

# ISO/IEC ISO/IEC 23090-4 MPEG-I Immersive Audio

Schuyler Quackenbush  
Audio Research Labs

***ARL***

---

audio research labs

# What is Audio for Immersive Experiences?

- 6 DoF is (x, y, z) and (yaw, pitch, roll)
  - User physical movement directs virtual experience
  - Within limitations of physical space
- This includes audio presentations for
  - Virtual Reality
  - Augmented Reality
- Presentation via HMD and headphones
- Position virtual sources in VR or AR world
- Render with
  - Localization, Occlusion, Reverberation, Directivity and Diffraction

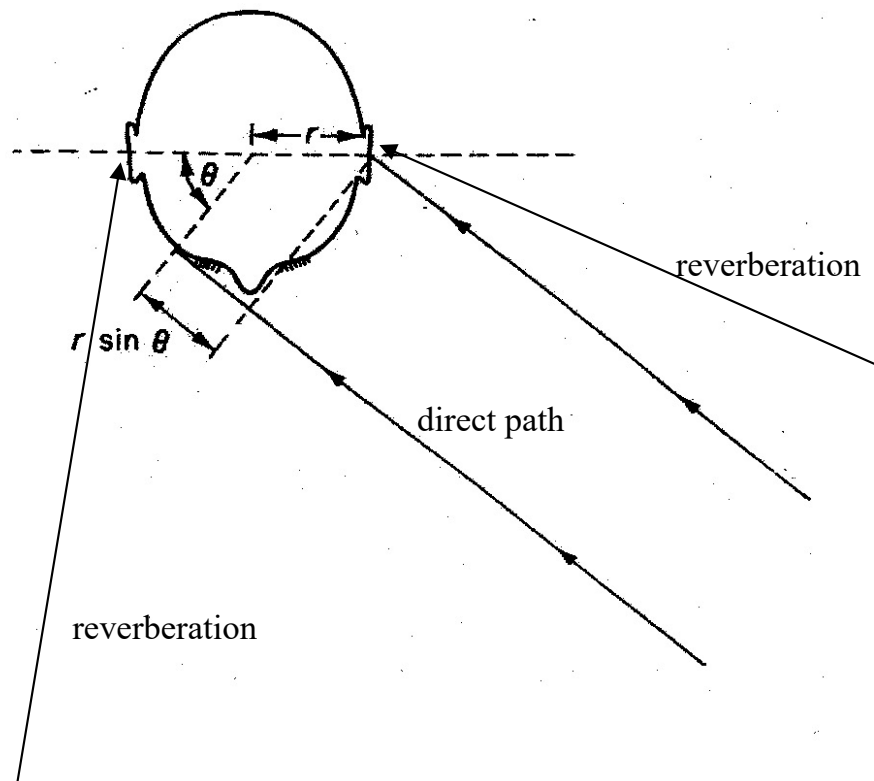
**ARL**

---

audio research labs

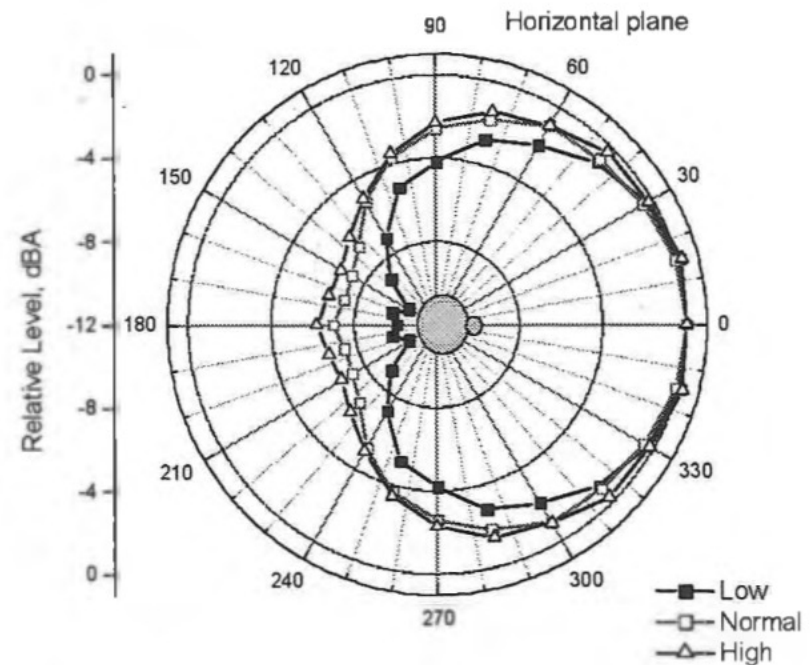
# Localization of Virtual Sources

- Use Head-Related Transfer Function (HRTF)
  - From virtual source to L and R ears
- Realistic rendering of spatial position due to perceptual cues
  - ITD – time of arrival differences
  - ILD – level difference
  - IC – coherence, due to reverberation differences



