

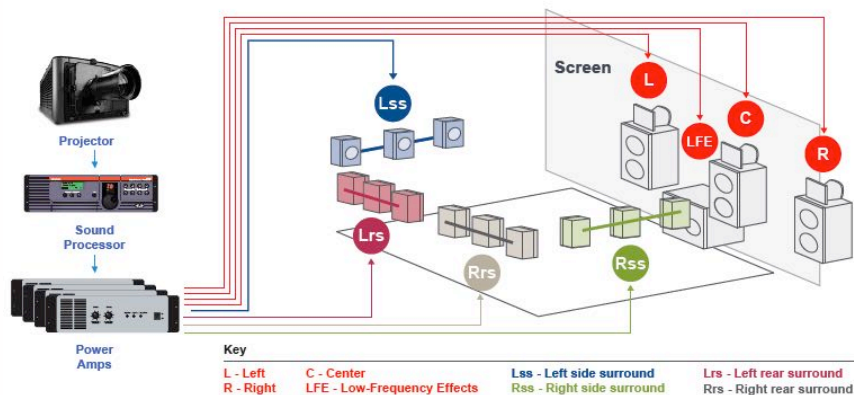
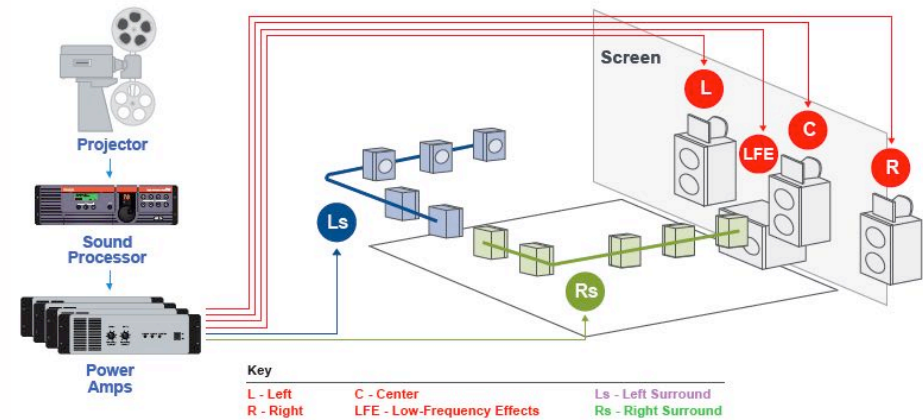
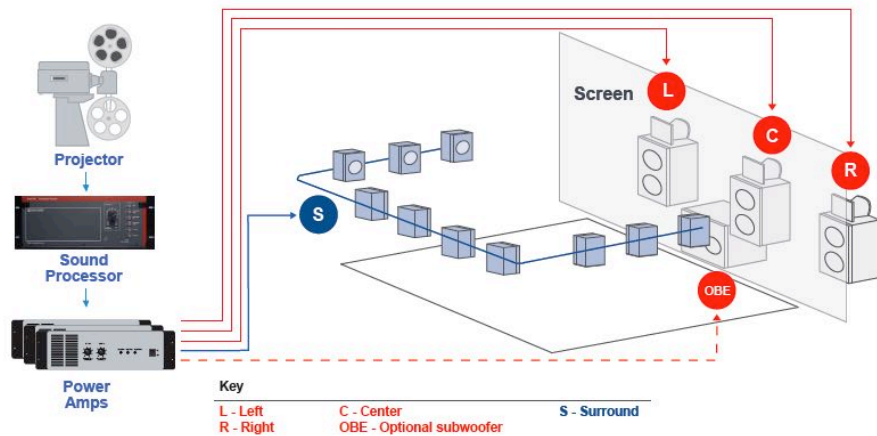


Dolby Atmos Immersive Audio From the Cinema to the Home

Nick Engel
Senior Director
Consumer Entertainment Technology

AES Melbourne Section Meeting
11 December 2017

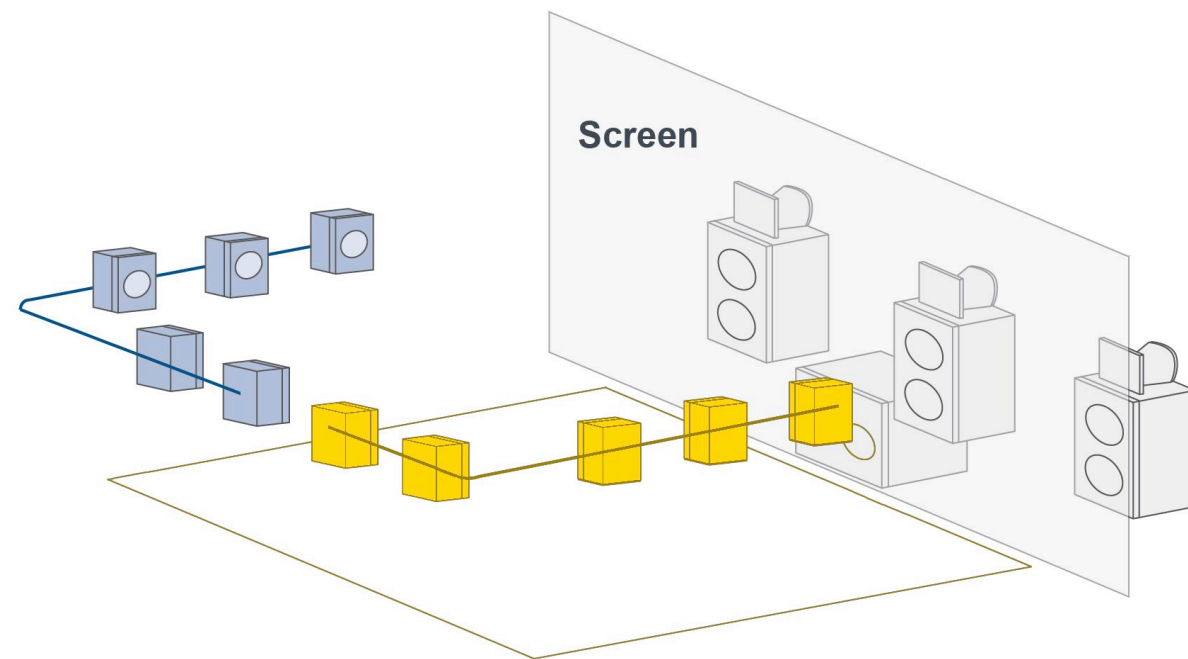
History of Cinema



- 1980s - Dolby SR noise reduction (Analog)
- 1990s - Dolby Digital Surround EX™ (5.1)
 - 2010 - Dolby Surround 7.1

