

NEW IMMERSIVE AND OBJECT-BASED MULTI-CHANNEL AUDIO FORMATS FOR CINEMA, ENTERTAINMENT AND CINEMATIC VR

Jean-Marc Jot
Xperi Corp. / DTS

April 2017



IN THIS HOUR...

Commercial entertainment experiences can now include high-resolution video presentation covering the full visual field (angular + depth).

Recent developments in cinema and Blu-Ray: audio *with height*.

What are suitable formats and workflows for immersive audio?
... and what about audio-only immersive experiences?



A turning point – immersive *media*

Audio and video scenes no longer spatially "disjoint", but "*congruent*"

Movies, live performance, user-generated content, VR...

... a new era in media experience expectation.



IN THIS HOUR...

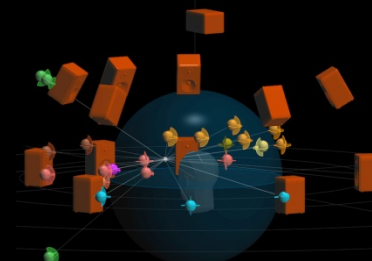
Early developments

Immersive audio creation, encoding and rendering
Approaches to format-agnostic audio



New formats for *immersive linear audio*

Creation, workflow, cinema
Distribution, broadcast, streaming



Cinematic VR

Binaural 3D audio and Ambisonic techniques – *the return!*

Pending issues, perspectives

Q&A



IMMERSIVE AUDIO

Non-linear – interactive | computer generated

Video games | simulation | interactive VR

Live performance: music | multimedia | dance | theater | DJ

Also... creation (mixing) of *linear* content...

Linear – scripted | recorded

Recording: music | radio drama | movie soundtrack | cinematic VR

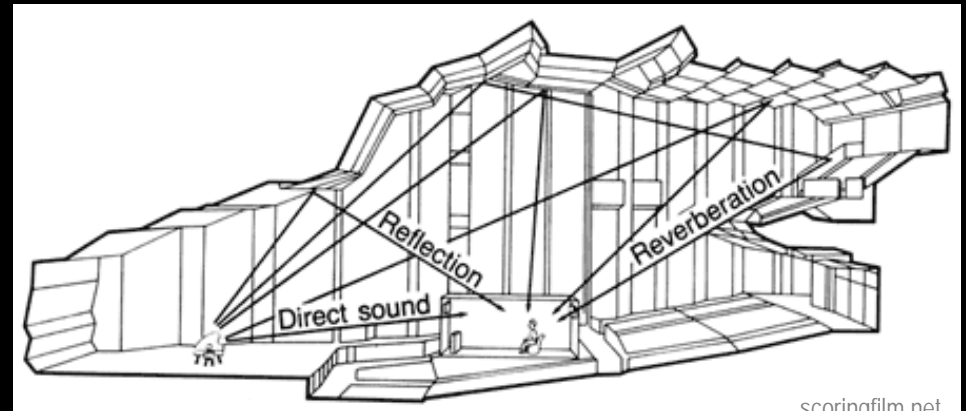
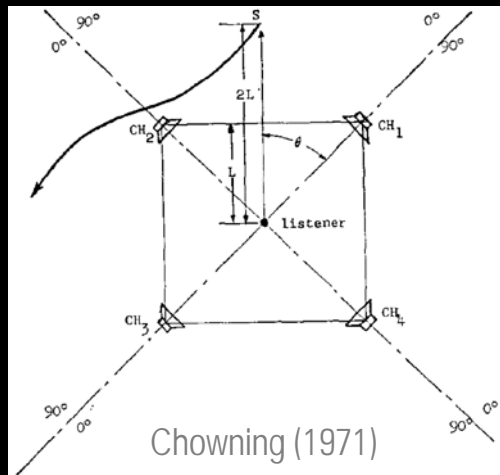
Content creation: computer-assisted mixing + automation

Also... live recording of *non-linear* content production or performance

NON-LINEAR OBJECT-BASED IMMERSIVE AUDIO

Workflow based on game authoring/rendering technology (late 90's – now)

Origins: computer music, concert hall acoustics research (since 70's)



Recent progress accelerated by rebirth of VR technology
supported by progress in computing hardware performance.



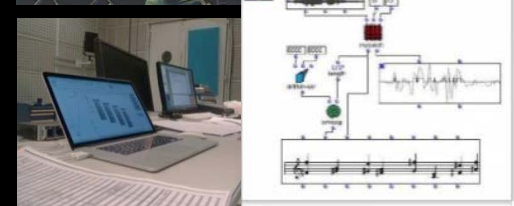
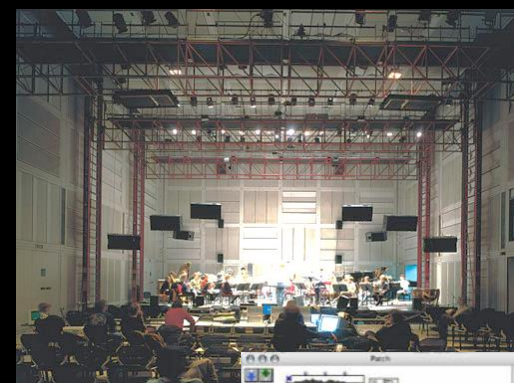
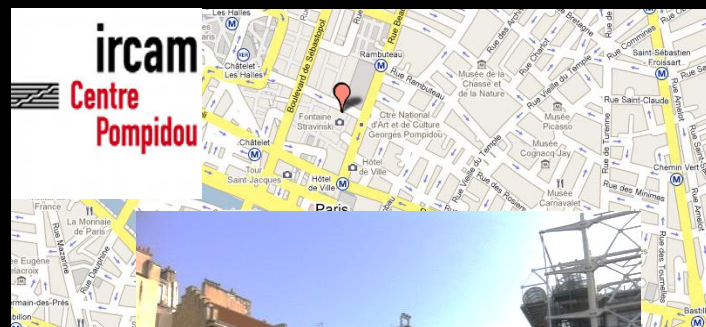
Beginnings...

Telecom Paris
1989-1992

FDN (Feedback Delay Network) artificial reverberation.

IRCAM
1993-1998

Spat – room acoustics and spatial audio for computer music.



FLUX:: IRCAM Spat (2010)

www.fluxhome.com/products/plug_ins/ircam_spat-v3 | forumnet.ircam.fr/product/spat/

Source Reverb Setup

Selected source: 1 On Solo

Perceptual Factors

Source Presence: 72 Source Warmth: 30 Source Brilliance: 30

Room Presence: Running Reverb: 34 Envelop.: 25

Acoustical criteria

Rt60: 0.37s EDT: 0.68s Es: -13.73 dB
Rev: -24.00dB ASW: -11.99dB Dsh: 0.00dB
RtHi: 0.50 RtLow: 1.00 Desl: 0.00dB

Options

Target Reverb: 1

Doppler: Air Absorption: Clear Solo:

Drop Mode: Log2 Lin/Log Drop: 6.00 dB Radius: 1.0 m Pan_rev: 0.00 Early Width: 10.0 deg

Radiation

Distance: 1.54 m Azimuth: -55 deg Yaw: 0 deg Elevation: 0 deg Pitch: 0 deg Aperture: 80 deg

Frequency Response

Low Freq: 177 Hz High Freq: 5657 Hz

Low Gain: 0.00 dB Med Gain: 0.00 dB High Gain: 0.00 dB Global Gain: 0.0 dB

ircam Tools

FLUX::

