Spatial Audio: From Screen-Based to World-Based with AR and VR

Jean-Marc Jot

Founder and Principal, Virtuel Works LLC
linkedin.com/in/jmmjot

An example of silent multiplayer AR demo (2022)

Sennheiser - Ambeo AR app, Magic Leap spatial audio (2018)

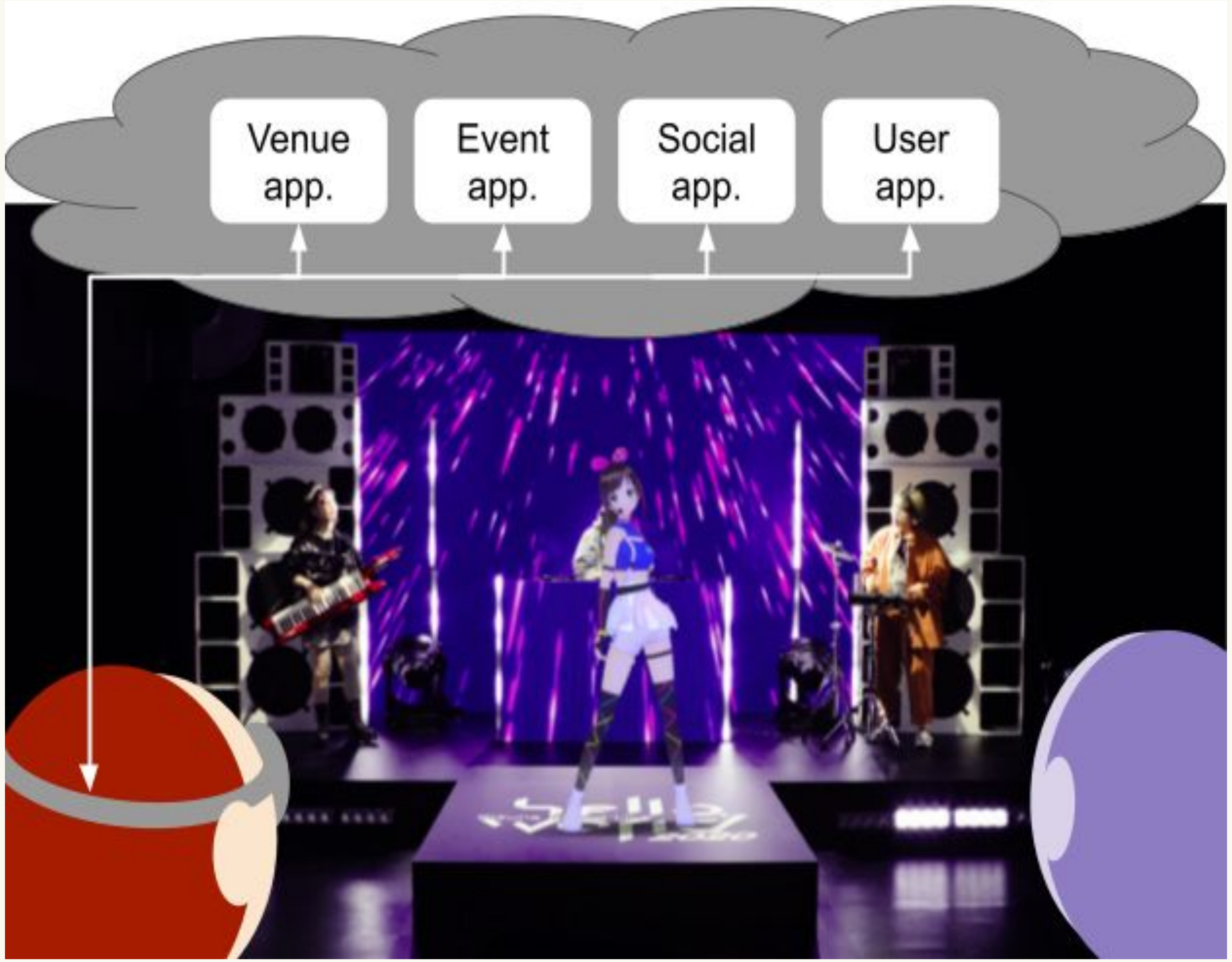
Example of a future "Metaverse" audio experience

6-Degree-of-Freedom listener navigation and placement of sound objects.

Real and/or virtual performers, audience members/avatars, and environment elements.

Wearable AR display device:
wireless, acoustically transparent, low audio latency, low power/heat.

