Coding and system approaches for next-generation audio

Mike Ward

Director, Consumer Entertainment Technology

Dolby Laboratories, Inc.

July 31, 2018

Next Generation Audio

Accessible

Descriptive Audio, Dialog Enhancement, Multi-Language

Personalized

Modify presentation to listeners preference

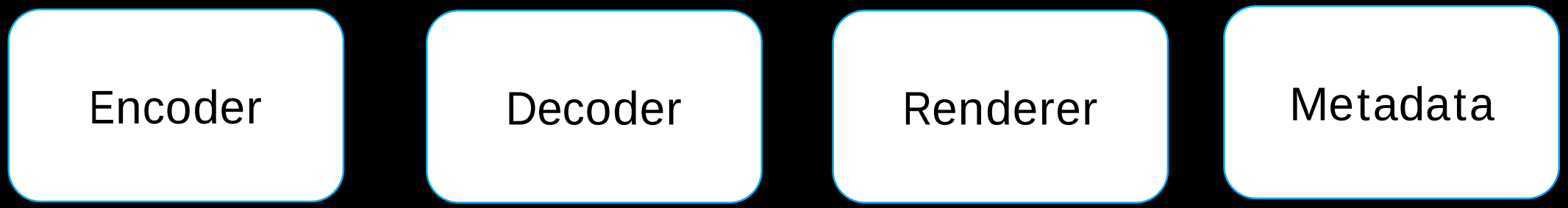
Immersive

Next-generation Living Room & Mobile Playback

Adaptable

Optimal playback on every device

Key Functional Components of Next-Gen Audio Systems



~~Next~~ Now-generation audio:
Immersive / adaptable audio available today

Coding for next-generation audio

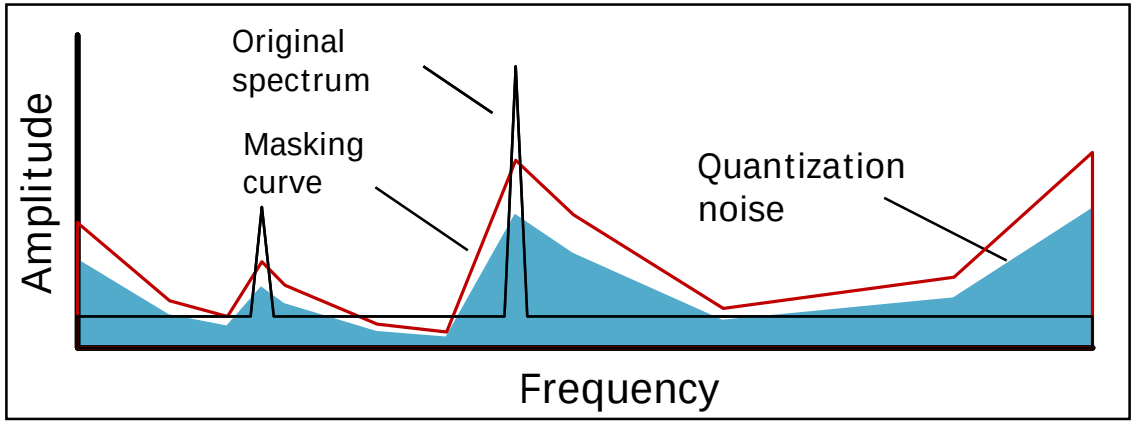
Perceptual audio coding basics

Perceptual codecs leverage two properties of audio signals

1. Redundancy – similarity in signal characteristics across time or frequency
   - Many audio signals are “stationary” over 20-40 ms time intervals
2. Irrelevancy – some signal components have no audible consequence

Removing redundancy and irrelevancy reduces data transmission requirements while maximizing audio quality

Next-gen codecs also trade-off spatial resolution (stereo image / immersiveness).



Lossy codecs create differences in the audio signal which show up as quantization noise. A psychoacoustic model is used to shape the noise for minimum audibility

Time/Frequency Analysis (Filterbank)

The most efficient perceptual audio codecs operate in the frequency domain.

- Yields a more compact signal representation (reduces redundancy).
- Allows quantization that matches limitations of the human auditory system.
    - Human ears already operate in the frequency domain.

Filterbanks decompose complex PCM audio signals into simpler frequency band components for subsequent analysis and encoding.

- Filterbank must have both good frequency AND time resolution.

Common time/frequency transforms used in audio codecs:

- MDCT (Modified Discrete Cosine Transform): Waveform and parametric coding
- QMF (Quadrature Mirror Filter): Parametric coding

Waveform coding + parametric coding

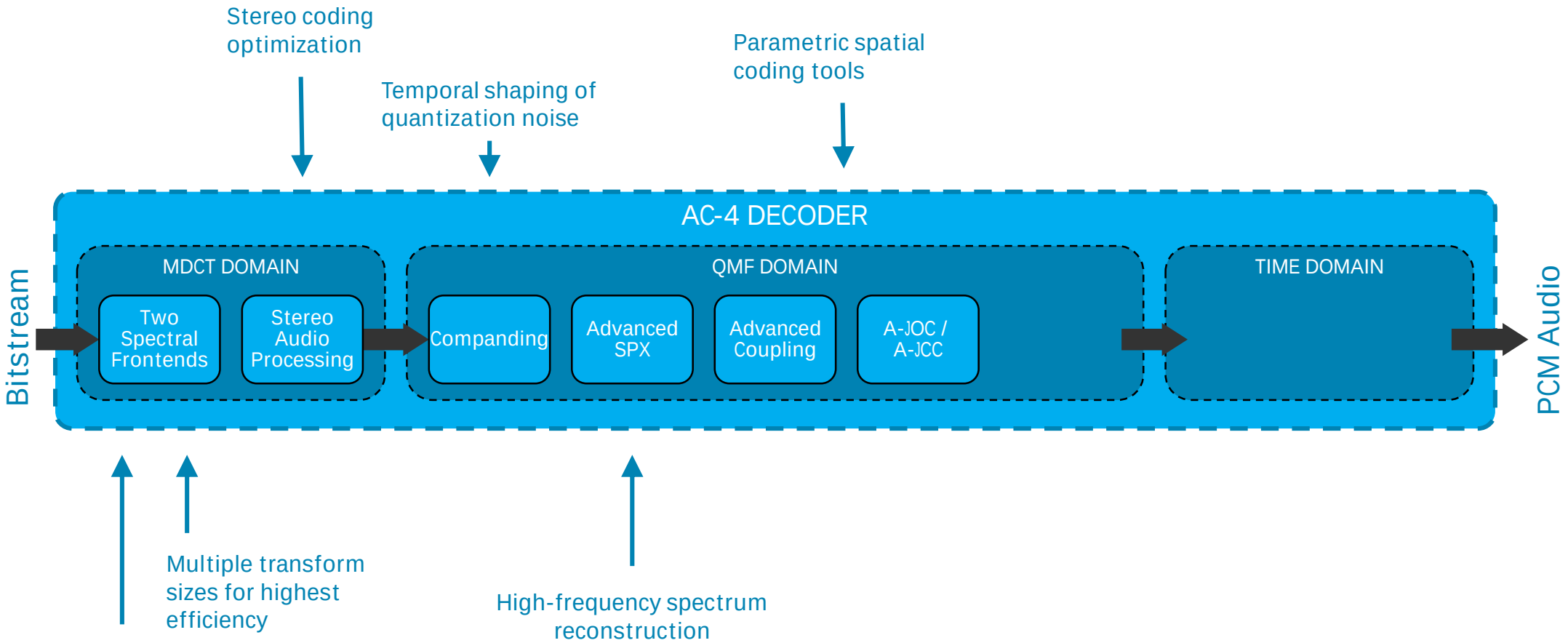
The most efficient perceptual audio codecs use a combination of waveform and parametric coding.