Input: 0.00 dB Output: 0.00 dB

SPAT v3

Save Recall Copy B Copy A Recall Save

automation

FLUX:: IRCAM Spat v3 — demo

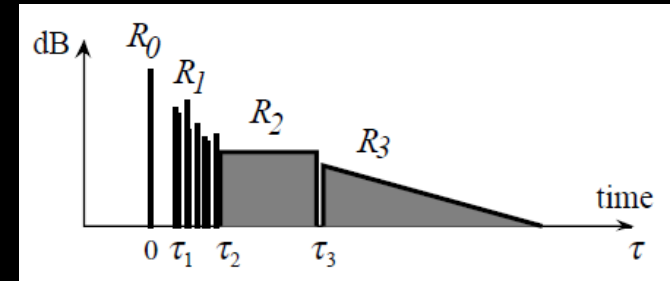
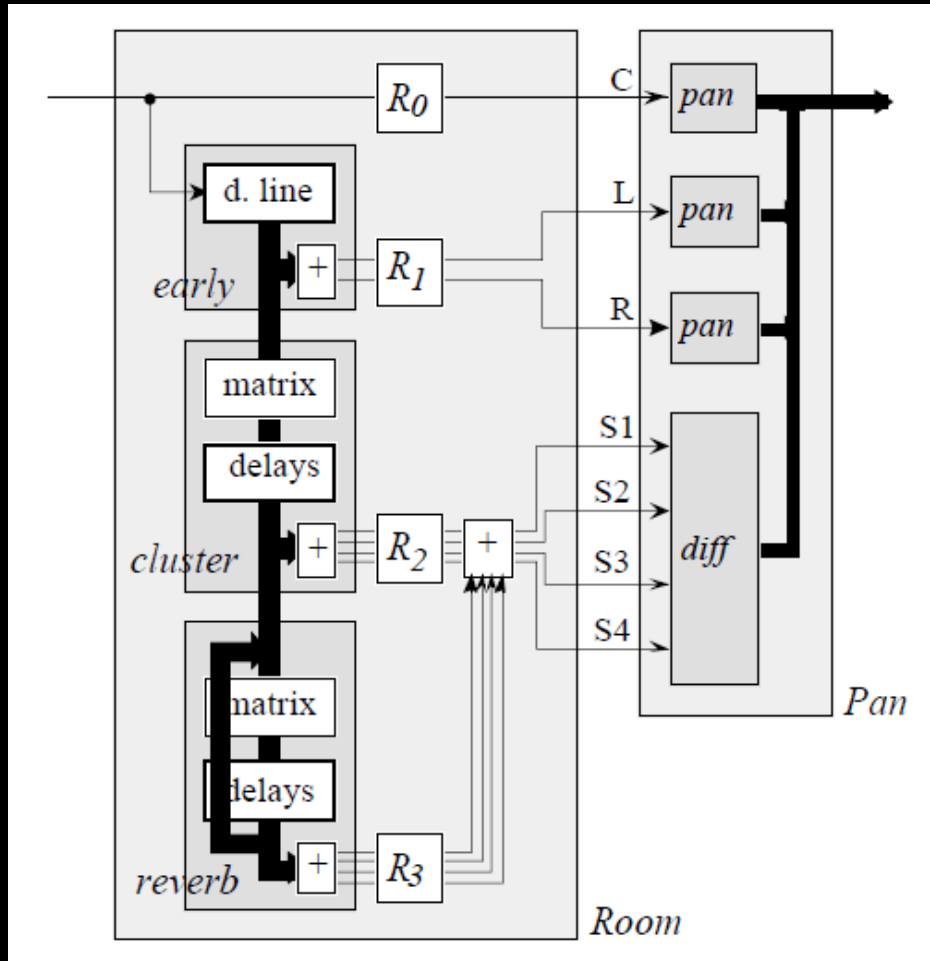
www.youtube.com/watch?v=XPLSrY4xLRw

... controlling reverberation & source parameters: Distance / proximity, Yaw (orientation), Aperture (directivity)

The screenshot displays the Pro Tools software interface with the FLUX:: IRCAM Spat v3 plugin open. The interface is divided into several sections:

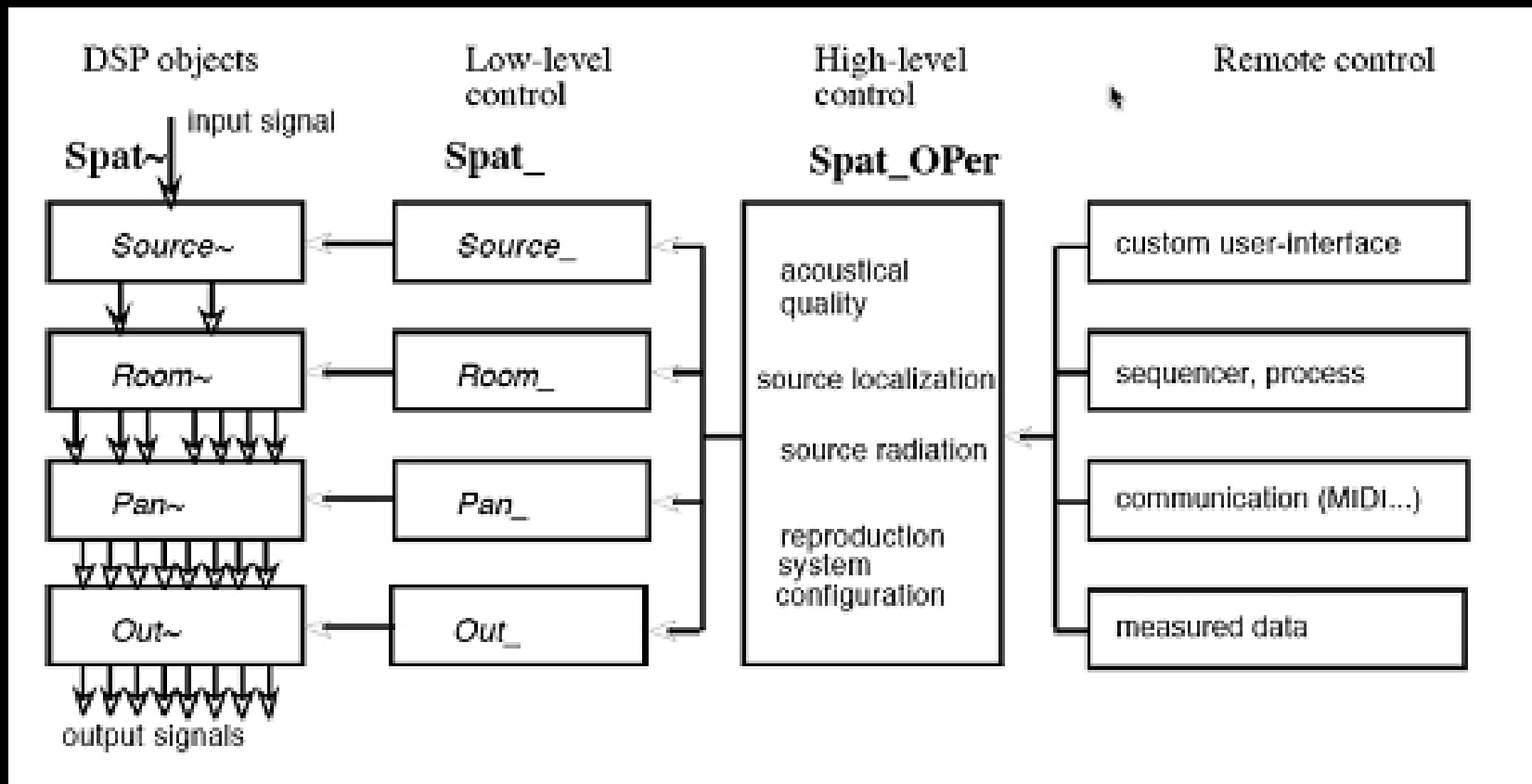
- Top Bar:** Shows the timecode at 1:00.885, a cursor at 0:19:105, and various transport controls. The track name is "Edit: Flux Spat v3 [Test]".
- Tracks:** On the left, a list of tracks is visible, including "02CnrySngs", "22 Bob_Ld1", and "Rec Bus".
- Plugin Interface:** The main window shows the "ircam Tools" interface for "FLUX::". It includes:
 - Perceptual Factors:** Sliders for Source Presence, Source Warmth, and Source Brilliance.
 - Room Presence:** Sliders for Reverb and Envelop.
 - Acoustical criteria:** A small display showing parameters like RIR, RT60, and ABR.
 - Radiation:** A central circular diagram with a pointer indicating the source direction. Below it are sliders for Distance, Azimuth, Yaw, Elevation, and Pitch.
 - EQ Section:** A frequency response graph with sliders for Low Freq (177 Hz) and High Freq (5657 Hz), and gain controls for Low, Med, High, and Global.
 - Gain Section:** Sliders for Input and Output gain.
- Bottom Bar:** Shows "SPAT" and "at. absorption" controls.

Spat — generic per-source processing architecture



- Directional early reflections
- Diffuse reverberation
- Per-source perceptual controls
- Format agnostic representation
- Library of "Pan" modules
- Extensible: immersion with height.

Spat — generic per-source processing architecture



FLUX:: IRCAM Spat Revolution (2017)

<http://www.spatrevolution.com/>

ircam by **FLUX::** tools

Setup Room 1 Effects

SOURCES

Search for sources or groups

- 1 Bass Mic 0.0 dB
- 2 Bougarabou 0.0 dB
- 3 Couple 0.0 dB
- 4 Farfi 1 0.0 dB
- 5 Farfi 2 0.0 dB
- 6 Farfi 3 0.0 dB
- 7 Guitar 0.0 dB
- 8 Guitar XY 0.0 dB
- 9 Guitar Rete 0.0 dB
- 10 Drum OverHead 0.0 dB
- 11 Drum Kick 1.2 dB

REVERB

7100 ms 15.0

OUTPUT M 0.0 dB

Speaker Alpha: 100.00% Shininess: 20% Lightness: 15% Back Color: Black

Search for anything about sources and their properties...

Send **Position** **Barycentric** **Radiation** **Perceptual Factors** **Options** **Spectral Axis** **Spectral Omni**

74.0 dB LFE

0.01 Pos X 0.13 Pos Y

0.00 deg Rotation X 0.00 deg Rotation Y

0.00 deg Rotation Z 1.00 % Width

0.70 deg Azimuth 13.00 deg Elevation 0.44 m Distance

13.31 deg Yaw 0.00 deg Pitch 70.75 deg Aperture

72 Presence 10 Warmth 30 Brilliance

16 R. Presence 34 Ass. Res. 75 Envelop.

Relative dir.