Interoperability of apps & devices
deployed by independent vendors.

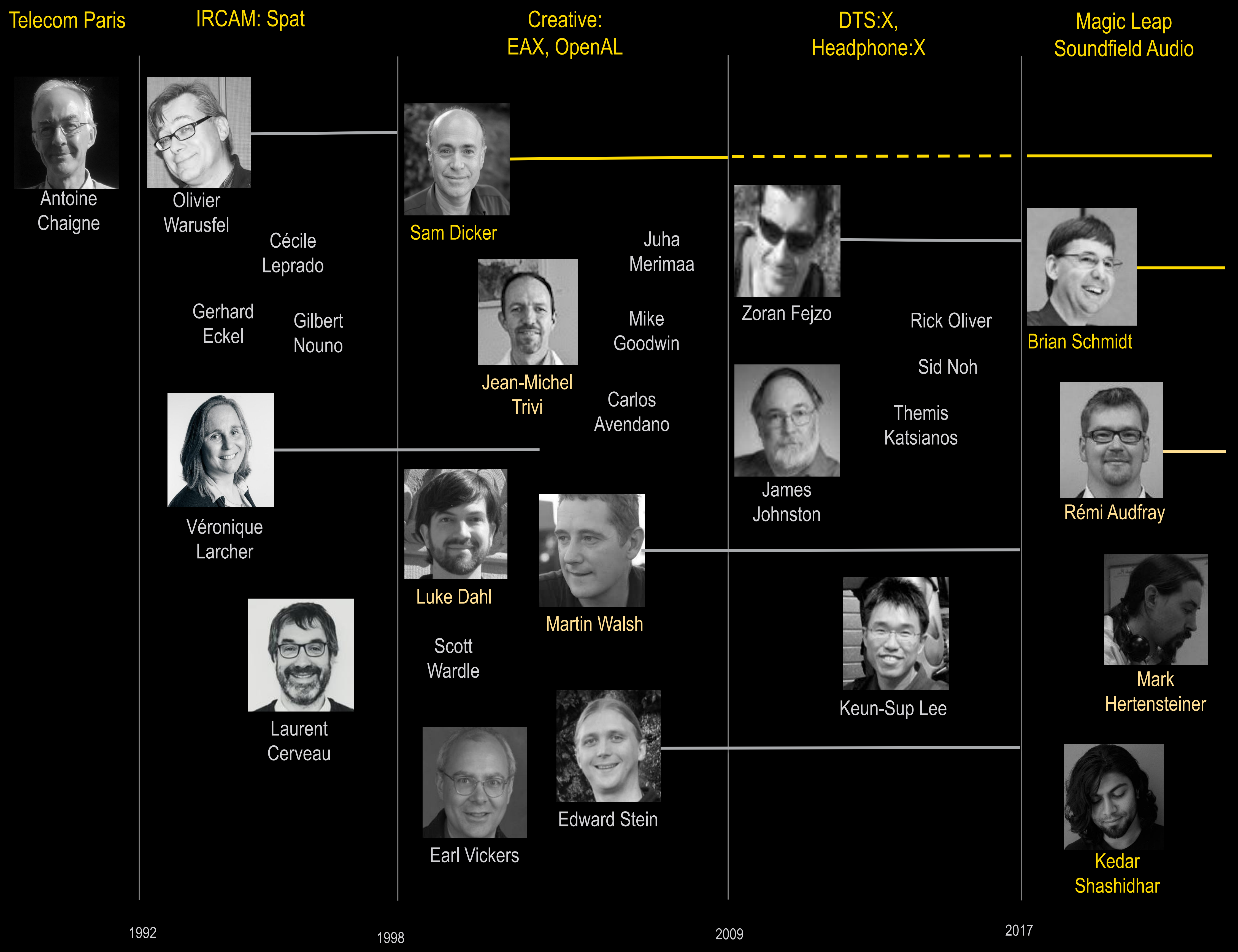


Some questions addressed in this talk…

- Opportunities enabled by spatial audio technology
  - in screen-based vs. world-based applications
  - in AR vs. VR applications
  - for diegetic sounds vs. music.
- What level of naturalness and immersion do devices and engines enable today?
  - What are some technological challenges that persist?

