 ISO/IEC JTC 1/SC 29/WG 6 N0055

**ISO/IEC JTC 1/SC 29/WG 6**

**MPEG Audio Coding Convenorship: US**

**Document type:** General

**Title:** MPEG-I Immersive Audio Augmented Reality Listener Space Description Format

**Status:** Approved

**Date of document:** 2021-04-30

**Source:** ISO/IEC JTC 1/SC 29/WG 6

**Expected action:** None

**Action due date:** None

**No. of pages: 4**

**Email of Convenor:** srq @ audioresearchlabs . com

**Committee URL:** https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg6

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG6, MPEG AUDIO CODING**

**ISO/IEC JTC1/SC29/WG6 N0055**

**April 2021, Virtual**

|  |  |
| --- | --- |
| **Title** | **MPEG-I Immersive Audio Augmented Reality Listener Space Description Format** |
| **Source** | **WG 6, MPEG Audio Coding** |
| **Status** | **Approved** |
| **Serial Number** | **20447** |

# Introduction

This document specifies the MPEG-I 6DoF Listener Space Description Format (LSDF). Its purpose is to describe the listening space for MPEG-I 6DoF audio AR evaluation. In AR, where content is augmented on top of real-world objects and spaces, knowledge of the geometry of the listening space is important for realistic evaluation. LSDF provides a mechanism to provide the listening space environment information directly to the renderer for MPEG-I 6DoF audio AR evaluation.

Figure 1, below, describes the overall concept of the LSDF. The scene information contained in the EIF [1] is available to the render as bitstream. However, for AR evaluation, information about the listening space is provided directly to the renderers.

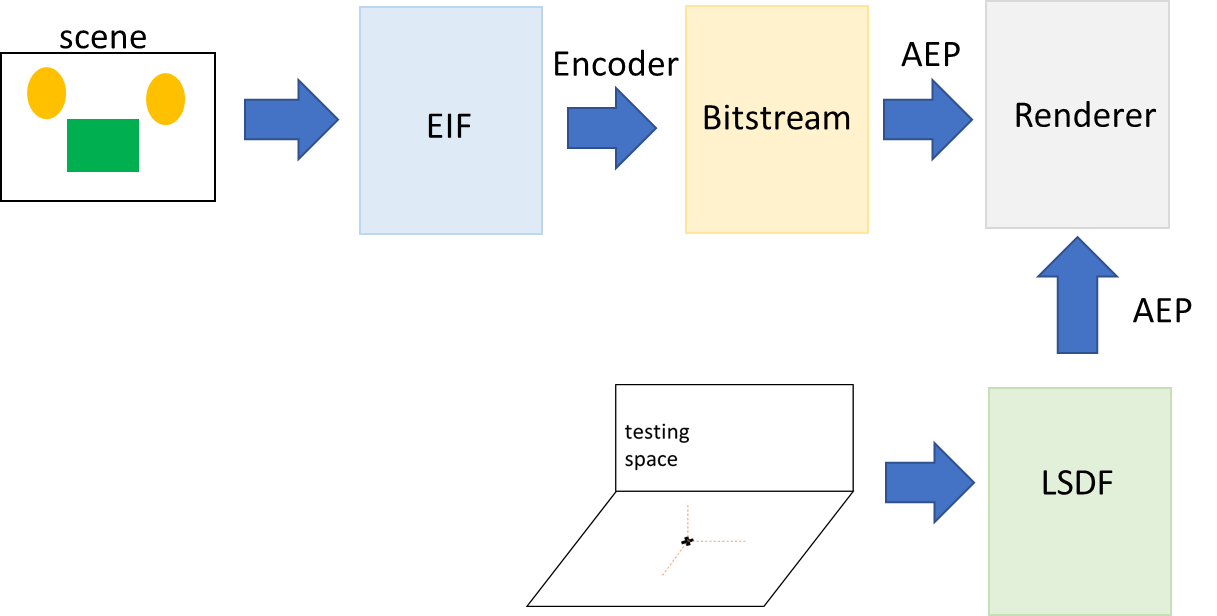


Figure 1. The LSDF is used to provide the information about the listening space to the renderers.

The LSDF includes a subset of elements of the MPEG-I 6DoF Audio Encoder Input Format [1] with restrictions. The elements are used to describe the walls, ceiling and floor of the listening space. Furthermore, the LSDF describes anchors for aligning elements in the scene EIF to their positions in the listening space. The elements used in the LSDF are listed in Section 2.

One LSDF file is needed for each of the sites performing AR evaluation. Once created, the file is placed in a specific folder inside the MPEG-I audio evaluation platform [2]. The AEP provides the path to the Max External plug-ins containing the proponent renderers. See the AEP documentation for more information [2].

# LSDF elements

Table 1 summarizes the elements that are used in the LSDF for describing the listening space. The elements are explained in more detail below. Since the LSDF does not go through a encoder, but rather directly to the renderers (via AEP), some restrictions have been added to the <Mesh> element to keep the listening space geometry simple.

|  |  |
| --- | --- |
| **Element** | **Purpose** |
| <AudioScene> | Listening space audio scene |
| <Mesh>, <Face>, <Vertex>, <AcousticMaterial> | Describing walls, floor and ceiling |
| <AcousticEnvironment>, <AcousticParameters>, <Frequency> | RT60, ddr |
| <ARAnchor> | Aligning EIF elements to real-world objects |