Doppler Air Absorption

Drop Mode

4.00 dB Drop Factor 0.50 Panner 10.0 deg Early Width 1.0 m Radius

177 Hz 6437 Hz

0.0 dB Amp L, Gain 0.0 dB Amp M, Gain 0.0 dB Amp H, Gain 0.0 dB Amp Gain

1.7 dB Over L, Gain 0.0 dB Over M, Gain 0.0 dB Over H, Gain -14.1 dB Over Gain

Provide feedback

device: None @ 48000 Hz, 1024 imp./block

DIRECTIONAL AUDIO ENCODING AND RENDERING

Designing the elementary "pan" module

Recording vs. "panning"

Common framework: pan law \leftrightarrow microphone directivity

Criteria. Re-recording principle. Psychoacoustics.

Binaural reproduction – microphones = ears

Dummy head. BRIR, HRTF measurements. Head-tracking. Cross-talk cancellation.

Performance limitations, challenges (more later re. VR audio). Cognitive factors!

Ambisonics – microphones = spherical harmonics (or linear combinations thereof)

First-order Ambisonics (FOA) – 4 channels

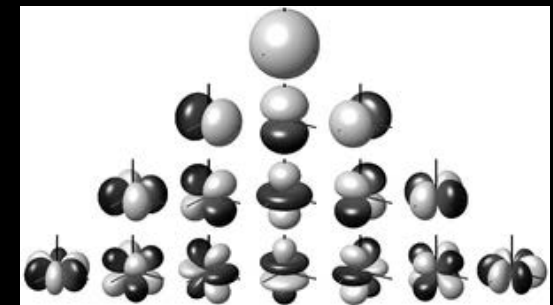
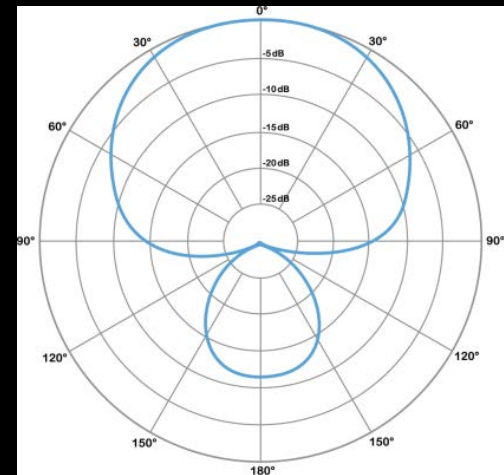
LF and HF decoder solutions

Gerzon localization vectors

Global interpolation over all speakers.

High-order Ambisonics (HOA) – more channels: 9, 16...

Linearly extend sweet spot size vs. order/frequency ratio.



DIRECTIONAL AUDIO ENCODING AND RENDERING

Designing the elementary “*pan*” module

Amplitude panning – optimizing localization “discreteness”

Local interpolation: panning weights given by centroid of nearest speaker localization vectors.

Egocentric: e.g. “vector based” (VBAP, VBIP)

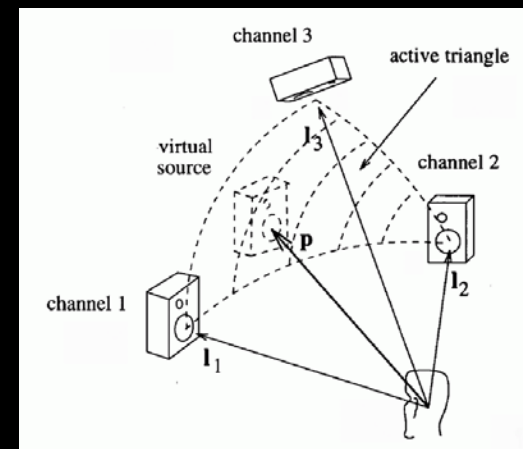
Allocentric: e.g. “distance based” (DBAP)

Holography – pressure field reconstruction over extended area

Wave-field synthesis (WFS), delay-based panning

Direction vs. localization, audio vs. visual

Theoretical equivalence with HOA for increasing order.



Challenges (for “surround” or “immersive” audio)...

No “one-size-fits-all” solution for consistent experience in all listening conditions

=> select the most effective rendering technique for given conditions

Rendering near-field or spatially extended sounds (incl. reverb)

Listening system calibration, device or room effect compensation.

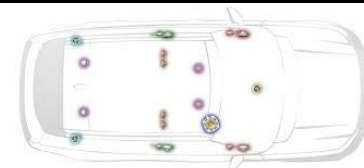
Off to the US...

Creative Ltd.
1998-2008

MPEG-4, EAX, OpenAL –
environmental audio for games.
Spatial audio “post-processing”.

DTS Inc.

DTS:X, Headphone:X –
consumer audio technology.
Immersive multi-channel audio



TWENTY YEARS OF IMMERSIVE AUDIO PROCESSING FOR GAMES / VR



APPROACHES TO FORMAT-AGNOSTIC AUDIO

Interactive object-based audio

MPEG-4 AABIFS, WFS. EU LISTEN project...

IRCAM, IoSono, Sonic Emotion, Astro Spatial Audio...

Game audio – EAX, OpenAL EFX (similar to Spat except in room effect control model)

Approach ok for non-linear, but not best for ubiquitous linear media content

Frequency-domain format conversion, parametric approaches

Examples: DirAC (Directional Audio Coding), SASC (Spatial Audio Scene Coding)

General approach: direct-diffuse decomposition, localization vectors

Why frequency domain: sparsity of representation, analogy with human hearing model

Metadata-assisted unmixing / informed source separation

MPEG SAOC (Spatial Audio Object Coding)

Informed Source Separation.

LINEAR IMMERSIVE AUDIO

Real-time rendering implies trade-off on fidelity.

Linear content production allows offline rendering => more MIPS per frame
... for both image and sound (e.g. computer generated animation or video).

For linear immersive audio content archiving and presentation, we need...

*Multi-channel recording format that faithfully encodes spatial audio cues
... but agnostic to loudspeaker configuration in the theater;*

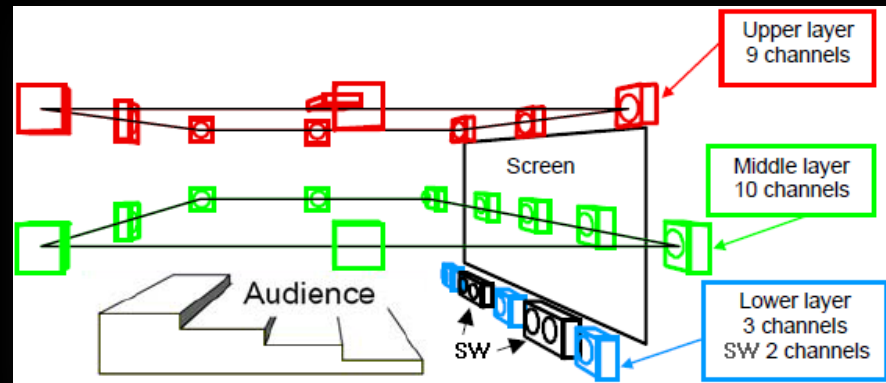
*Efficient delivery + faithful rendering in consumer environments
... flexible for playback in home, mobile, headphone, automotive scenarios.*

Create Once, Play Everywhere.

FROM SURROUND TO IMMERSIVE AUDIO FORMATS

Channel-based / scene-based *fixed* audio formats

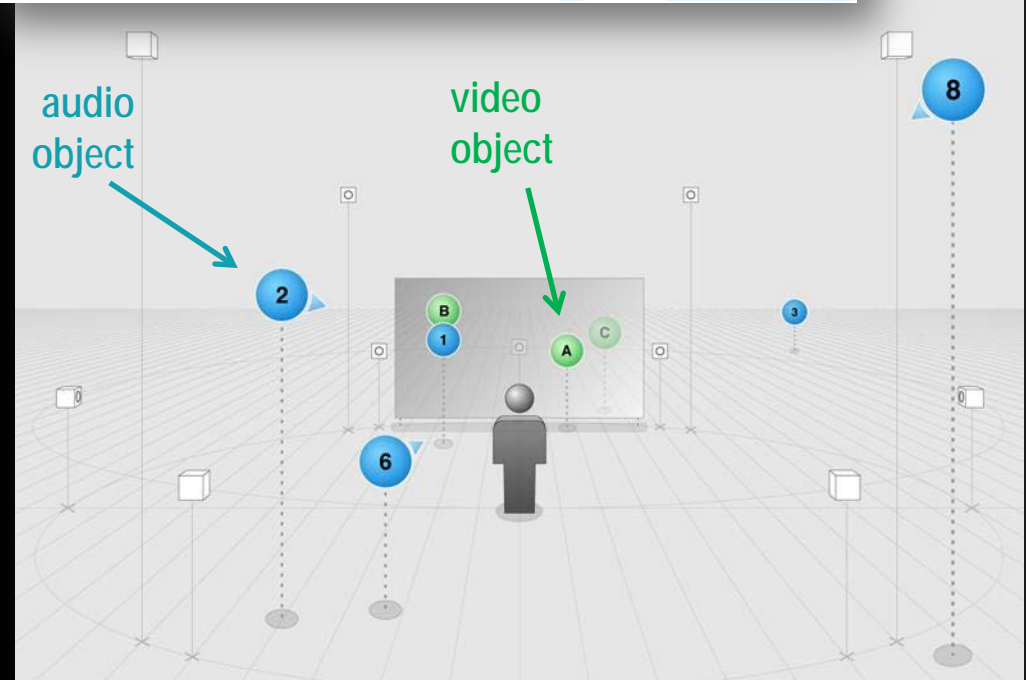
Add discrete "height" channels
High-order Ambisonics (HOA)
... "Baked" audio mix.



