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

Schuyler Quackenbush

Audio Research Labs

Talk Overview

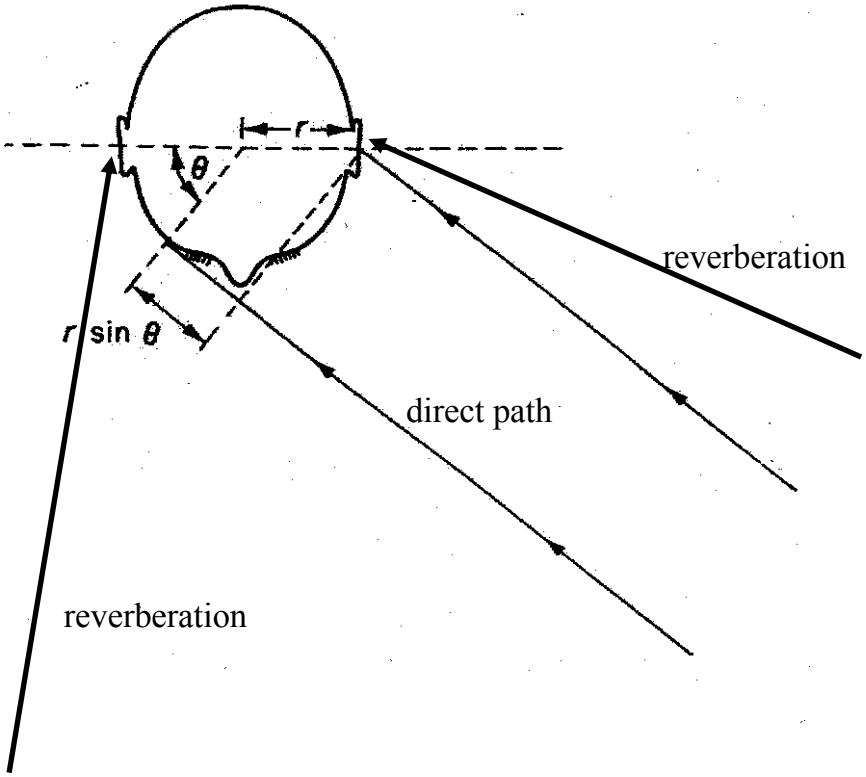
- What is 6DoF Audio?
- Architecture for 6DoF Audio standard
- Requirements
- Evaluation Platform for CfP
- Test Material
- Timeline for the work

What is 6DoF Audio for Immersive Experiences?

- The good news is that 6DoF Audio is relatively easy to do!
- 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, Directivity and Reverberation

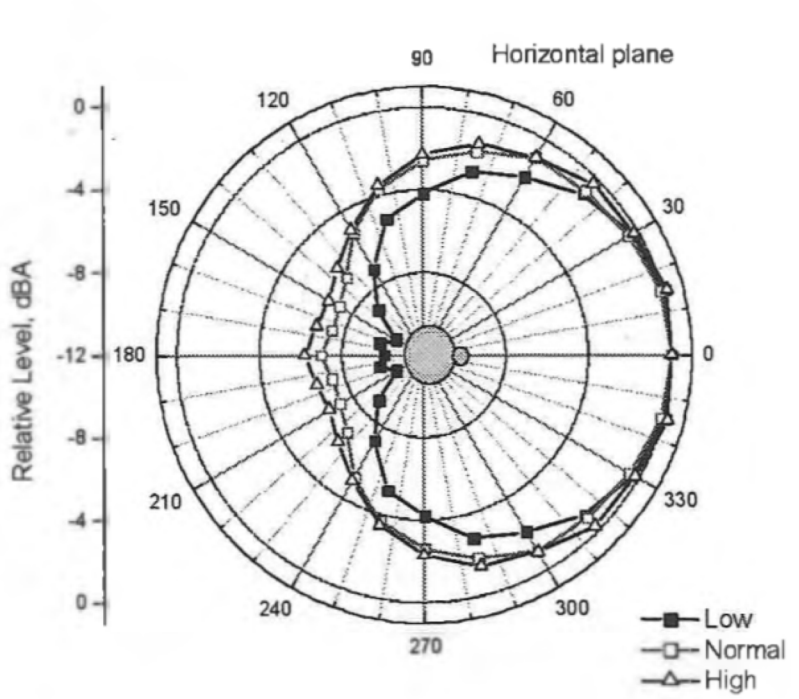
Position of Virtual Sources

- 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 moved 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
  - Going “around the corner” from a sound object

