Dolby Atmos
Immersive Audio
From the Cinema to the Home

Nick Engel
Senior Director
Consumer Entertainment Technology

AES Melbourne Section Meeting
11 December 2017
History of Cinema

• 1980s - Dolby SR noise reduction (Analog)
• 1990s – Dolby Digital Surround EX™ (5.1)
  • 2010 - Dolby Surround 7.1
Spatial Audio Description: Channel Based
Spatial Audio Description: Channel Based
Spatial Audio Description: Object Based
Spatial Audio Description: Object Based

• Position
Spatial Audio Description: Object Based

- Position
- Size
Spatial Audio Description: Object Based

- Position
- Size
- Diffusion
Spatial Audio Description: Object Based

Object
• audio essence + metadata

Metadata
• Spatial metadata
  • Position
  • Size
  • Diffusion
• Gain
• Object distance
• Zone
• Width
• Screen Scaling
• Snap
Spatial Audio Description

Benefits of Channel Based

• Established work-flow, tools and techniques
• Easy to render & monitor
• Direct artistic control over key loudspeakers (e.g. Center, LFE)
• Efficient storage (with scene simplification)
Spatial Audio Description

Liabilities of Channel Based

• Spatial Resolution Limited by Channel Count
Spatial Audio Description

Liabilities of Channel Based

• Spatial Resolution Limited by Channel Count
Spatial Audio Description

Liabilities of Channel Based

• Spatial Resolution Limited by Channel Count
• Not Scalable or Adaptable
Spatial Audio Description

Liabilities of Channel Based

- Spatial Resolution Limited by Channel Count
- Not Scalable or Adaptable
Spatial Audio Description

Benefits of Object Based

• Scalable
Spatial Audio Description

Benefits of Object Based

• Scalable
Spatial Audio Description

Benefits of Object Based

• Scalable
Spatial Audio Description

Benefits of Object Based

• Scalable
Spatial Audio Description

Liabilities of Object Based

• Less efficient for complex soundscapes

• One object authored, distributed and rendered for every cricket or raindrop?

• Learning curve for mixers
Hybrid Channel-based + Object-based Audio

Base Audio Bed
Complex Audio Textures

Sound “Objects”
Discrete and Moving Sounds
Audio plus Positional Data

Dolby Atmos
Unlimited Creative Freedom
Extremely Flexible
Translates to any speaker configuration

- Audio Bed contain stationary ambient sounds – similar to a channel based mix
- Objects consist of audio plus positional information
e.g. explosions, animals, vehicles, bullets, dialogue, etc.
- Objects move throughout the three dimensional listening space as content is played back
Loudspeaker Layout
Loudspeaker Layout
Loudspeaker Layout
Loudspeaker Layout
Loudspeaker Layout

Screen Loudspeaker

Surround Loudspeaker
Loudspeaker Layout
Loudspeaker Layout
Loudspeaker Layout
Loudspeaker Layout
Reference Level
Mixer Survey

All of the Dolby field engineers around the world and over 60 mixers were interviewed for their opinions...
Experimental Mixes

Incredibles
- Skywalker Sound
- Michael Semanick
- AMS NEVE Console DFC
- Screen and Surround Height

Avatar
- Skywalker Sound
- Gary Summers
- 3D Audio Panner
- Screen and Surround Height, and Surround Point Source (surround direct)
Double Blind Subjective Tests
WORLD OF ATMOS

Content creation → Distribution → Playback Devices
BRINGING IMMERSIVE AUDIO TO THE HOME

- TVs
- HOME THEATER
- SOUND BARS
- GAMING
- CINEMA
- BLU-RAY™
- AVRs
- MOBILE DEVICES
- VIRTUAL REALITY
- PCs
Bringing Atmos to the Home

Creation

Delivery

Playback

Spatial Coding

Audio Coding

Connected Consumer Devices

Built In Speakers

Home Theatre Devices
DOLBY ATMOS AUDIO

An Atmos scene is made up of audio objects.

Audio objects consist of an audio stream and associated metadata.

Up to 159 discrete objects can be represented during content creation.