Object-Based Audio – *the return!*

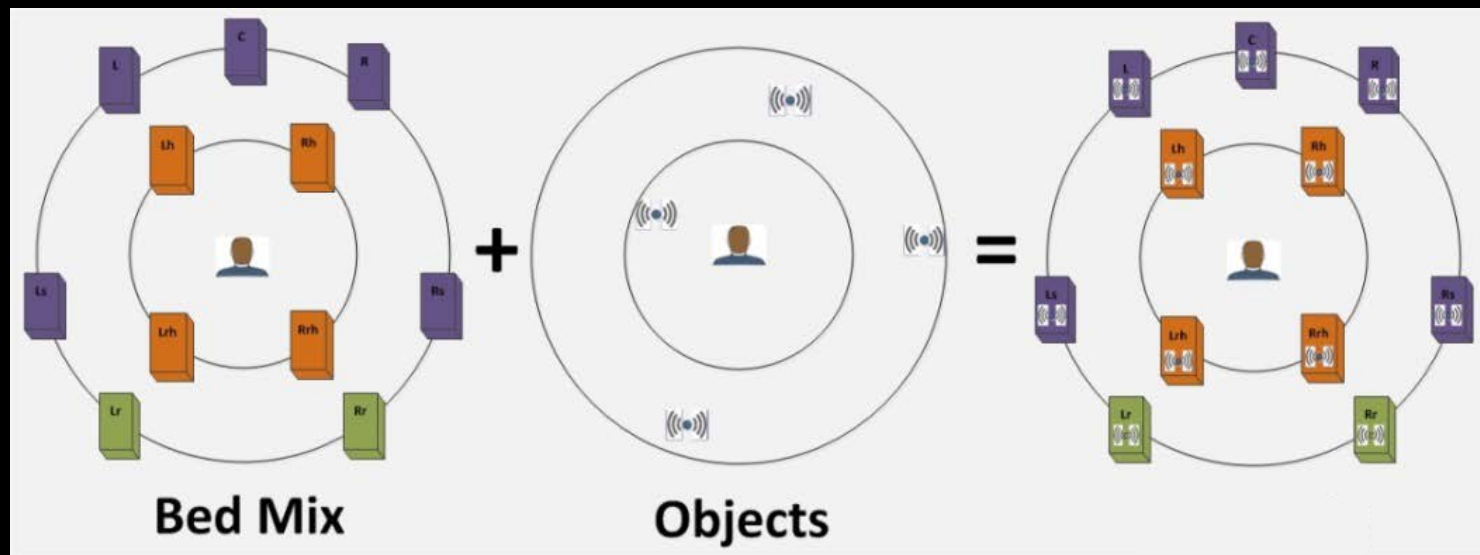
Audio essence tracks "rendered"
into mix at playback time

Scene description metadata
agnostic to playback configuration
Compromise-free object rendering
Audio/video spatial congruence.



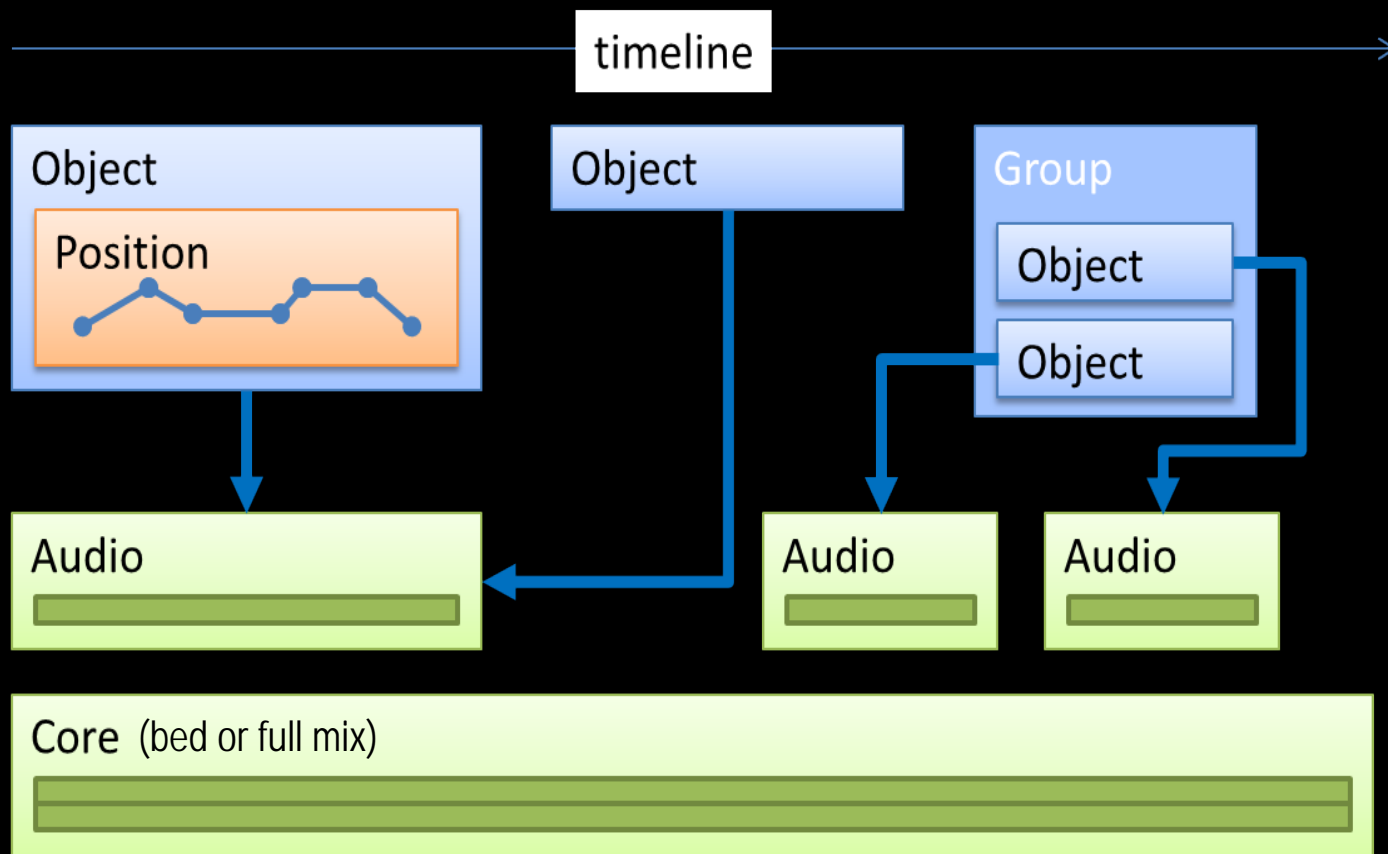
OBJECT-BASED SCENE DESCRIPTION

Production / theatrical: Auro, Atmos, MDA



OBJECT-BASED SCENE DESCRIPTION

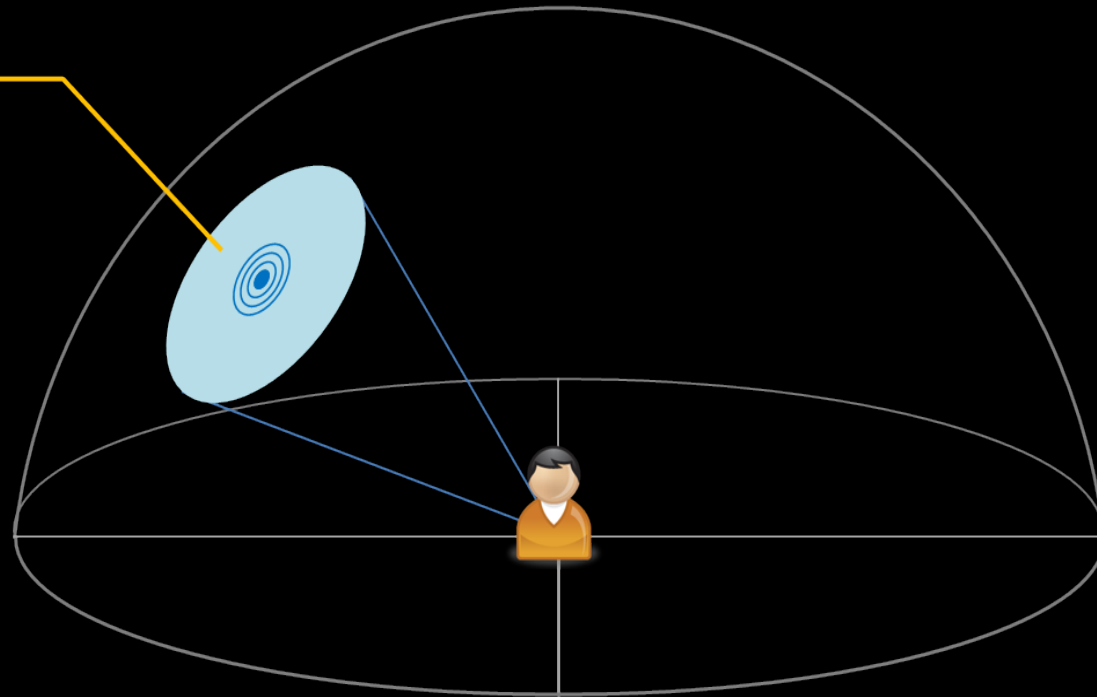
Delivery / disks, downloads, broadcast, streaming: DTS:X/MDA, Dolby AC4/Atmos, MPEG-H/ADM



AUDIO OBJECT PROPERTIES — “moving virtual loudspeaker”

Note: reverberation is encoded in the bed mix

Identifier
Asset Type (e.g. 'dialog')
Gain
Position
Spatial extent
Render exceptions
Priority
Loudness control data
DRC data
Dialog control data
...