Rather than a quantized representation of the original waveform, parameters describe the characteristics of the original waveform which the decoder then uses for reconstruction.

Two prevalent parametric coding methods:

1. High Frequency Reconstruction - a.k.a. spectral band replication (SBR), spectral extension (SPX)
    - Reproduce perceptual high frequency cues using parameters describing harmonic relationships with lower frequencies.
2. Spatial Channel Coding (Coupling, Joint Channel Coding, Joint Object Coding)
    - Reproduce the spatial cues of a high number of channels using parameters that describe spatial relationships with a lower number of channels.
    - Note: Recreating the ambience of a higher number of channels from a lower number of channels relies on signal decorrelation during decoding.

Next-gen audio codec design: AC-4 coding features



NGA: Compression Efficiency

Designed for scalability from mobile to the living room.

Example data rates for AC-4 shown.

FORMAT	BIT RATE RANGE (typical)
2.0	32 - 96 kbps
5.1	96 - 192 kbps
7.1.4	144 - 384 kbps

(note: the formats listed are not exhaustive and the bitrate ranges are just a snapshot as of today)

Next-gen audio system design: immersive and adaptive

Next-generation audio: post production v live production

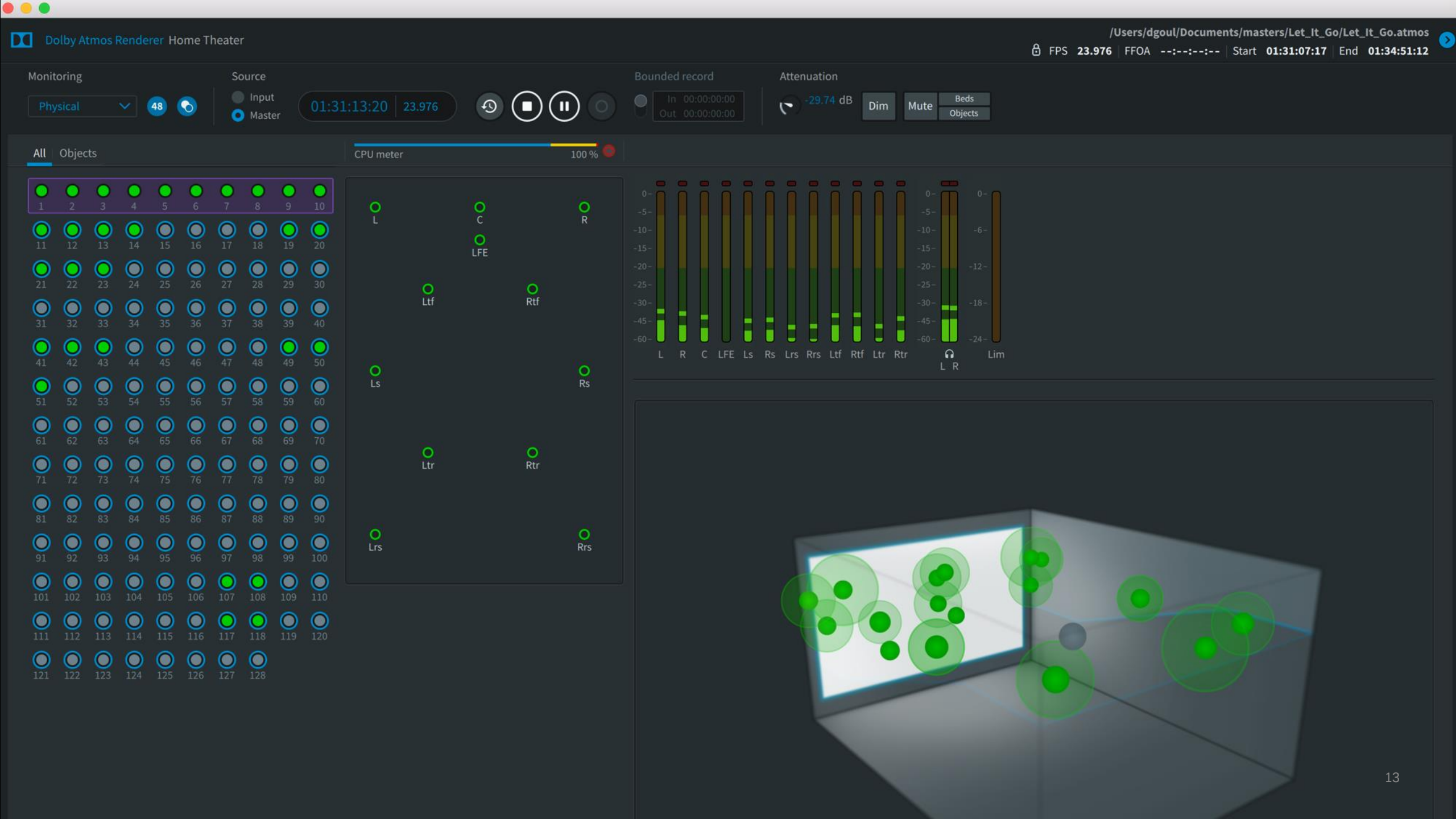
The live mixing environment is very different from mixing post produced content

Post-production:

- Able to manage a 9.1-ch bed and up to 118 objects, and export an entire mix for delivery to the consumer.
- File-based workflows are not constrained by real-time requirements or infrastructure limitations, just by file size (still way smaller than video!)

Live:

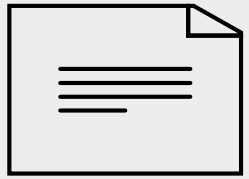
- Managing mic sources, levels and panning in real-time.
- Compatibility with existing real-time distribution infrastructure and latency constraints is required.
- Mix is primarily based on a bed (typically 5.1.4), with a small number of (usually) static objects for specific audio elements (e.g. dialogue / commentaries, descriptive audio).


Dolby Atmos Renderer Home Theater
/Users/dgoul/Documents/masters/Let_It_Go/Let_It_Go.atmos
FPS 23.976 | FFOA --:--:--:-- | Start 01:31:07:17 | End 01:34:51:12
Monitoring
Physical
48
Source
Input
Master
01:31:13:20
23.976
Bounded record
In 00:00:00:00
Out 00:00:00:00
Attenuation
-29.74 dB
Dim
Mute
Beds
Objects
All
Objects
CPU meter
100 %
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50
51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70
71 72 73 74 75 76 77 78 79 80
81 82 83 84 85 86 87 88 89 90
91 92 93 94 95 96 97 98 99 100
101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120
121 122 123 124 125 126 127 128
L
C
R
LFE
Ltf
Rtf
Ls
Rs
Ltr
Rtr
Lrs
Rrs
0 -5 -10 -15 -20 -25 -30 -45 -60
L R C LFE Ls Rs Lrs Rrs Ltf Rtf Ltr Rtr
L R
0 -6 -12 -18 -24
Lim


MADI input
1-16
17-32
33-48
49-64
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
0
-9
-18
-60
1
2
3
4
Monitored presentation
5.1.4
1 2 3 4 5 6 7 8 9 10
L R C LFE Ls Rs Lfh Rfh Lrh Rrh
Production outputs
5.1
11 12 13 14 15 16
L R C LFE Ls Rs
Stereo
17 18
L R
HP
Stereo
19 20
L R
Loudness
Monitored presenta
-20
E
S
Spanish in Dolby Atmos
Atmos
5.1
E English
S Spanish
0
5.1 Bed
5.1
MADI 1..6
Tl
Mono
MADI 7
Tr
Mono
MADI 8
TLs
Mono
MADI 9
TRs
Mono
MADI 10
English
Mono
MADI 11
Spanish
Mono
Routing
Metadata
Gai
+15
0
-49
X
50
Y
0
Z
0