Table 1. Summary of elements in LSDF

## Audio scene

The <AudioScene> element declares the AR listening space. It is similar to the <AudioScene> described in the EIF [1], but with changes to the list of allowed child node entities and their counts. A single <Mesh> and a single <AcousticEnvironment> are allowed. A new child node, <ARAnchor>, has been added.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <AudioScene> | | | | |
| Declares an audio scene. Each scene description XML file must have exactly one AudioScene node. This node marks the root node of the static description of the audio scene. It contains all entities of the scene as children. | | | | |
| Child node | Count | Description | | |
| <Mesh> | 1 | Mesh (see below) | | |
| <AcousticMaterial> | >=0 | Acoustic material (see below) | | |
| <AcousticEnvironment> | 1 | Acoustic environment (see below) | | |
| <ARAnchor> | >=0 | AR anchor (see below) | | |
| Attribute | Type | Flags | Default | Description |
| id | ID | R |  | Identifier |
|  |  |  |  |  |

## Mesh

The <Mesh> element is used to describe the walls, ceiling and floor of the listening space.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <Mesh> | | | | |
| Declares a triangle mesh. A mesh consists of a list of vertices (3D coordinates) and a number of triangular faces (i.e. the indices of three vertices). Meshes can be used to describe arbitrary geometry. Any <Mesh> node has to have one or more <Vertex> or <Face> child nodes, defining points and triangles. The mesh is required to be a manifold mesh. | | | | |
| Child node | Count | Description | | |
| <Vertex> | >=1, <= 36 | Vertex (see EIF [1]) | | |
| <Face> | >=1, <= 48 | Face (see EIF [1]) | | |
| Attribute | Type | Flags | Default | Description |
| id | ID | R |  | Identifier |
| position | Position | O | (0, 0, 0) | Position (origin of the mesh) |

Vertex and Face are as specified in the EIF [1].

A manifold mesh can be checked for by checking that all vertices are manifold by verifying [2]:

* Every edge is shared by exactly two faces
* Every vertex has a single, complete loop of triangles around it

**Acoustic Environment**

<AcousticEnvironment>, <AcousticParameters> and <Frequency> are the same as specified in the EIF [1].

**Anchors**

An <ARAnchor> is used to indicate a real-world position for the EIF to reference.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <ARAnchor> | | | | |
| Indicates the position and orientation of a real-world anchor. <AnchorObject> in EIF may refer to this object for correct placement of scene elements. | | | | |
| Attribute | Type | Flags | Default | Description |
| id | ID | R |  | Identifier |
| position | Position | R |  | Position of the anchor object w.r.t to origin |
| orientation | Orientation | R |  | Orientation of the anchor object w.r.t to origin |

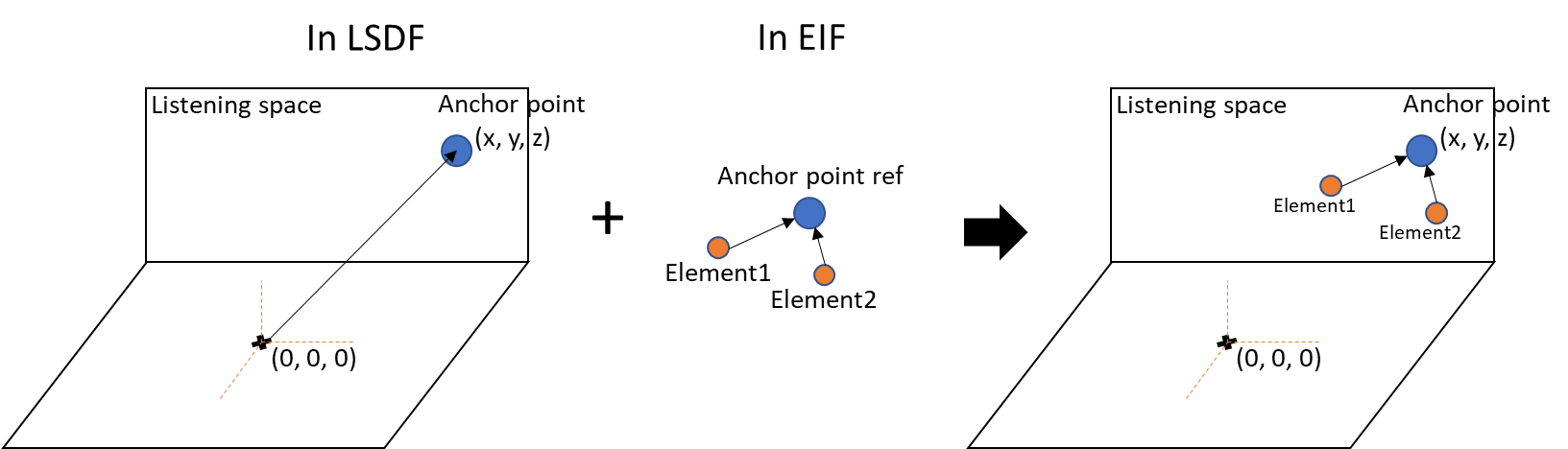


Figure 2. Anchor mechanism. <ARAnchor>s in the LSDF describe points in the listening space which may be referred to in the EIF using <AnchorObjects> ([1]).

# LSDF creation

Each test site prepares an LSDF file for their own listening space. Creation of the LSDF file starts by defining an origin in the listening space, usually a spot on the floor close to the center of the listening space. The room geometry is then measured w.r.t the origin and this information is inserted to the single mesh in the LSDF file. All anchors required by the different AR scenes to be tested at a given site are then defined in the LSDF file, again w.r.t the chosen origin. Description of the required anchors are to be provided by the scene authors along with the scenes. Further information (RT60, material coefficients, etc.) of the listening space may be added using the element described in Section 2.

# References

[1] N0054, ‘MPEG-I Immersive Audio Encoder Input Format’

[2] ‘MPEG-I Immersive Audio Documentation for the Audio Evaluation Platform’