MDA CREATOR IN PRO TOOLS

The screenshot displays the MDA Creator interface within a Pro Tools session. The central focus is a 3D visualization of a speaker array for a 'ucsd-24NEW.vbap [private]' track. The array is circular with 16 speakers arranged in a grid. Parameters for the array include Aperture (0.0°), Divergence (0.0°), Elevation (0.0°), and Azimuth (10.8°). A 'Send' button is visible at the bottom of the 3D view.

Surrounding the 3D view are various control panels:

- Monitoring:** Shows 'Standard' and 'Near-Field' options. The 'Near-Field' option is selected. Parameters include Speaker Configuration (ucsd-24NEW.vbap [private]), Monitor Reference (0), and Interactive Dialog Volume.
- Rendering:** Shows 'Render' and 'Export MDA' buttons. The 'Render' button is highlighted. The 'Export To' field is set to 'Unspecified'.
- Speaker Assignment:** A table showing speaker assignments for the 'MDA 7.1 Bed' track. The table is as follows:

Speaker	Assignment
A (0)	1
B (1)	2
C (2)	3
D (3)	4
E (4)	5
F (5)	6
G (6)	7
H (7)	8

At the bottom of the interface, there is a 3D city street scene rendered in a blue-tinted, wireframe style. The scene shows a perspective view of a city street with buildings and a sky with a sunset or sunrise. The DTS logo is visible in the bottom right corner of the interface.

MDA TOOLS FOR OBJECT-BASED, IMMERSIVE AUDIO CONTENT CREATION

Suite of software tools for object based, immersive content production

Easy integration into existing production workflows and professional environments

Efficient and flexible solution for today's professional content production needs

Support for a vast variety of immersive speaker layouts

No dedicated rendering hardware required

Full DTS:X ecosystem feature support (Object preservation, Interactive Dialog Control)



DTS HEADPHONE-X MONITOR FOR MDA CREATION

The image displays two side-by-side screenshots of the DTS Headphone-X Monitor software interface, showing different audio configurations.

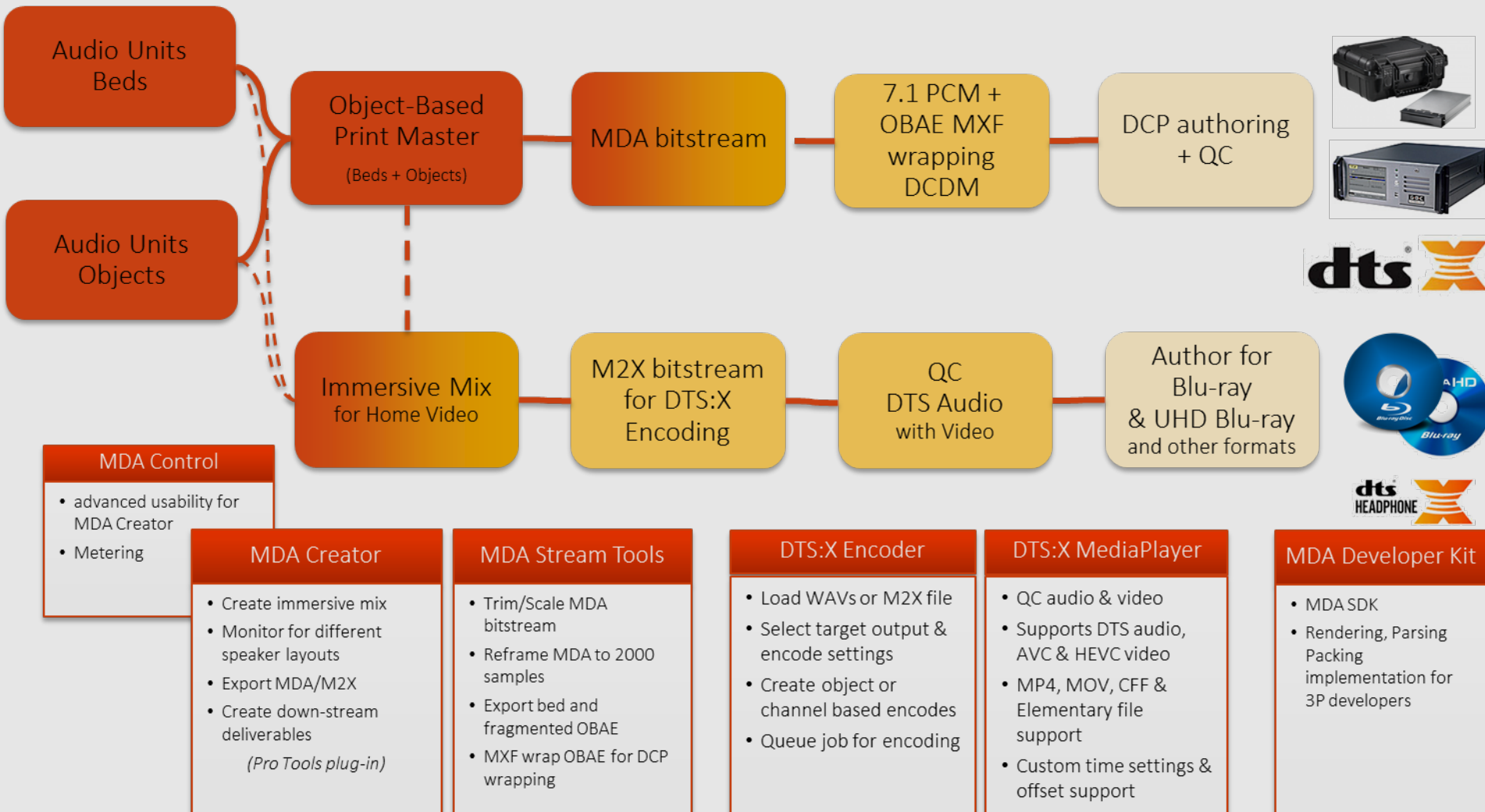
Left Screenshot (7.1 Standard - NEAR):

- Track:** 7.1 Out
- Preset:** <factory default>
- Auto:** BYPASS
- Room Profile:** 7.1 Standard - NEAR
- Surround Config:** 7.1
- Speaker Channels:** Lss, L, C, R, Rss, Lsr, Rsr, Lh, Rh, Lhr, Rhr, LFE
- Output:** Surround, Headphone:X (selected), Sennheiser HD650
- Input Format:** 7.1

Right Screenshot (4.0 Height Standard - NEAR):

- Track:** Height Quad Out
- Preset:** <factory default>
- Auto:** BYPASS
- Room Profile:** 4.0 Height Standard - NEAR
- Surround Config:** 4.0 High
- Speaker Channels:** Lhr, Rhr, Lh, Rh, Lsr, Rsr, L, C, R, Lss, Rss, LFE
- Output:** Surround, Headphone:X (selected), Sennheiser HD650
- Input Format:** Quad

DTS:X CONTENT CREATION SOLUTIONS



DTS:X CONTENT CREATION SOLUTIONS

Digital Cinema

Content Production

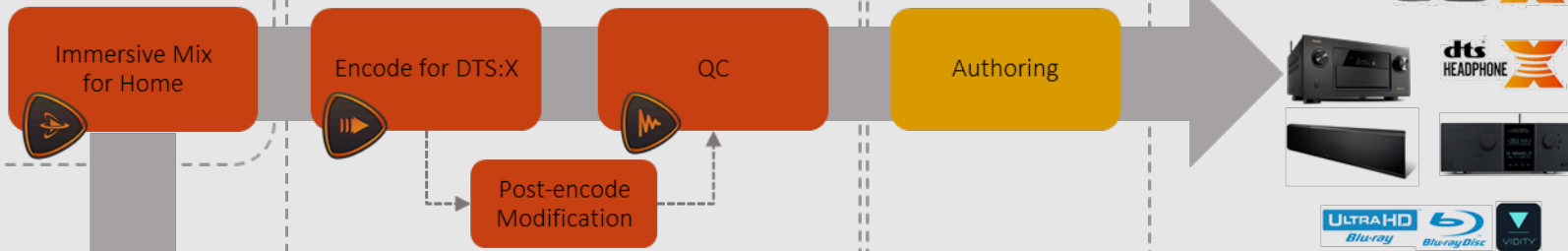
Elementary Stream Creation

Mux/Package



Digital Media

Sourced from Cinema



Streaming

Sourced from Cinema



WORLDWIDE ADOPTION – DTS:X & MDA TECHNOLOGY

Future-proof content mezzanine / archiving format

Published standard specs: ETSI TS 103 223, ...