Interchange Standards for NGA

Post-production



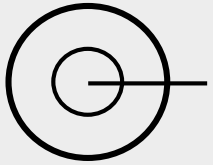
Production/Distribution

ITU-R BS.2076-0 Audio Definition Model
ITU-R BS.2088-0 Broadcast Wave

SMPTE IMF (in development)

PCM + METADATA

Live Real-time



Production/Distribution

Audio metadata via SMPTE 2109

ITU Serialized ADM (in development)

PCM + METADATA

AC-4

ETSI

TS 103 190-1
TS 103 190-2

ATSC 3.0

ATSC
ADVANCED TELEVISION SYSTEMS COMMITTEE

A/342 Part 1 & 2

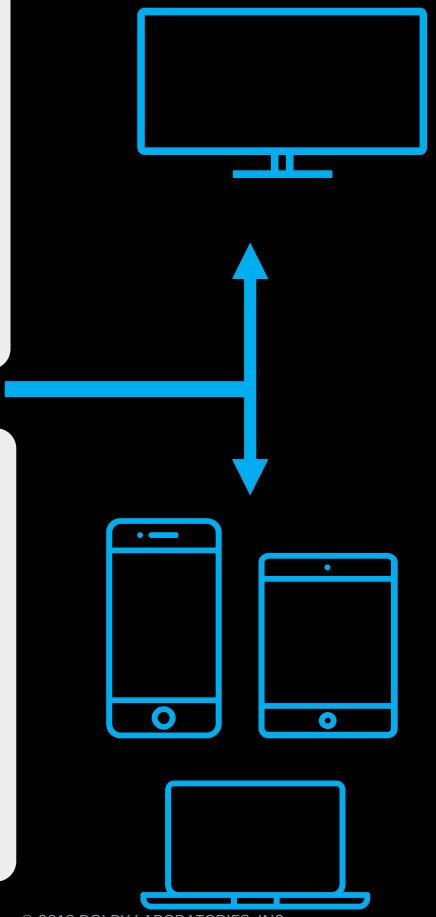
OTT

MPEG DASH

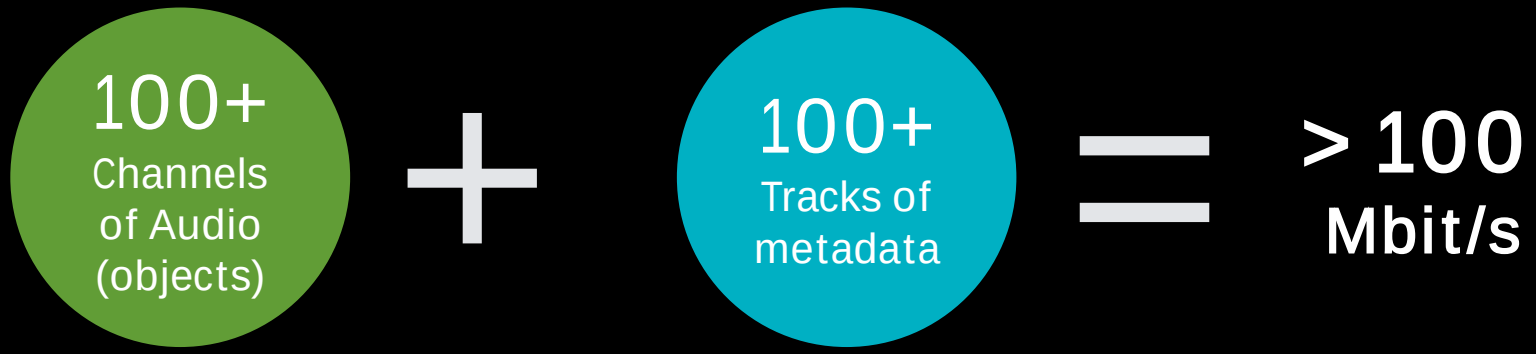
CMAF

HLS

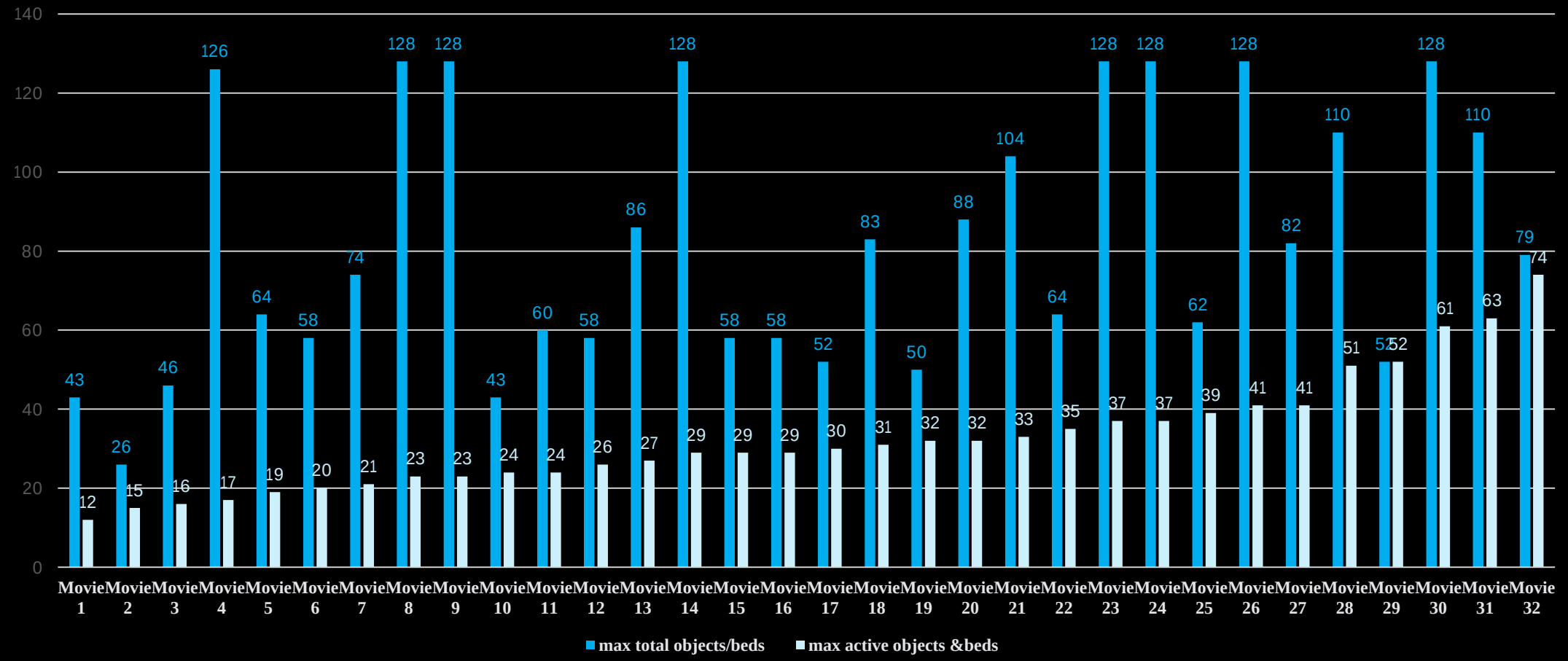
Playback Devices



How to get post-produced next-gen audio to the consumer?