In addition to objects, an Atmos scene can be described in terms of channel beds or Intermediate Spatial Format (ISF).
HOME CINEMATIC CONTENT CREATION

Creation

Near-field Mix Pro Tools

Dolby RMU

Dolby Atmos Master File .atmos

DMG

2.0 and 5.1 or 7.1 Mix

DMP

Pcm

Dolby

Encoding

TrueHD

DD+JOC

MP4 w/ DD+JOC and H264

DMD

DUP

3rd party muxer or DMG

Delivery

Blu-ray

DD+ JOC bitstream in ES, TS or MP4 with video

(Optional) Non-Atmos

ProRes Video
LPCM 5.1 or 7.1 and 2.0 Mix in a .mov container

AVR/DP580
EXAMPLE LIVE WORKFLOW

Stadium

Audio Console ➔ DP591 Object Audio Encoding (ED2) ➔ Contribution Encoder

<table>
<thead>
<tr>
<th>Monitoring</th>
<th>Metadata</th>
</tr>
</thead>
</table>

DP590 Object Audio Authoring ➔ DP580 Object Audio Monitoring (ED2) ➔ Fiber or Satellite ➔ Contribution Decoder

Playout

DP591 Object Audio Encoding (DD+ IOC) ➔ Transmission Encoder

DP580 Object Audio Monitoring (DD+ IOC)

Home

STB

© 2016-17 DOLBY LABORATORIES, INC.
Bringing Atmos to the Home

Creation
- Content

Delivery
- Spatial Coding
- Audio Coding

Playback
- Connected Consumer Devices
  - Built In Speakers
- Home Theatre Devices
The Challenge: Delivering 5.1

**Creation**
- 5.1 Content
- 4,608,000 bps

**Delivery**
- Spatial Coding
- Audio Coding

**Playback**
- 5.1 Content
- 192,000 bps
- Compression Ratio: 23.4x
The Challenge: Delivering Atmos

**Creation**
- ATMOS Content
  - 150,000,000 bps

**Delivery**
- Spatial Coding
- Audio Coding

**Playback**
- Atmos Content
  - 384,000 bps
  - Compression Ratio: 390x
Immersive content requires new thinking!

<table>
<thead>
<tr>
<th>Format</th>
<th>Bitrate</th>
<th>Peak-bitrate</th>
<th>File Size (90 minute)</th>
<th>AES3 Pairs Required</th>
<th>Metadata Channel</th>
<th>Metadata Sync Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM+Metadata</td>
<td>Constant</td>
<td>115.2Mb/s</td>
<td>77.76GB</td>
<td>50</td>
<td>OOB</td>
<td>Extremely High</td>
</tr>
</tbody>
</table>
DATA REDUCTION PART 1 – SPATIAL CODING

- Screen

- Bed Object
- Dynamic Object

Spatial coding
DATA REDUCTION PART 2 – AUDIO CODING

- Dolby Digital (AC-3): Traditional transform codec
- Dolby Digital Plus (E-AC3): Supports Dolby Atmos (JOC)
- Dolby TrueHD: Lossless codec, supports Dolby Atmos
- Dolby AC-4: Dolby Atmos, Accessibility, Personalization

- MAT: Container format for Audio over HDMI
Dolby Digital Plus JOC

DD+ 5.1
Dolby Digital Plus Audio Frame

DD+ 5.1
JOC with metadata
Dolby Digital Plus Audio Frame

192kbps
384kbps
DD+ JOC Encoder: Object Based Immersive (OBI) content
DD+ JOC Decoder (DD+ JOC bitstream, 5.1-channel core)
Dolby AC-4 is the new audio coding system that improves today’s stereo and 5.1 experiences and enables efficient delivery of the next generation audio experiences like immersive and personalized
AC-4 - State of the art coding tools

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

MDCT DOMAIN
- Two Spectral Frontends
- Stereo Audio Processing

QMF DOMAIN
- Companding
- Advanced SPX
- Advanced Coupling
- DRC & DE

TIME DOMAIN
- SRC
- Limiter

Bitstream
PCM Audio

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
AC-4 - Flexible Support for Multiple Device Types

**Full decoding**

- Dolby playback system (support for current generation AVR)
  - Broadcast
  - Bitstream formatter
  - AC-4 decoder
  - Dolby MAT formatter
  - via HDMI
  - Advanced renderer

- Dolby AVR or soundbar

**Core decoding**

- Dolby playback system (e.g. TV)
  - Broadcast
  - Bitstream formatter
  - AC-4 decoder
  - Simple renderer
  - PCM (2.0 ch)
  - PCM (up to 5.1 ch)