Royalty-free in professional content industry



Worldwide deployment in authoring facilities, theaters and homes (as of Nov 2016)

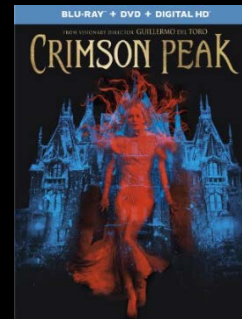
Over 35 post-production and authoring facilities world-wide

130+ DTS:X theatres

All major AVR manufacturers

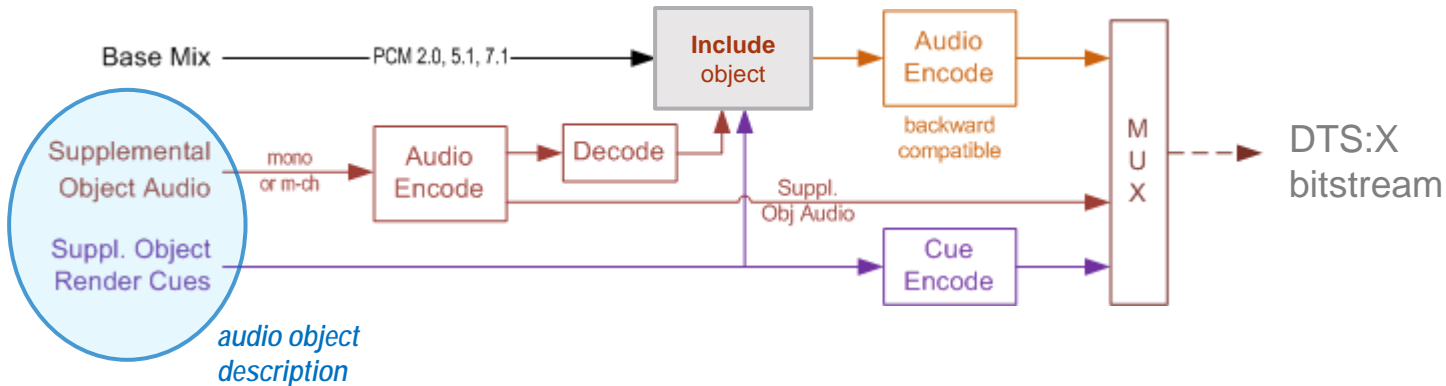


More than 50 international titles released

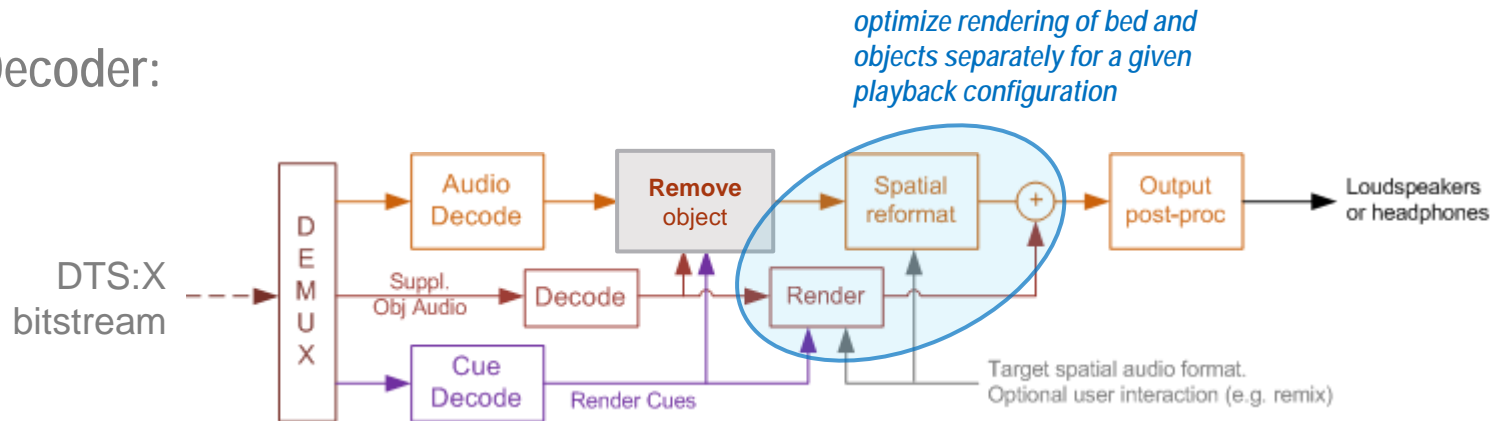


LEGACY DECODER COMPATIBILITY

Encoder:



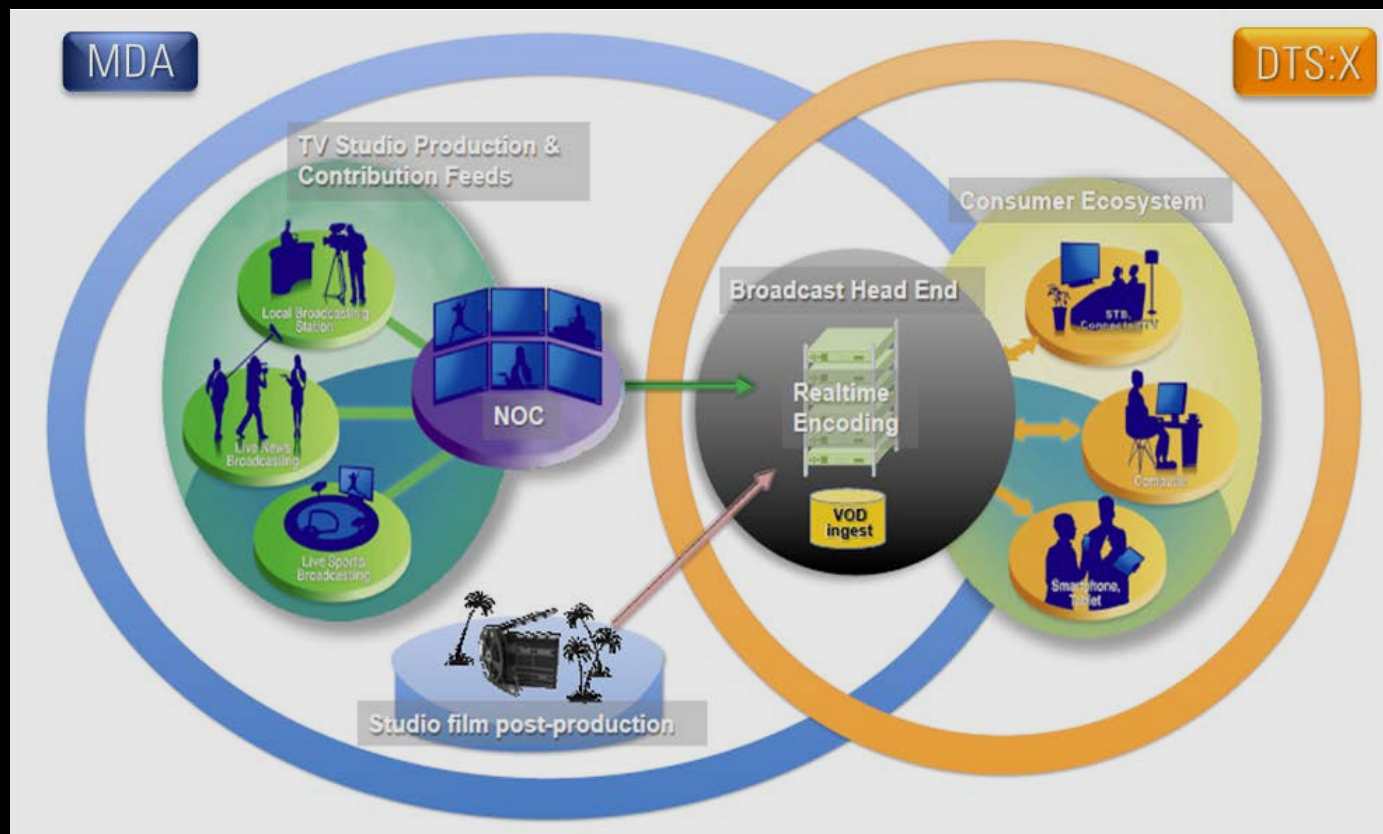
Decoder:



NEXT...

More theaters, homes, film & Blu-ray releases...