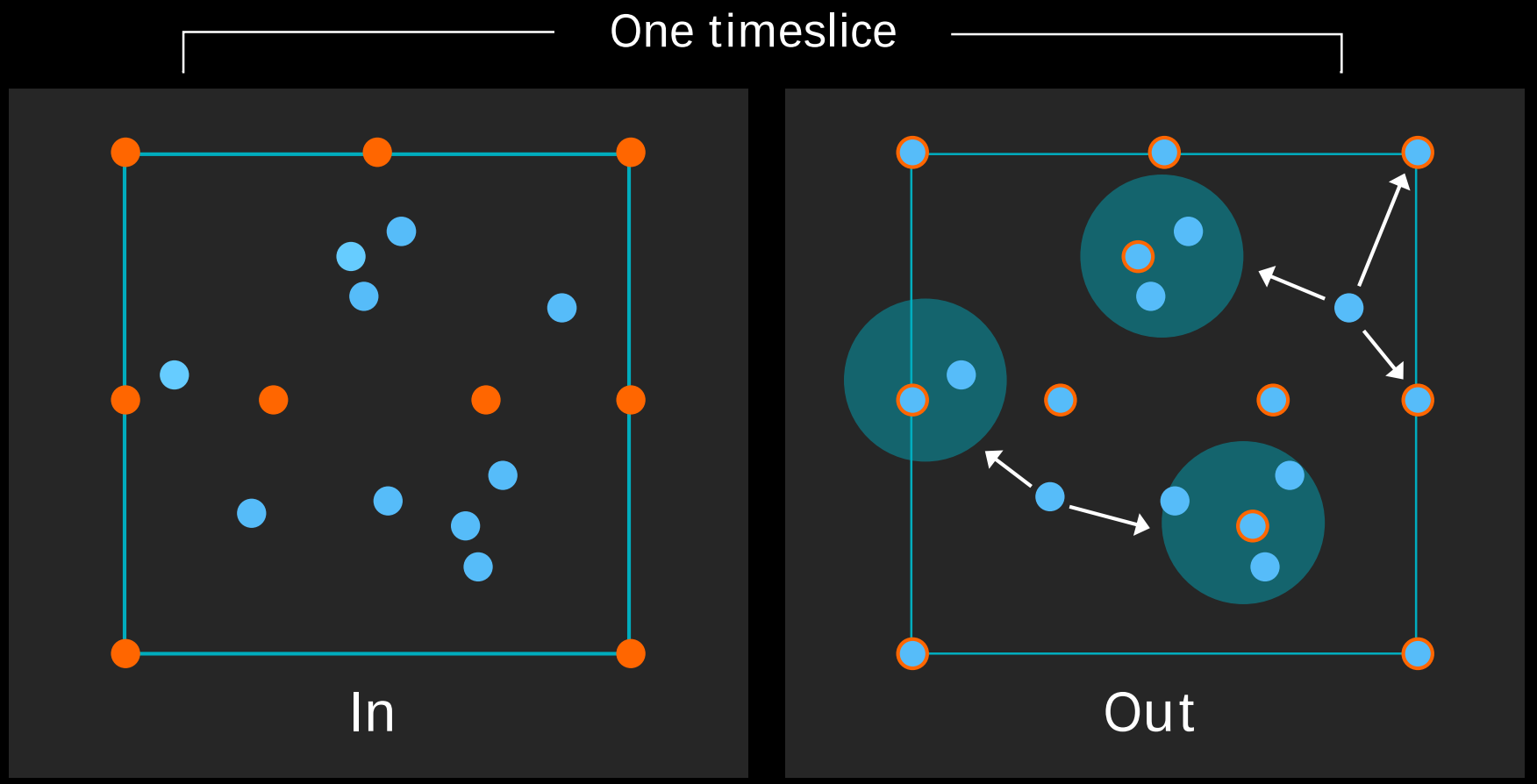
Step 1 (Easy): Not all beds/objects are simultaneously active



Step 2: (Harder) Spatial Coding: cluster objects by position and loudness

Each timeslice ‘scene’ is analyzed and clustered objects are generated

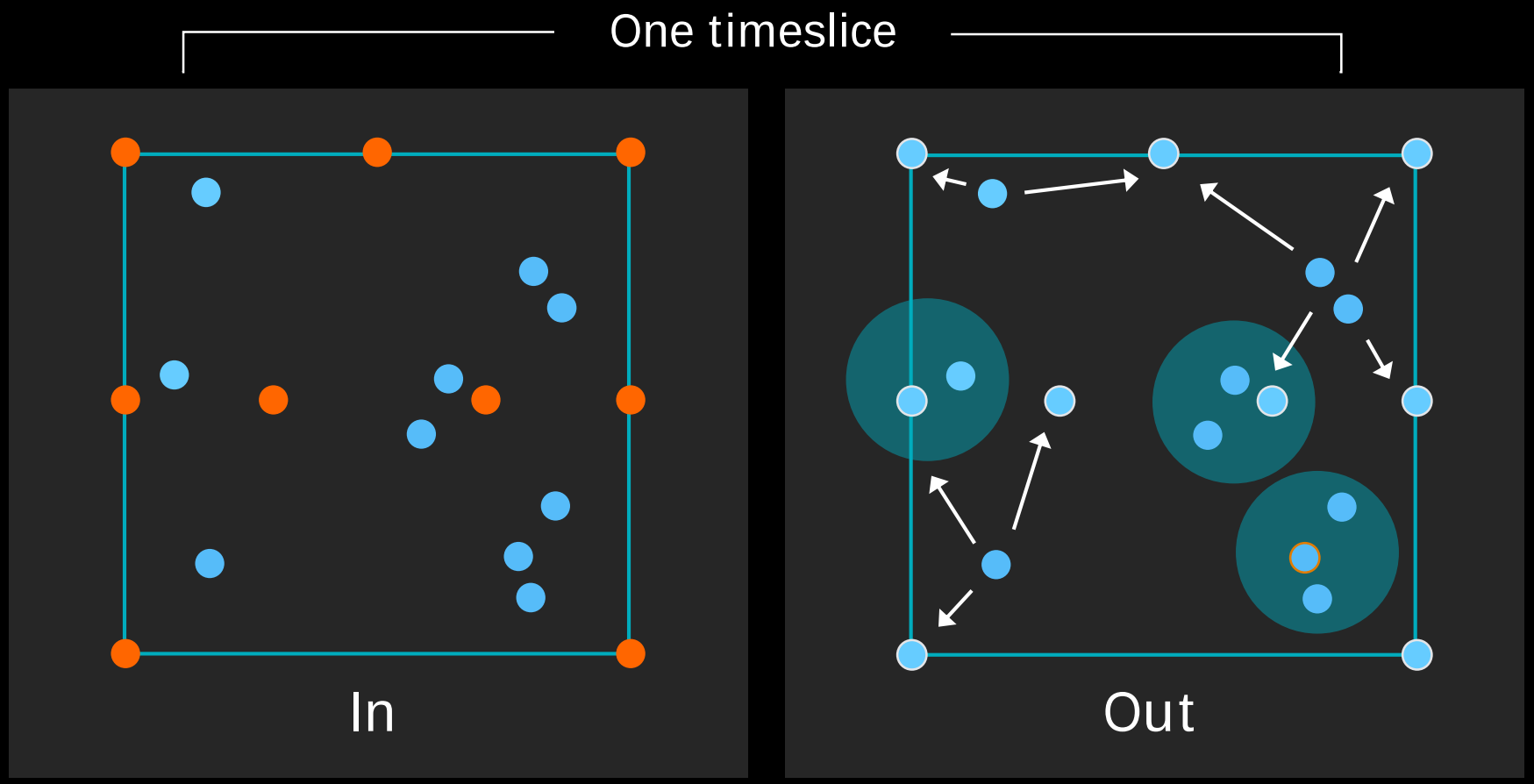
The spatial resolution contained in 100+ individual objects can be fully represented by a smaller number for interchange / encoding