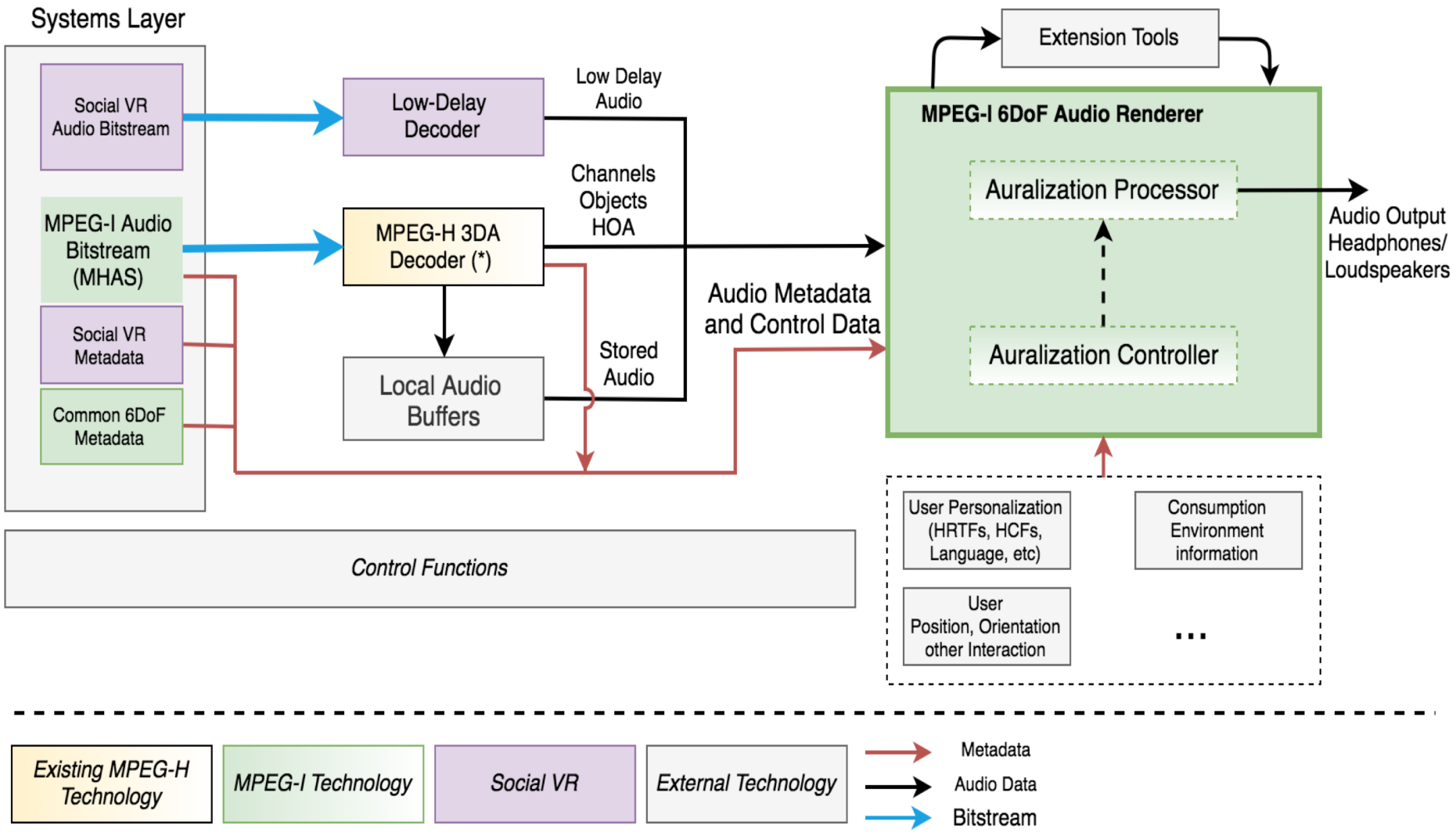
Architecture

- MPEG-I Immersive Audio includes
  - Coding of audio sources
  - Coding of meta-data (e.g. source directivity or room acoustic properties)
  - Rendering of audio presentation for headphones
- MPEG-H 3D Audio is already a 3DoF presentation technology
  - Will use for coding of audio sources

New Technology in MPEG-I Audio

- Bitstream format, to include both
  - MPEG-H 3D Audio
  - MPEG-I additional meta-data
- Audio rendering technology
  - Can be acoustic “ray-tracing”
  - Can be parametric model (e.g. RT-60, describes envelope of reverberation decay)

MPEG-I Audio Architecture



(*) **MPEG-H 3DA Decoder** as defined in this document.