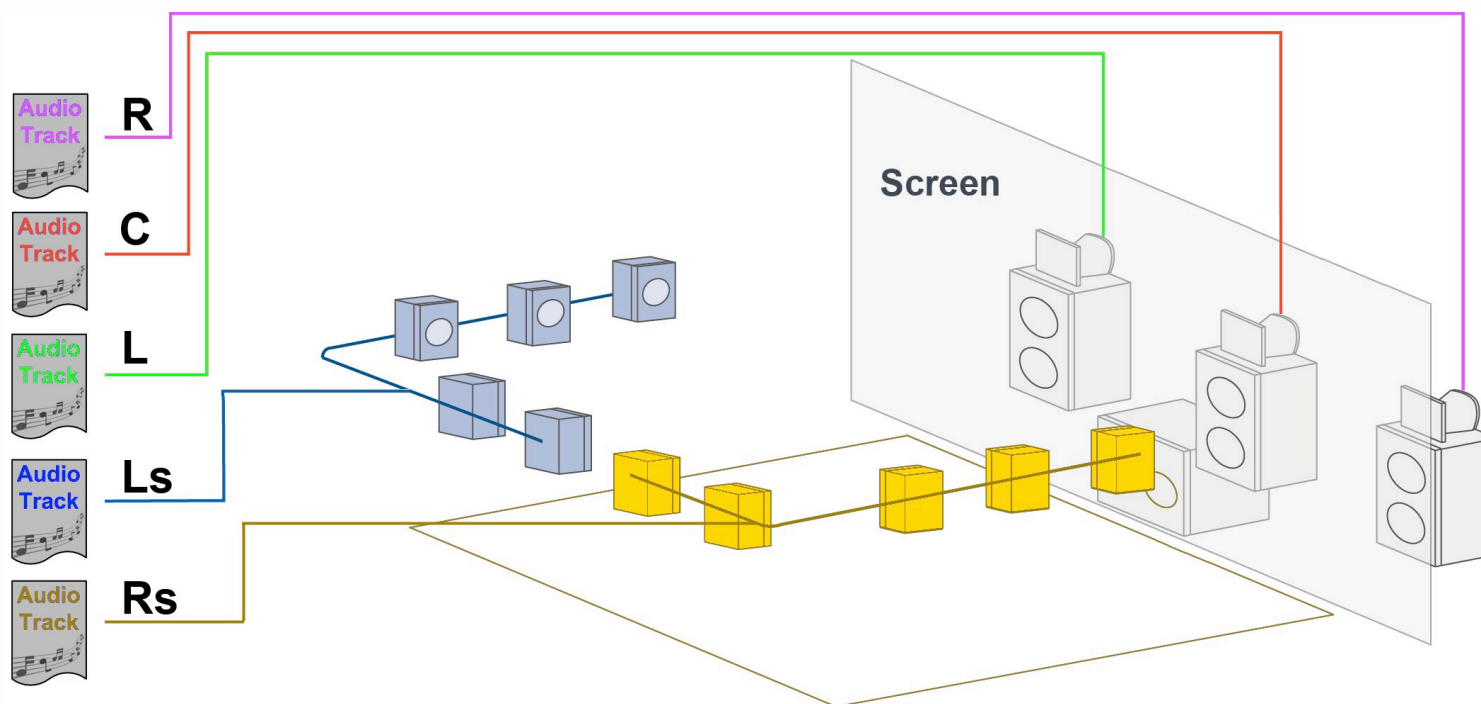


Spatial Audio Description: Channel Based

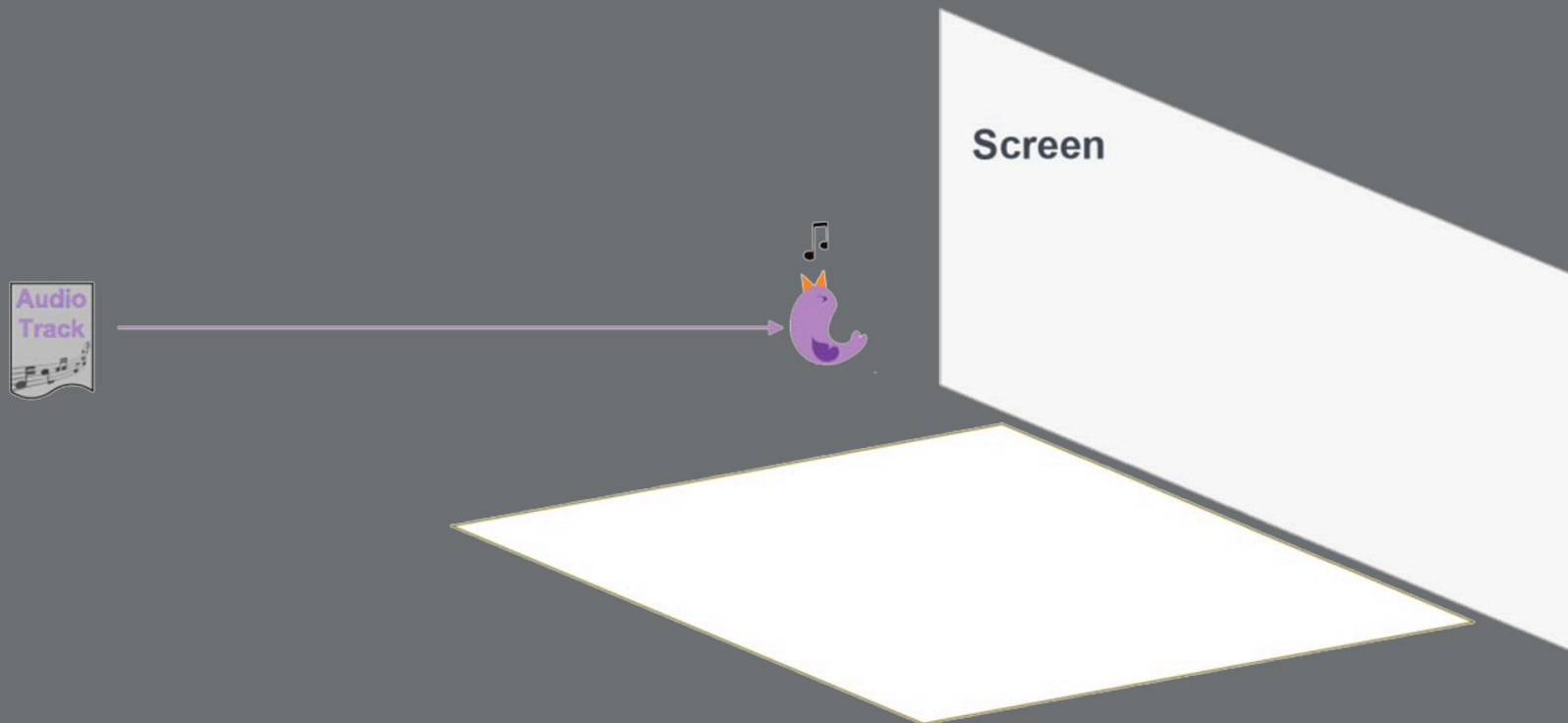




Spatial Audio Description: Channel Based

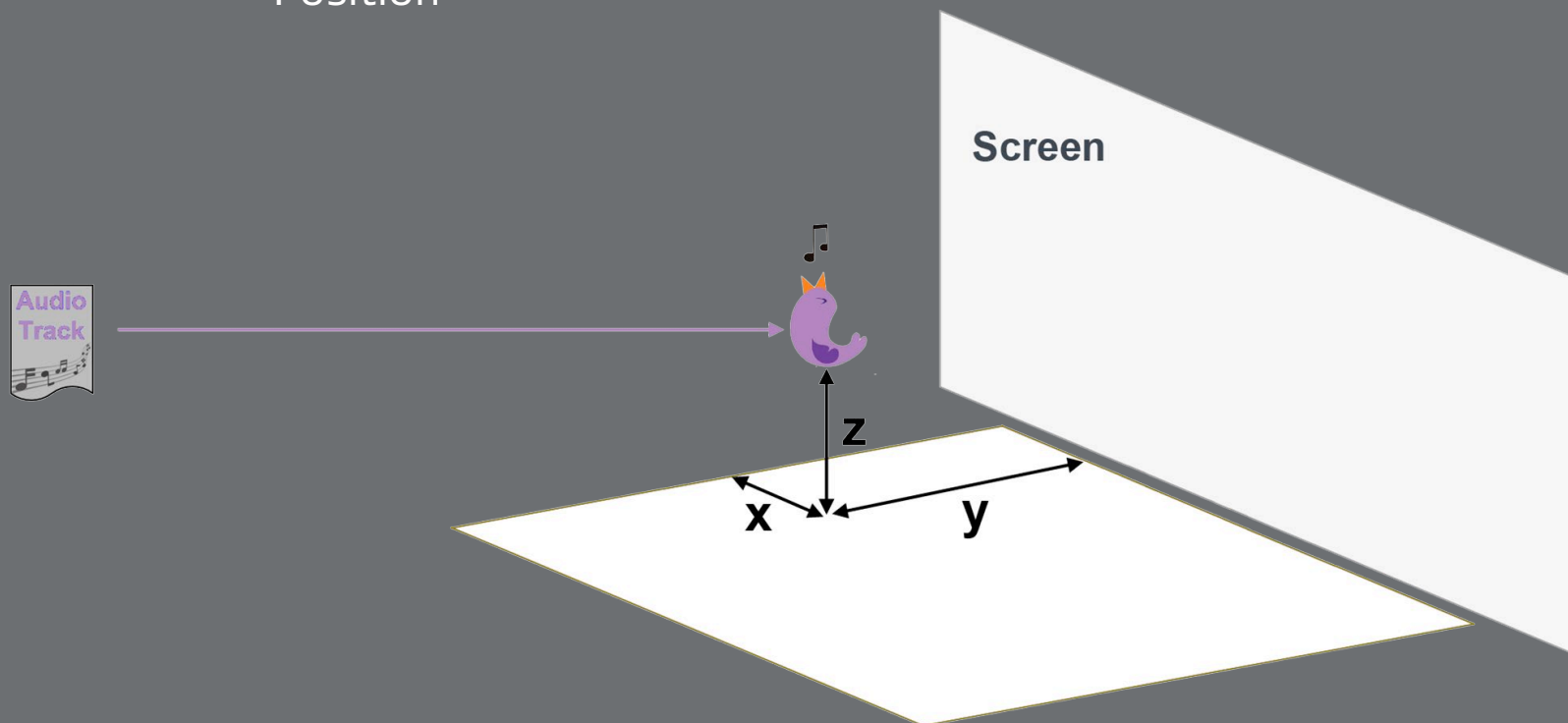


Spatial Audio Description: Object Based



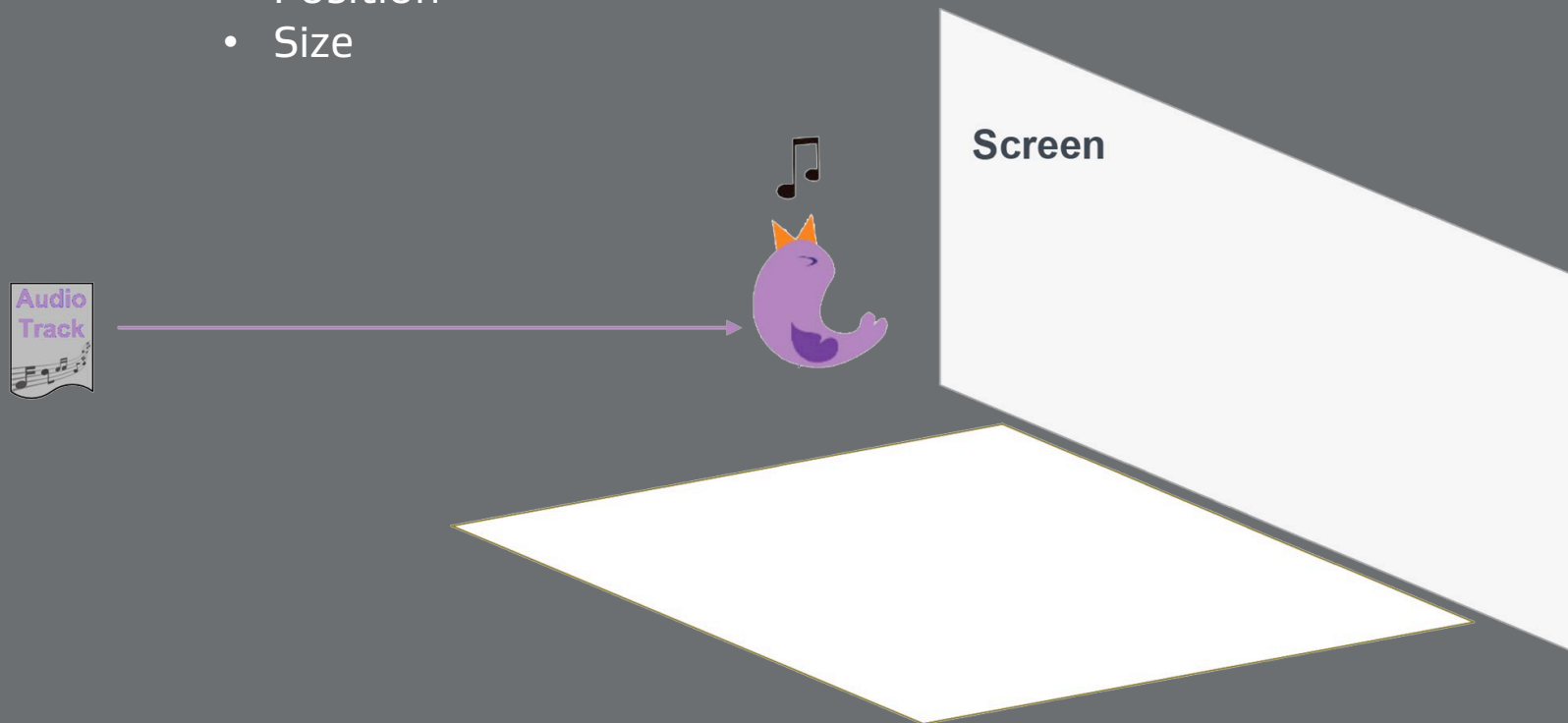
Spatial Audio Description: Object Based

- Position



Spatial Audio Description: Object Based

- Position
- Size



Spatial Audio Description: Object Based

- Position
- Size
- Diffusion



Screen

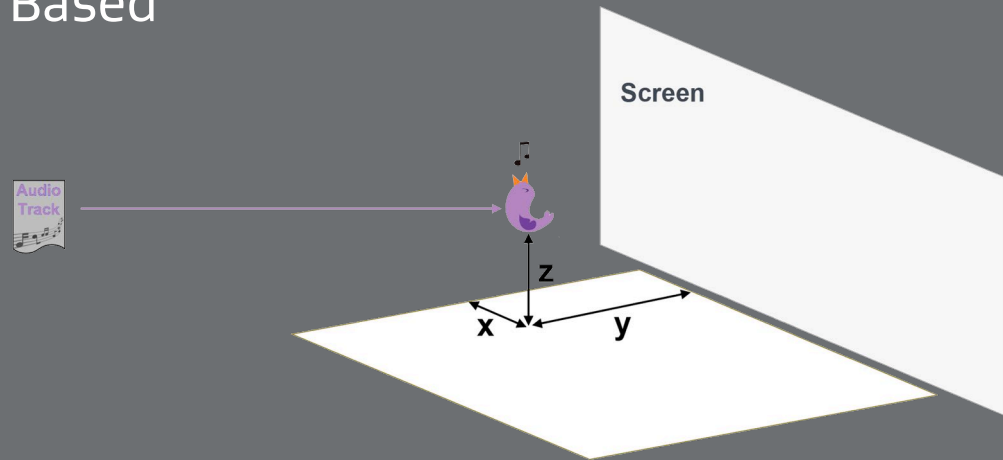
Spatial Audio Description: Object Based

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

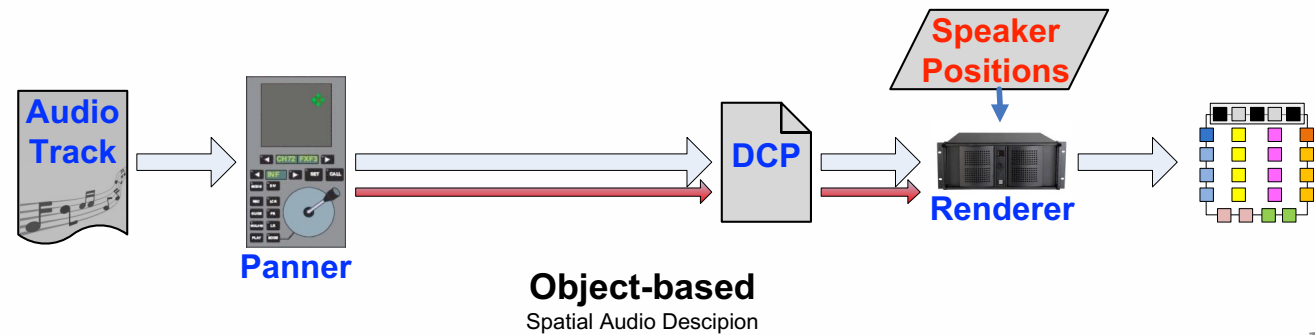
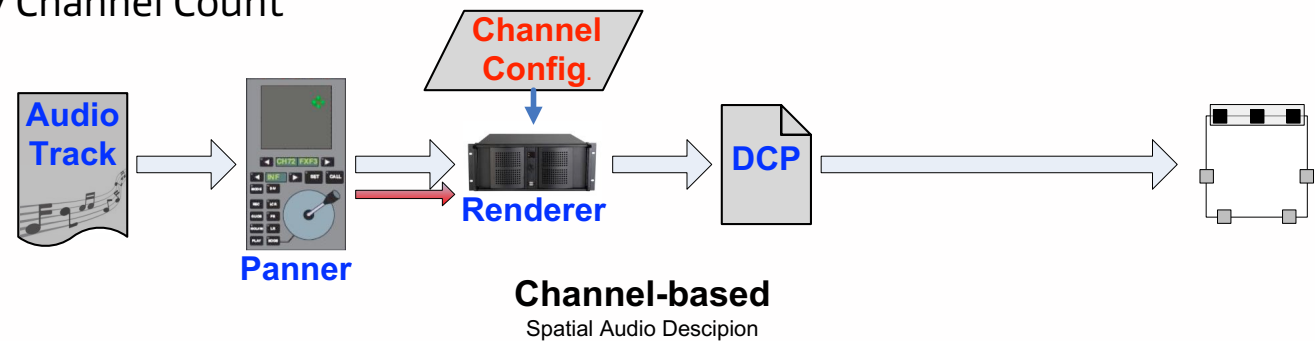
Liabilities of Channel Based

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

Spatial Audio Description

Liabilities of Channel Based

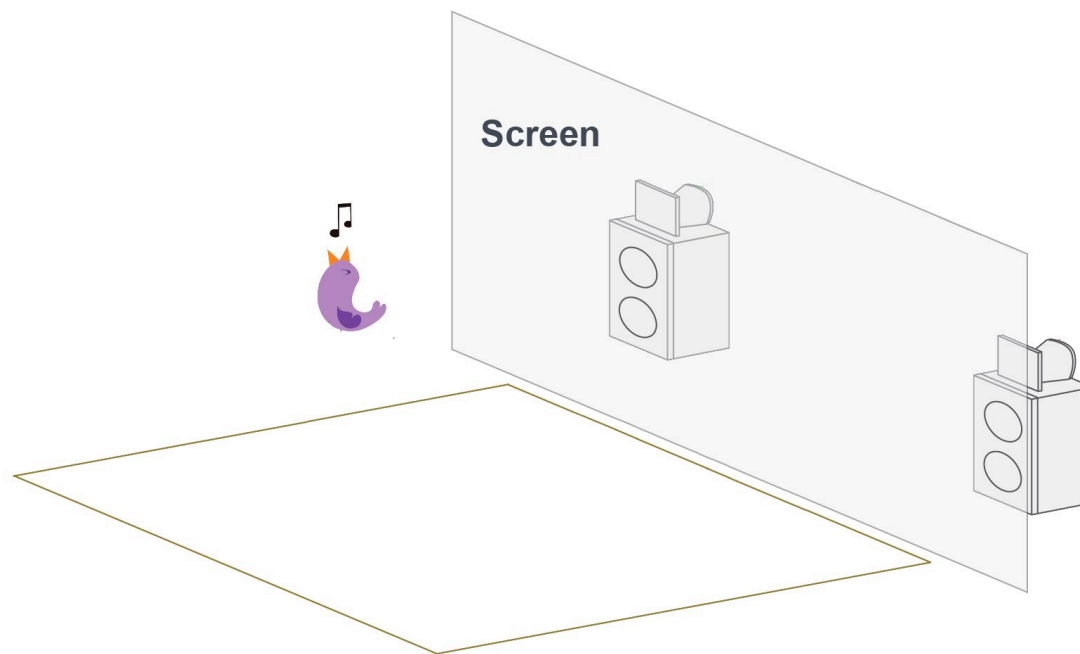
- Spatial Resolution Limited by Channel Count
- Not Scalable or Adaptable