About me…

Virtuel Works Founder, Principal Consultant	9 mo	***Immersive & XR audio/music*** *Technology IP creation/strategy.*
iZotope/Soundwide VP Research, Chief Scientist	20 mo	***Audio/music production SW*** *Audio processing innovation/research.*
Magic Leap VP, Head of Audio & Media	3.5 y	***Augmented Reality HW & OS*** *Spatial audio, electroacoustics (ML1).*
DTS/Xperi Sr. VP, R&D	~8 y	***Audio SDK for consumer devices*** *TV, mobile, immersive audio (DTS:X).*
Creative Labs Audio Research Fellow	~10 y	***Computer & mobile peripherals*** *Game 3D audio standard (EAX).*
IRCAM + PhD studies Audio Research Scientist	~10 y	***Computer music creation SW*** *Spatial audio, room acoustics (Spat~).*


Telecom Paris
IRCAM: Spat
Creative:
EAX, OpenAL
DTS:X,
Headphone:X
Magic Leap
Soundfield Audio
Antoine
Chaigne
Olivier
Warusfel
Cécile
Leprado
Gerhard
Eckel
Gilbert
Nouno
Véronique
Larcher
Laurent
Cerveau
Sam Dicker
Jean-Michel
Trivi
Juha
Merimaa
Mike
Goodwin
Carlos
Avendano
Luke Dahl
Scott
Wardle
Martin Walsh
Earl Vickers
Edward Stein
Zoran Fejzo
James
Johnston
Rick Oliver
Sid Noh
Themis
Katsianos
Keun-Sup Lee
Brian Schmidt
Rémi Audfray
Mark
Hertensteiner
Kedar
Shashidhar
1992
1998
2009
2017

Context of audio technology evolution – from 1990s to 2020s…

- **From "surround" to "immersive" audio** (new audio content formats)
- **From "stereo speakers" to immersive wireless headphones, soundbars, car audio systems**
- **The coming of age of Ambisonics and the "unleashing" of binaural 3D audio experiences**
- **From game audio to XR audio** (Virtual and Augmented Reality)
- **Scope of interactive spatial audio technology: from podcast and movies to games and XR**
- **Binaural 3D audio rendering is a required feature in head-worn display devices.**

Panning
HRTF

Directivity
Orientation
Distance
Doppler

Occlusion
Obstruction
Air absorption
Reverb “zones”

Dynamic reflections

Reverb “morphing”

Multiple reverbs, aux. FX

Insert & output FX

Own-voice reverb

Physics-based rendering

1998 — **2000** — **2005** — **2010** — **2015**

PC *“HW acceleration”* — **Console** — **Mobile** — **Web** — **VR**

VRML2 MPEG-4 WebAudio

I3DL1 I3DL2 OpenAL OpenAL EFX

Google

NVidia

DS3D *XAudio* *FMOD* *Wwise*

Unreal *Unity*

VisiSonics *AMD*

A3D1 *A3D2*

Two Big Ears → Facebook

Impulsonic → Valve

EAX1 *EAX2* *EAX3* *EAX4* *EAX5*

Considerations about spatial audio in interactive applications – similarities and differences

- “Screen-based” vs. “world-based”
- VR vs. AR/MR
- Diegetic and non diegetic sound elements.

