**Source:** InterDigital Inc.

**Title: MPEG Scene Description for MeCAR**

**Document for** Approval

**Agenda item:** 9.5 - MeCAR (Media Capabilities for Augmented Reality)

# Introduction

The MeCar permanent document in section 3.7.2 includes a brief mention of Scene Description, and references section 4.4.2 of 3GPP TR 26.998[1] which provide an overview of MPEG Scene Description.

This contribution describes in more details the features of MPEG Scene Description, which are essential for realizing MeCar use-cases such as:

* AR gaming, e.g. as defined by use cases 21 in TR 26.998
* AR conferencing, e.g., as defined by use case 22 in TR 26.998
* Spatialization of streamed media, e.g. as defined by use case 18 in TR 26.998

The AR gaming demonstration presented at this meeting (described in S4-230206 and shown in the attached video) highlights key MPEG Scene Description functionalities. The demonstration corresponds to the use case 21 called AR Gaming defined in TS26.998 [1], in its annex A.6.



# MPEG Scene Description

NOTE: As part of the working assumptions MPEG-SD functionalities will be evaluated, as a candidate solution, against the device capabilites and requirements of MeCAR.

## overview

As described in section 4.6.5 of TR 26.998 [1], a key technology in enabling immersive 3D user experiences is scene description. Scene description is used to describe the composition of a 3D scene, referencing and positioning the different 2D and 3D assets in the scene. The information provided in the scene description is then used by an application to render the 3D scene properly, using techniques such as Physically-Based Rendering (PBR) that produce realistic views

To address the needs of immersive applications, MPEG Scene Description (SD) specified a collection of extensions to glTF to support scene description. glTF 2.0 [3] provides a solid and efficient baseline for exchangeable and interoperable scene descriptions.

A white paper providing an overview of MPEG Scene Description is available [9].

Khronos glTF 2.0 specification was released in 2017. Khronos allows to extend the glTF specifications through an extension mechanism [8]. The process of registering MPEG extensions as Khronos Vendor extensions is now in its final stage [14].

MPEG SD defines the following extensions:

* the First Edition of MPEG-I Scene Description (ISO/IEC 23090-part14) is a set of extensions providing support for audio, timed media, and scene update.
* Amendment 1 [4] provides support for immersive media (MIV and V-PCC, V3C standards [10][11][12]).
* Amendment 2 [5] provides support for immersive audio ([13]), anchoring, interactivity, lighting, avatar, haptics.

In the MPEG SD Technologies under Consideration (TuC) document [6], support for advanced interactivity and for real environment are under study for an Amendment 3.

## MPEG glTF extensions

### The anchoring extension

The anchoring extension enables the sharing of a common spatial reference among multiple users by specifying trackables and anchors.

Trackable is an element of the real world which features can be extracted. This element can then be detected and tracked over time.

An anchor is a virtual object defining a local reference space for the virtual scene.

The anchor is attached to this trackable. The anchor position is then relative to the trackable it is attached to.

The MPEG Scene Description anchoring extension is defined in Amd2 [5] clause 8.1. It includes all the necessary syntax to define the trackable, the anchor, the relative position of the anchor in relation to the trackable.

### The interactivity extension

MPEG-I Scene Description Amd2 [5] defines an interactivity framework for the description of allowed runtime interactions between a user with the virtual scene objects and between the virtual objects themselves. The framework relies on the definition of behavior, composed of a logical combination of generic triggers and of the related actions to be launched sequentially or in parallel.

The trigger types are :

* Visibility trigger, allowing to detect if a 3D object in a virtual scene is visible from the user viewpoint,
* Proximity trigger, allowing to detect the proximity of the user to a 3D object of a virtual scene,
* Collision trigger, allowing to determine a collision between 3D objects,
* User-input trigger, allowing to detect a specific user input.

The physical parameters to animate the 3D objects according to physical laws are provided through interactivity extension.

The interactivity extension is defined in Amd2 [5], clause 8.2.