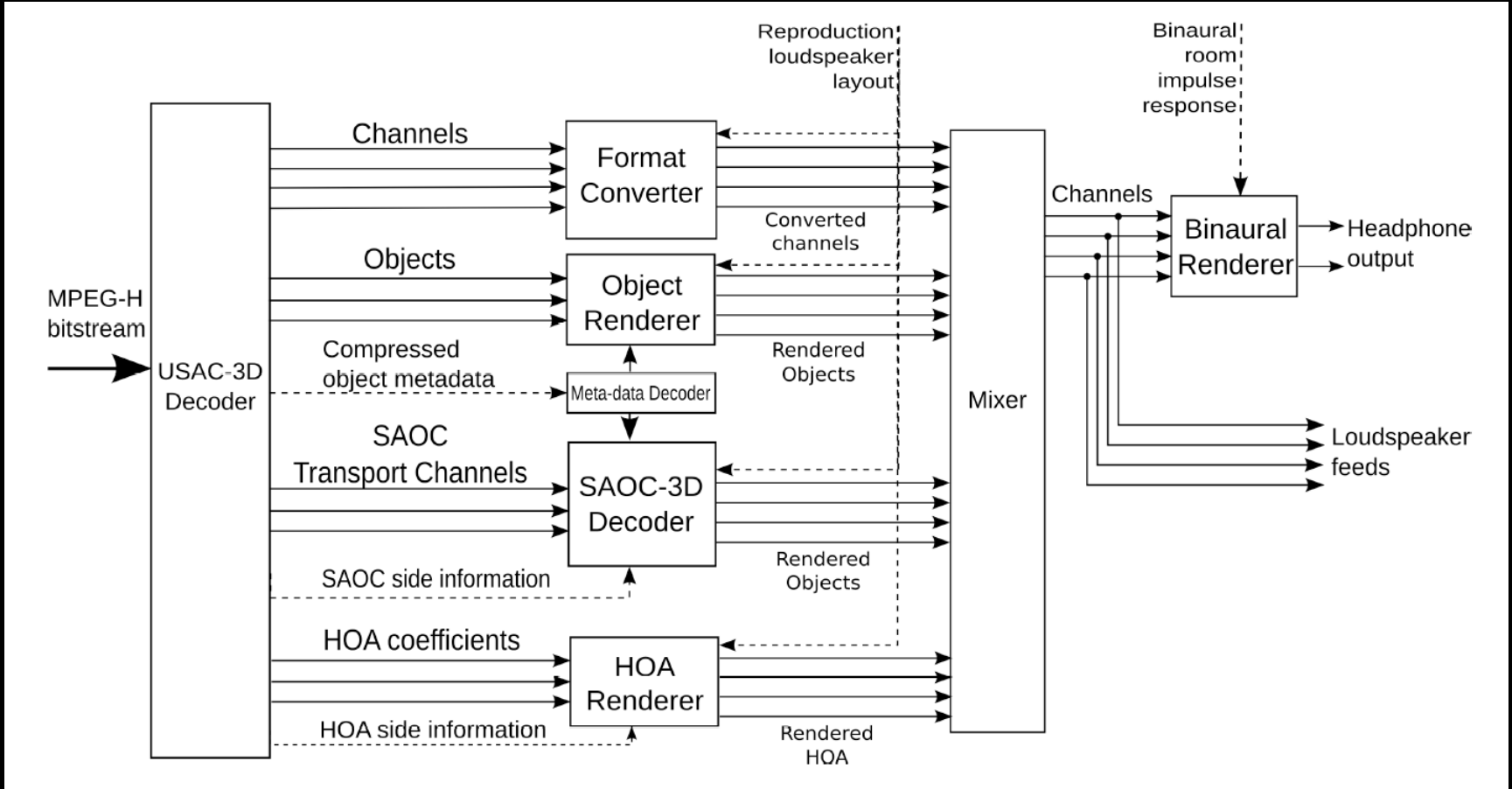
Broadcast, streaming. Mobile devices. Virtual reality.



SCALABILITY FROM LOW BIT RATE TO LOSSLESS

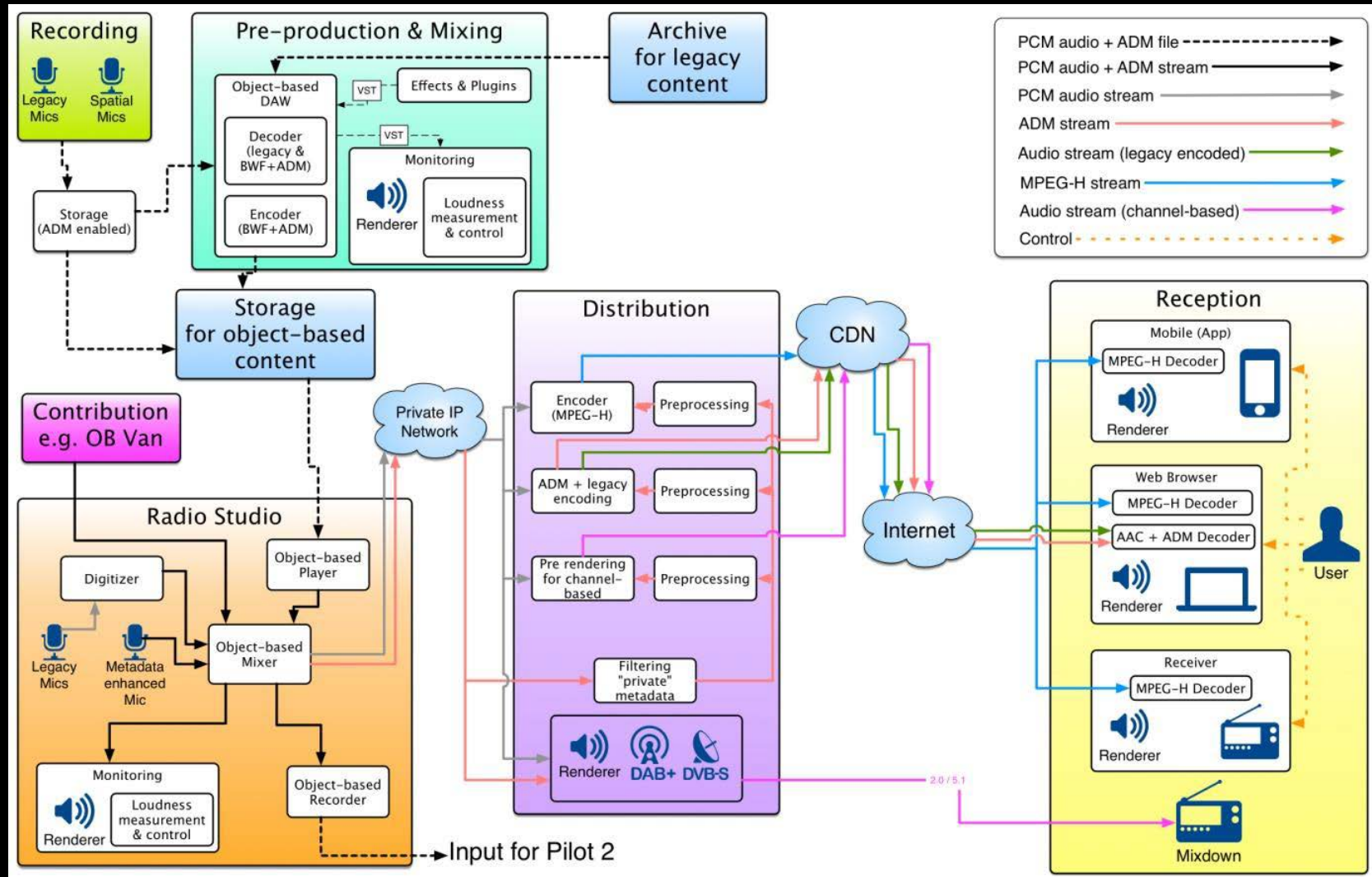
Channels	Bitrates [kbps] (Operational Range)	Bitrates [kbps] (Nominal range for broadcast quality)
Mono	24 - Lossless	48 - 64
Stereo	32 - Lossless	80 - 128
5.1	80 - Lossless	224 - 300
7.1	128 - Lossless	256 - 340
11.1	192 - Lossless	288 - 384
22.2	320 - Lossless	512 - 640

MPEG-H DECODER



ORPHEUS – OBJECT-BASED AUDIO EXPERIENCE

<https://orpheus-audio.eu/>



CONSUMER EXPERIENCE

... as enabled by object-based audio

Immersion

Elevation effects, realistic diffuse sounds/ambiences

Enable optimal spatial audio fidelity

Flexibility

Mobile, Home, Car...

Non-standard loudspeaker layouts

Ease of setup

Personalization

Dialog intelligibility enhancement

Dynamics Control

Ability to "change the mix"

... to the extent permitted by author



DTS:X PRACTICAL REPRODUCTION AT HOME



DTS:X PRACTICAL REPRODUCTION AT HOME



DTS:X PRACTICAL REPRODUCTION AT HOME



DTS:X PRACTICAL REPRODUCTION AT HOME — DTS VIRTUAL:X



How can virtual elevation cues work although they are (primarily) monaural?
... suggesting research on dynamic and differential elevation cues.