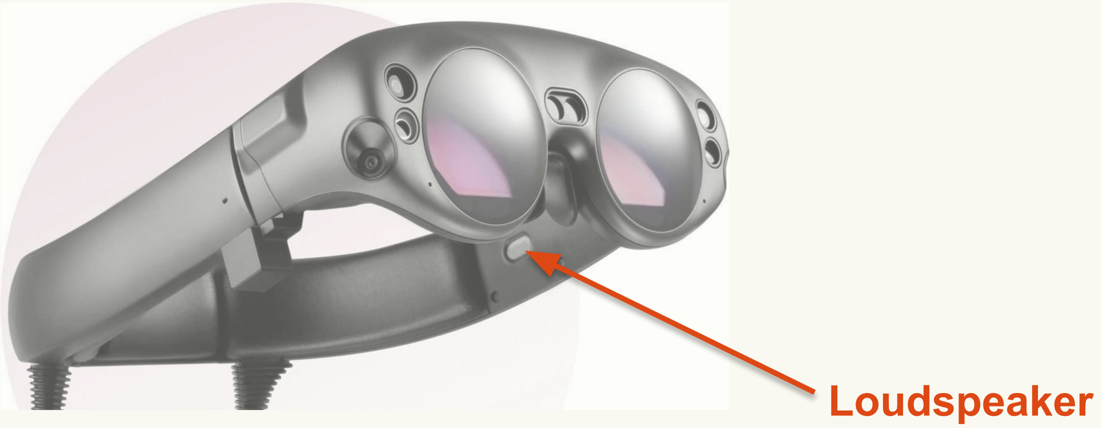
From 2D to 3D interactive audio

Video presentation	2D Screen (legacy console, PC)	Tethered VR (recent console, PC)	Untethered XR (mobile)
Listening experience	• Fixed (sitting) • Sounds stick to screen and speakers.	• 3-DoF: respond to head rotation (while sitting) • Binaural rendering is required • Sounds follow global scene rotations.	• 6-DoF: rotation + navigation (walk around) • Binaural rendering is required • Sounds must stick to identifiable objects.
Spatial audio compute	• MIPS: ~15% max • Memory: tight.	• MIPS: ample resource • Memory: loose.	• MIPS: 10%~15% • Memory: tight.
Mixing workflow	• Submix FX: OK • Low channel count.	• Binaural or multi-channel submixes.	• Multi-channel submixes • Submix FX: costly.

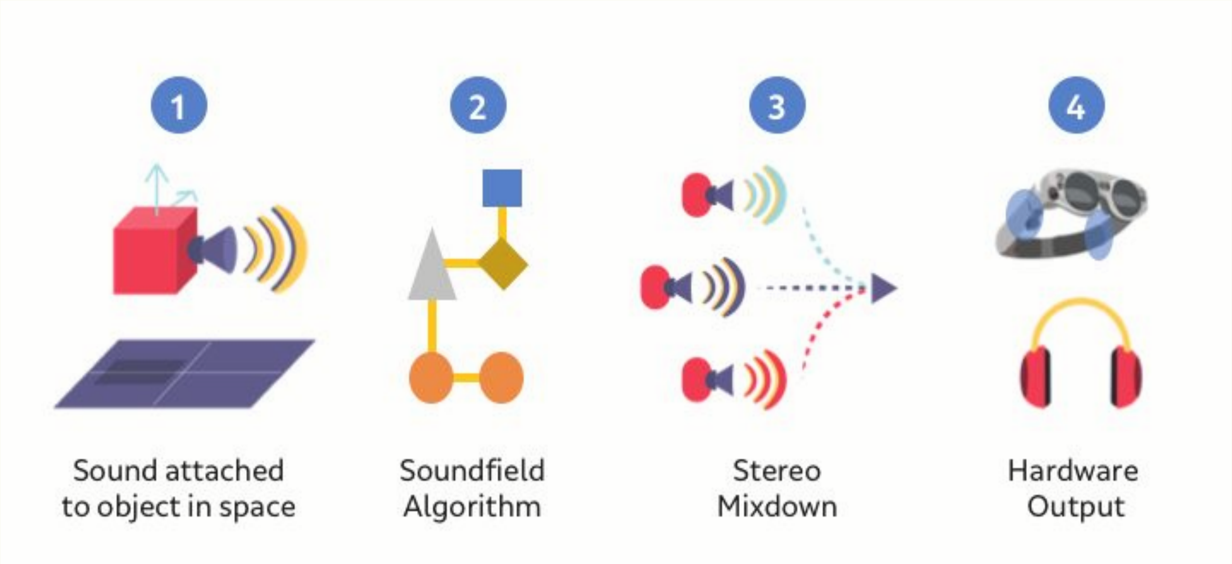
Magic Leap Soundfield Audio (MSA) engine and API

developer-docs.magicleap.cloud/docs/guides/unity/soundfield-audio/soundfield-components

In ML1:

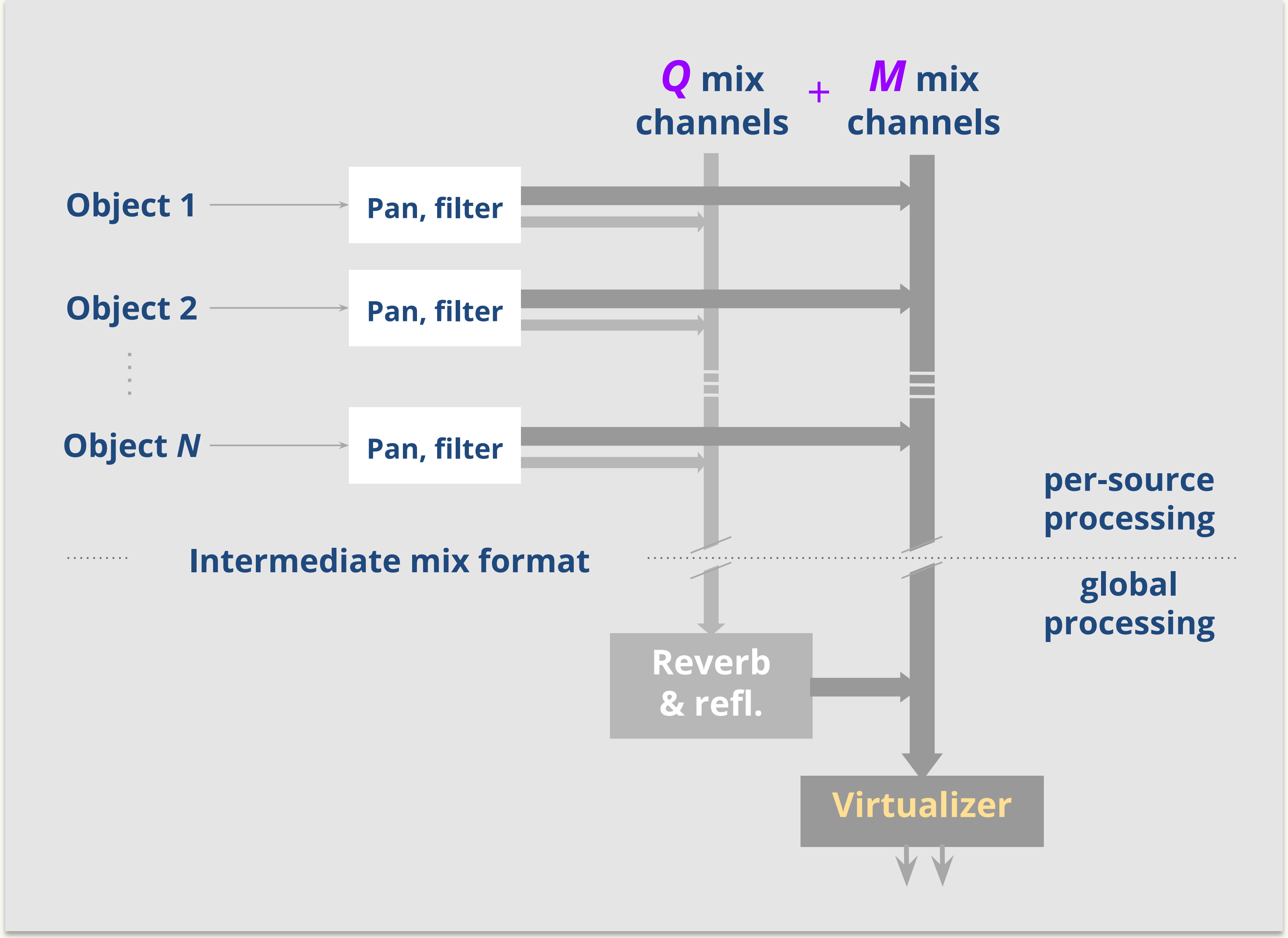


- Object-based binaural rendering
- 3-D and near-field positional audio
- Perceptual room model
- Source radiation model
- Physics-based distance/depth model



- Spectator motion tracking
- Video and audio rendering according to relative object-vs-spectator positions and orientations

From 2D to 3D interactive audio



2D Audio

- Virtualizer *was not needed*
- M + Q = 3 ~ 6 mix channels
  => processing a submix was cheaper than processing per object.