### Additional extensions

Several other MPEG SD extensions increase the quality of the AR user experience, notably:

* Avatar extension: Annex H of Amd2 [5] provides a description of a reference topology for the representation of humanoid avatars.
* Lighting extension (clause 8.4 of Amd2 [5]): this extension enables a support of timed lighting.
* Haptic extension (Annex G3 of Amd2): this extension defines a support for haptic signals. An example may be to add vibrations to scene objects.
* Support for Volumetric Video (V3C) codecs (Annex G1 of Amd1 [4]): this extension allows to support V3C compressed objects in MPEG-I Scene Description.

### MPEG SD extensions in the TuC:

Extensions described in the TuC document [6] include, but are not limited to:

* Advanced interactivity: this extension addresses the need to support scene updates generated at runtime from the activation of triggers. The advanced interactivity extension is typically required in the case of a common virtual scene, shared between multiple users, each of them interacting and modifying the scene at runtime.

* Real environment: this extension aims at providing scene understanding by supporting a real-world representation in scene description to achieve a seamless integration of the virtual into the real world (an example may be the interaction of virtual objects falling on a real surface).

# Impact on MeCAR capabilities

MeCar Permanent Document [7] in its introduction chapter mentions, among its objectives: “*Define media types and formats produced and consumed by the AR device, including basic scene descriptions,..*”.

The discussion in the previous chapter shows that basic scene descriptions can be achieved by MPEG Scene Description including First edition, Amendment 1, and Amendment 2.

# Proposition

It is proposed to include section 2 as a subclause of chapter 3 (“Working Assumption”) of the MeCar permanent document.

# References

1. TR 26.998 “Support of 5G Glass-type Augmented Reality / Mixed Reality (AR/MR) devices”.
2. MPEG-I Coded representation of immersive media – Part14: Scene Description for MPEG media, <https://dms.mpeg.expert/doc_end_user/current_document.php?id=80994&id_meeting=188>; https://www.iso.org/standard/80900.html
3. glTF 2.0 specification <https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html>
4. ISO/IEC 23090-14 CD Amd1 Support for immersive media codecs in Scene Description <https://dms.mpeg.expert/doc_end_user/current_document.php?id=85103&id_meeting=192>; https://www.iso.org/standard/84769.html
5. ISO/IEC 23090-14 CD Amd2: Support for Haptics, Augmented Reality, Avatars, Interactivity and Lighting <https://dms.mpeg.expert/doc_end_user/current_document.php?id=86321&id_meeting=193>; <https://www.iso.org/standard/86439.html>
6. Technologies under Consideration (TuC) in Scene Description <https://sd.iso.org/documents/ui/#!/doc/1f2ddef0-cfc3-4731-b083-53f61e28f8f3>
7. MeCar Permanent Document <https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_121_Toulouse/Docs/S4-221567.zip>
8. Khronos glTF extensions for glTF 2.0 <https://github.com/KhronosGroup/glTF/blob/main/extensions/README.md>
9. MPEG-I Scene Description, White Paper <https://www.mpeg.org/wp-content/uploads/mpeg_meetings/140_Mainz/w22138.zip>
10. ISO/IEC 23090-5 Information Technology — Coded Representation of Immersive media — Part 5: Visual Volumetric Video-Based Coding (V3C) and Video-Based point Cloud Compression (V-PCC).
11. ISO/IEC 23090-10 Information Technology — Coded Representation of Immersive media — Part 10: Carriage of Visual Volumetric Video-Based Coding Data.
12. ISO/IEC 23090-12 Information Technology — Coded Representation of Immersive media — Part 12: MPEG Immersive Video.
13. ISO/IEC 23090-4 Information Technology — Coded Representation of Immersive media — Part 4 MPEG-I Immersive audio
14. Registration of MPEG Extensions for first edition of ISO/IEC 23090-14 by haudiobe · Pull Request #2203 · KhronosGroup/glTF (github.com)