|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligencewww.mpai.community |

**Public Document**

|  |  |
| --- | --- |
| **N1661** | 2024/03/20 |
| **Source** | Marina Bosi, Chair MPAI-CAE DC |
| **Title** | Use Cases and Functional Requirements - Six DoF Audio (CAE-6DA) WD0.1 |
| **Target** | MPAI Members |

# Introduction

MPAI has developed the specification of Audio Scene Descriptors enabling the description of the Geometry of an Audio Scene and the identification of the Audio Objects in it (MPAI-CAE V2.1).

This document is a working draft collecting Use Cases and Functional Requirements for the digital representation of a sound scene, i.e., the Audio Scene, that enables the perceptually veridical reconstruction of an acoustical environment comprising real, synthetic or a mixture of sound sources, for an end user at an arbitrary user-selected point in the Audio Scene.

# Use Cases

Use Case 1 – Immersive Concert Experience (Music plus Video)

User is sitting at home listening to a concert rendered in the space of his/her living room. User can go to a point in his/her living room area and have the audio experience of a human in the equivalent position in the concert hall.

Use Case 2 – Immersive Radio Drama (Speech plus Foley/Effects)

User is sitting at home listening to an immersive radio drama rendered in a space of his/her living room. User can go to a position in his/her living room to get closer to voice actors/actresses in the drama, experiencing the same play from a different auditory perspective.

Use Case 3 – Virtual lecture (Audio plus Video)

User is sitting at home and using an immersive audiovisual display while

1. watching 360° videos captured at different points in a lecture theatre
2. Listening to a lecturer teaching

User can go to a position in his/her living room to experience the same lecture from another fixed perspective that corresponds to the position where the 360° video is captured.

Use Case 4 – Immersive Opera/Ballet/Dance/Theatre experience (Music, Drama plus Video/Volumetric)

User is sitting at home and using an immersive audiovisual display while:

1. watching 360° videos, a volumetric dynamic model of a captured scene (for example, obtained using dynamic NeRF or dynamic Gaussian splatting), or a combination of both, also including singers/actors within and,
2. Listening to immersive opera or theatre recorded during an actual performance, in the space of his/her living room.

User can move within his/her living room to experience the same scene from another arbitrary perspective, for example at the orchestra pit, sitting in a seat at the stalls, or among the singers/actors.

# Functional Requirements

1. The Functional Requirements apply to the Audio experience and to the impact of visual conditions on the Audio experience. For instance:
	1. Audio-Visual Contract, alignment of audio scenes with visual scenes.
	2. Effects of locomotion on a human audio-visual perception.
	3. Orientation response turning toward a sound source of interest.
	4. Distance perception: visual and auditory modalities affect each other.
2. The rendering space need not have the same characteristics as the captured space:
	1. Shape and dimensions:
		1. Not larger than the captured space.
	2. Acoustic ambient characteristics:
		1. Early decay time (T10).
		2. Frequency mode density.
		3. Echo density.
		4. Clarity index.
		5. Reverberation time (T60).
		6. Energy decay curve characteristics.
		7. Critical distance.
		8. Background noise.
3. The User movement in the rendering space may be the result of actual or virtual locomotion or orientation.
	1. Actual locomotion/orientation of the User as tracked by sensors.
	2. Virtual locomotion/orientation is actuated by controlling devices.
4. The spatial extent of the rendered content shall not exceed the boundaries of the rendering space.

# Definitions

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Boundary | A surface that bounds the volume of an enclosure |
| Captured space | The space where content is generated. |
| Rendering space | The physical space where content is consumed. |
| Virtual space | The digital representation of the captured space. |