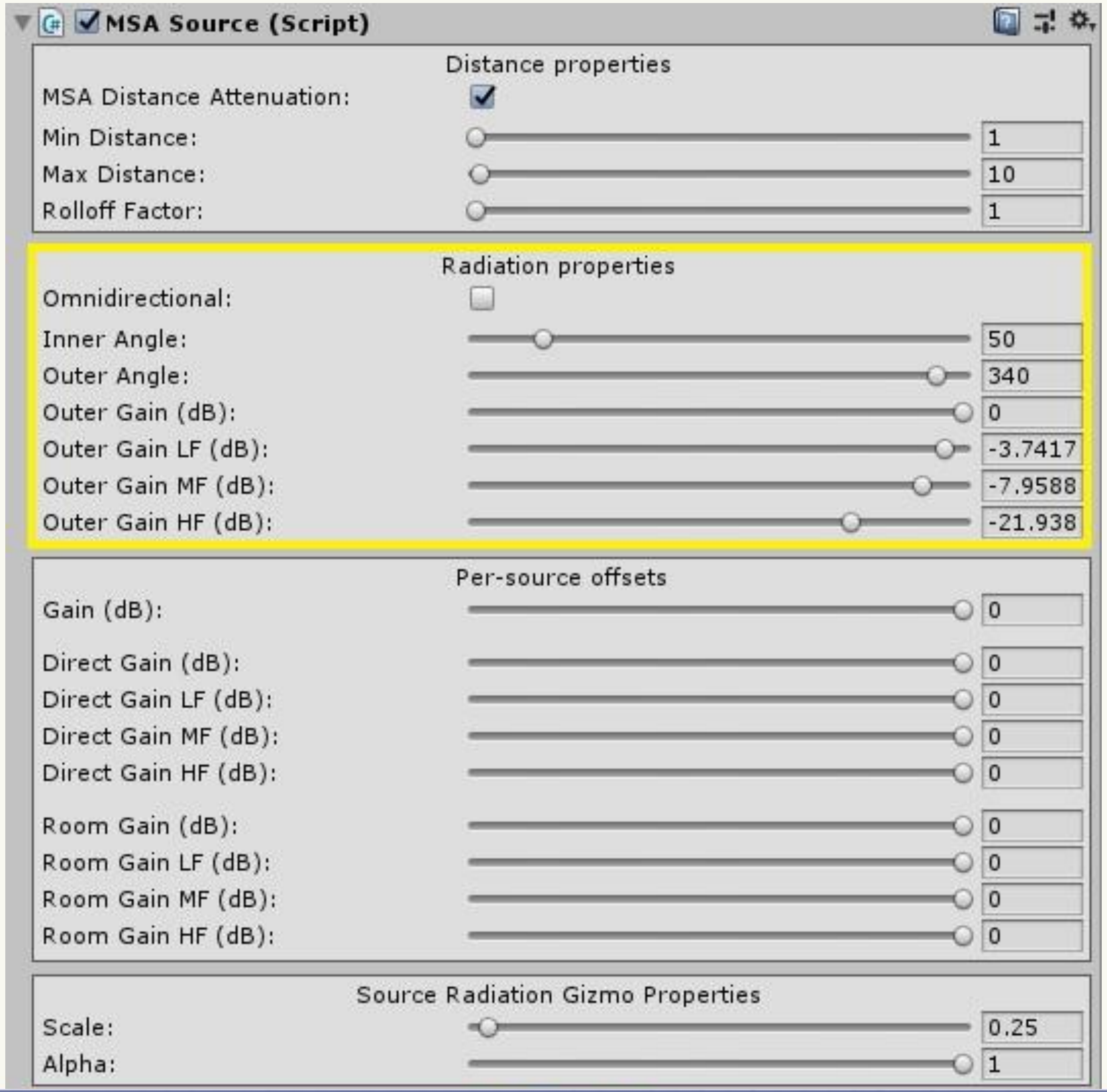
3D Audio

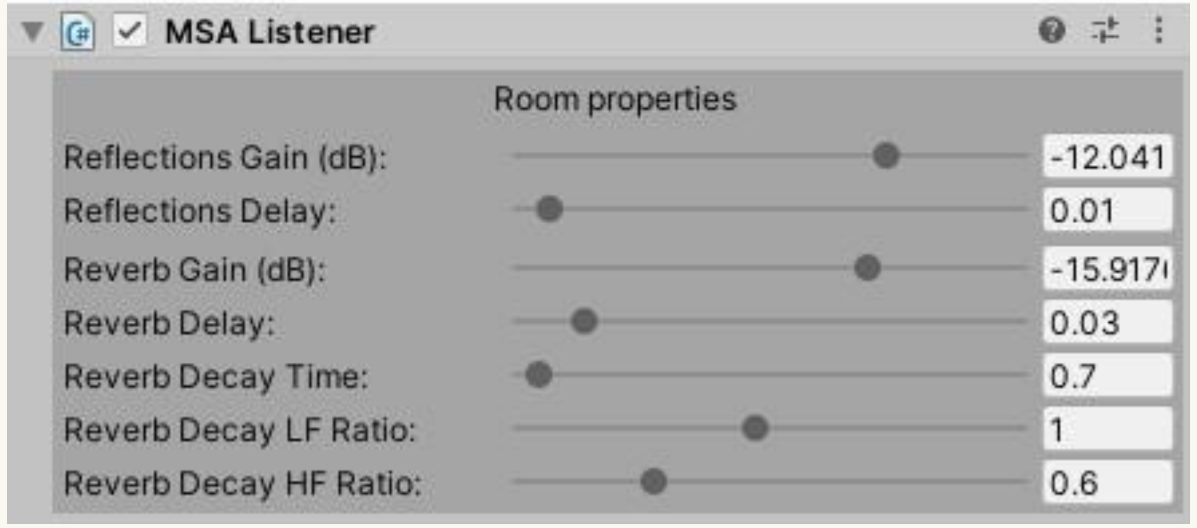
- Virtualizer is now *required*
  => additional baseline compute cost
- M + Q > 16 … "fat mix pipe"
  => processing per object is more efficient.