**Or**

- Full decoding

- Dolby playback system (e.g. Mobile device)
  - Broadcast
  - Bitstream formatter
  - AC-4 decoder
  - Virtualizing renderer
  - PCM for headphones virtualized
  - PCM (2.0 ch) virtualized

**Full decoding**

- Dolby playback system (e.g. AVR)
  - Broadcast
  - Bitstream formatter
  - AC-4 decoder
  - Advanced renderer
  - PCM
  - (Flexible number of speaker feeds)
AC-4 – Immersive Stereo – Overview

- Low Bitrate
- Low complexity upon playback
- Rendering flexible in encoder
- Adjustment to listener playback condition
- LoRo compatible
AC-4 – Immersive Stereo – Playback

Upon playback the decoder provide 3 different outputs to best suit the current playback condition.

Deliver LoRo

Metadata

Decode LoRo

Post-Process

Stereo

Atmos Experience Headphone

Atmos Experience For Mobile Speakers
AC-4 – Immersive Stereo – Playback

Upon playback the decoder provide 3 different outputs to best suit the current playback condition:

- **Delivery LoRo**
- **Post-Process**
- **Atmos Experience**
- **Headphone**
- **Atmos Experience For Mobile Speakers**

Works for offline playback:
One stream serves three portable playback scenarios.
Bringing Atmos to the Home

Creation

Delivery

Playback

Spatial Coding

Audio Coding

Connected Consumer Devices

Built In Speakers

Home Theatre Devices

© 2016-17 DOLBY LABORATORIES, INC.
Device Playback Workflow

Consumer

- TS MP4 Smooth HLS DASH
- DD+ JOC or TrueHD Pass-thru via HDMI
- DD+ 7.1/5.1 Bed Decode
- DD+ JOC Bed + Objects Renderer
- ATMOS capable AVR
- Non Atmos
- Atmos capable
- Dolby enabled Speaker
- 5.1/7.1

Legacy AVR
Dolby Atmos – Speaker Playback Examples

5.1.2 AVR

3.2.1 Soundbar

5.1.4 AVR

5.1.4 Soundbar
Dolby Atmos – Stereo Speakers and Headphones

PC
TV
Mobile
Gaming
Virtual Immersive Audio

... *More than just panning a source between the channels*

**Speaker Virtualisation**
- Creating the impression that missing speakers are present.
- Give the impression that sounds are coming from a bigger sound stage – off the screen, above you.

**Headphone Virtualisation**
- Creating the impression that you’re listening to speakers rather than on headphones.
- Relieve ‘stereo pressure.’
- Provide directional sound.
Why / how is it possible to virtualise sound sources?

Provocation: Humans only have two ears, yet we can localise sound from any direction.

Generally Humans can distinguish relative azimuth, elevation and distance of a sound source.

The primary cues for localisation are:

- Interaural Time Difference (ITD)
- Interaural Intensity Difference (IID) (Interaural Level Difference (ILD))

Additional cues include:

- Pinna filtering
- Interaural Cross-Correlation (IACC)
- Direct-to-Late Reverberation ratio (DLR)
Localising Sound

Interaural Time Difference (ITD) is the difference in a sound’s arrival time at each ear and provides a strong azimuth cue.

Interaural Intensity Difference (IID) is the amplitude difference of a sound at each ear – mainly due to head shadowing.

ITD and IID are used together to determine direction of arrival and interact differently at low and high frequencies.

• At low frequencies (f < 800 Hz) the dimension of the head and the wavelength of the soundwaves are such that IID is reduced, and the auditory system can determine the phase differences introduced by ITD.
• At high frequencies (f > 1600 Hz) IID is more pronounced but the phase relationship between the two ears becomes too difficult to resolve. The group delay of impulsive sounds can still be used to estimate ITD.
Localising Sound

ITD & IID aren’t sufficient to distinguishing between the front and back, or any direction of arrival which can result in the same delay and intensity differences between ears – this is called a ‘cone of confusion’

Sound localisation – cone of confusion
Localising Sound

The Pinna – the flappy bit of cartilage stuck to the side of your head which people like to decorate.

The shape of the pinna causes incoming sound to be filtered based on their direction of arrival.