Spatial Audio Description

Benefits of Object Based

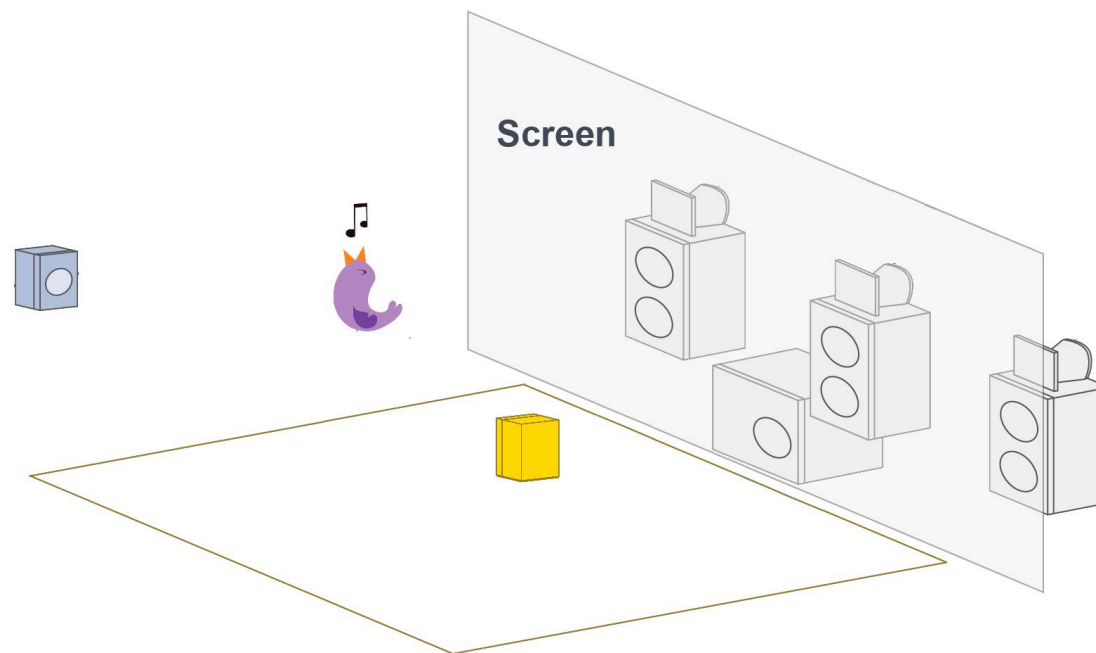
- Scalable



Spatial Audio Description

Benefits of Object Based

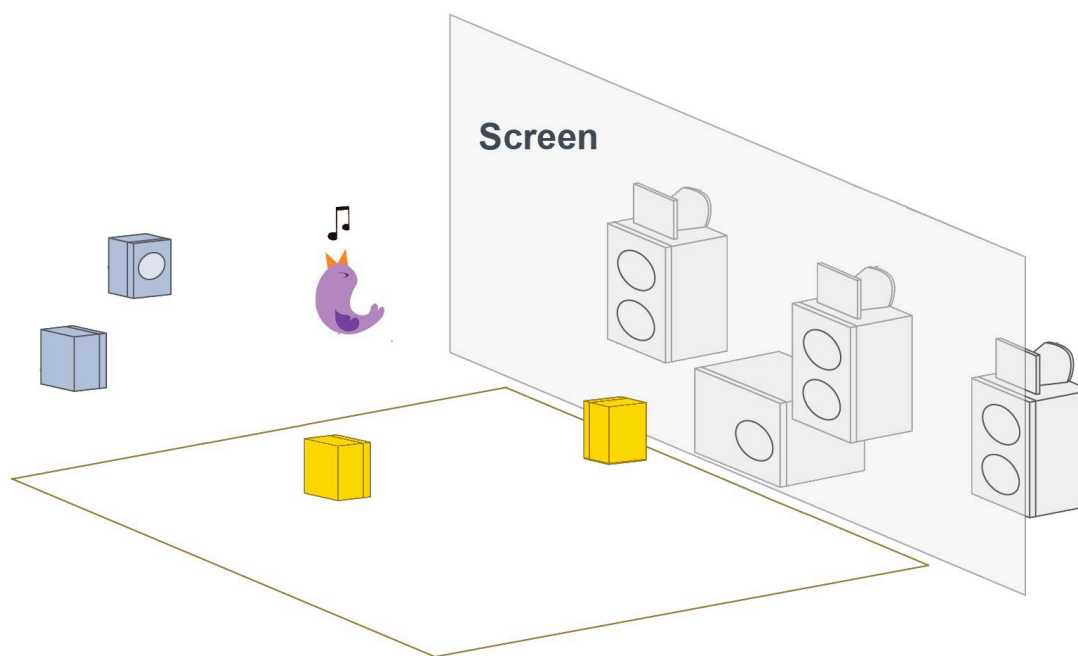
- Scalable



Spatial Audio Description

Benefits of Object Based

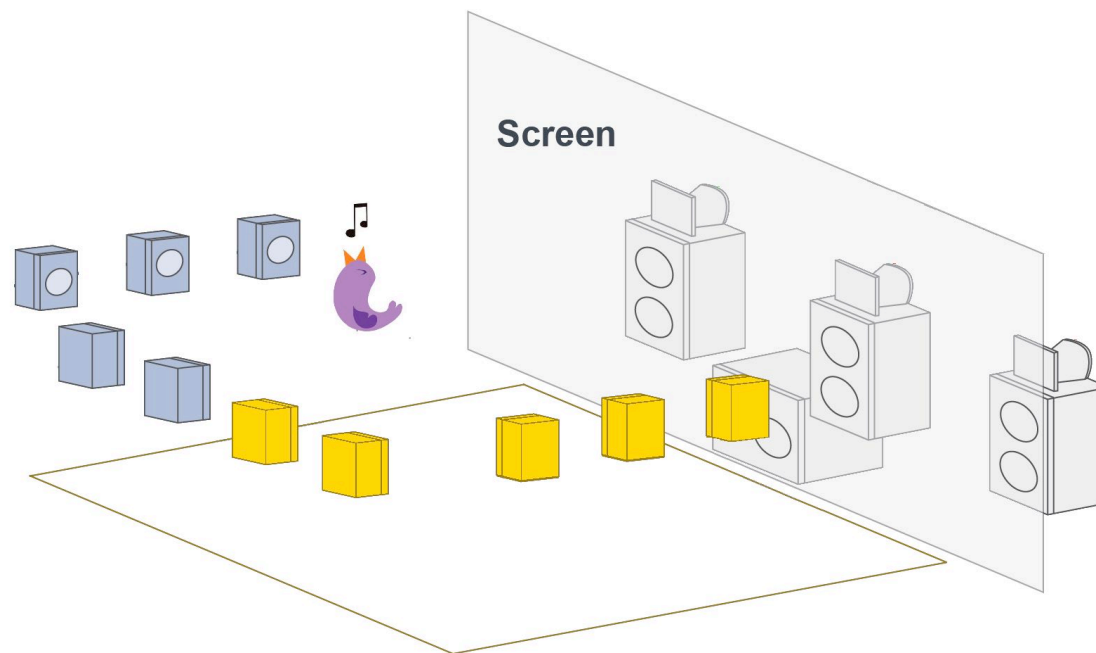
- Scalable



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



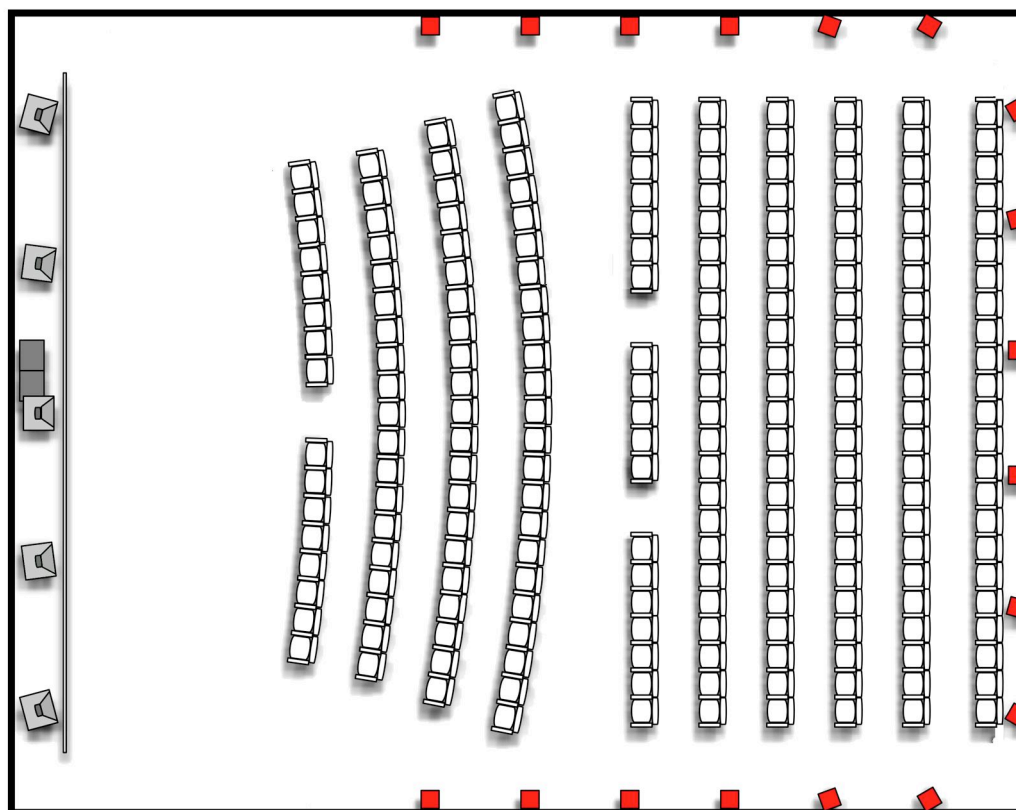
Dolby Atmos

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

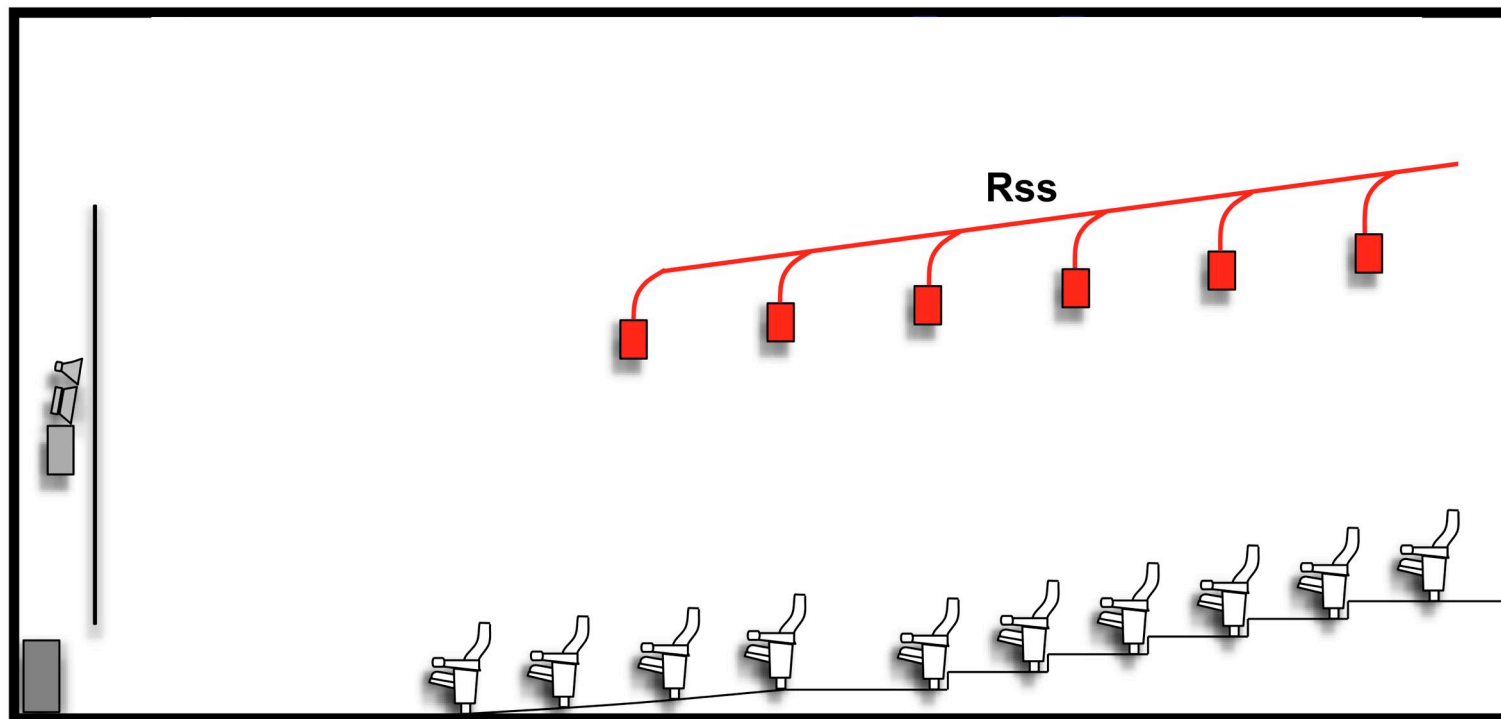


Loudspeaker Layout



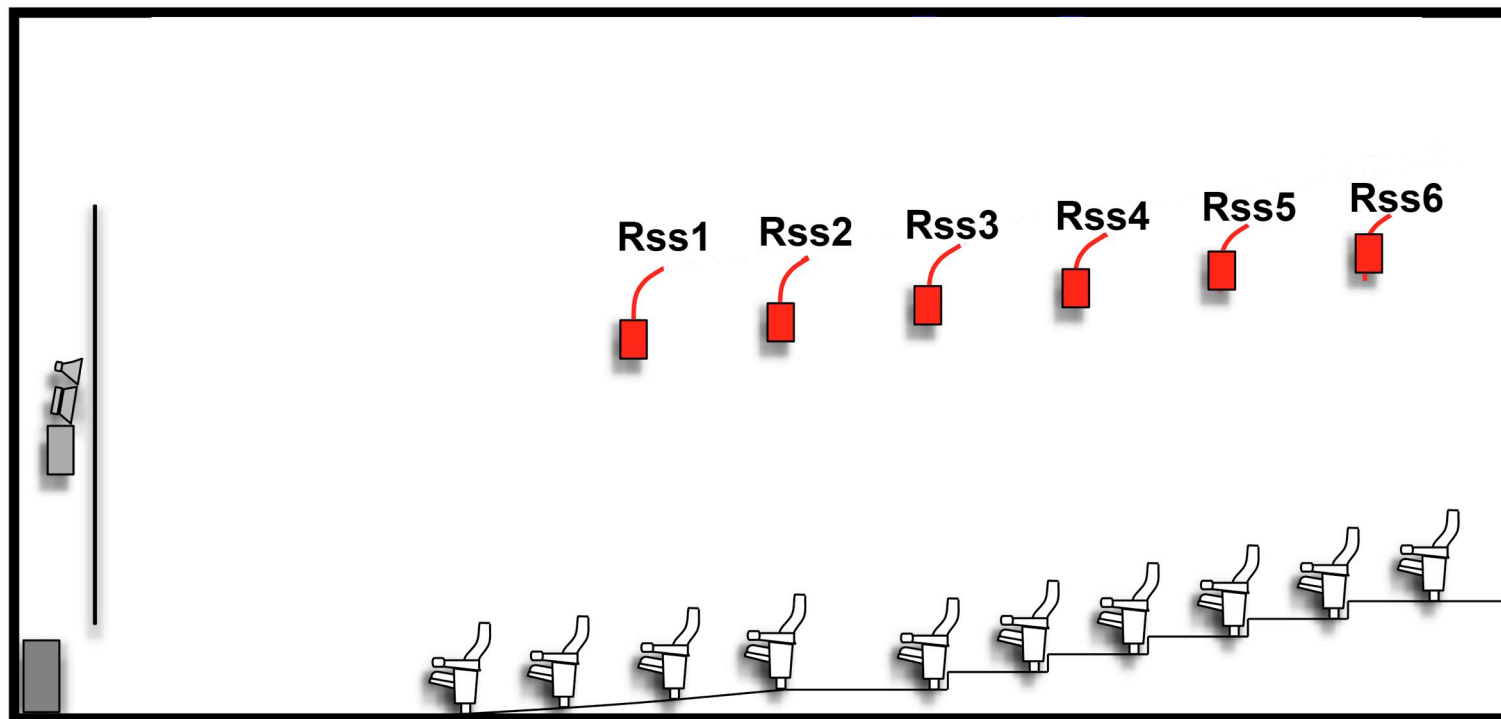


Loudspeaker Layout



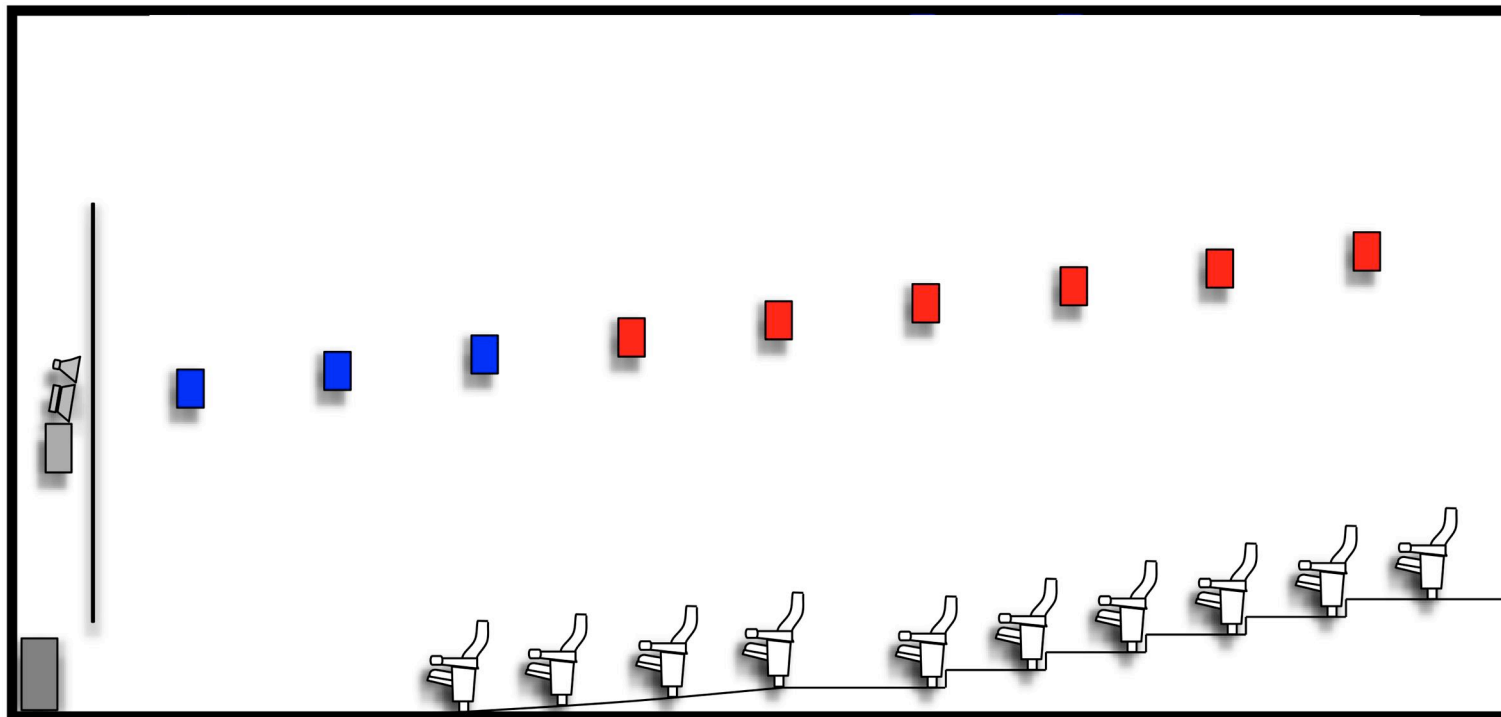


Loudspeaker Layout





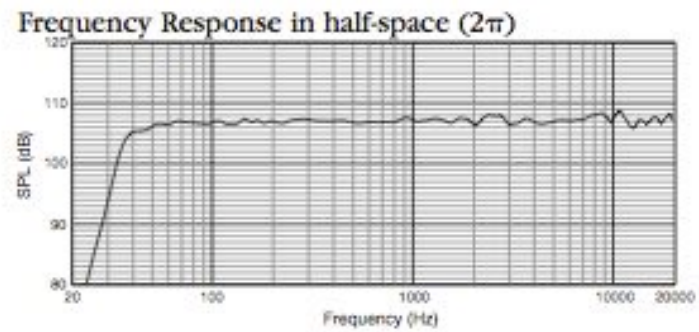
Loudspeaker Layout



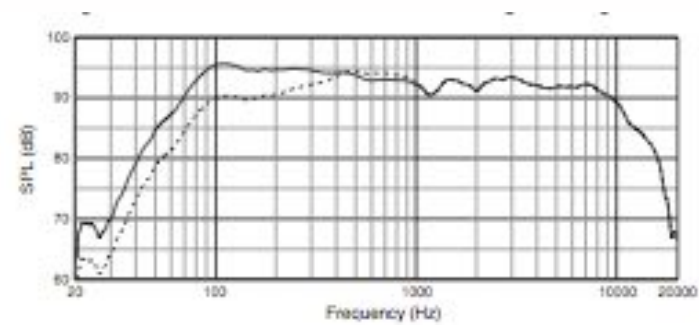


Loudspeaker Layout

Screen Loudspeaker

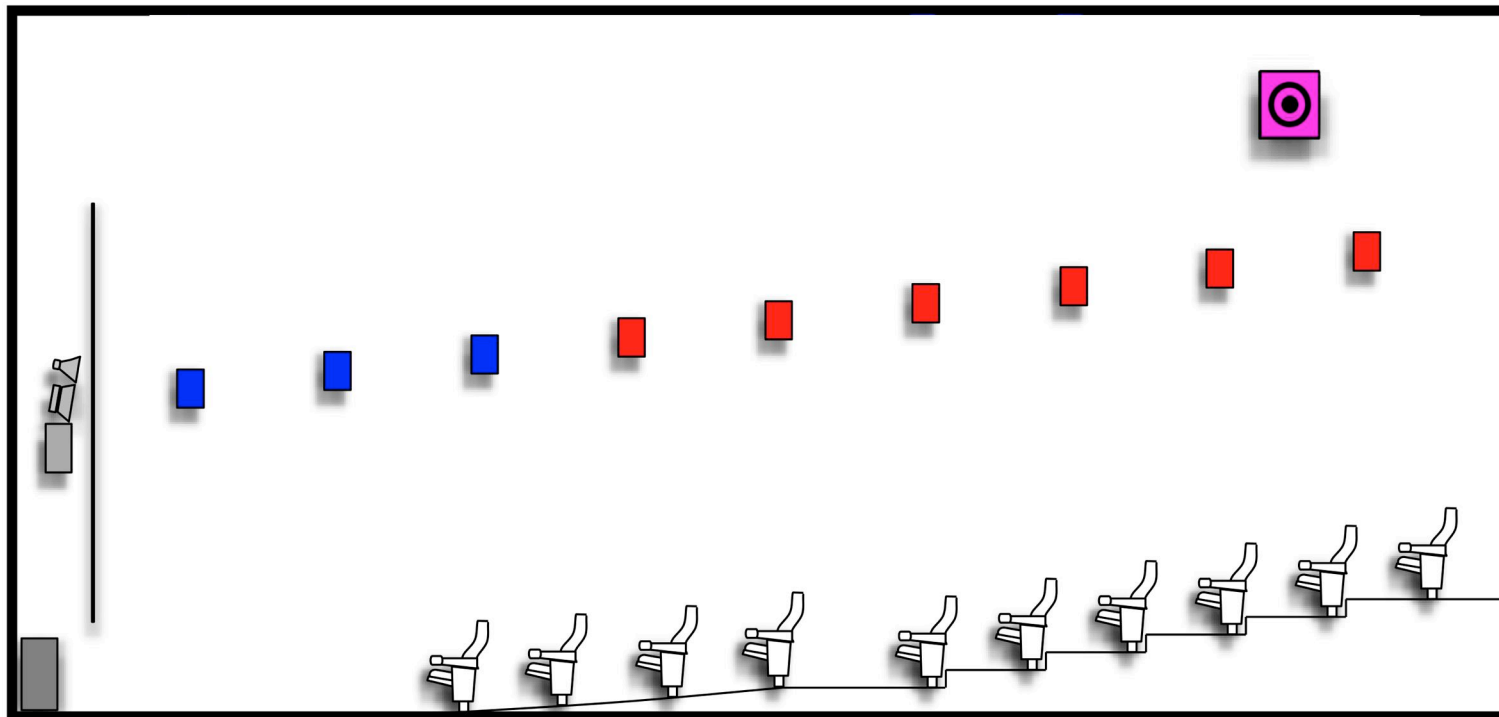


Surround Loudspeaker



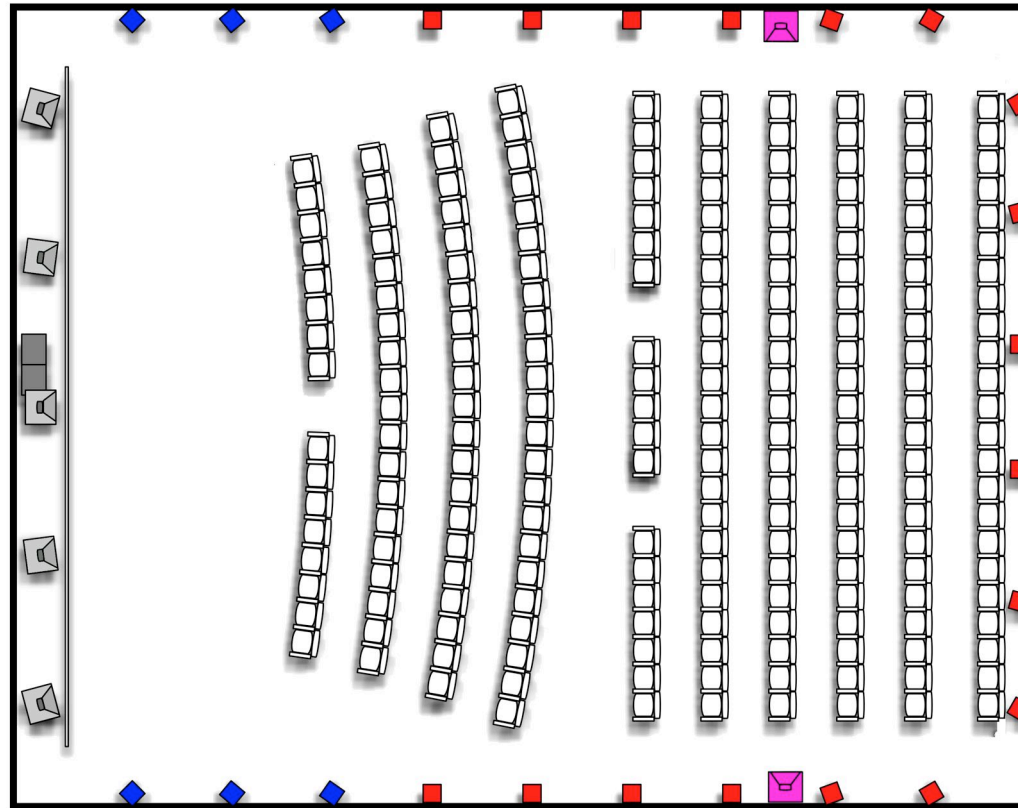


Loudspeaker Layout



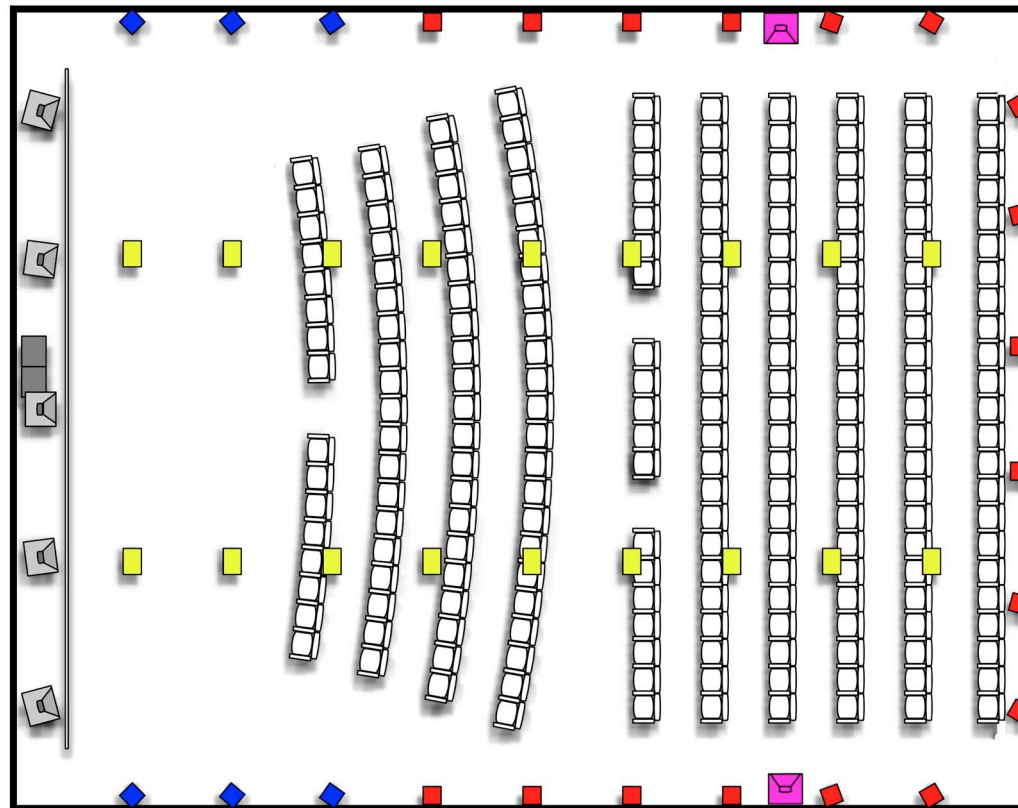


Loudspeaker Layout



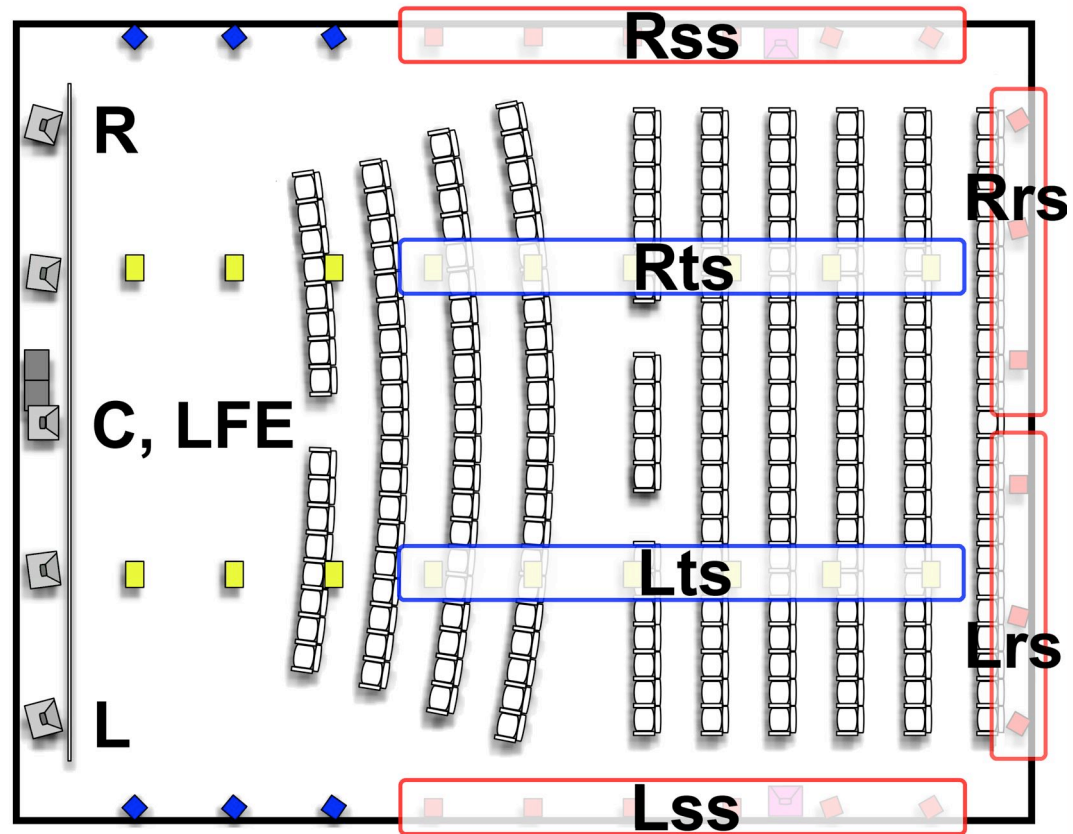


Loudspeaker Layout



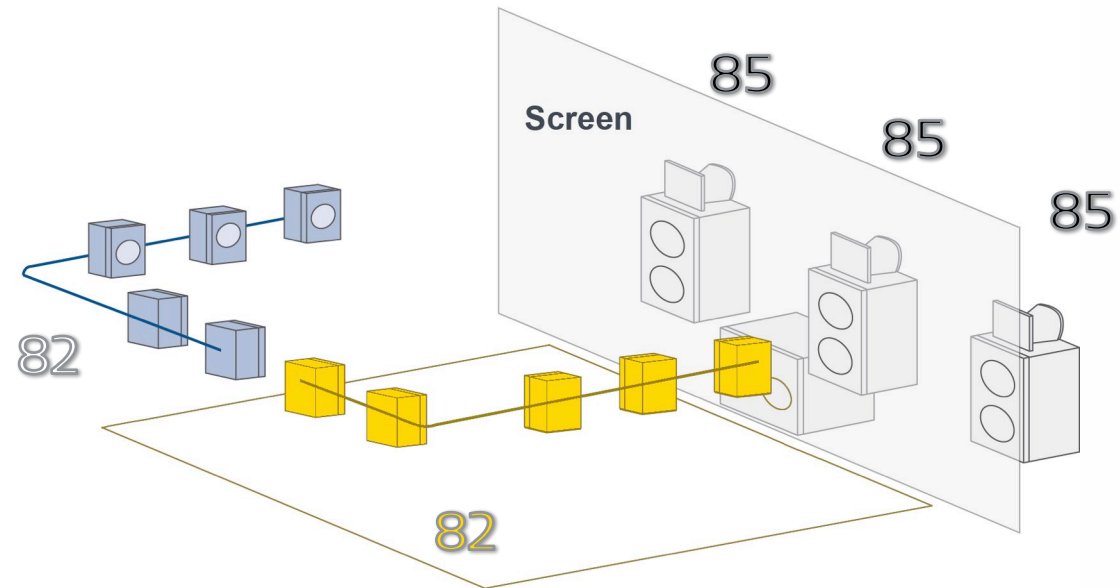


Loudspeaker Layout



