Dolby AC-4
Next-Generation Audio

Jeffrey Riedmiller
Dolby Laboratories
jcr@dolby.com
Improving Today

**Improve Reach**
Efficient delivery to all devices

**Simplify Operations**
Intelligent loudness, A/V frame alignment

**Enhance Accessibility**
Descriptive audio, dialogue enhancement

**Bridge to the Future**
Deliver today’s audio and add new features later
Enabling Tomorrow

Scalable

Personalized

Immersive

End-to-End
Immersive. We Took a Fresh Look at Life-like Sound

Base Audio Bed
Static audio elements, with stirring base tones and rich audio textures

Sound “Objects”
Moving around and above you in multi-dimensional space...adds a **height dimension** not ever before present!

Real Sound Simulation
Multi-dimensional audio that fully immerses you in the scene!

A New Approach To Audio Mixing And Playback
Which Revolutionized the Cinema Experience

- **2014 Oscar winners** — featured Dolby Atmos sound
  “Gravity” (Best Director, Best Sound Mixing, Best Sound Editing and Best Original Score), “Frozen” (Best Animated)

- **Dolby Atmos** won the Cinema Audio Society Award for Post Production Tech Achievement and a similar award from the Hollywood Post Alliance

- **Major Directors** including Alfonso Cuarón, JJ Abrams, Ang Lee, Mark Andrews, Peter Jackson, Danny Boyle, and Joseph Kosinski, among other have used Dolby Atmos

- **1250+ Screens** installed or Committed to Dolby Atmos

- **190+ Titles** mixed in Dolby Atmos

- **150+ Exhibitor Partners** present in Dolby Atmos

- **100+ Mixing Facilities** equipped with Dolby Atmos

- **40+ Countries** with Dolby Atmos screens
Now Dolby has Brought Immersive Audio to the Home

- Upward firing speakers do a remarkable job of introducing the dimension of height
- Exceptional opportunities for product innovation
- In market now
The object-based paradigm = the foundation for immersive audio
Object-based Audio = Audio Elements with a “Flight Plan”

Objects require **dynamic metadata**

- Time stamp, e.g. number of samples/seconds from program start
- Object position – X-axis, Y-axis, Z-axis
- Rendering Mode(s)
  - Speaker zone masks
  - Speaker snap mode
- Size → width (spread) + decorrelation
- Divergence
- Update interval typically 512 – 1536 samples @ 48kHz

Personalization requires presentation + interactivity metadata

Cartesian Coordinates (room normalized)
Object Playback – Requires a “Renderer”

Definition:

An audio renderer converts a set of audio signals with associated metadata to a different configuration of audio signals – e.g. speaker feeds, based on the metadata, and a set of control inputs derived from the rendering environment and/or user preference.

Object(s) w/ room normalized coordinates (x,y,z) 

Renderer

Speaker and/or Environment configuration

Speaker Feeds e.g. 7.1.4, 5.1, 2.0, etc.
Rendering Implementations & Philosophies

Directional/VBAP

**Pros:** discrete, good timbre

**Cons:** limited to objects on a sphere

Balance-based (current practice)

**Pros:** boundary and inside the room panning, good timbre, independent control of surround and overheads

**Cons:** speakers need to be organized in 2D layers (e.g. base, ceiling, floor)

Distance-based

**Pros:** smoother pans, more stable inside the room, very general solution

**Cons:** worse in terms of discreteness, timbral issues
Object-based Audio Rendering – Design Considerations

Positional **not just directional**
- Renders position and direction
- Renders objects inside the room

Matches current practice – for traditional surround
- Today’s mixing consoles render position – joystick panning *into* the room is supported

Smooth pans
- Small variations in angles/position result in small changes in gains

Discrete
- Uses a minimal amount of speakers to render a ‘point source’ (size= 0)
- Improves localization and timbral quality of complex mixes

Sync between metadata and audio
- Maintain tight control of E2E jitter on the order of 32 samples max
Example – Computing Rendering Gains for Stereo

A simple indicative example for computing a one-dimensional (stereo) set of rendering gains utilizing an audio object’s x-coordinate at 0.0 (center) in [-1, 1] could be derived as follows:

\[
\begin{align*}
G_{\text{left}} &= \cos\left(\frac{x+1}{2.0} \cdot \frac{\pi}{2}\right) \\
G_{\text{right}} &= \sin\left(\frac{x+1}{2.0} \cdot \frac{\pi}{2}\right)
\end{align*}
\]

where,

\[
\begin{align*}
0.707 &= \cos\left(\frac{0.0+1}{2.0} \cdot \frac{\pi}{2}\right) = \text{left_gain} \\
0.707 &= \sin\left(\frac{0.0+1}{2.0} \cdot \frac{\pi}{2}\right) = \text{right_gain}
\end{align*}
\]

then,

\[
\begin{align*}
\text{Spkr_bus}_L &= \text{object_signal} \times 0.707 \\
\text{Spkr_bus}_R &= \text{object_signal} \times 0.707
\end{align*}
\]

