the foundation of the immersive soundtrack mix

**Target Environment**
Specific set of conditions that is present in the playback environment

Bitstream Structure

- **Bitstream Frame**
- **Preamble**
- **IAFrame**
  - **Bed 1**
  - **Object 1**
  - **Object 2**
  - **Audio Data**
  - **Sub Blocks**
    - **Frame-Level**
    - **Sub Block Level**

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)
Basic Bitstream Element

<table>
<thead>
<tr>
<th>IAElement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElementID</td>
</tr>
<tr>
<td>ElementSize</td>
</tr>
<tr>
<td>(Contents – based on ElementID)</td>
</tr>
</tbody>
</table>

- **ElementID** – Identifies the element and its syntax. Decoders may skip unrecognized IDs
- **ElementSize** – Can be used to skip an unknown element
- The combination of these two items allows extensibility
- All bitstream elements are based on this basic structure

IAFrame

<table>
<thead>
<tr>
<th>IAFrame</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElementID</td>
</tr>
<tr>
<td>ElementSize</td>
</tr>
<tr>
<td>BitDepth</td>
</tr>
<tr>
<td>FrameRate</td>
</tr>
<tr>
<td>MaxRendered</td>
</tr>
<tr>
<td>SubElementCount</td>
</tr>
</tbody>
</table>

- ST 2098-2 specifies a single IAFrame without constraint on adjacent bitstream frame content
- Each IAFrame is independently decodable
- Applications can constrain bitstreams to contain IAFrames with consistent parameters

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)
Bed Structure (simplified)

- Forms the basis of the mix
- ChannelID determines routing destination
- ChannelCount and ChannelID not constrained
- Each Channel References an AudioData element

Object Structure (simplified)

- Is not routed, but ‘rendered’
- Each ObjectDefinition element references audio essence carried as PCM or DLC using an AudioDataID
- Supports rendering modifications: Spread (size), Snap, Zone exclusion and Decorrelation
- Metadata can be updated on a sub block basis (typically 8 per frame or around every 5ms)
Object Dynamic Metadata (PanInfo)

- Object Gain – Defines gain for a sub block
- ObjectPos – Defines position intent for the sub-block
- ObjectSnap – Defines importance of Timbre vs Position
- ObjectZoneControl – Defines extent of use of zones
- ObjectSpread – Defines object ‘size’ intent
- ObjectDecor – Defines decorrelation effect intent

- At least one sub block per frame
- Decoder uses previous PanInfo if current block has none

Object Coordinate System and Frame of Reference

- Positional metadata is Allocentric and uses a Cartesian coordinate system
- Positional reference is to walls of a unit cube
- The origin is at the intersection of the front (screen) wall and the left wall at the height of the Base Layer
**Cube and reference points**

- ‘Midpoint line’ is at Z=0 and is height of Base Layer speakers
- Left and Right are relative to a viewer facing the screen
- All walls (except floor) are located at 0 or 1

**Mapping location to a cinema**

- Figure on left shows idealized location
- Figure on right shows mapping example
- Cinema layout affects precise location, but mapping preserves intent

Position \([x,y,z] = [0.25, 0.33, 0]\)
Bed Remap Structure (simplified)

- Allows user to create optional re-mix of existing bed to be used under a certain playback condition (Target Environment)
- Remap is created using a matrix with each new output channel created by summing the existing bed channels with a gain parameter for each.

New Elements

ObjectZoneDefinition19 – Defines zones for use with the Auro speaker configs. If used, it only replaces the existing ObjectZoneControl metadata
AuthoringToolInfo – Carries a URI to identify the authoring tool
UserData – Carries unspecified data whose syntax is indicated by the UserID (not used in Cinema)
Backward Compatibility

- Many elements based on existing bitstreams (Dolby Atmos®)
- No new required elements
- No change to syntax for existing elements (although some new codes were added)
- Decoders can skip unrecognized elements
- Result: Existing bitstreams are compliant, but new bitstreams may not work properly on existing decoders

Element Hierarchy

- IAFrames are composed of the shown elements
- All elements are optional
- Some elements can contain sub-elements
- Upper-level elements have Parent-Child relationship with the lower-level element
- The Child will replace all or a portion of the Parent, if used
Conditional Bitstream Elements

- Allows bitstream creator to supplement the bitstream for alternate Target Environments
- Always associated with a UseCase code which represents a specific playback condition
- A conditional element can be a Child or a Top-Level element
- A conditional Child element replaces the Parent (or a part)
- The renderer will determine the UseCase based on its configuration, which is a one-time setup
- Only rendered if renderer configuration matches UseCase

Conditional Bitstream Elements

- Conditional Bed
  - Replaces parent Bed with a new mix (new audio essence)
- Conditional Object
  - Replaces parent Object with a substitute Object (may use new audio essence)
- BedRemap
  - Remixes parent bed channels to create alternate mix
Examples - BedRemap

- Diagram shows element relationships and hierarchy: Lower elements are contained in upper (parent) elements
- RemapUseCase = 1 (5.1). Note SourceChannel count and DestinationChannel count.
- BedRemap is used IFF the renderer configuration indicates a 5.1 playback environment (not an immersive audio system)

Examples – Typical Stream

- Typical element structure of Bed plus Object(s)
- Active elements are rendered simultaneously
- Neither element is conditional, so both will always be rendered
- MaxRendered is the total of bed channels plus objects
Examples – Conditional Bed

- Conditional Child Beds will replace the parent under a specific UseCase
- Child Beds carry additional essence i.e., a different mix intended for the Target Environment
- The bitstream standard constrains the number of child elements
  - Does not allow Children of Children
  - MaxRendered cannot exceed 128 (at 48kHz) or 64 (at 96kHz)

Coding Efficiency

- Metadata and essence only present in a frame if needed in that frame. Objects need not persist if they are not used
- Sub Blocks are only used when needed to carry metadata that changes over a frame
- Audio Essence can be losslessly coded (2:1 typical)
- In each frame, for each channel or object, silence can be carried as metadata (AudioDataID = 0)
- Anecdotally: IAB track file approx. equal to 14-chan Main Audio
Cinema Packaging (in progress)

<table>
<thead>
<tr>
<th>I A Bitstream</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>MXF Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitstream Frame</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ST 249-18 Immersive Audio Track File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Header</td>
</tr>
</tbody>
</table>

IMF application (in progress)

<table>
<thead>
<tr>
<th>I A Bitstream</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
<th>Bitstream Frame</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>MXF Clip</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ST 2067-201 IAB Level 0 Plug-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Header</td>
</tr>
</tbody>
</table>

Based on amended bitstream

Contains index table for IAB frame access

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)
Bitstream Standard Summary

• Efficient, flexible, extensible standard that supports the features needed to convey immersive audio essence and rendering intent
• Backward compatible with existing immersive audio content (Dolby Atmos®)
• File transport will soon be supported by SMPTE standards for D-Cinema Packaging and IMF

Questions Please!
(Verbal Questions Take Precedence)

Stan Cossette
Senior Staff Engineer
Dolby Laboratories