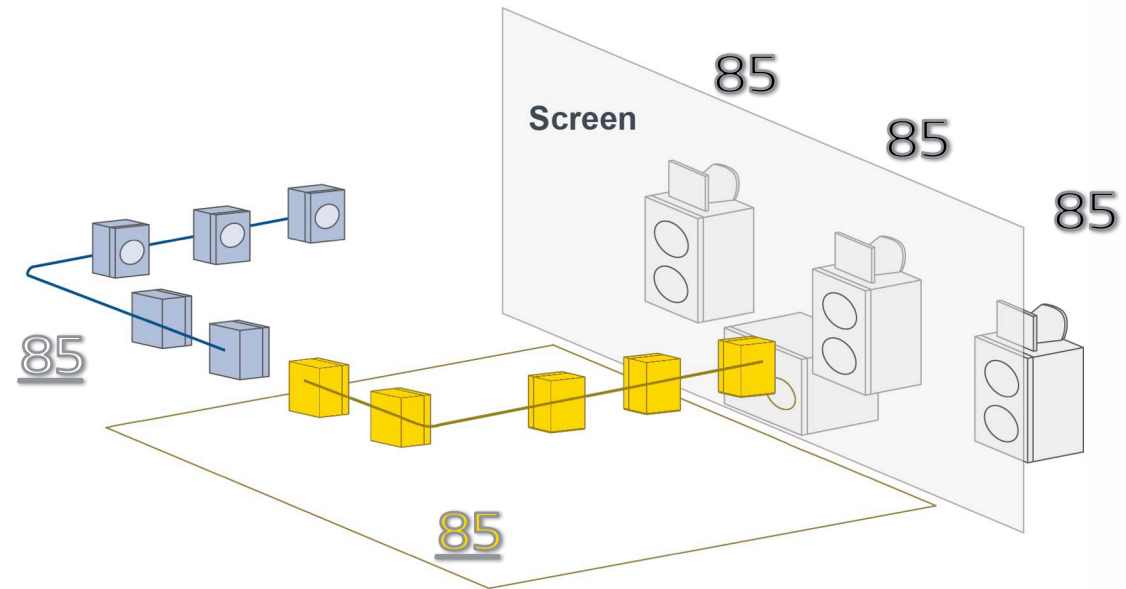


Reference Level





Reference Level





Mixer Survey

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



322

© 2016-17 DOLBY LABORATORIES, INC.

Experimental Mixes

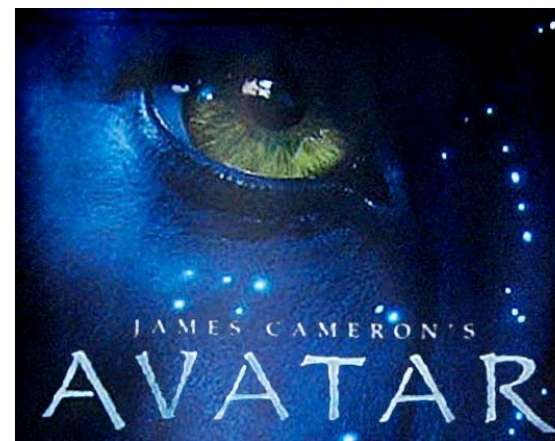
Incredibles

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

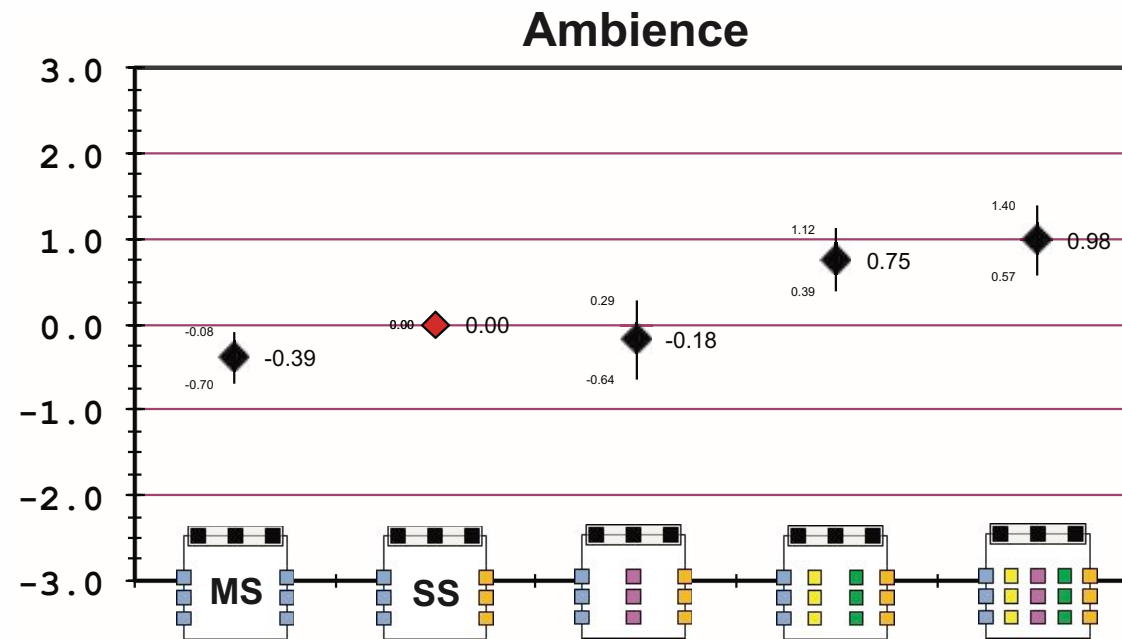


Avatar

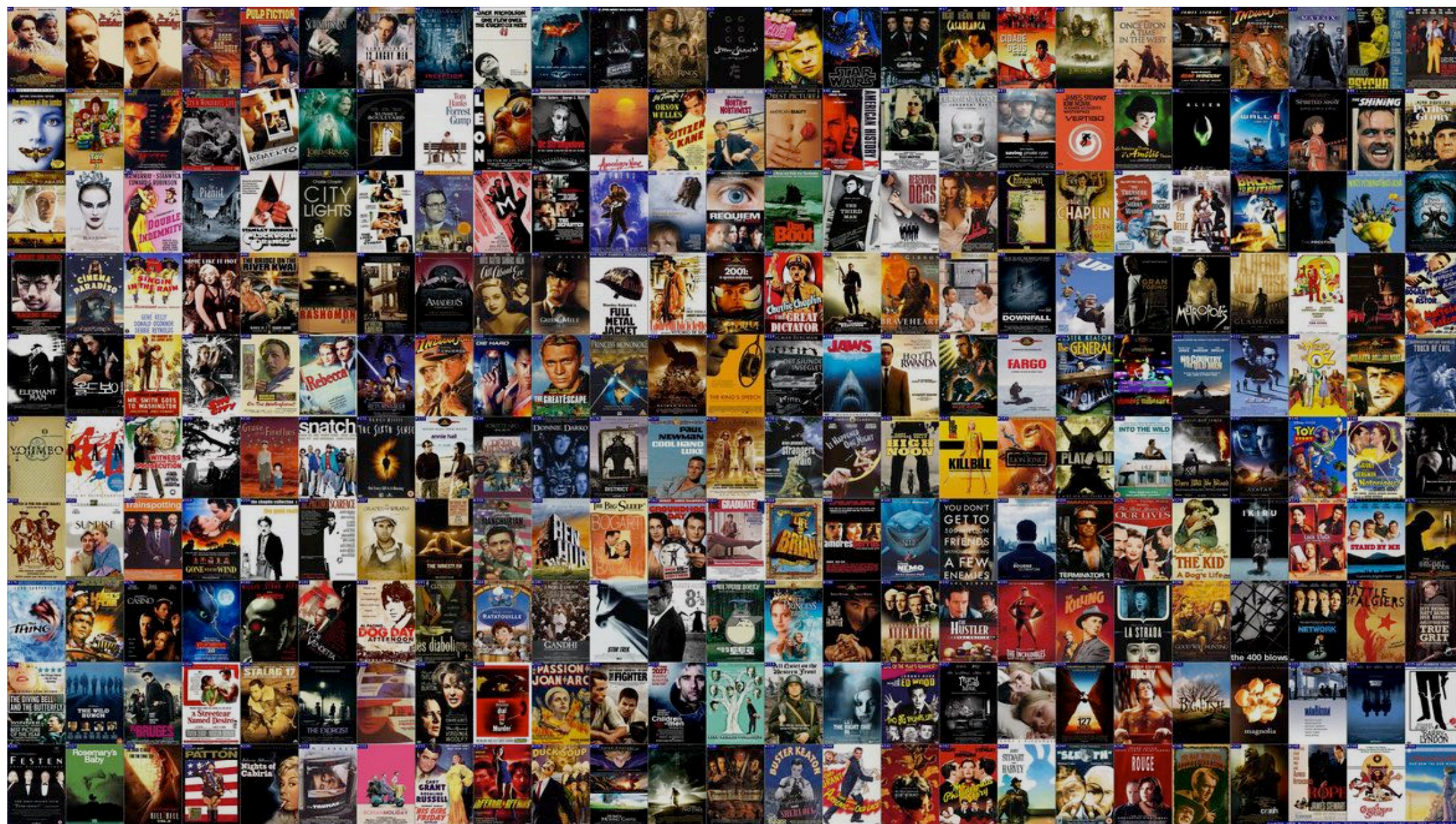
- Skywalker Sound
- Gary Summers
- 3D Audio Panner
- Screen and Surround Height, and Surround Point Source (surround direct)



Double Blind Subjective Tests

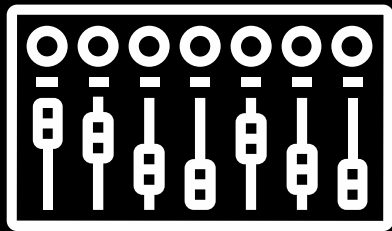




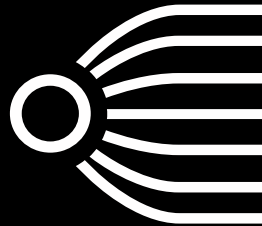




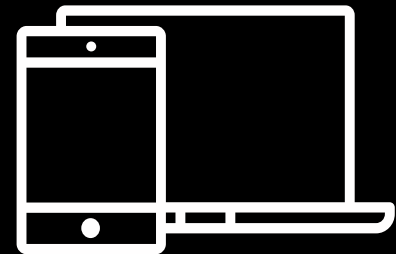
WORLD OF ATMOS



Content creation



Distribution



Playback Devices



BRINGING IMMERSIVE AUDIO TO THE HOME



TVs



CINEMA



PCs



HOME THEATER



BLU-RAY™



MOBILE DEVICES



SOUND BARS



AVRs

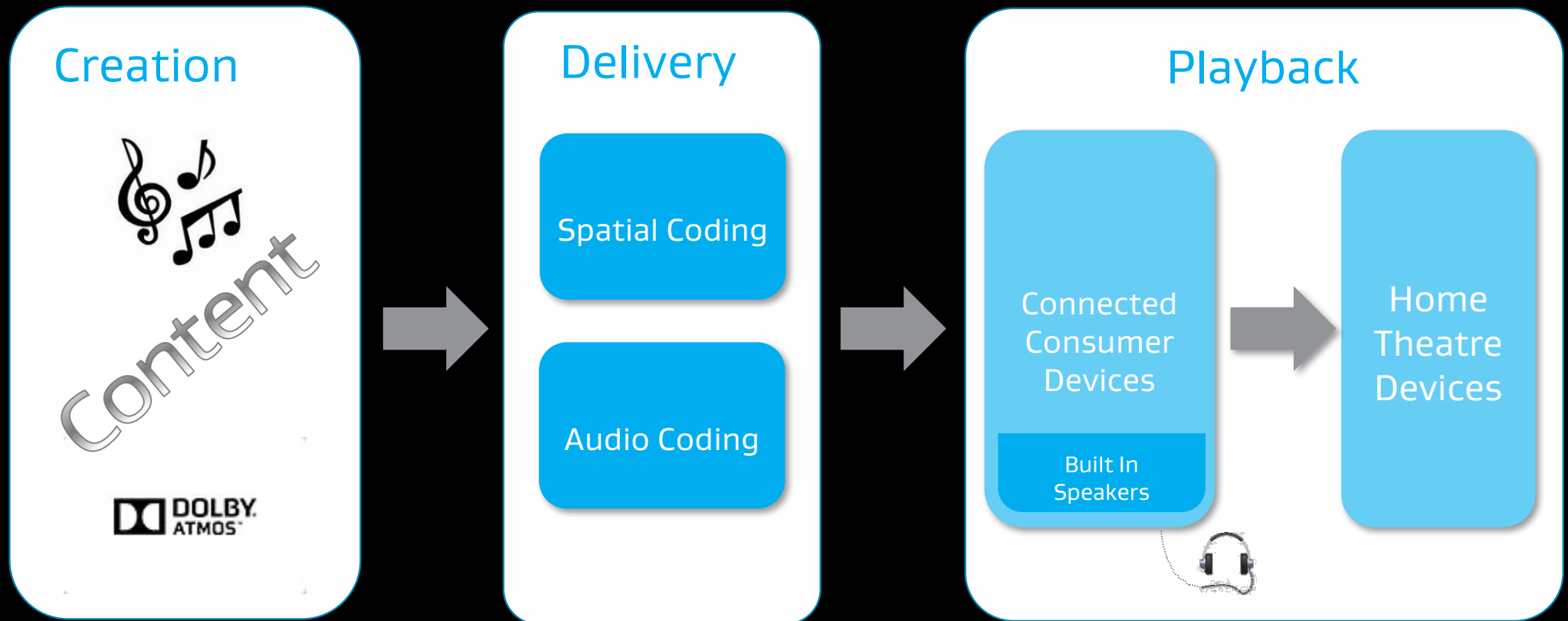


VIRTUAL REALITY



GAMING

Bringing Atmos to the Home



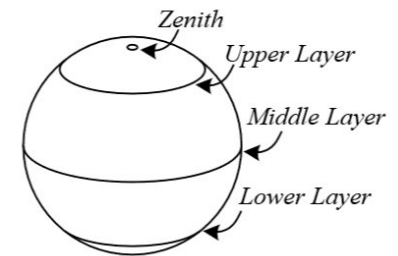
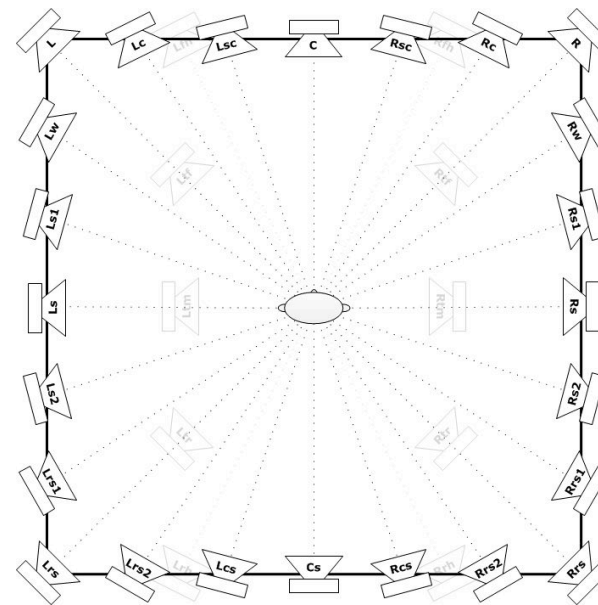
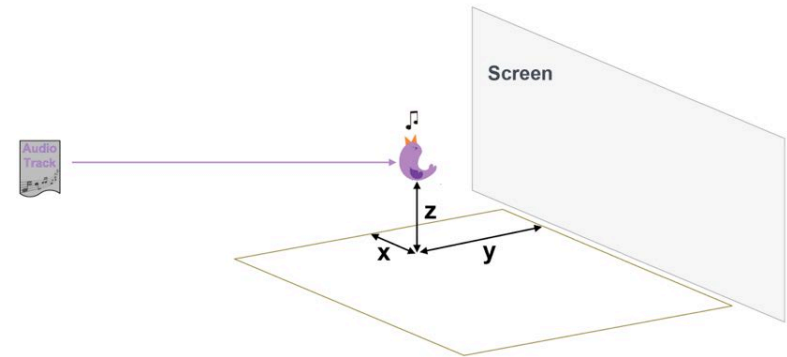
DOLBY ATMOS AUDIO