# Directivity of Sources

- Object perceived loudness changes as user moves around object
  - Louder in front
  - Softer in back



# Ambience and Reverberation

- Almost all spaces impose some reverberation on sound sources
  - Need to simulate this
    - Have model for virtual reality
    - Need to estimate model for augmented reality
- Also need to simulate occlusion and diffraction
  - Going “around the corner” from a sound object

# Technology in MPEG-I Audio

- We have audio compression engine
  - MPEG-H 3D Audio (23008-3:2019)
  - MPEG-I adds new meta-data in bitstream
- Audio rendering technology
  - Localization, Directivity
  - Could use parametric model of reverberation, other effects

# MPEG-I Encoder Input Format (EIF)

- Specifies format of all information needed by proponent to respond to MPEG-I Audio CfP
  - Audio signals (objects, channels, HOA)
  - Metadata for signals (position, orientation, directivity)
  - Room information (walls, acoustic reflectivity)
  - Animation (moving objects)
- Hierarchical scene description, like a scene graph
- Expressed in XML

# Example

## Audio Object in EIF

- Trumpet
  - Position (x, y, z)
  - Orientation (y, p, r)
  - Directivity
  - Gain
  - mode=“Continuous”
- Streaming sound

```
<AudioScene>
  <AudioStream id="signal:trumpet"
    file="armstrong.wav"
    mode="continuous" />
  <SourceDirectivity id="dir:trumpet"
    file="trumpet.sofa" />
  <ObjectSource id="src:trumpet"
    position="2 1.7 -1.25"
    orientation="30 -12 0"
    signal="signal:trumpet"
    directivity="dir:trumpet"
    gainDb="-2"
    active="true" />
</AudioScene>
```



# Room Acoustic Model

- Define a room
  - Geometric model from triangle mesh
  - Triangle faces have acoustic attribute label (“upholstery”)
- Define Frequency-dependent properties (“upholstery”)

```
<Mesh id="room_surface">  
  <Vertex index="0" position="1 2 3" />  
  <Vertex index="1" position="0.2 -1 0" />  
  <Vertex index="2" position="2 2 1" />  
  <Vertex index="3" position="1 0.5 -0.1" />  
  
  <Face vertices="0 1 2" material="upholstery" />  
  <Face vertices="0 2 3" />  
</Mesh>  
  
<AcousticMaterial id="upholstery">  
  <Frequency f="125.0" r="0.5" />  
  <Frequency f="250.0" r="0.42" />  
  <Frequency f="500.0" r="0.30" />  
  <Frequency f="1000.0" r="0.21" />  
  <Frequency f="2000.0" r="0.15" />  
  <Frequency f="4000.0" r="0.06" />  
</AcousticMaterial>
```

# Scene Updates

- Updates are *atomic*

```
<Update time="0.2">  
  <Modify id="engine" position="2.2 1.7 -1.25"  
  <Modify id="tire1" position="2.2 1.7 0.75" />  
  <Modify id="tire2" position="2.2 1.7 -0.95" />  
</Update>
```



```
<Update time="0.4">  
  <Modify id="engine" position="2.4 1.7 -1.20" />  
  <Modify id="tire1" position="2.4 1.7 0.70" />  
  <Modify id="tire2" position="2.4 1.7 -0.95" />  
</Update>
```

**ARL**

# User Interaction Updates

- User can open the door
  - Door handle has position
  - mode=“event”
  - Sound effect is local (cached)
  - Playout triggered by “update” message
    - From Unity to Max



```
<AudioStream id="sig:doorHandle1" file="doorHandle1.wav" />
```

```
<ObjectSource id="src:doorHandle" position="1.0 2.0 3.0" signal="sig:doorHandle1" mode="event" />
```

```
<Update id="upd:doorOpen">
```

```
  <Modify id="src:doorHandle" play="true" />
```