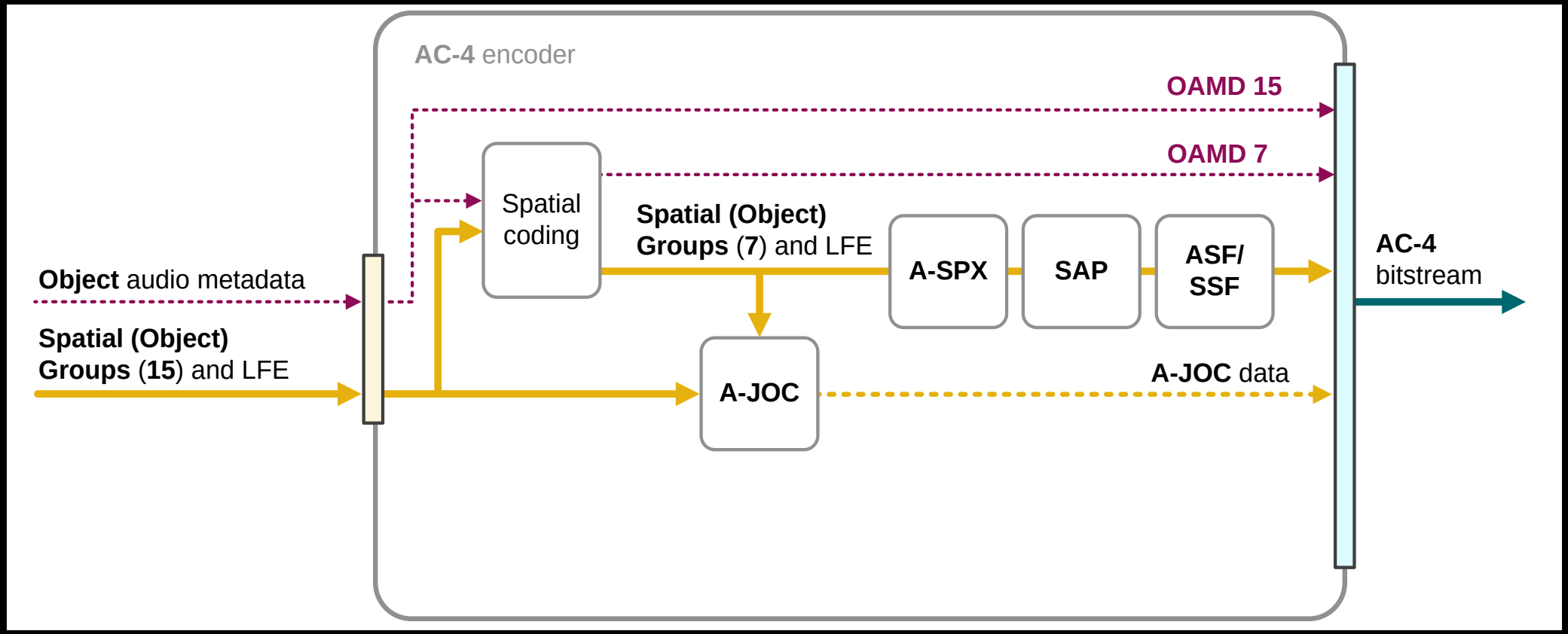
Step 2: (Harder) Spatial Coding: cluster objects by position and loudness

Each timeslice ‘scene’ is analyzed and clustered objects are generated

The spatial resolution contained in 100+ individual objects can be fully represented by a smaller number for interchange / encoding

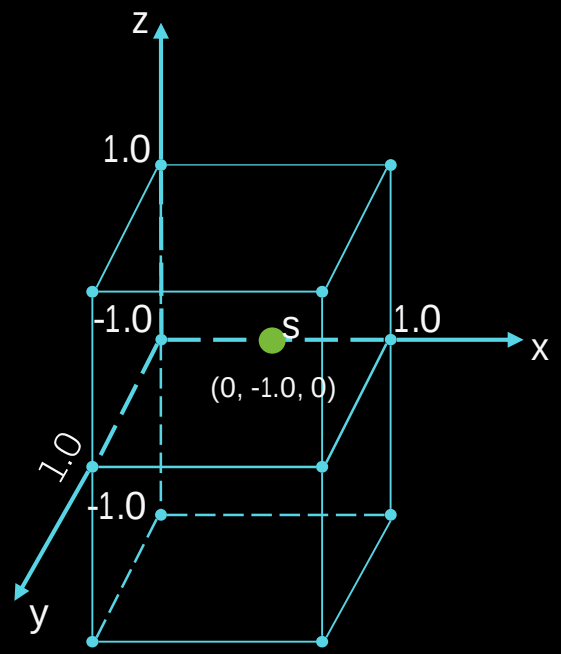


Object-based Immersive Content Encoding



Object audio = PCM + time-varying metadata

- Time stamp, e.g. seconds from program start
- X/Y/Z position (typically normalized to a room)
- Render Mode
  - Speaker zone masks
  - Speaker snap mode
- Size / Decorrelation
- Divergence
- Update interval typically 512 – 1536 samples @ 48kHz



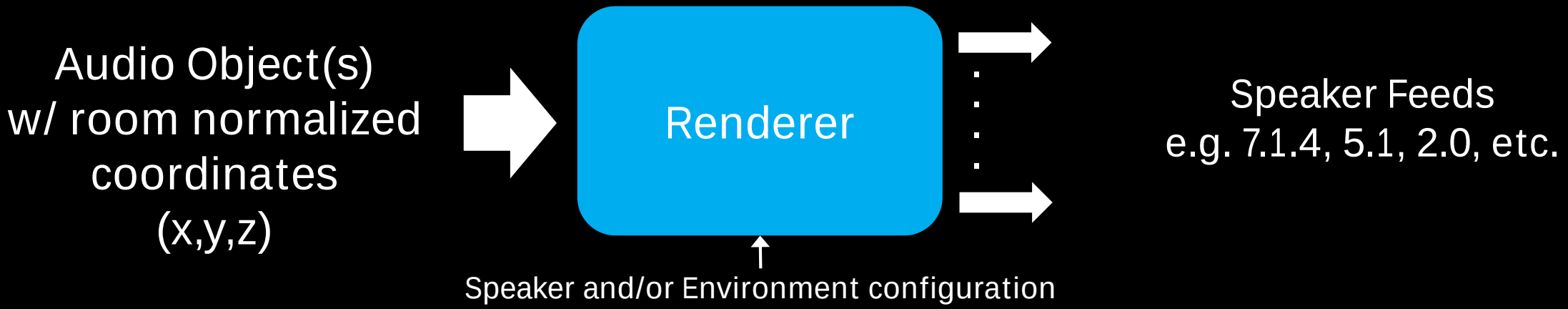
Cartesian Coordinates
(room normalized)