An Atmos scene is made up of audio objects

Audio objects consist of a audio stream and associated metadata

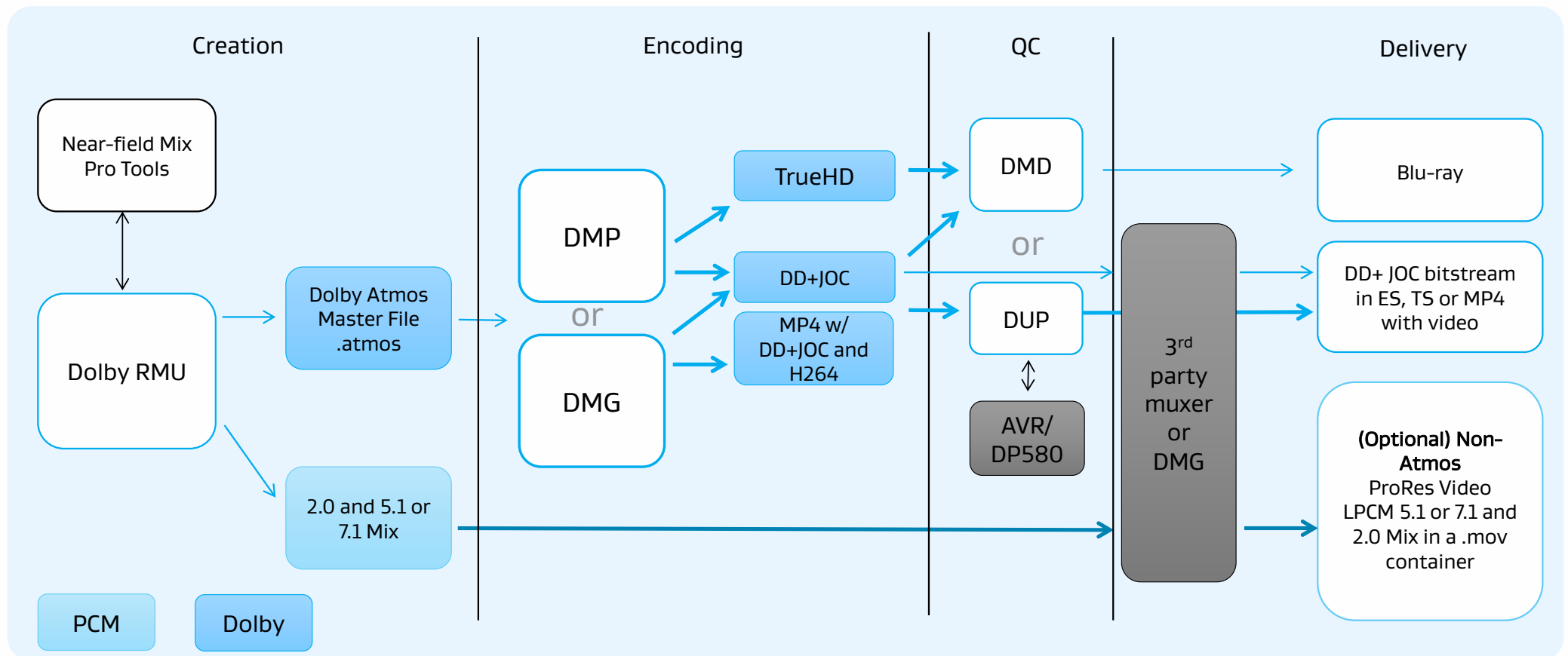
Up to 159 discrete objects can be represented during content creation

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



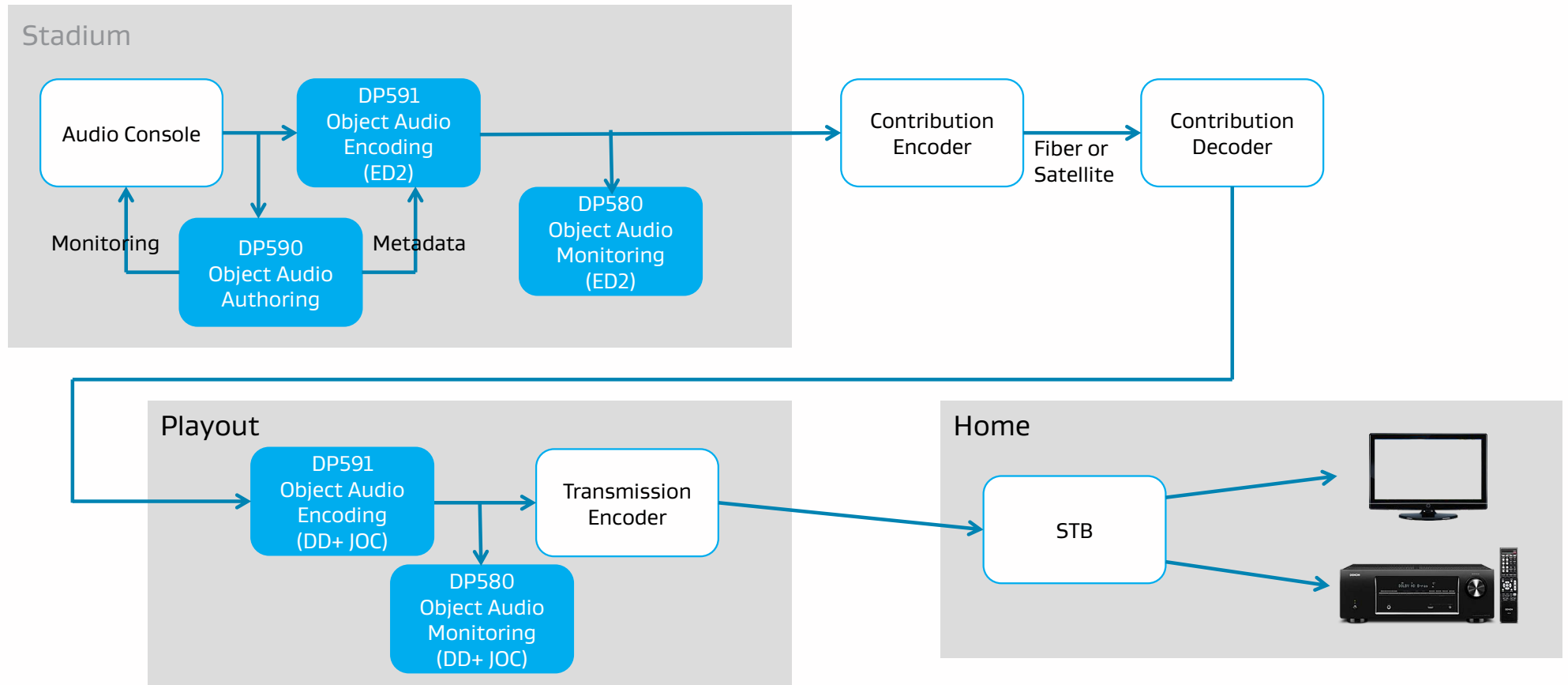


HOME CINEMATIC CONTENT CREATION

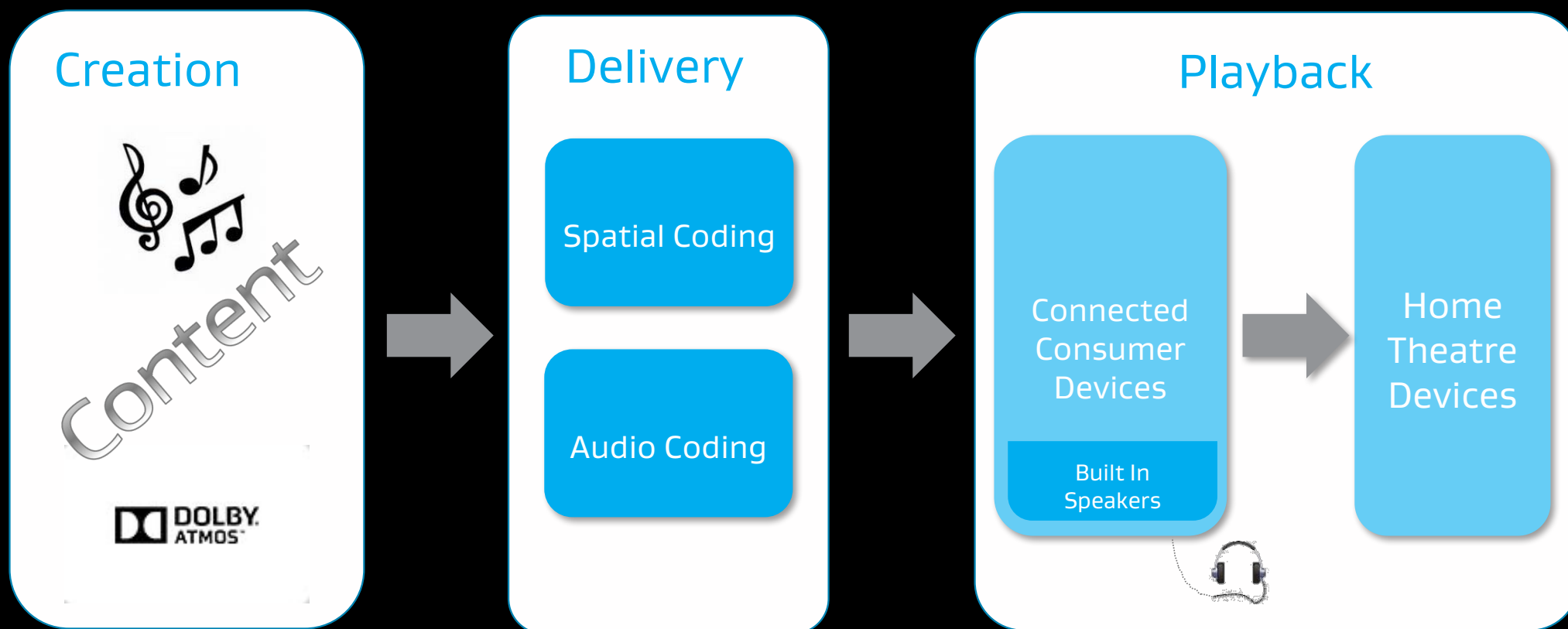




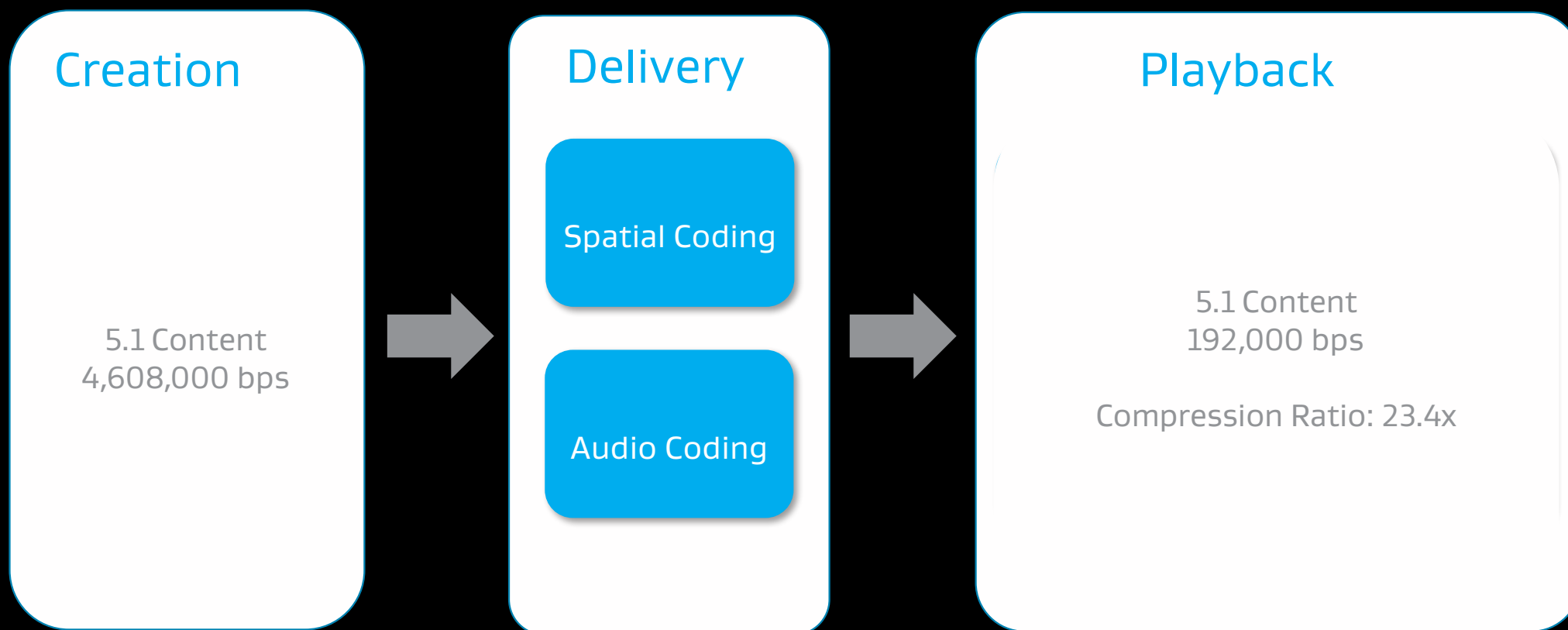
EXAMPLE LIVE WORKFLOW



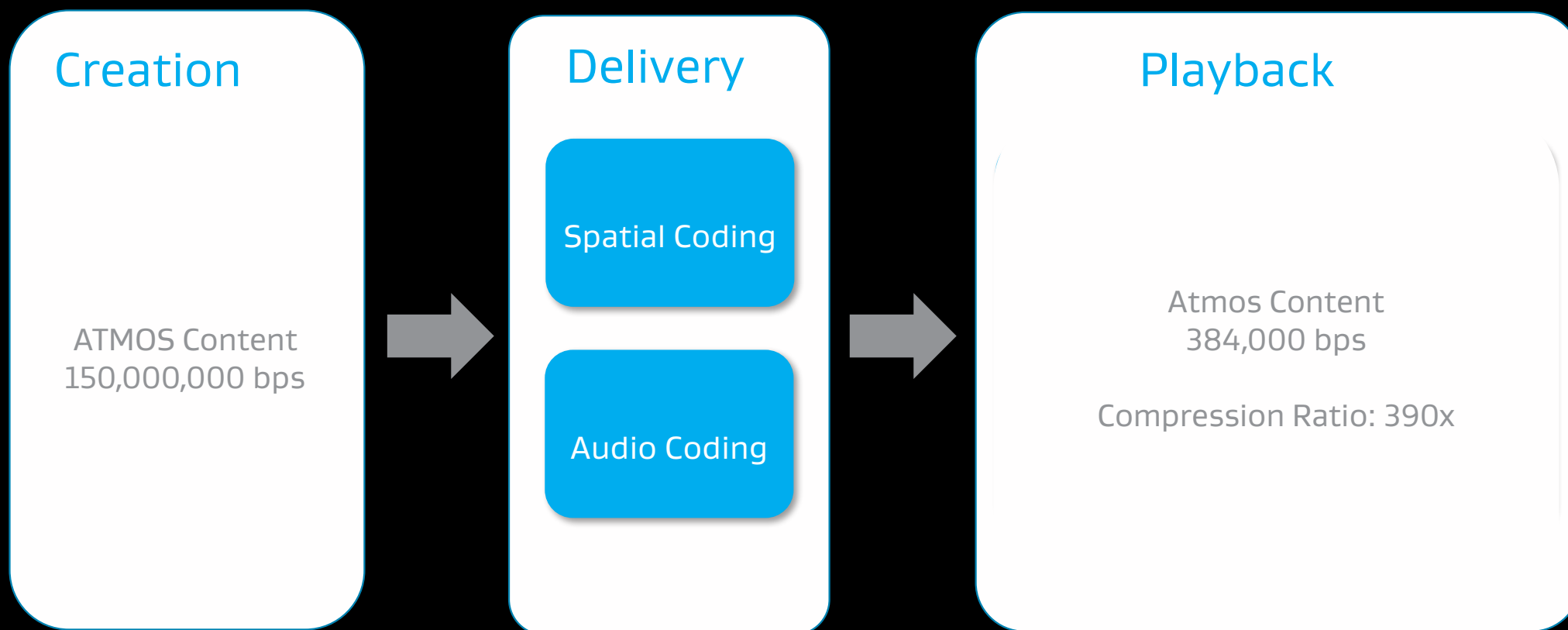
Bringing Atmos to the Home



The Challenge: Delivering 5.1



The Challenge: Delivering Atmos





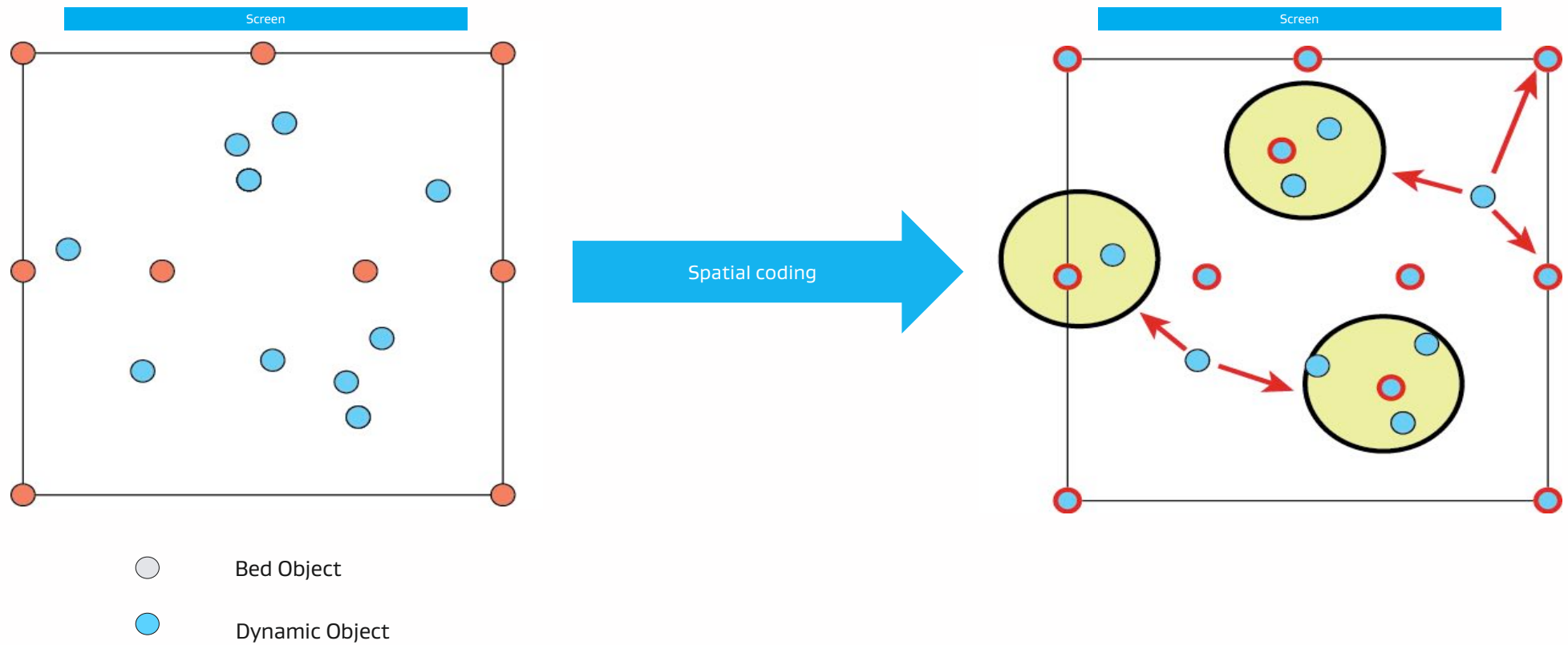
DATA REDUCTION PART 1

90-minute Object-based (OBI) Movie – 100 Object-Channel/Track Stats

Format	Bitrate	Peak-bitrate	File Size (90 minute)	AES3 Pairs Required	Metadata Channel	Metadata Sync Complexity
PCM+Metadata	Constant	115.2Mb/s	77.76GB	50	OOB	Extremely High

Immersive content requires new thinking!

