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/
FLUX:: IRCAM Spat v3 — demo

www.youtube.com/watch?v=XPLSrY4xLRw

... controlling reverberation & source parameters: Distance / proximity, Yaw (orientation), Aperture (directivity)
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/
DIRECTIONAL AUDIO ENCODING AND RENDERING

Designing the elementary “pan” module

Recording vs. “panning”

Common framework: pan law ↔ microphone directivity

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

1995 2000 2005 2010

PC “HW acceleration”

Game Console

Mobile

VR

Occlusion
Obstruction
Air absorption

Reverb “zones”
Dynamic reflections
Reverb “morphing”

Multiple reverbs, aux. FX

Own-voice reverb

Insert & output FX

Increasing complexity!

Panning
HRTF

Directivity
Distance
Doppler

Physics-based rendering

DirectX

VRML2

I3DL1

OpenAL

MPEG-4

I3DL2

Unreal

EAX2

EAX3

EAX1 (Creative)

EAX2

EAX3

EAX4

EAX5

EAX1 (Creative)

EAX2

EAX3

EAX4

EAX5

Two Big Ears ➔ Facebook

NVidia

AMD

Impulsonic ➔ Valve

VisiSonics

Unreal

XAudio

Wwise

Unity

OpenAL

EAX1

EAX2

EAX3

EAX4

EAX5

I3DL2 (Aureal)

FMOD

I3DL1

A3D2 (Aureal)

DirectX

A3D1

EAX1 (Creative)
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
...

![Diagram showing audio object properties with a virtual loudspeaker moving in a virtual space.](image-url)
MDA CREATOR IN PRO TOOLS
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
DTS:X CONTENT CREATION SOLUTIONS

Audio Units
- Beds

Object-Based
Print Master
(Beds + Objects)

MDA bitstream

7.1 PCM +
OBABE MXF
wrapping
DCDM

DCP authoring
+ QC

Audio Units
- Objects

Immersive Mix
for Home Video

M2X bitstream
for DTS:X
Encoding

QC
DTS Audio
with Video

Author for
Blu-ray
& UHD Blu-ray
and other formats

MDA Control
- advanced usability for MDA Creator
- Metering

MDA Creator
- Create immersive mix
- Monitor for different speaker layouts
- Export MDA/M2X
- Create downstream deliverables
  (Pro Tools plug-in)

MDA Stream Tools
- Trim/Scale MDA bitstream
- Reframe MDA to 2000 samples
- Export bed and fragmented OBAE
- MXF wrap OBAE for DCP wrapping

DTS:X Encoder
- Load WAVs or M2X file
- Select target output & encode settings
- Create object or channel based encodes
- Queue job for encoding

DTS:X MediaPlayer
- QC audio & video
- Supports DTS audio, AVC & HEVC video
- MP4, MOV, CFF & Elementary file support
- Custom time settings & offset support

MDA Developer Kit
- MDA SDK
- Rendering, Parsing Packing implementation for 3P developers
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:

Base Mix → PCM 2.0, 5.1, 7.1 → Include object → Audio Encode → Decode → DTS:X bitstream

Supplemental Object Audio

Suppl. Object Render Cues

audio object description

Decoder:

DTS:X bitstream → DEMUX → Audio Decode → Remove object → Spatial reformat → Output post-proc → Loudspeakers or headphones

Include object → Suppl. Obj Audio → Decode → Render → Target spatial audio format. Optional user interaction (e.g. remix)

Remove object

Optimize rendering of bed and objects separately for a given playback configuration
More theaters, homes, film & Blu-ray releases...
### Scalability from Low Bit Rate to Lossless

<table>
<thead>
<tr>
<th>Channels</th>
<th>Bitrates [kbps] (Operational Range)</th>
<th>Bitrates [kbps] (Nominal range for broadcast quality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mono</td>
<td>24 - Lossless</td>
<td>48 - 64</td>
</tr>
<tr>
<td>Stereo</td>
<td>32 - Lossless</td>
<td>80 - 128</td>
</tr>
<tr>
<td>5.1</td>
<td>80 - Lossless</td>
<td>224 - 300</td>
</tr>
<tr>
<td>7.1</td>
<td>128 - Lossless</td>
<td>256 - 340</td>
</tr>
<tr>
<td>11.1</td>
<td>192 - Lossless</td>
<td>288 - 384</td>
</tr>
<tr>
<td>22.2</td>
<td>320 - Lossless</td>
<td>512 - 640</td>
</tr>
</tbody>
</table>
MPEG-H DECODER

- MPEG-H bitstream
- USAC-3D Decoder
- Compressed object metadata
- SAOC Transport Channels
- SAOC side information
- HOA coefficients
- HOA side information

- Channels
- Objects
- Format Converter
- Object Renderer
- SAOC-3D Decoder
- HOA Renderer
- Meta-data Decoder
- Converted channels
- Rendered Objects
- Reproduction loudspeaker layout

- Binaural room impulse response
- Binaural Renderer
- Headphone output
- Loudspeaker feeds
- Mixer

Information on this slide is confidential property of DTS. Any unauthorized copying is strictly prohibited. © 2013 DTS, Inc.
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
How can virtual elevation cues work although they are (primarily) monaural?  
... suggesting research on dynamic and differential elevation cues.
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

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

www.youtube.com/watch?v=9RamLHvdFMs
CINEMATIC VR AUDIO – Ambisonics and HRTF-based virtualization

VR Headset or Smartphone → Orientation sensor data

Rotator → 8 Virtual Speakers

Spatial media → 4-channel audio (AmbiX format)

8 Virtual Speakers → Rotated ambisonic streams for each virtual speaker

Stereo Out → Binaurally-rendered stereo streams

Stereo Out → 2-channel audio

Headphones
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 (1\textsuperscript{st} order, 3\textsuperscript{rd} order)
Facebook

Perspectives
Limitation of Ambisonic rendering: frequency vs. order for size of head (ear positions)
Extension to 6DOF: Ambisonic bed + separate objects...
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...