Personalization requires presentation + interactivity metadata

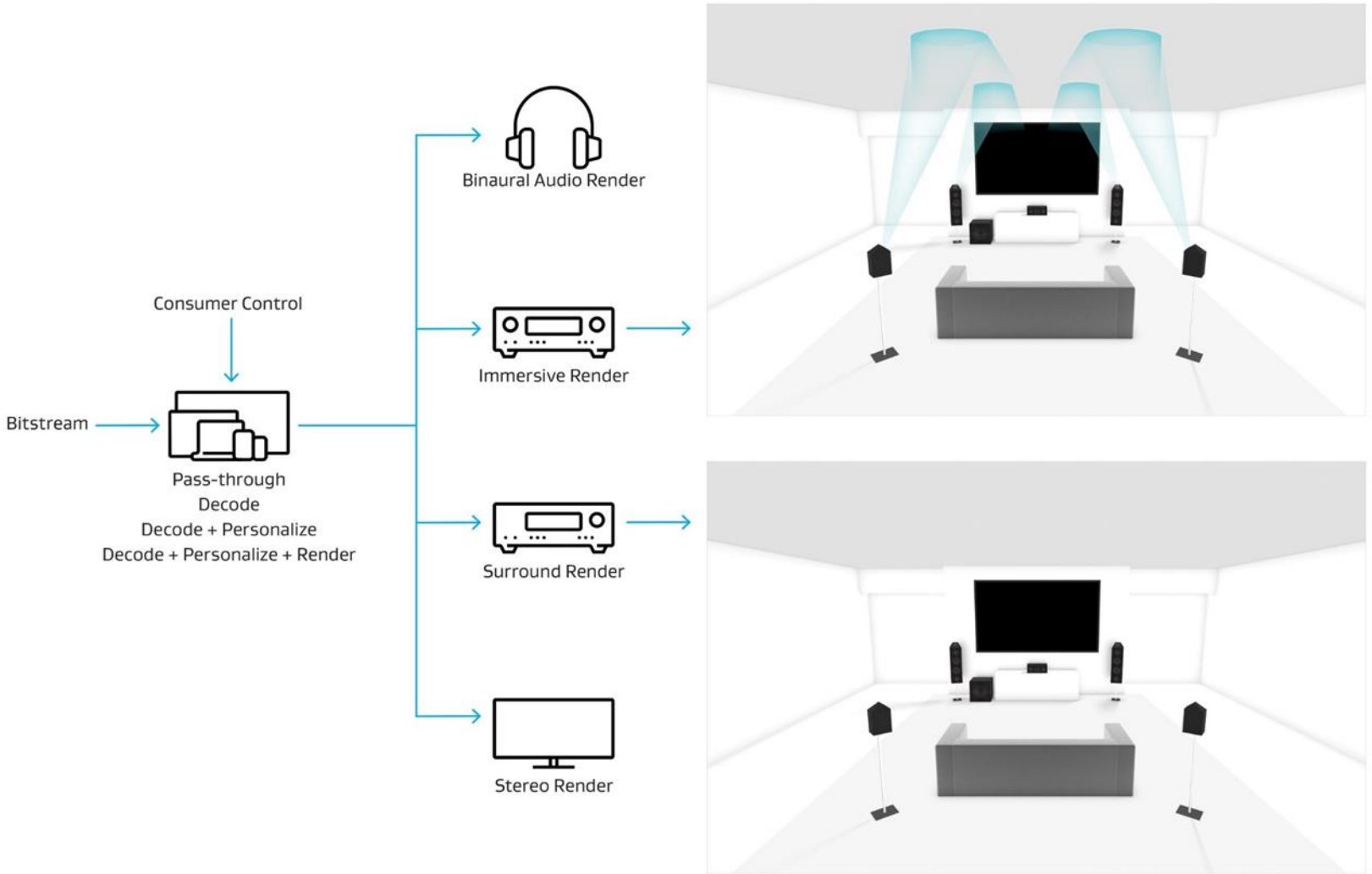
Object Playback – Requires a “Renderer”

Definition:

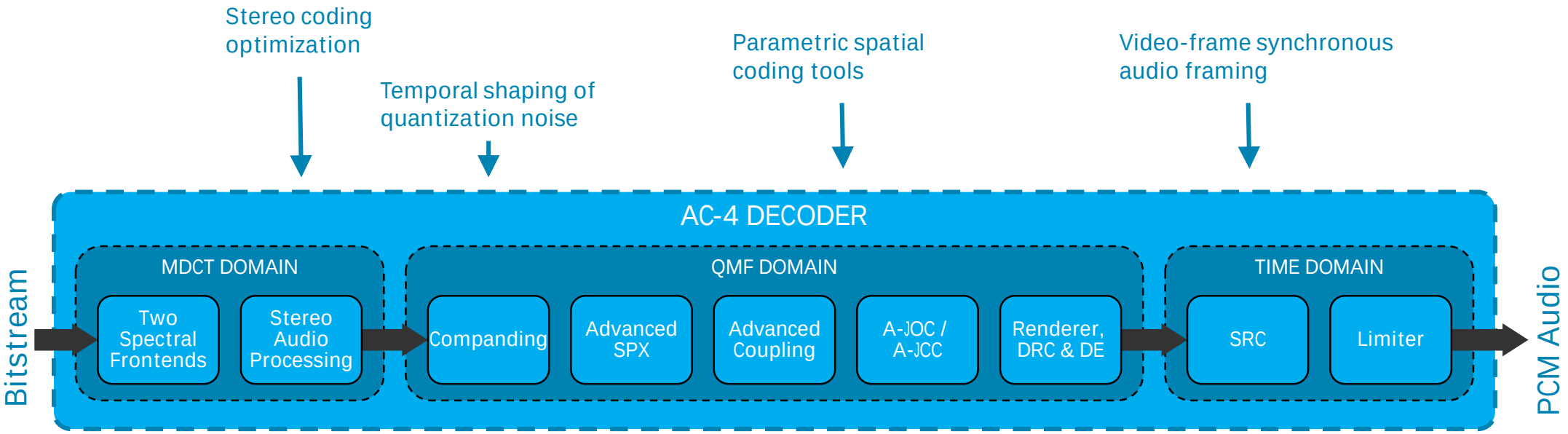
- An audio renderer converts a set of audio signals with associated metadata to a different configuration of audio signals – e.g. speaker feeds, based on the metadata, and a set of control inputs derived from the rendering environment and/or user preference.



Adaptable: NGA single bitstream input, optimized output



Next-gen audio codec design: AC-4 system features



Multiple transform sizes for highest efficiency

Two inverse quantization methods for optimal speech performance

High-frequency spectrum reconstruction

Renderer to adapt to playback environment
Multi-band and multi-channel DRC
Built in dialogue enhancement

Next-gen audio system design: accessible and personalized

Next Generation Audio is Personalized and Accessible

Enables efficient accessibility

- Delivers high-quality multiple-language playback
- Supports audio description
- Support for native dialogue enhancement

Choose between different audio experiences: alternate languages, team-biased sports commentary, or director's commentary

- E.g. can change the audio balance to focus on the announcer or listen to the ambient noise at the stadium.