DATA REDUCTION PART 1 – 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





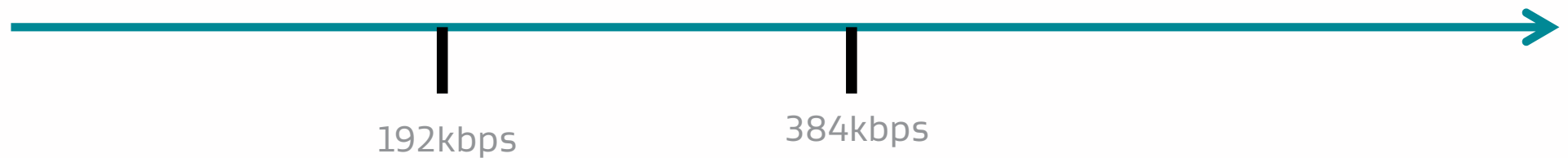
Dolby Digital Plus JOC



Dolby Digital Plus Audio Frame

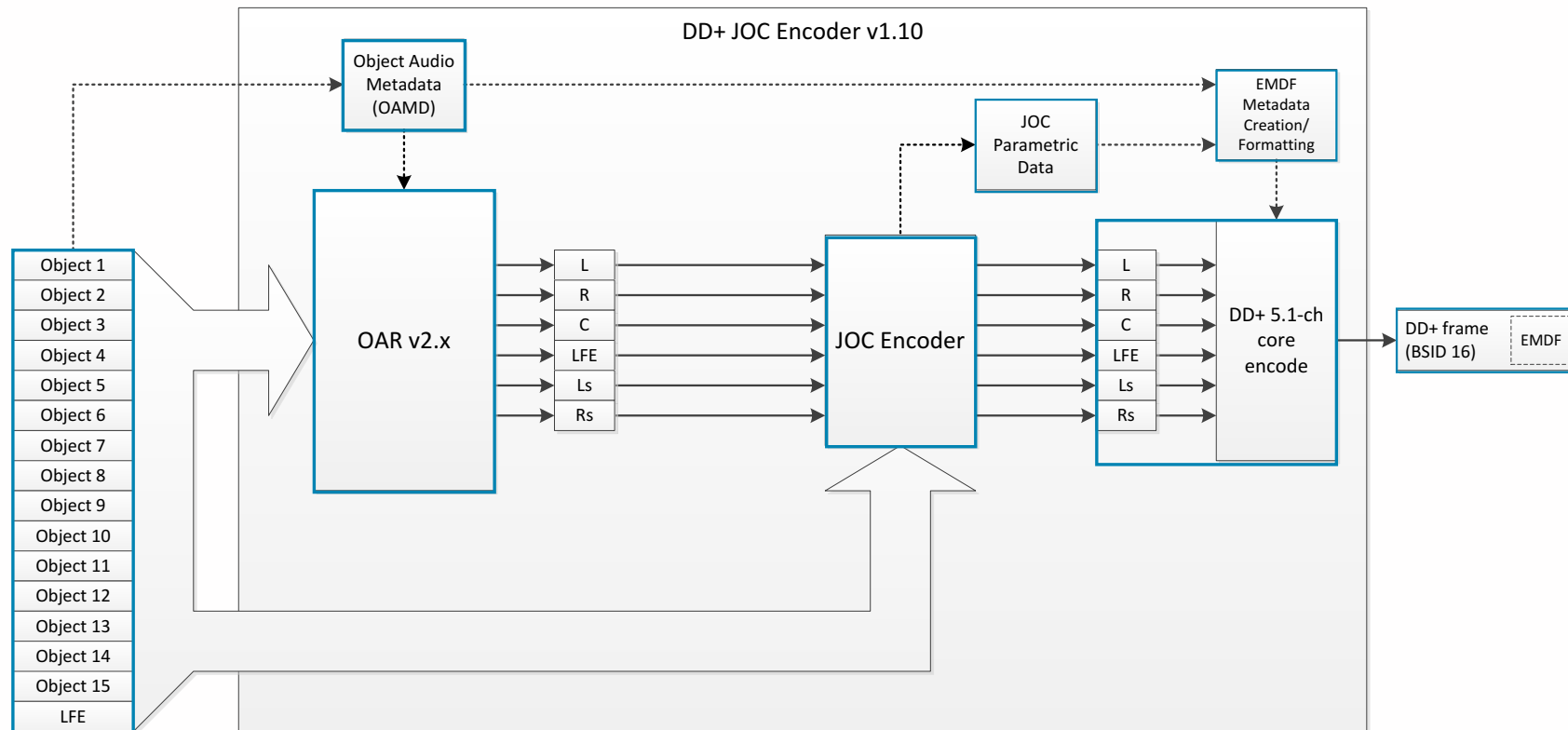


Dolby Digital Plus Audio Frame





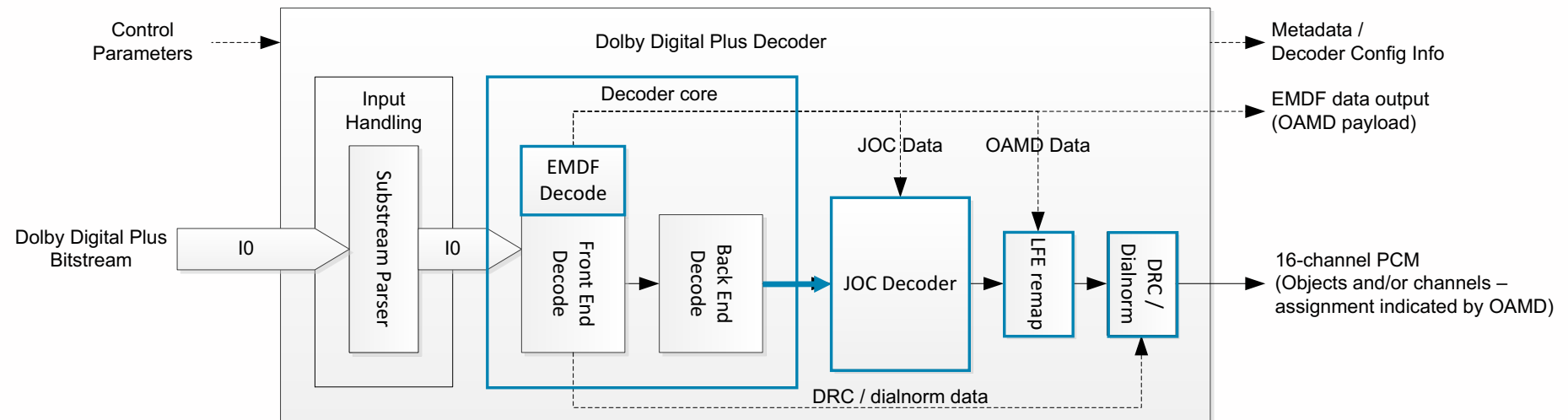
DD+ JOC Encoder: Object Based Immersive (OBI) content



CONFIDENTIAL INFORMATION



DD+ JOC Decoder (DD+ JOC bitstream, 5.1-channel core)

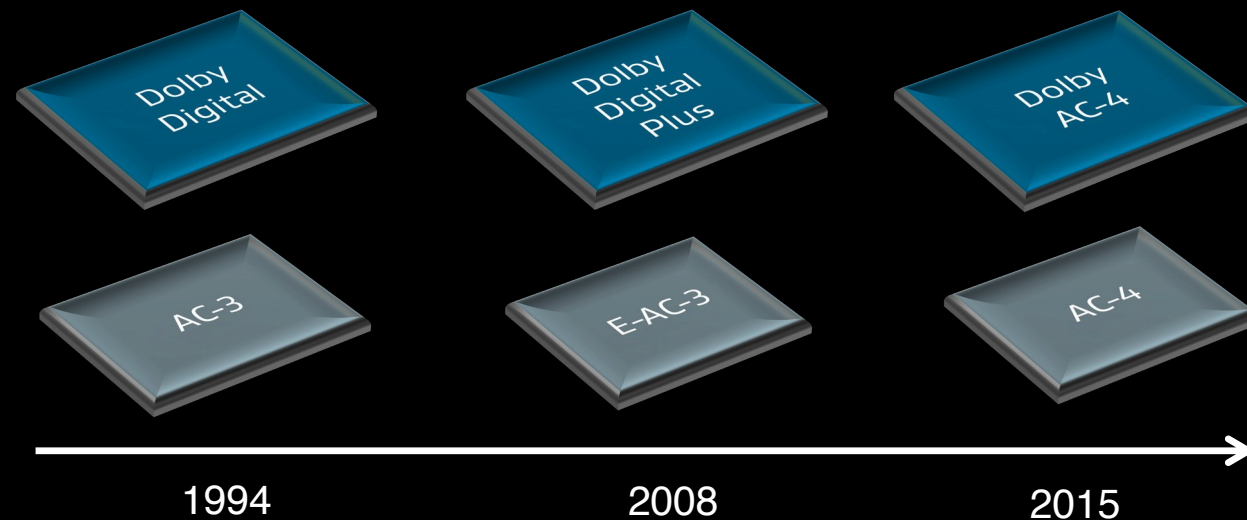


CONFIDENTIAL INFORMATION

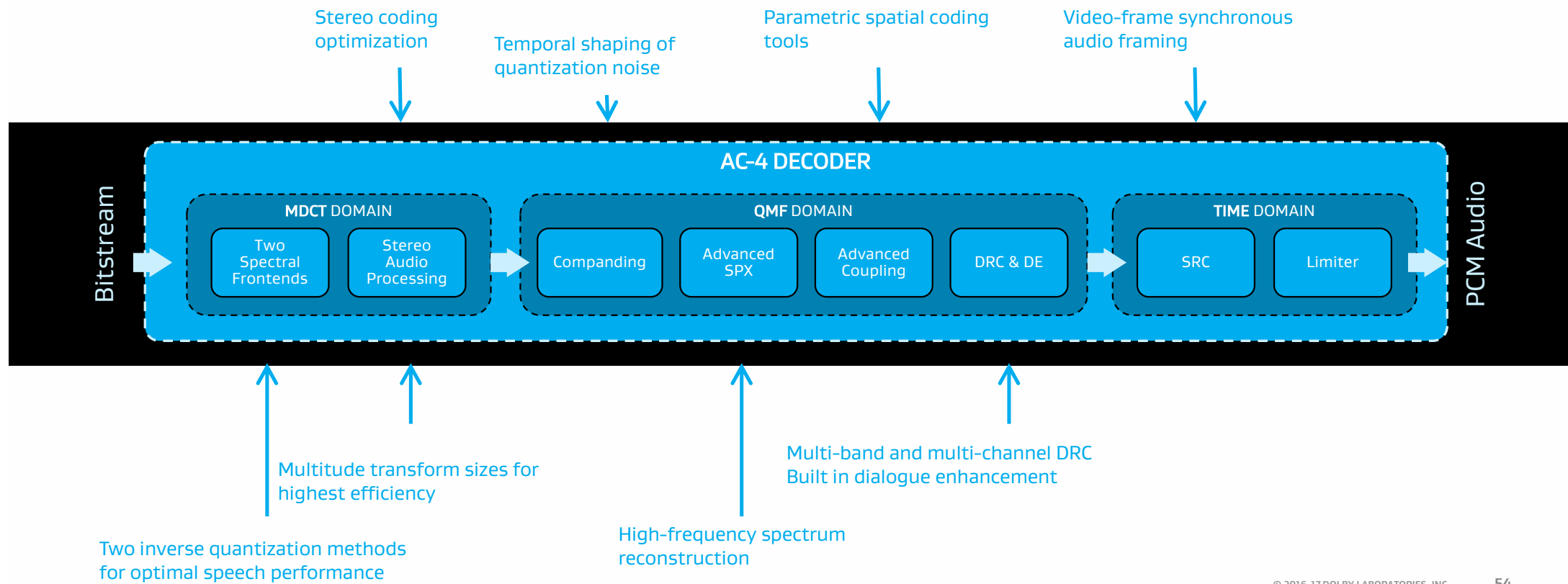


DELIVERY – DOLBY AUDIO CODECS

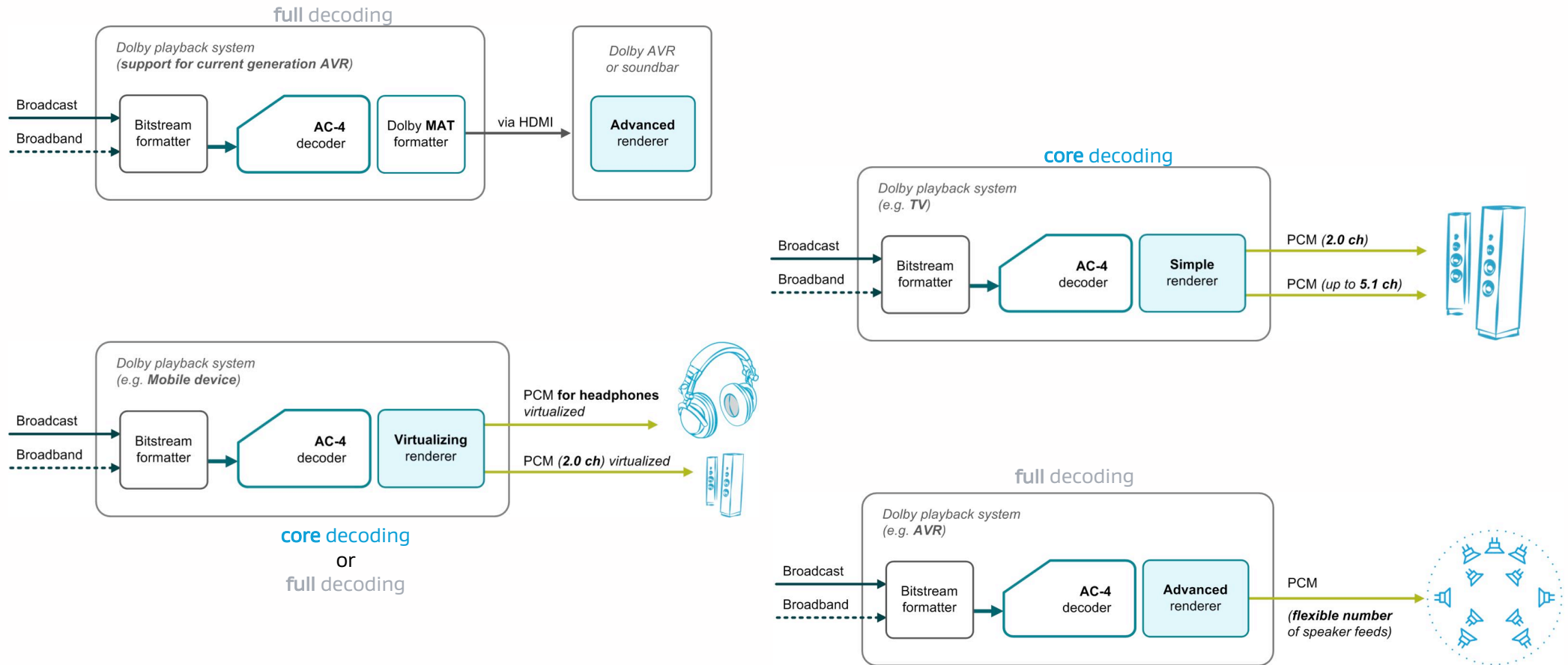
Dolby AC-4 is the new audio coding system that improves today's stereo and 5.1 experiences and enables efficient delivery of the next generation audio experiences like immersive and personalized



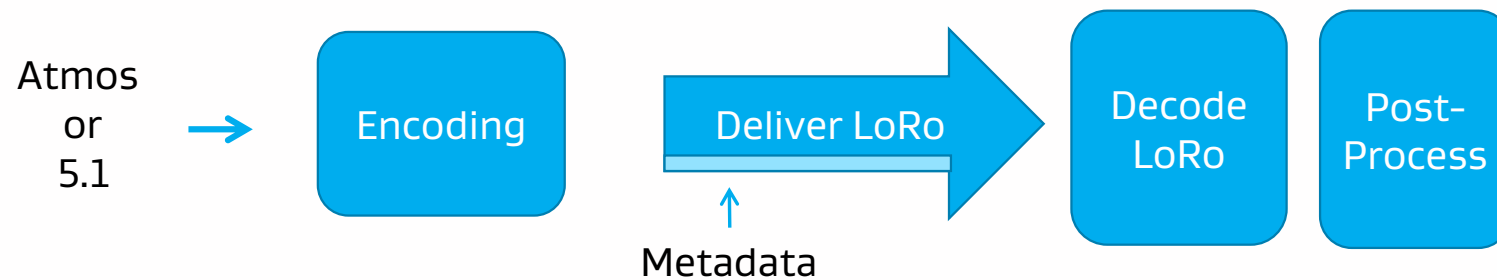
AC-4 – State of the art coding tools



AC-4 – Flexible Support for Multiple Device Types



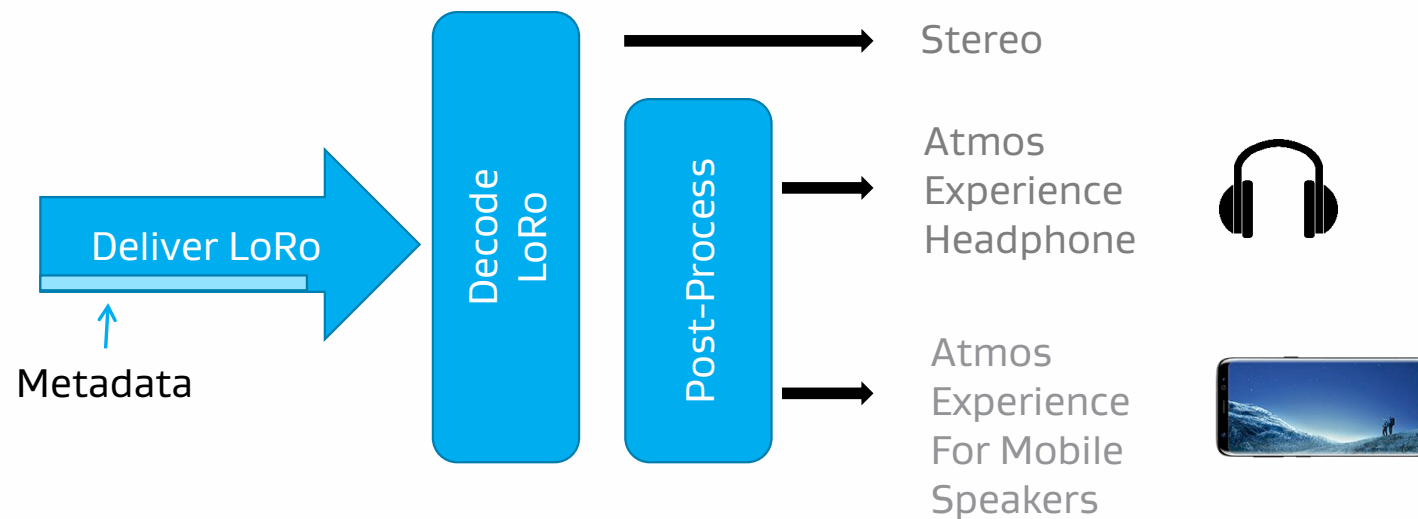
AC-4 – Immersive Stereo – Overview



- Low Bitrate
- Low complexity upon playback
- Rendering flexible in encoder
- Adjustment to listener playback condition
- LoRo compatible