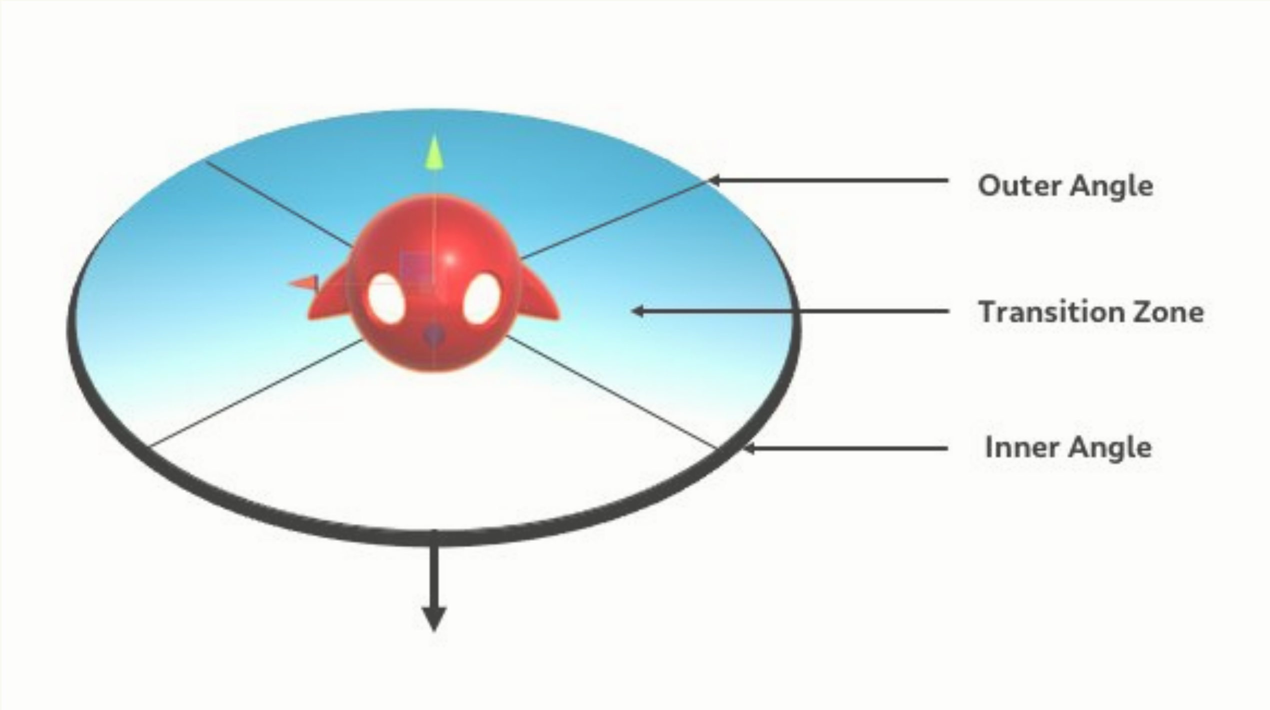
Magic Leap Soundfield Audio (MSA) engine and API

developer-docs.magicleap.cloud/docs/guides/unity/soundfield-audio/soundfield-components

ML1:



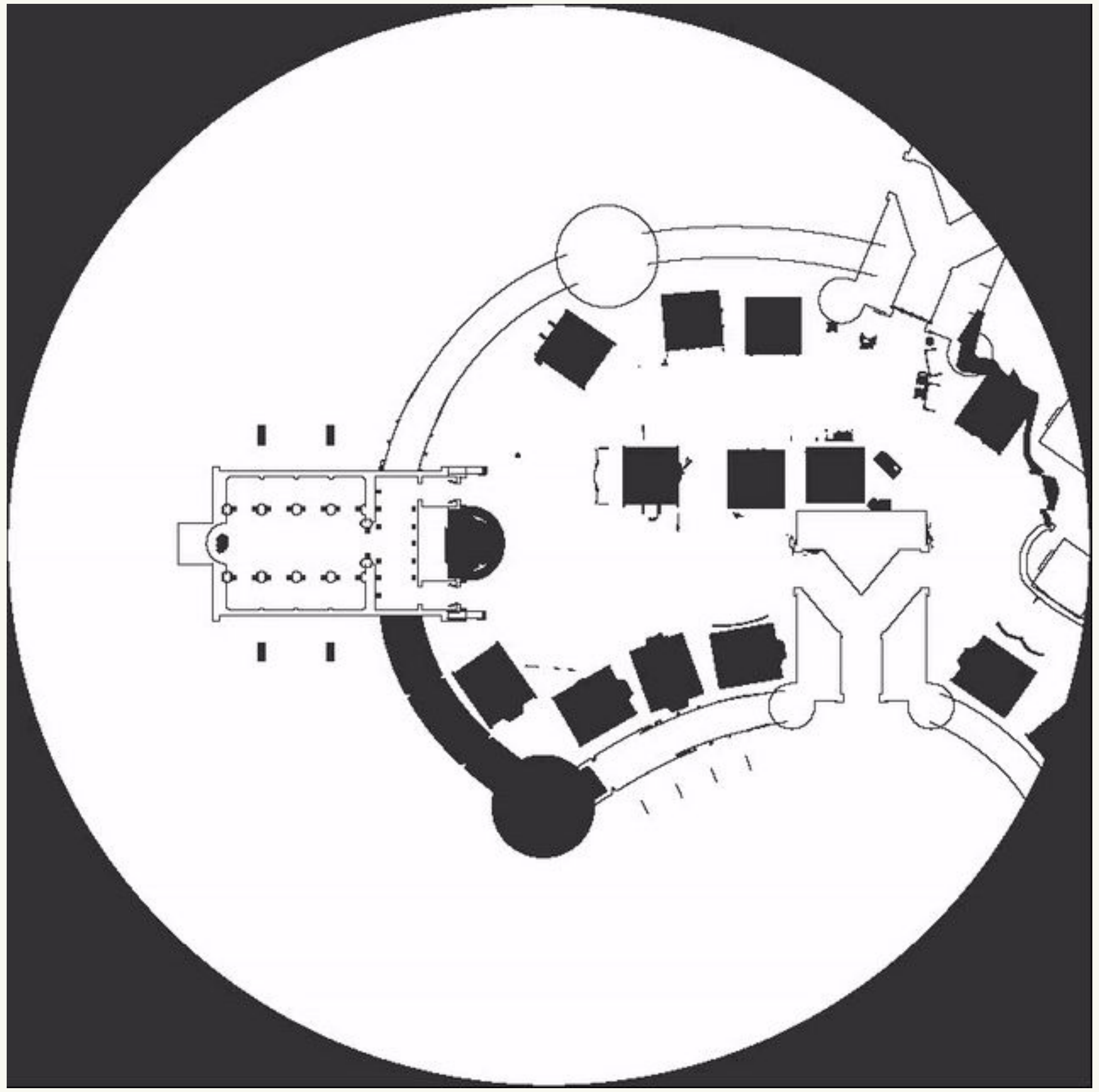




Technological challenges and opportunities

- **AR audio rendering adaptive to runtime physical environment**
- **Simulating acoustic propagation…**
- **Binaural rendering technology…**

docs.microsoft.com/en-us/gaming/acoustics/what-is-acoustics



Binaural rendering technology

- **Simplification of HRTF models** (equalization, low-frequency extension, near field effects...)
- **Binaural rendering from Ambisonics** (efficient implementation, 3-DoF head tracking)
- **Individual customization of head-related transfer functions**
- **Binaural multi-channel object-based audio** (preserving the ITD cue)
- **Binaural externalization processing** (web demo)

Recap: questions addressed in this talk…

- Opportunities enabled by spatial audio technology
  - in screen-based vs. world-based applications
  - in AR vs. VR applications.
- Level of naturalness and immersion enabled today by devices and engines
  - Some of the technological challenges pending to address…