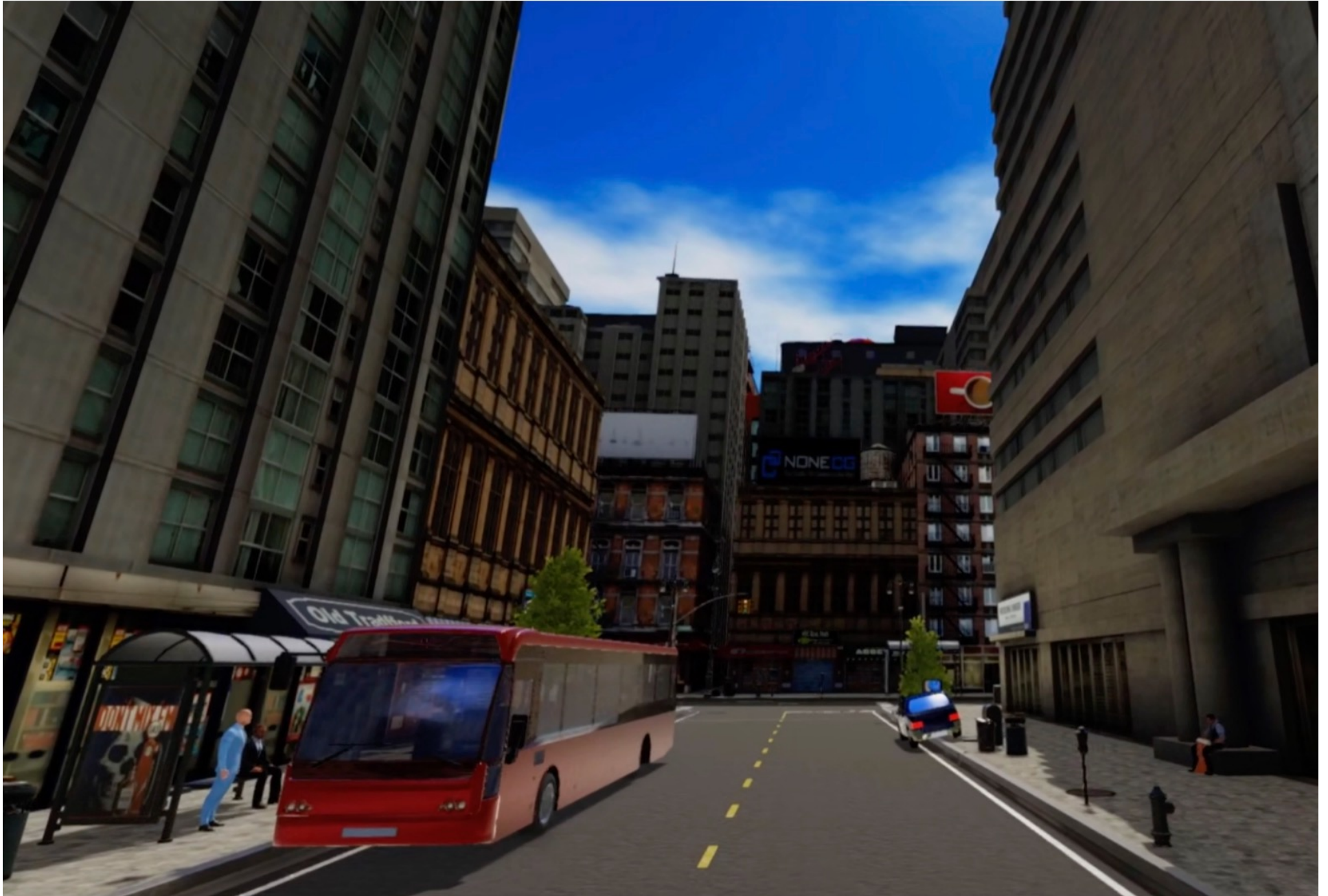
```
</Update>
```

**ARL**

# Evaluation of Technology

- Immersive VR world requires audio and visual presentations
  - Correctly perceiving virtual audio world without any visual cues is very difficult
- Hence, we will evaluate audio technology using a full, real-time audio-visual 6 DoF presentation (with full user body motion)
  - Head-Mounted Display for “Unity” visual presentation
  - Headphones and “Max 8” for audio presentation
  - Proponent technology runs in real-time in Max VST3 plugin

# Example from HMD



# Timeline for Standardization

Apr 2021	Call for Proposals
Jan 2022	Evaluation and Selection of Technology
Apr 2022	Working Draft
Jan 2023	CD
Apr 2023	DIS
Jul 2023	Verification Test complete
Oct 2023	FDIS

Thank You!

***ARL***

---

audio research labs