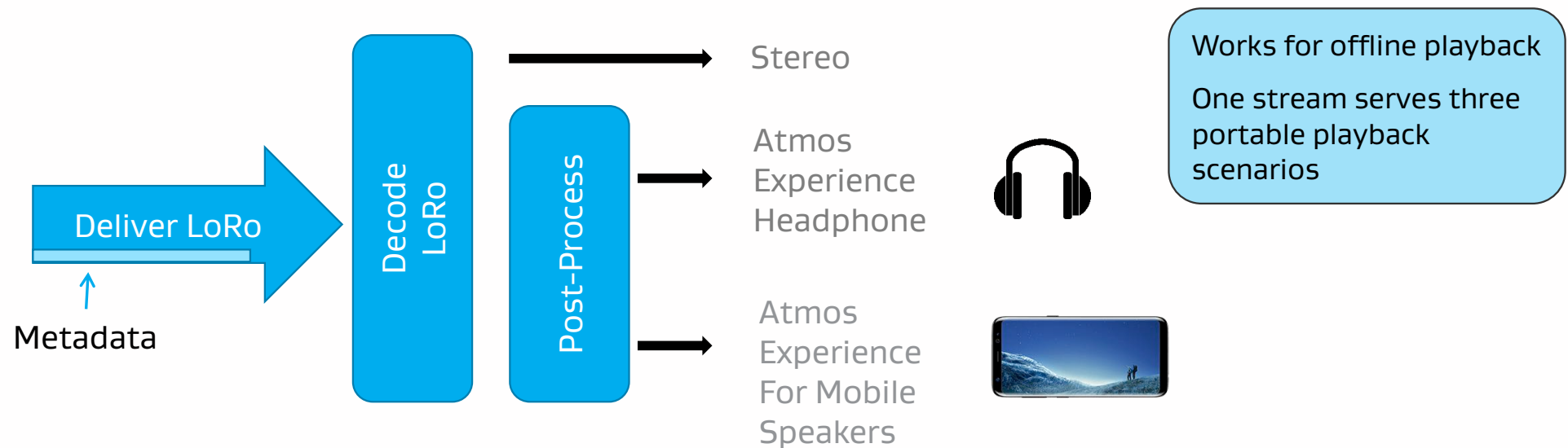
AC-4 – Immersive Stereo – Playback

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

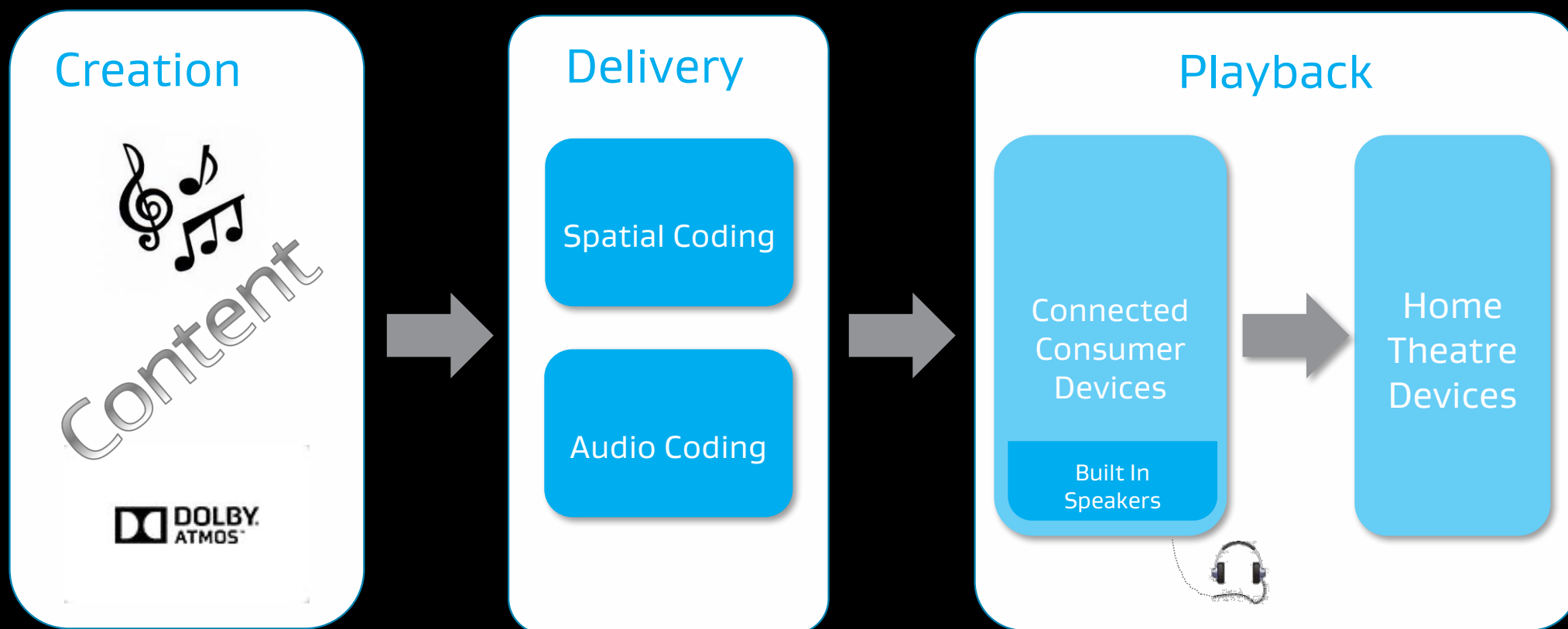


AC-4 – Immersive Stereo – Playback

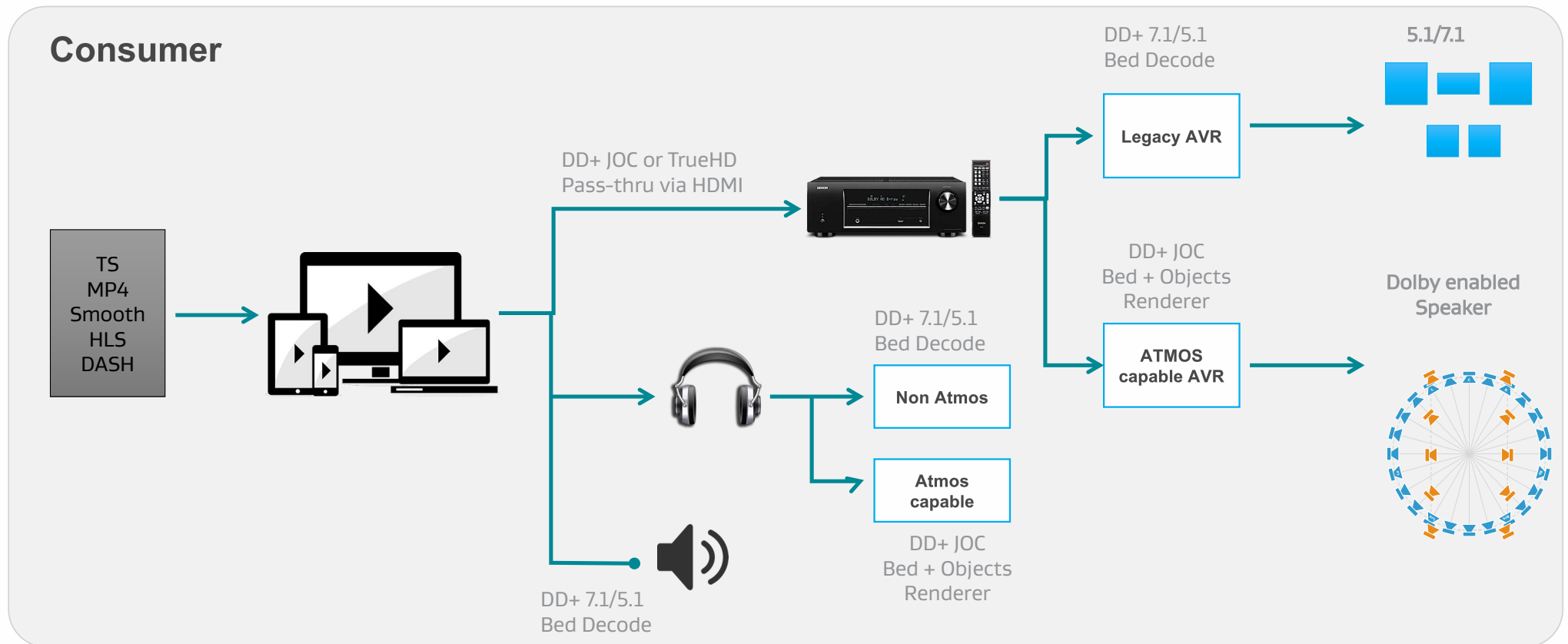
Upon playback the decoder provide 3 different outputs to best suit the current playback condition



Bringing Atmos to the Home

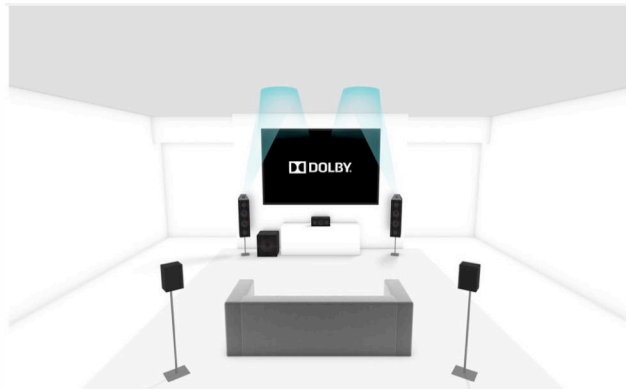


Device Playback Workflow



Dolby Atmos – Speaker Playback Examples

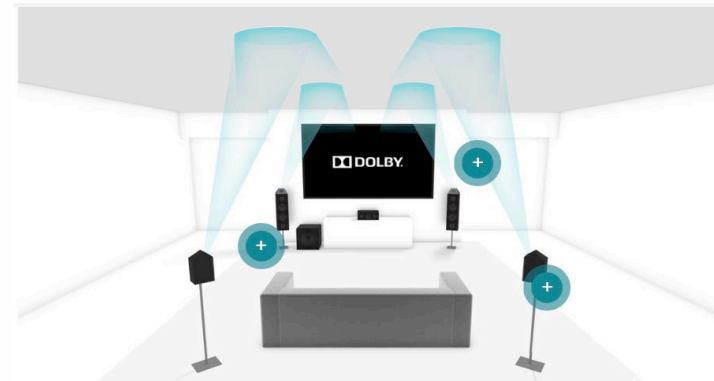
5.1.2 AVR



3.2.1 Soundbar



5.1.4 AVR

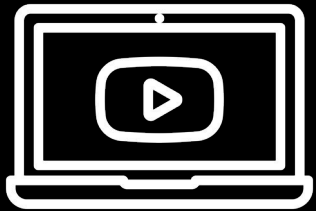


5.1.4 Soundbar





Dolby Atmos – Stereo Speakers and Headphones



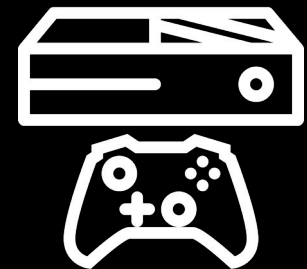
PC



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~~ ^{many} 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 sounds' arrival time at each ear and provides a strong azimuth cue.

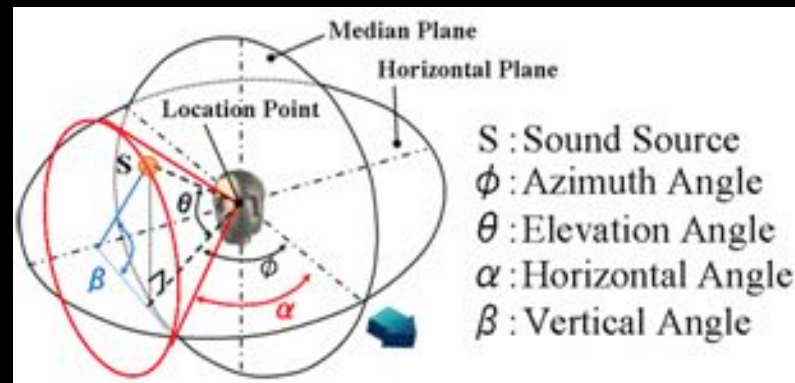
Interaural Intensity Difference (IID) is the amplitude difference of a sound at each ear – mainly due to head shadowing.

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'



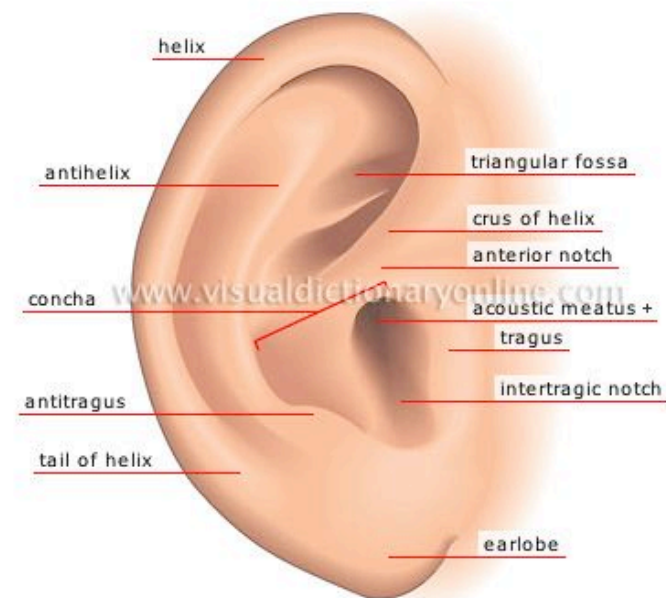
Sound localisation – cone of confusion
https://en.wikipedia.org/wiki/Sound_localization

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.

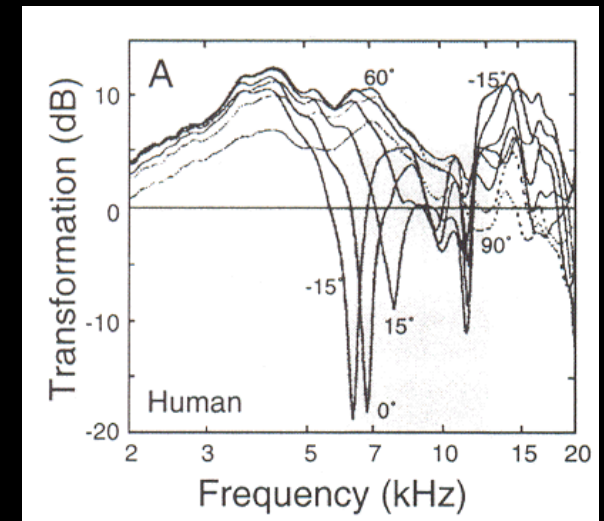
- HRTFs are a function of distance, azimuth and elevation
- Left and right ears are not usually symmetrical

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;

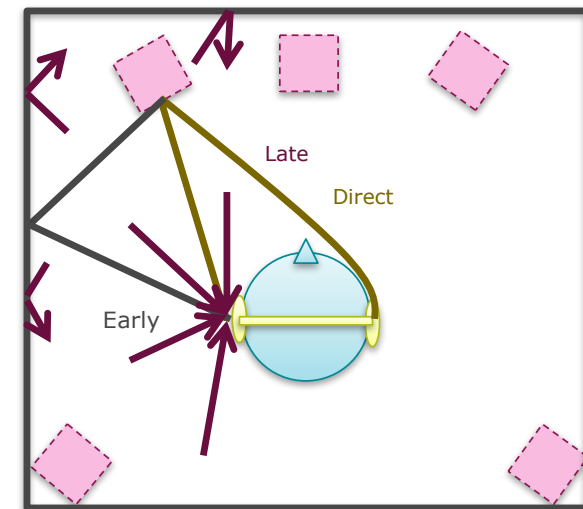
- the head geometry,
- the room geometry,
- the position and orientation of the head in the room, and
- the position and radiation pattern of the sound source.

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

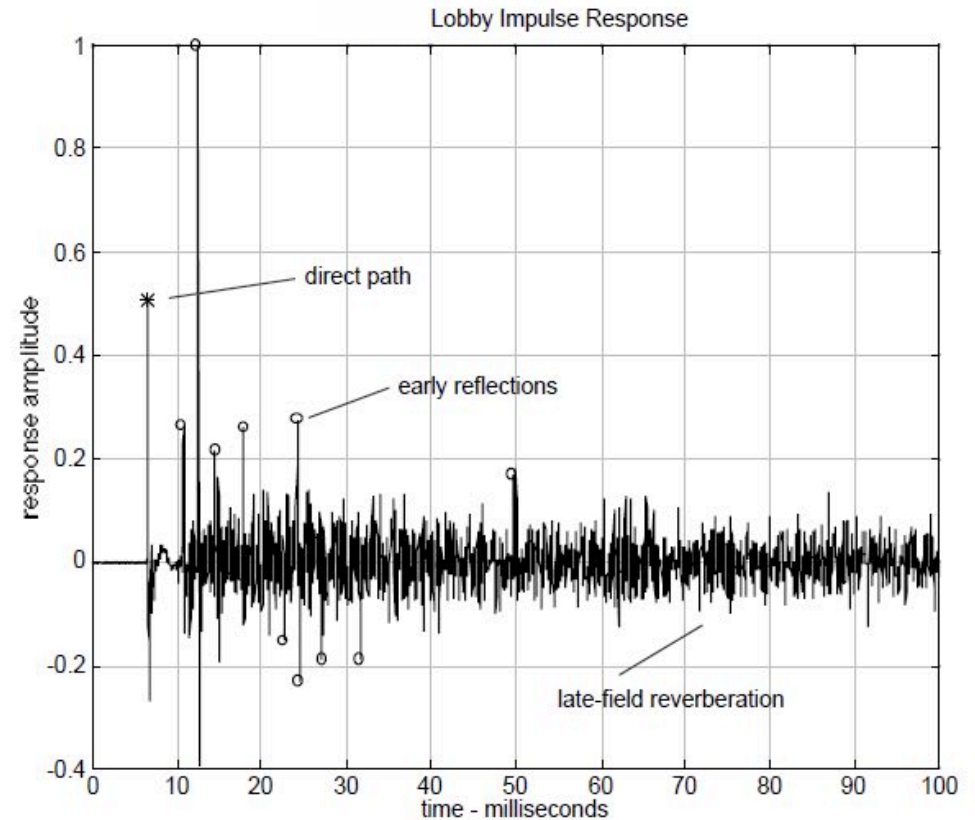


Room Impulse Response

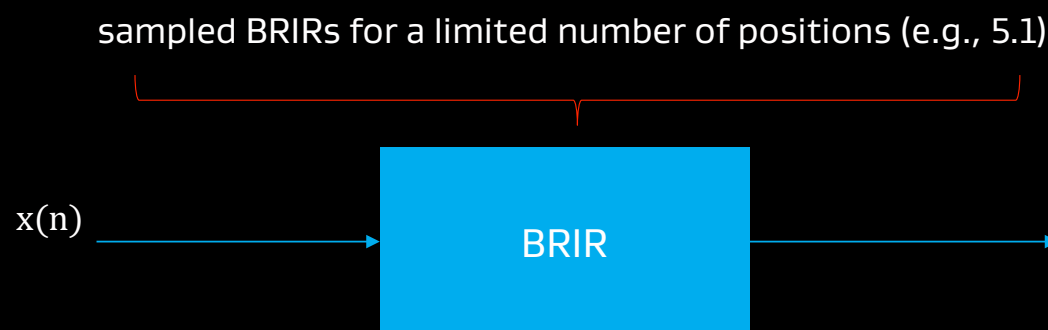
When the RIR includes the effect of a head, and is expressed with both ears it is referred to as a Binaural Room Impulse Response (BRIR)