Note: There is a lot of individual variation in pinnas.
Localising Sound - HRTFs

A Head-related Transfer Function (HRTF) is the result of treating the head as a linear system and characterising the ITD, IID and pinna filtering.

- HRTFs are a function of distance, azimuth and elevation
- Left and right ears are not usually symmetrical

Measuring HRTFs is tricky...

- For a true HRTF, you need an anechoic chamber
- You need to capture impulse responses from all directions of arrival and distances

HRTF databases are another source

Sometimes a ‘spherical head model’ is enough
What’s a room model, and why is it important

Humans can do more than just localise a sound source when listening, we can also deduce things about our environment – particularly how big a space we are in, and how ‘live’ it might be.

Binaural Room Impulse Responses (BRIRs) are a combine response of a head and a room. (i.e., what you would get if you captured an HRTF in a room, rather than an anechoic chamber)

- Importantly they include the echoes (reverberation) of the room.

BRIRs are a function of;

- the head geometry,
- the room geometry,
- the position and orientation of the head in the room, and
- the position and radiation pattern of the sound source.

To capture a full BRIR of a person in a room would be infeasible given the large number of dependencies, so some form of model is required.
Room Impulse Response

RIR can be broken into three parts:

- **Direct** path is the shortest, and first to arrive
- Distinct specular reflections make up the **Early** part of the response
- Many reflections merge into indistinct reverberations forming the **Late** part of the response
Room Impulse Response

When the RIR is includes the effect of a head, and is expressed with both ears it is referred to as a Binaural Room Impulse Response (BRIR)

Some qualities of BRIRs of interest include:
- RT60 – the rate of the reverberations decay (how ‘live’ is the room?)
- Direct-to-Late Ratio (DLR) – the ratio of energy in the direct path to the reverberations
- Interaural cross-correlation (IACC)
High level look at headphone virtualiser

sampled BRIRs for a limited number of positions (e.g., 5.1)

\[ x(n) \rightarrow \text{BRIR} \]
High level look at headphone virtualiser

- Anechoic HRTF: \( h_l(az, el, r), \quad h_r(az, el, r) \)

- BRIR model without the direct path
High level look at headphone virtualiser

\[ x(n) \rightarrow \text{Short BRIR (HRTF+ early)} \rightarrow + \rightarrow \text{Late reverb model} \]
High level look at speaker virtualiser

Headphone virtualisers are a matter of simulating how we localise sound ... relatively simple as the processed signal is delivered directly to each ear.

Speaker virtualisation adds the complication that sound for each speaker ‘leaks’ into both ears.
High level look at speaker virtualiser

Stereo mix of 5.1

C -> L -> + -> Lo

Ls -> R -> + -> Ro

Le <- Lo

Re <- Ro
High level look at speaker virtualiser

Virtualised mix of 5.1

C
L
Ls
Rs
R
HRTF
X-talk canceller
+
+
Lo
Le
Ro
Re
Dolby Atmos – Stereo Speakers and Headphones
IMMERSIVE AUDIO STANDARDS

- ITU-R
  - BS.2076 BWAV ADM
  - Reference Renderer

- SMPTE
  - Interchange, Metadata
  - Immersive Audio for Cinema

- Emission
  - ETSI TS 102 366 v1.3.1 [E-AC-3]
  - ETSI TS 103 190 [AC-4]

- ATSC 3.0

- DVB NGA
Where can I ... Atmos?
Experience Atmos

Experience Atmos
... Mix Atmos

Soundfirm Melbourne Facility
Mixing and Mastering Stage
Only facility in Australia
AVID PRO TOOLS & Dolby Atmos

Avid Pro Tools with their 12.8 release offers native Dolby Atmos mixing.

Pro Tools enhancements include support for Dolby Atmos 7.1.2 audio bussing and built-in fold-down logic.
DOLBY ATMOS MASTERING SUITE

• FULL DOLBY RMU CAPABILITY
• ATMOS CONVERSION TOOL
• 3 PRODUCTION SUITE LICENSES

DOLBY ATMOS PRODUCTION SUITE

• SOFTWARE-BASED SOLUTION
• MIXING AND MASTERING FOR LESS COMPLEX PROJECTS
• FULL DOLBY ATMOS MASTER FILE QC ABILITY
• COMPLETE ATMOS VR PRODUCTION SOLUTION