DTS:X PRACTICAL REPRODUCTION AT HOME — DTS VIRTUAL:X



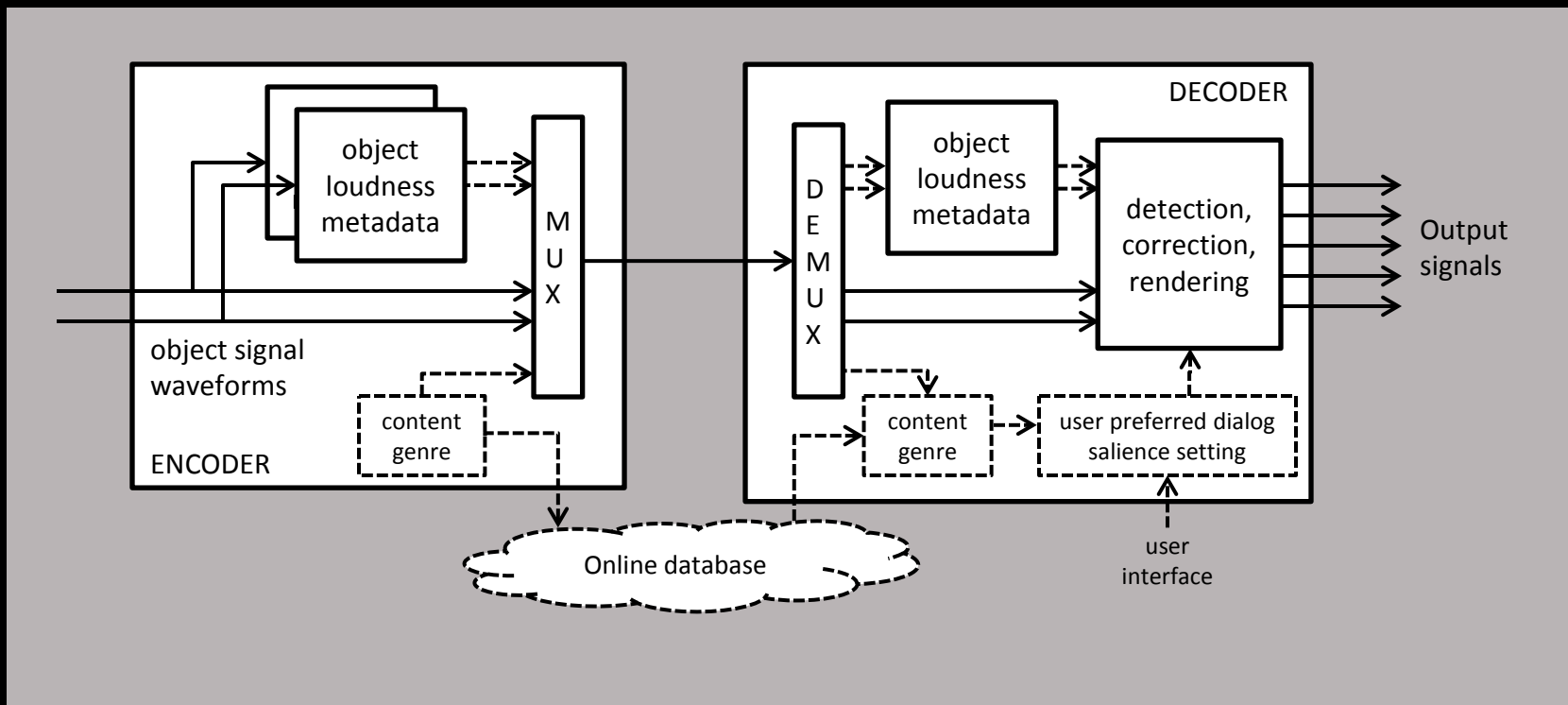
How can virtual elevation cues work although they are (primarily) monaural?
... suggesting research on dynamic and differential elevation cues.

OBJECT-BASED DIALOG CONTROL

Encode: include loudness metadata

Global programme loudness, global dialog loudness/salience measures

Optionally, short-term dialog loudness/salience measure sequence.



CINEMATIC VR AUDIO — demo

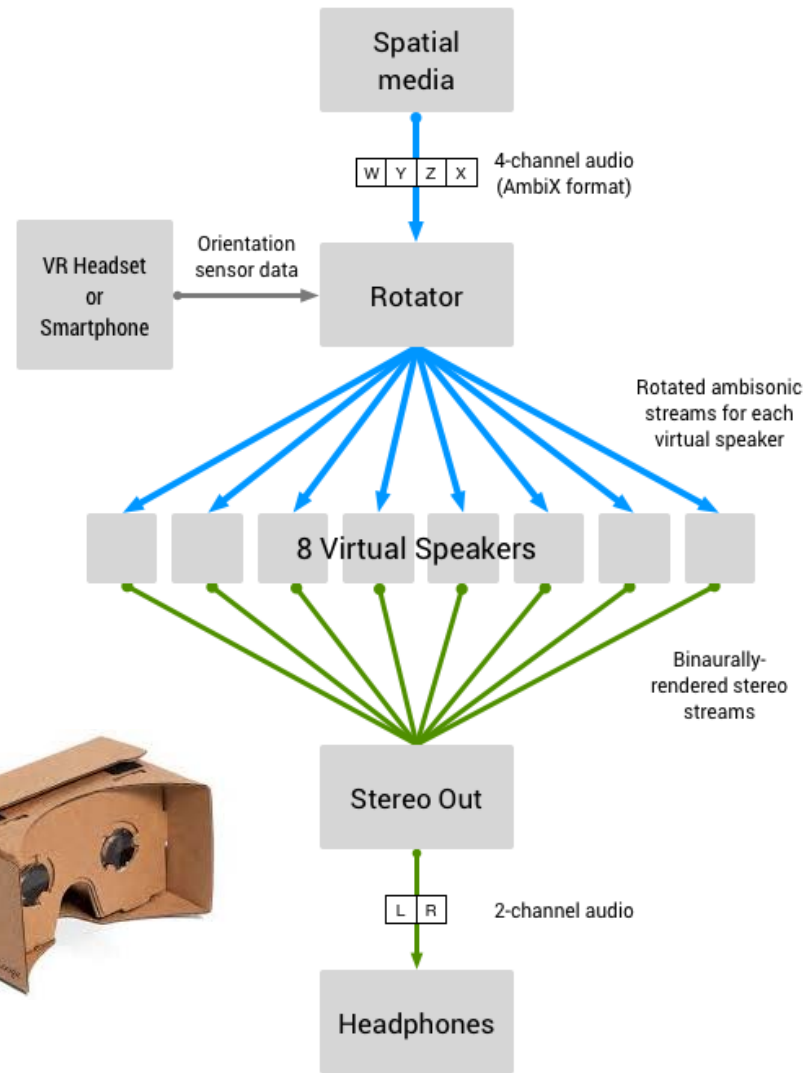
www.youtube.com/watch?v=9RamLHvdfms

Amp Fiddler - "Fiddler on tha Roof" - Output, Brooklyn - 6/24/16 - 360° video with spatial audio

Amp Fiddler - "Fiddler on tha Roof" - Output, Brooklyn - 6/24/16 - 360° video with spatial audio



CINEMATIC VR AUDIO — Ambisonics and HRTF-based virtualization



CINEMATIC VR AUDIO — Ambisonics and HRTF-based virtualization

Principles

Chris Travis 1996 AES paper “Virtual Reality Perspective on Headphone Audio”

Head-tracking at decode (for *diegetic* sounds)

Any source format compatible in principle by dynamic speaker virtualization

Current prevalent internet streaming formats

Google: Opus low-bit-rate codec + Ambisonics (1st order, 3rd order)

Facebook

Perspectives

Limitation of Ambisonic rendering: frequency vs. order for size of head (ear positions)

Extension to 6DOF: Ambisonic bed + separate objects...

MORE PERSPECTIVES ...

Natural-sounding audio scenes accompanying our experience of the environment

Teleporting into another (virtual) world – “you are there” experience.

With or without image. Linear or non-linear.

Success: effortless to tune into virtual scene + tune out real world (by occluding it, for instance)

What next? – Extending our physical world!

AR (non-diegetic sounds). MR (diegetic sounds).

MR in a dark/silent room is equivalent to interactive VR.

Success: minimize “cognitive effort”: listening/visual fatigue, attention conflict

Examples: telepresence, situational awareness (navigation, alerts...)

MORE PERSPECTIVES ...

Technological implications

Blurring distinctions between currently disparate media applications/industries

Music, movies, games, communication, performance, collaboration, travel...

Critical role of immersive audio technology: spatial congruence, naturalness

Minimize cognitive effort

Evolve the notion of audio object in immersive audio formats

Today: producer/broadcaster engineering perspective

Future: neuroscientist/composer perspective

Psychoacoustic notion of "sound event," "audio stream," "audio emitter"

Some previous attempts: IRCAM Spat. EAX, OpenAL...