Some qualities of BRIRs of interest include:

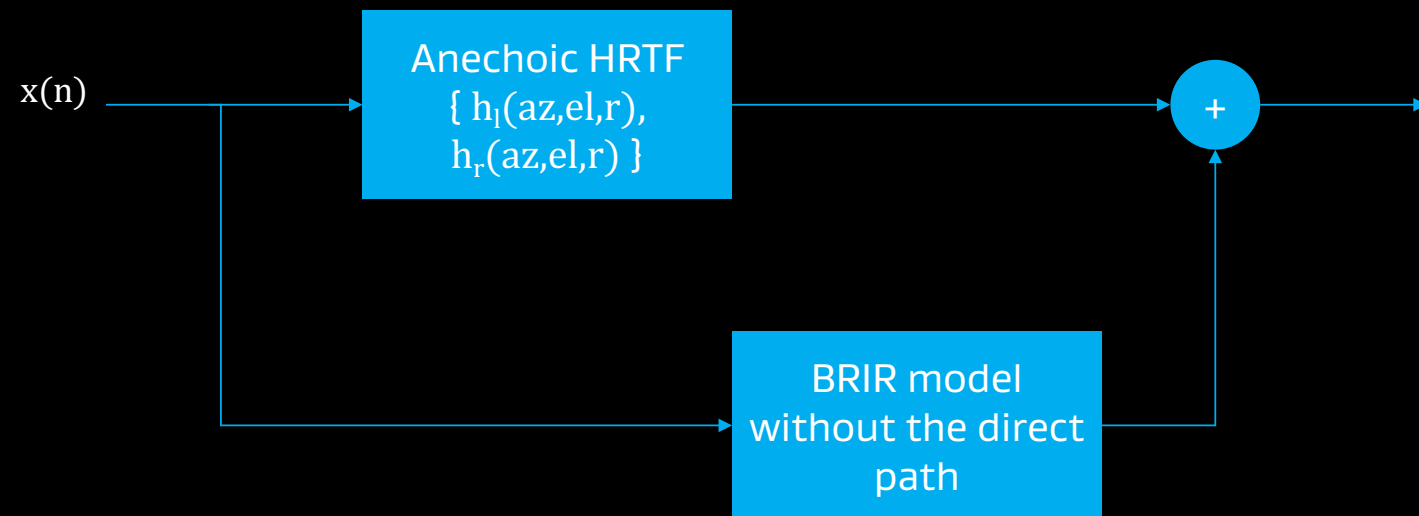
- RT60 – the rate of the reverberations decay (how 'live' is the room?)
- Direct-to-Late Ratio (DLR) – the ratio of energy in the direct path to the reverberations
- Interaural cross-correlation (IACC)



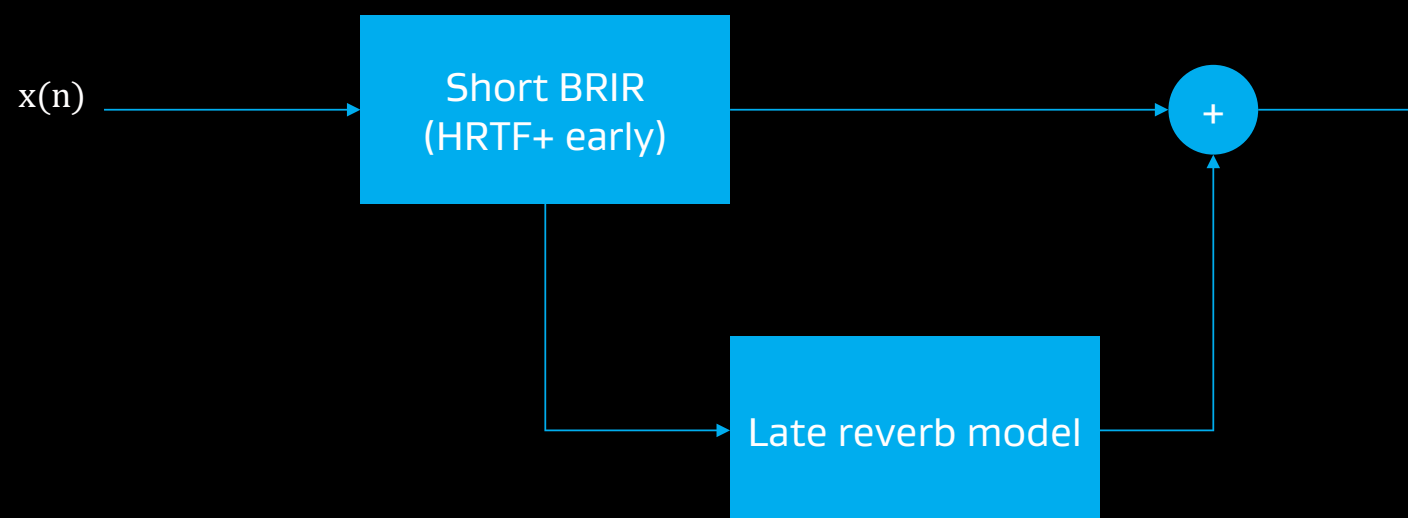
High level look at headphone virtualiser



High level look at headphone virtualiser



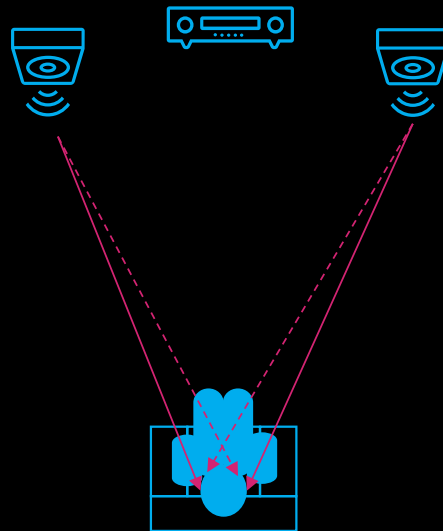
High level look at headphone virtualiser



High level look at speaker virtualiser

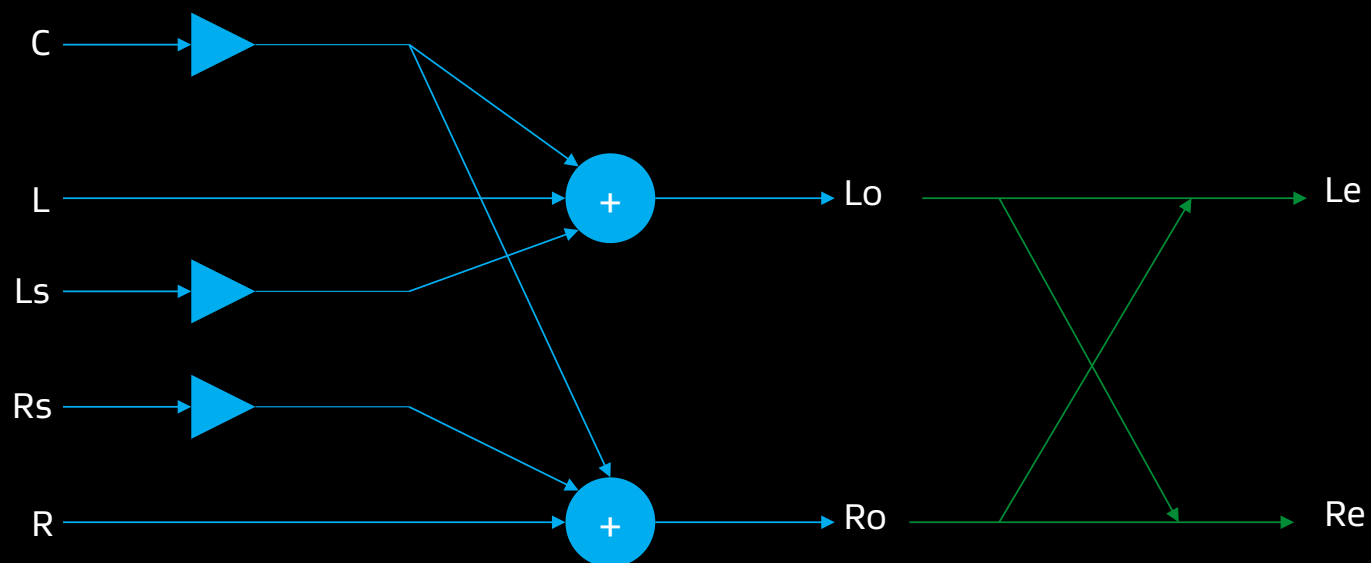
Headphone virtualisers are a matter of simulating how we localise sound . . . relatively simple as the processed signal is delivered directly to each ear.

Speaker virtualisation adds the complication that sound for each speaker 'leaks' into both ears



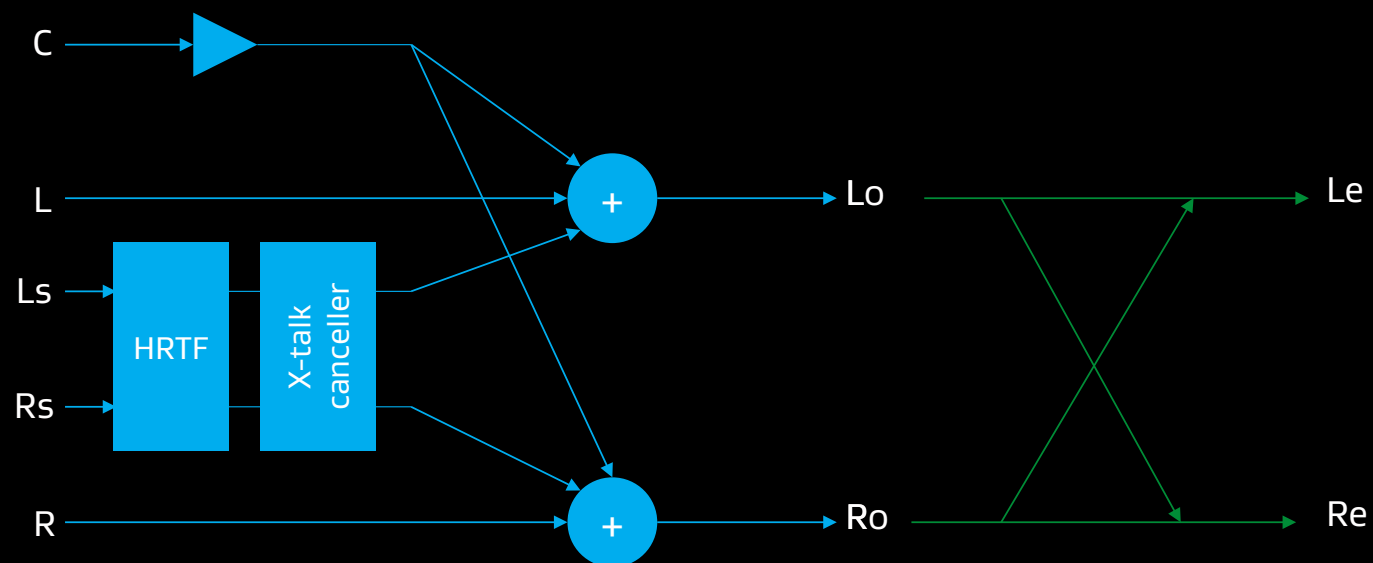
High level look at speaker virtualiser

Stereo mix of 5.1



High level look at speaker virtualiser

Virtualised mix of 5.1





Dolby Atmos – Stereo Speakers and Headphones



IMMERSIVE AUDIO STANDARDS

- ITU-R
 - BS.2076 B-WAV 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

<https://www.dolby.com/us/en/find-a-movie-theatre.html>

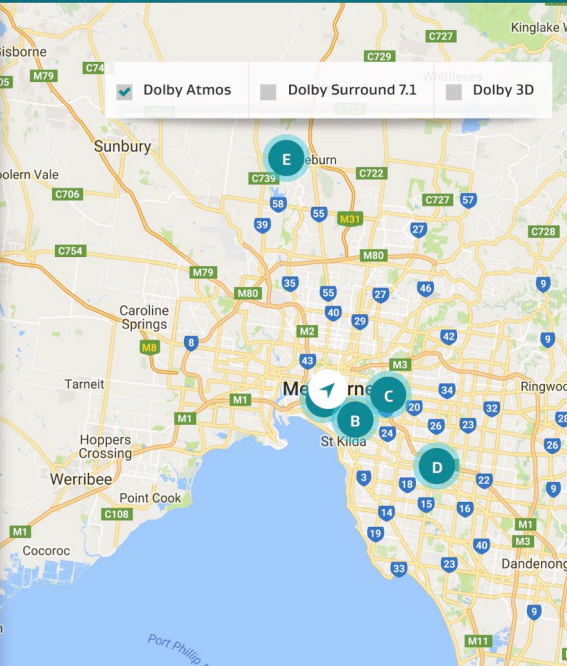
Movie Theatre Locator

Enter a city, state, or ZIP code

Melbourne City, AU

☒ Dolby Atmos ☐ Dolby Surround 7.1 ☐ Dolby 3D

- A. Village Cinemas Crown**
Crown Melbourne, 8 Whiteman Street, Southbank
Melbourne Victoria 3006
Dolby Atmos | Dolby Surround 7.1
[Directions](#) | [Visit Site](#)
- B. Village Cinemas Jam Factory**
Village Cinemas Jam Factory, 500 Chapel Street, South Yarra
Melbourne Victoria 3141
Dolby Atmos | Dolby Surround 7.1
[Directions](#) | [Visit Site](#)
- C. Lido Cinemas Hawthorn**
675 Glenferrie Road
Hawthorn Victoria 3122
Dolby Atmos | Dolby Surround 7.1
[Directions](#) | [Visit Site](#)
- D. Hoyts Chadstone**
Chadstone Shopping Centre, 1341 Dandenong Rd
Chadstone Victoria 3145
Dolby Atmos | Dolby Surround 7.1
[Directions](#) | [Visit Site](#)





... Experience Atmos



NETFLIX

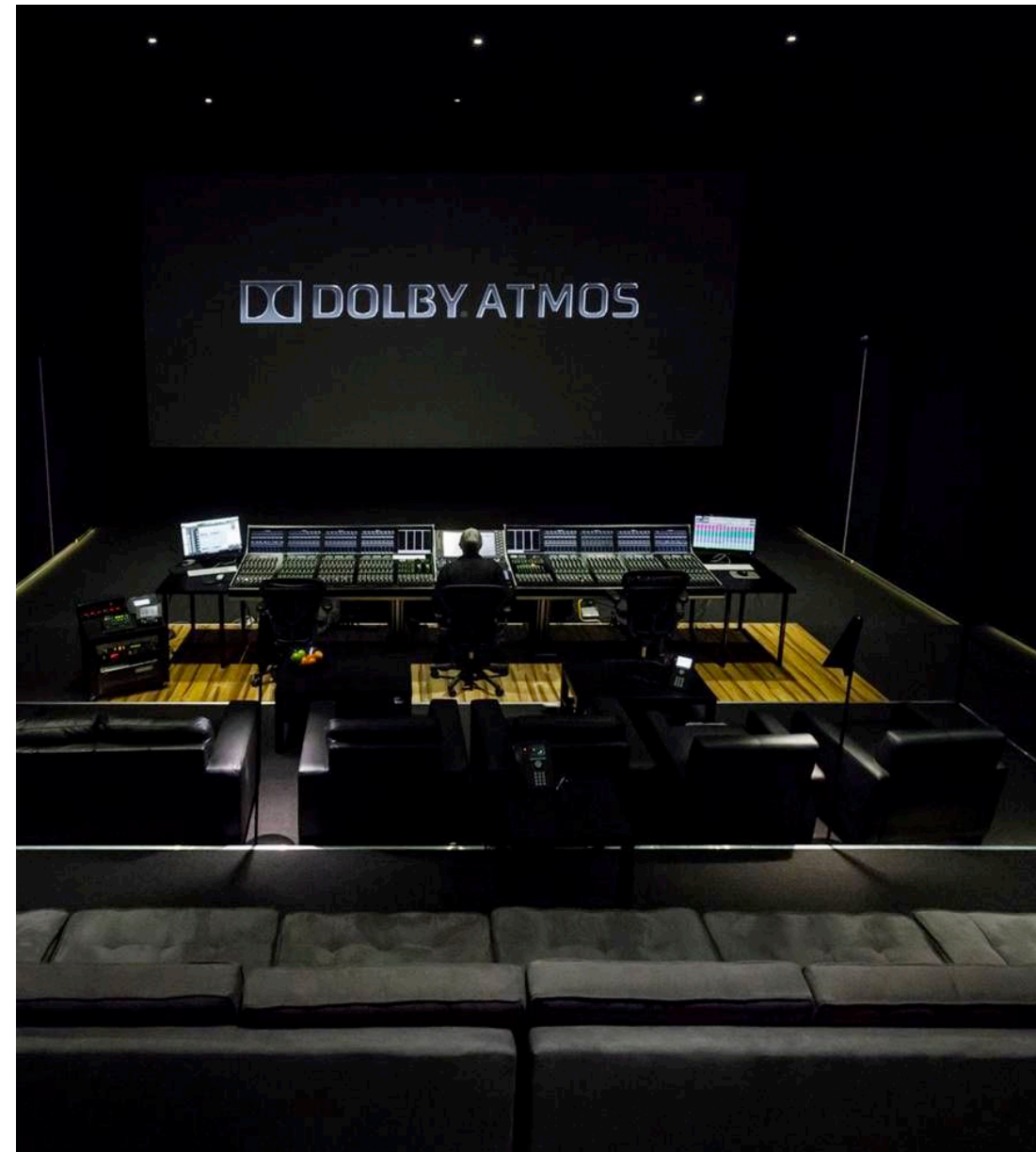


... Mix Atmos

Soundfirm Melbourne Facility

Mixing and Mastering Stage

Only facility in Australia



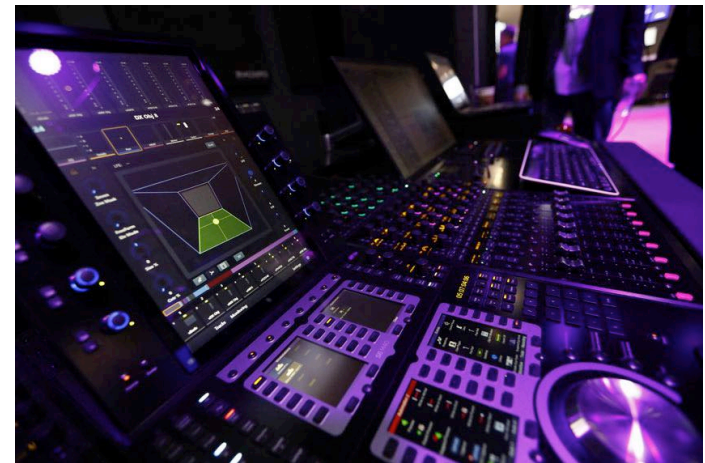
AVID PRO TOOLS & Dolby Atmos

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[®]





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