Note: \(\pi/2\) is expressed in radians!
Personalization: Create Your Own Experience

- Dialog Control
- Crowd Ambience
- English Comm.
- German Comm.
- Star Player A
- Star Player B
- Comedy Comm.
- Goal Flash

- Default
- MY choice
- "THE ENFORCER EXPERIENCE"

Brought to you by AirCanada
Dolby AC-4 – Better Taste with Half the Calories

MORE DOLBY GOODNESS

HALF THE BIT-RATE
Better Taste...

**AC-4 CORE VALUES**
- Reliable, managed ecosystem
- Broad device compatibility
- Always audio out
- Consistent consumer experience

**AC-4 ENHANCED FEATURES**
- End-to-end automated loudness management
- Dialog enhancement
- Dynamic range controls
- Accessibility

**AC-4 NEW CAPABILITIES**
- Rich object audio support
- Next-gen experiences
- A/V frame alignment
- Multi-stream delivery
- Multiple outputs
- Scalable decode
Half the Calories...

AC-4 Bitrate Performance

<table>
<thead>
<tr>
<th>Format</th>
<th>Bit Rate Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>40 - 96 kbps</td>
</tr>
<tr>
<td>5.1</td>
<td>80 - 192 kbps</td>
</tr>
<tr>
<td>7.1.4</td>
<td>144 - 288 kbps</td>
</tr>
</tbody>
</table>

(good through excellent quality ranges)
Platform for Experience Innovation

- Personalized
- Stereo
- Mono

Dolby Digital
Dolby Digital Plus
Dolby AC-4
Dolby Digital Plus Atmos
One Codec for All Content

- Stereo coding optimization
- Temporal shaping of quantization noise
- Parametric spatial coding tools
- Video-frame synchronous audio framing

**AC-4 DECODER**

- **MDCT DOMAIN**
  - Two Spectral Frontends
  - Stereo Audio Processing

- **QMF DOMAIN**
  - Compinging
  - Advanced SPX
  - Advanced Coupling
  - DRC & DE

- **TIME DOMAIN**
  - SRC
  - Limiter

- ** PCM Audio**

- **Bitstream**
  - Multitude transform sizes for highest efficiency
  - Two inverse quantization methods for optimal speech performance

- **High-frequency spectrum reconstruction**
- **Multi-band and multi-channel DRC Built in dialogue enhancement**

Dolby AC-4

© 2016 Dolby Laboratories, Inc.

March 22, 2016
Dolby AC-4 Dialogue Enhancement

Pre-mixed content

Clean dialog tracks
Flexible syntax enable efficient multi-language and descriptive audio delivery
Enabling Video Frame Alignment

Audio and speech spectral frontends both operate in the MDCT domain.
Video Frame Alignment Simplifies Turn-Around

MVPD Turnaround with AC-3 Today

Audio is modified and original metadata is lost

Metadata prone to loss

Local Timebase (Vref)

MVPD Turnaround with AC-4

Audio & Metadata are passed through

AC-4 Bitstream

Local Timebase (Vref)

To Subscribers (viewers)
More Than An Audio Codec

Channel-based Audio Example
Dynamic Range Control (DRC)

Multi-band, multi-channel DRC with profiles for each playback device type
Specific Tools for Object-based Immersive Content

**core decoding**

- **AC-4 encoder**
  - Spatial coding
  - Spatial (Object) Groups (15) and LFE
  - A-JOC

- **AC-4 bitstream**
  - OAMD 15
  - OAMD 7

- **Spatial (Object) Groups (7) and LFE**

- **A-JOC data**

- **AC-4 decoder**
  - Spatial (Object) Groups (7) and LFE
  - Object audio metadata

- **SAP**
- **ASF/SSF**
- **A-SPX**

**full decoding**

- **AC-4 decoder**
  - Object audio metadata
  - Spatial (Object) Groups (15) and LFE

- **A-JOC**
Specific Tools For Channel-based Immersive Content
Single Bitstream, Optimized Output
AC-4 in Broadcast Standards