Delivering spatial object / channel groups

Spatial Groups

Object

Metadata

Music & FX

11-15 channels / clusters

Metadata

Dialogue Helper

3 channels / clusters

Metadata

Dialogue Language X

4 channels / clusters

Metadata

Commentary Language X

1 Object

Metadata

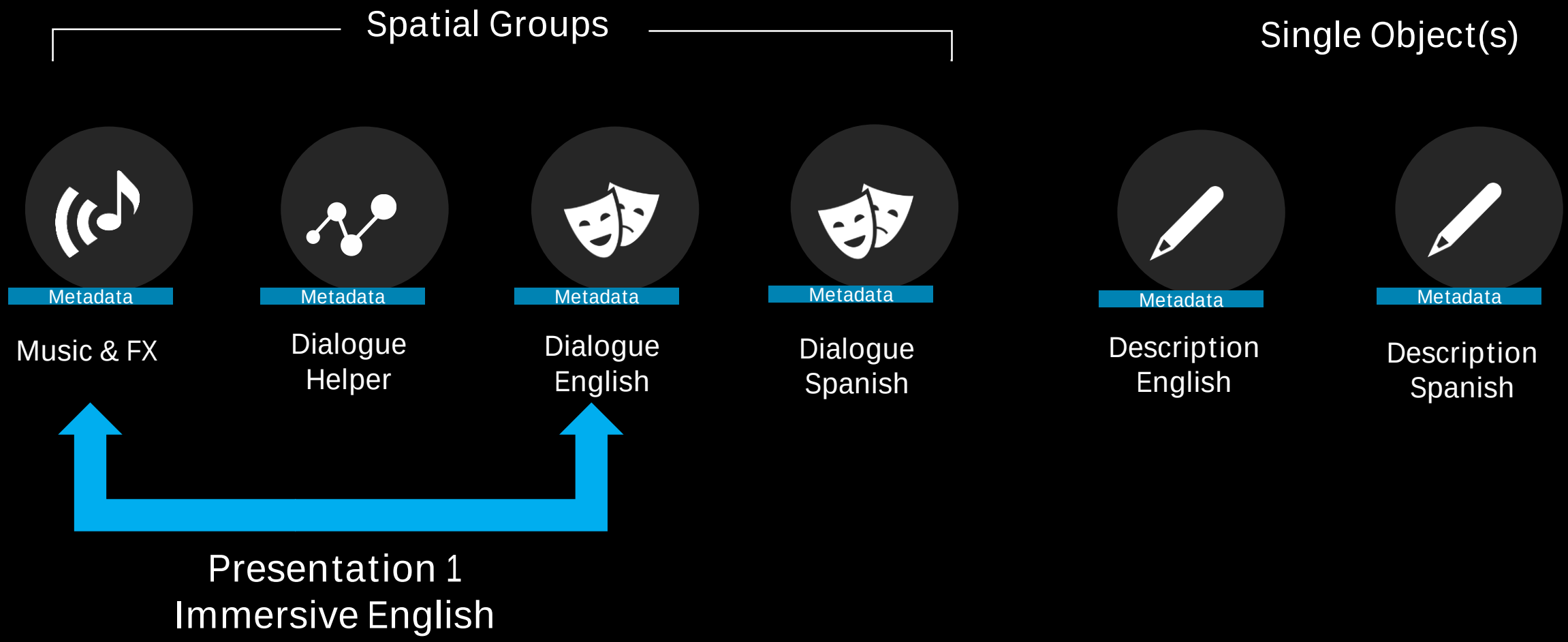
Description Language X

1 Object

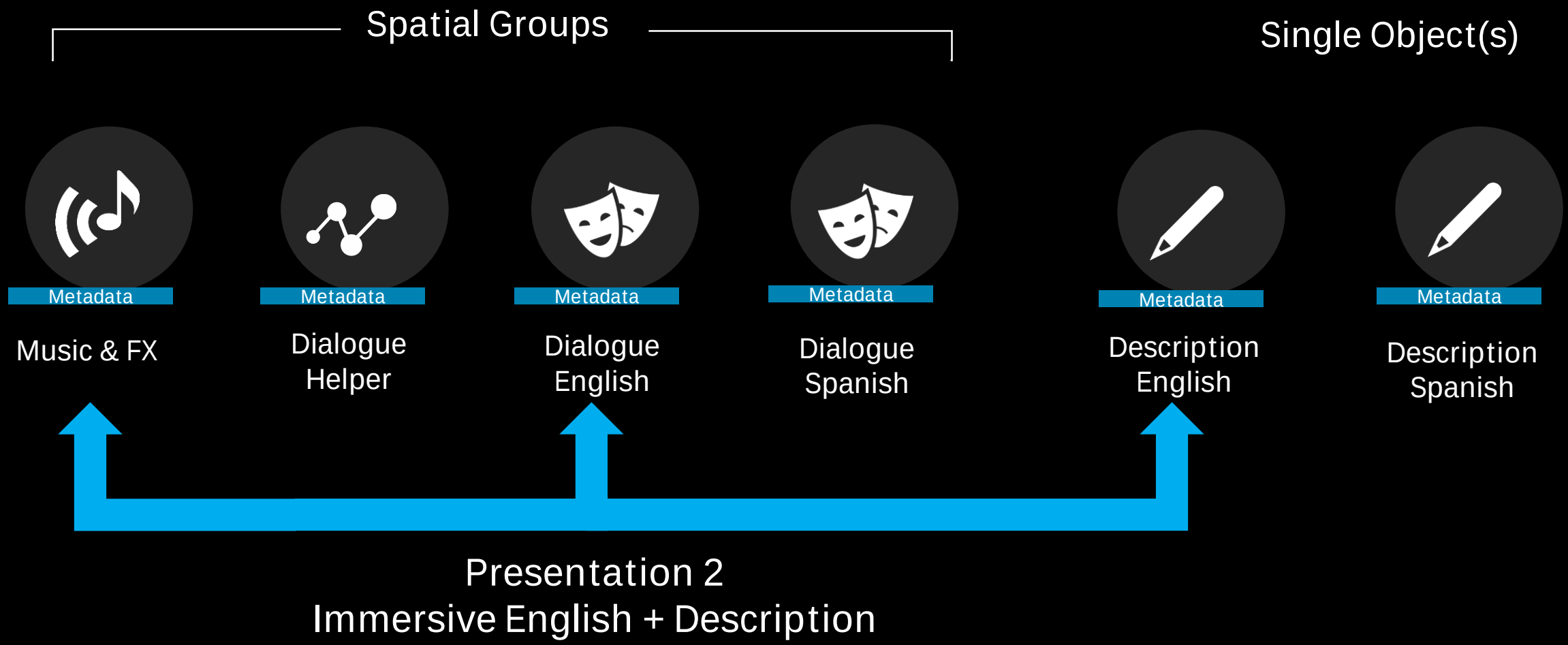
Spatial Object-based Program Building Blocks

These become substreams delivered via the codec

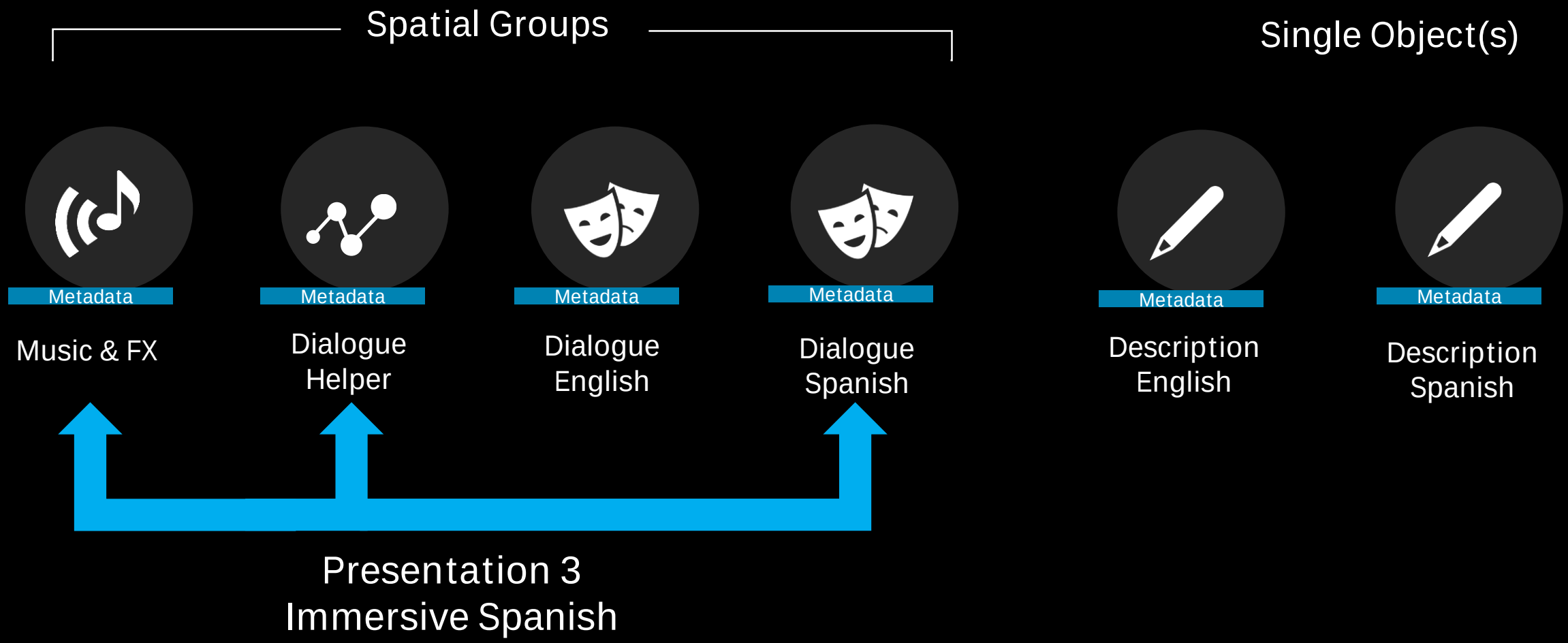
Personalization: Leverage the object building blocks