MPEG-I Audio Requirements

- Categories of Requirements
  - General (e.g. audio quality, perceived realism)
  - Rendering
  - Interfaces (e.g. user 6DoF)
  - Extensibility (i.e. “future-proof”)
  - Presentation modes (headphones but can be loudspeakers)
  - Social VR (e.g. two users in one virtual world)

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 presentation
  - Head-Mounted Display for “Unity” visual presentation
  - Headphones and “Max 8” for audio presentation
  - Proponent technology runs in real-time in Max VST3 plugin

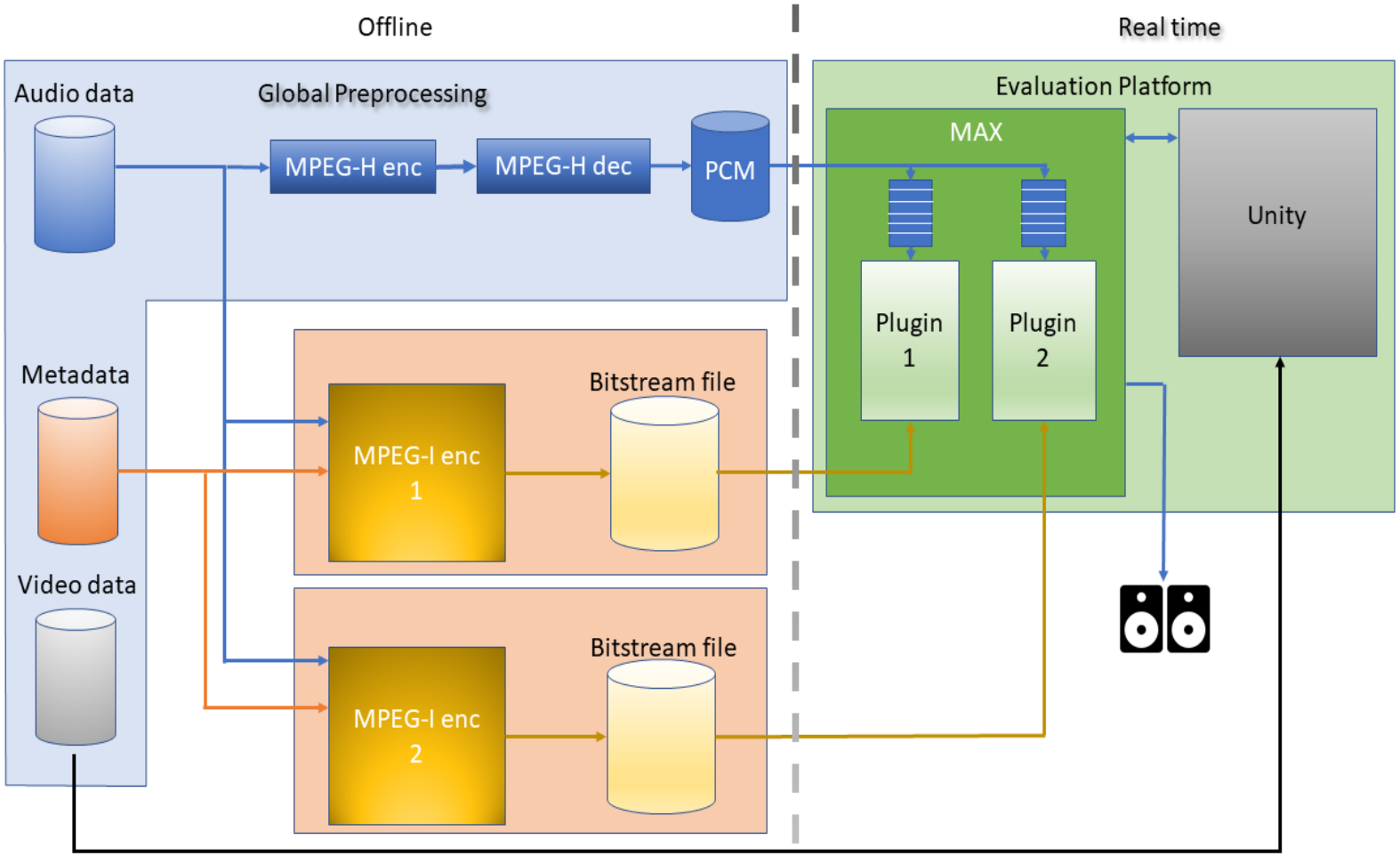
Example VR Presentation

- Outside scene with piano, fountain and birds
- [demo]

MPEG-I Evaluation Platform

- MaxMSP version 8 with VST3 plugins for proponent technology
- Automatically configured for each test
- Full randomization (since it is platform for subjective test)
- Signaling between Max and Unity to coordinate