<table>
<thead>
<tr>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS 103 190, I</td>
<td>TS 103 190, Part II</td>
<td>ATSC 3.0</td>
</tr>
<tr>
<td>Channels</td>
<td>Channels + Objects</td>
<td>Nov 2015 – S34-2</td>
</tr>
<tr>
<td>April 2014</td>
<td>March 2015</td>
<td>Recommended for U.S.</td>
</tr>
</tbody>
</table>

Interoperability Guidelines v3.1
July 2015
A Future Night of Content (PCM + Metadata) via SDI

<table>
<thead>
<tr>
<th>News</th>
<th>Drama</th>
<th>Sports</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo</td>
<td>5.1</td>
<td>5.1 M&amp;E + 6 Objects</td>
<td>7.1.4 Immersive</td>
</tr>
</tbody>
</table>

- Ch 1/2: Complete Main
- Ch 3/4: Complete Main
- Ch 5/6: Complete Main
- Ch 7/8: SAP / DVS
- Ch 9/10: Eng / Spa Commentary, Home / Away Commentary
- Ch 11/12: Height Objects
- Ch 13/14: SAP / DVS
- Ch 15/16: Metadata
Decoded Examples using Program Elements

Main + DVS
- 5.1 CM
- DVS

Alternate Language
- 5.1.2 M&E
- Eng
- Spa
- Home
- Away

Immersive / Personalized Arena Sports
- 5.1 M&E
- Eng
- Spa
- Home
- Away
- Height
- DVS

Personalized Motor Sports
- 5.1 M&E
- Comm 1
- Comm 2
- Team Radio 1
- Team Radio 2
- Team Radio 3
- Track Ann.
A Future Night of Content (Legacy + NGA) via SDI

<table>
<thead>
<tr>
<th>Channel</th>
<th>News</th>
<th>Drama</th>
<th>Sports</th>
<th>Movie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch 1/2</td>
<td>Stereo</td>
<td>5.1</td>
<td>5.1 M&amp;E + 6 Objects</td>
<td>7.1.4 Immersive</td>
</tr>
<tr>
<td>Ch 3/4</td>
<td>Complete Main</td>
<td>Complete Main</td>
<td>Complete Main</td>
<td>Complete Main</td>
</tr>
<tr>
<td>Ch 5/6</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
</tr>
<tr>
<td>Ch 7/8</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
</tr>
<tr>
<td>Ch 9/10</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
<td>SAP / DVS</td>
</tr>
<tr>
<td>Ch 11/12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch 13/14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch 15/16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Stereo**: For News.
- **5.1**: For Drama and Sports.
- **5.1 M&E + 6 Objects**: For Sports.
- **7.1.4 Immersive**: For Movie.
- **SAP / DVS**: For Ch 7/8.
- **Compressed NGA Mezzanine**: For Ch 13/14.
- **Compressed NGA Mezzanine**: For Ch 15/16.
Why *Dynamic* Metadata?

- Encoder can’t auto-detect between complete (legacy) and compositional mixes (next-generation)
- Signaling is necessary
  - In-band carried with audio
  - Out-of-band via automation

Note: time-synchronous dynamic metadata is *mandatory* for interchanging full *object-based* immersive experiences
Example AC-4 Audio Bitrates for ATSC 3.0

For ATSC 1.0 audio today:
5.1 Main + 2.0 SAP = ~576kbps

Rates for excellent quality are shown in this example
End-to-End Workflow

Production
- Audio Console
- Object Audio Contribution Encoder

Contribution
- Fiber Mix
- Contribution Encoder
- Contribution Decoder
- Object Audio Decoder
- Audio Processing

Studio
- Baseband Processing
- Object Audio Transmission Encoder
- Transmission Encoder

Transmission
- Set-top Box
- Table
- TV
- Legacy A/V Receiver
- Object Audio A/V Receiver

Playback

Channel-based Personalized or Immersive:
PCM + Metadata

Object-based Personalized + Immersive:
Full Dynamic object in bandwidth efficient Dolby ED2 – transitioning to PCM+Metadata over IP in future

AC-4

March 22, 2016
© 2016 Dolby Laboratories, Inc.
AC-4 Supports Taking Small Steps Toward Future

“Now”
- Dolby Digital
- Discrete Audio Description Mix
- Discrete Alternate Language Mix
- Loudness Issues

“Same, but Better”
- Better Efficiency
- Integrated Loudness Management
- Dialogue Enhancement

“Accessibility 2.0”
- Audio Description Receiver Mixing
- Clean Dialogue
- Language replacement

“Immersive”
- Height
- Object-based Audio
- Scalable rendering

“Immersive and Personalized”
- Interactivity
- Side-band audio delivery
Thank You