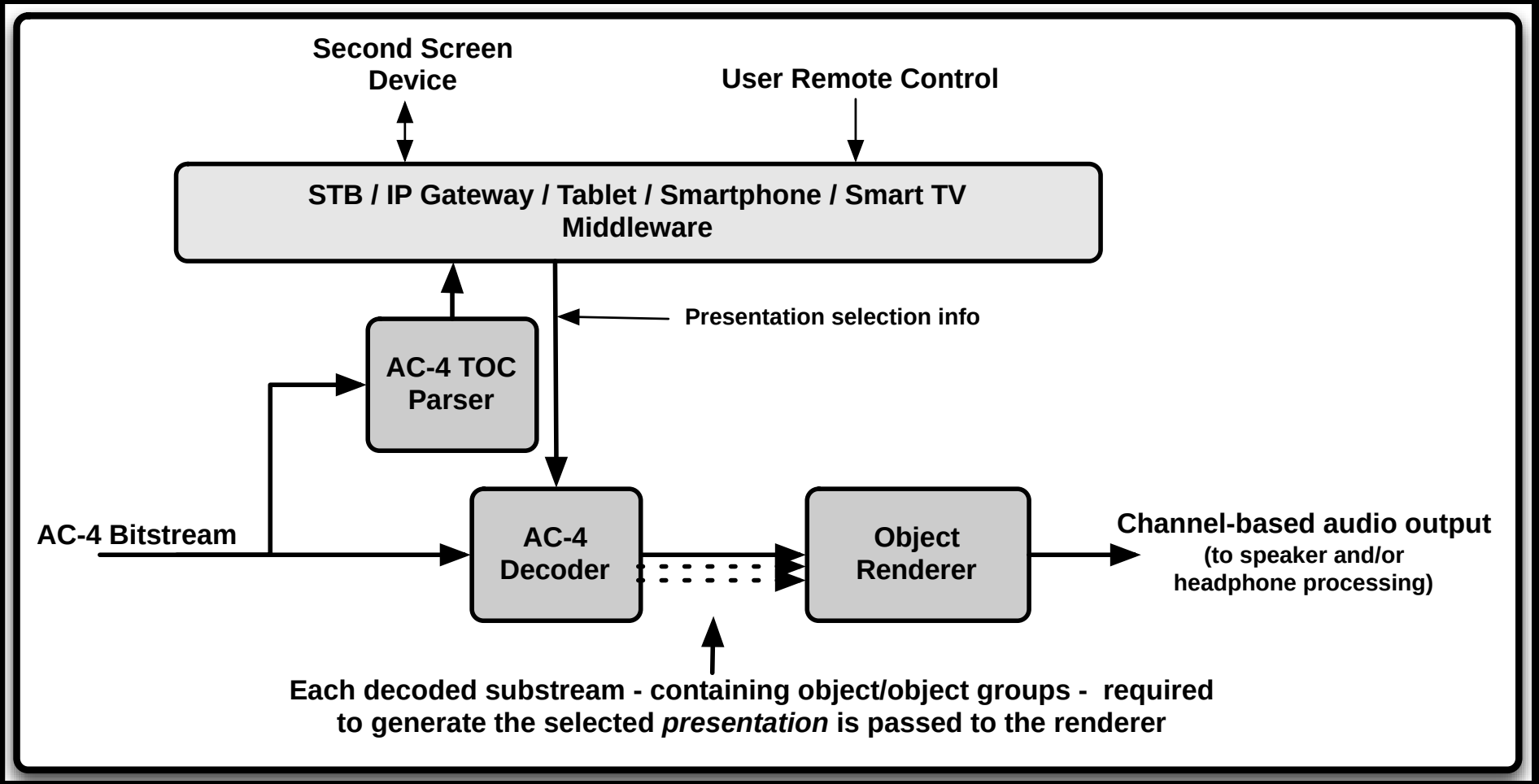
Personalization: Leverage the object building blocks



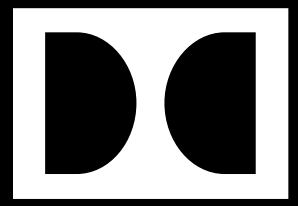
Personalization: Leverage the object building blocks



Personalized Playback Engine (AC-4 Example)



Thank you



Useful background information:

Dolby contributions, published specifications and papers on Next Generation Audio & Related Systems Aspects

ETSI specifications

- ETSI TS 102 366: Digital Audio Compression (AC-3, Enhanced AC-3) Standard - http://www.etsi.org/deliver/etsi_ts/102300_102399/102366/01.04.01_60/ts_102366v010401p.pdf
- ETSI TS 103 420: Backwards-compatible object audio carriage using Enhanced AC-3 (2016) - http://www.etsi.org/deliver/etsi_ts/103400_103499/103420/01.01.01_60/ts_103420v010101p.pdf
- ETSI TS 103 190-1: Digital Audio Compression (AC-4) Standard; Part 1: Channel based coding - http://www.etsi.org/deliver/etsi_ts/103100_103199/10319001/01.03.01_60/ts_10319001v010301p.pdf
- ETSI TS 103 190-2: Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio - http://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.02.01_60/ts_10319002v010201p.pdf
- ETSI TS 103 448: AC-4 Object Audio Renderer for Consumer Use (2016) - http://www.etsi.org/deliver/etsi_ts/103400_103499/103448/01.01.01_60/ts_103448v010101p.pdf

AES publications

- June 2016: “AC-4 – The Next Generation Codec” – http://www.aes.org/tmpFiles/elib/20180416/18190.pdf
- June 2016: “Immersive Audio Delivery Using Joint Audio Coding” - http://www.aes.org/tmpFiles/elib/20180416/18285.pdf
- May 2016: “ITU-R BS.1770 Based Loudness for Immersive Audio” - http://www.aes.org/e-lib/browse.cfm?elib=18199
- May 2017: “Parametric Joint Channel Coding of Immersive Audio” - http://www.aes.org/tmpFiles/elib/20180416/18616.pdf

IEEE Publications

- February 2017 BTS Journal Special Issue: “Delivering Scalable Audio Experiences using AC-4” - https://ieeexplore.ieee.org/document/7864321/

SMPTE Publications

- October 2014: “Immersive & Personalized Audio: A Practical System for Enabling Interchange, Distribution & Delivery of Next Generation Audio Experiences” - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7269346
- July/August 2015 Journal: https://ieeexplore.ieee.org/document/7306465/
- Nov/Dec 2016 Journal: “Recipes for Creating and Delivering Next-Generation Broadcast Audio” - https://ieeexplore.ieee.org/document/7803442/
- March 2018 Journal: “An Open, Standards-Based Framework for Audio Metadata Transport in Live Content Workflows” - https://ieeexplore.ieee.org/document/8304846/

ITU-R Recommendations

- ITU-R BS.1196-6: Audio Coding for Digital Broadcasting - https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1196-6-201712-I!!PDF-E.pdf
- ITU-R BS.2076-1: Audio Definition Model - https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2076-1-201706-I!!PDF-E.pdf
- ITU-R BS.2051-1: Advanced Sound System for Programme Production - https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2051-1-201706-I!!PDF-E.pdf
- ITU-R BS.1770-4: Algorithms to Measure Audio Programme Loudness and True-peak Audio Level - https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf

UHD Forum

- Phase A Guidelines v1.4 [August 25 2017] - https://ultrahdforum.org/wp-content/uploads/Ultra-HD-Forum-Guidelines-v1.4-final-for-release.pdf
- Phase B Guidelines v1.0 [April 7 2018] - https://ultrahdforum.org/wp-content/uploads/Ultra-HD-Forum-Phase-B-Guidelines-v1.0.pdf