MPEG-I Audio CfP Evaluation Platform

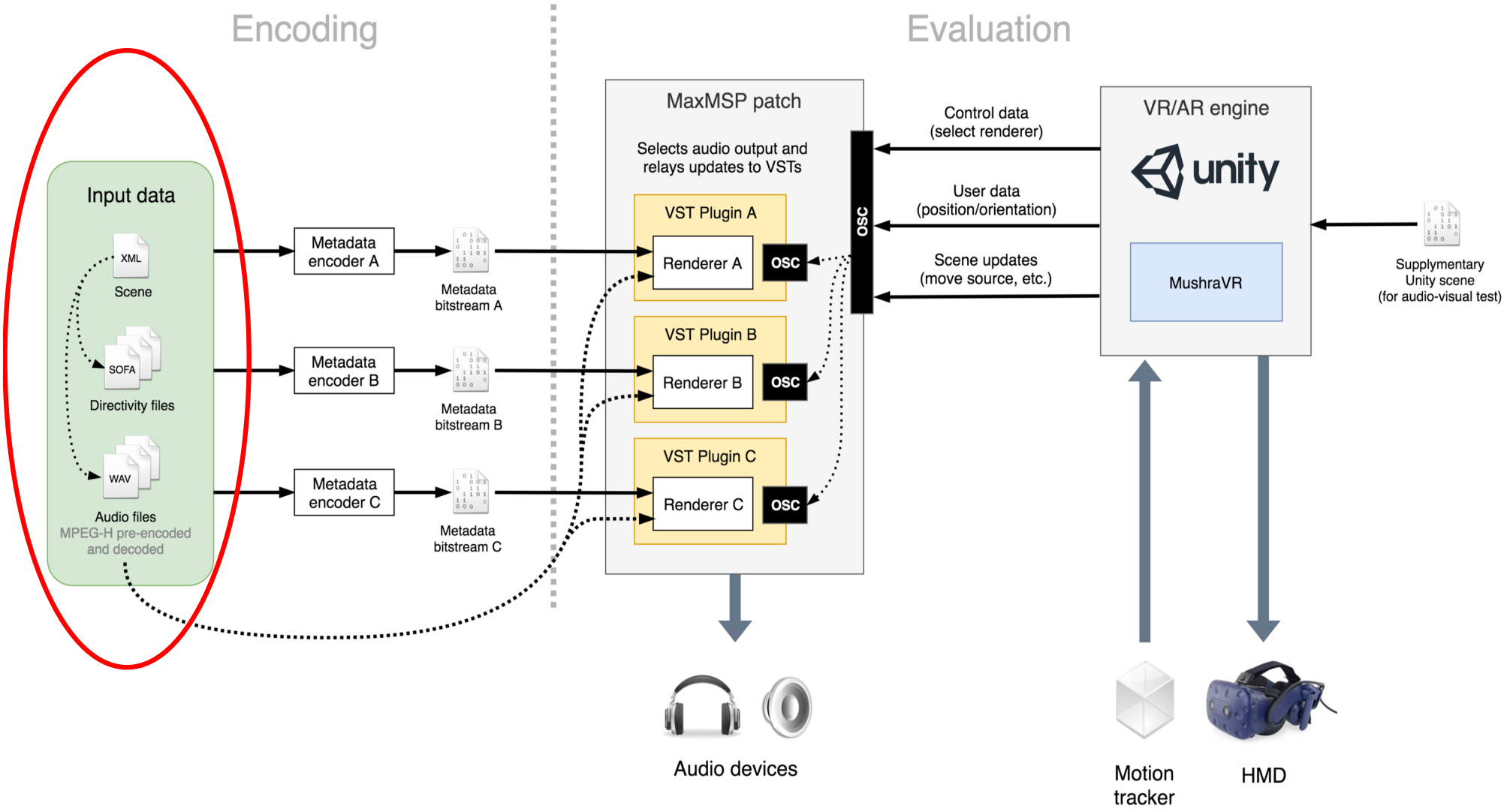


- Thanks to Philips for graphic

MPEG-I Encoder Input Format

- 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
- Expressed in XML

Use in Evaluation Platform



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>
```

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>
```

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>
```

MPEG-I Test Material

- Test Material expressed in Encoder Input Format
- Need richer test material that supports
  - Sound object localization
  - Sound object radiation patterns
  - Sound object extent or width
  - Occlusion of sounds
  - Reverberation of environment
  - Transition through “scene gateways” such as doorways

Timeline for Standardization

Jan 2020	Call for Proposals
Jun 2020	Evaluation and Selection of Technology
Oct 2020	Working Draft
Apr 2021	CD
Jul 2021	DIS
Oct 2021	Verification Test complete
Jan 2022	FDIS

Questions?