**3GPP TSG-S4 Meeting #116-eS4-211607**

**Online, 10th – 19th November 2021**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.1* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | 26.962 | **CR** | - | **rev** |  | **Current version:** |  |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME |  | Radio Access Network |  | Core Network |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | Guidelines on Overlays using MPEG-I Scene Description | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Qualcomm | | | | | | | | | |
| ***Source to TSG:*** | S4 | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | ITT4RT | | | | |  | ***Date:*** | | |  |
|  |  | | | |  | |  | | |  |
| ***Category:*** |  |  | | | | | ***Release:*** | | |  |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-15 (Release 15) Rel-16 (Release 16) Rel-17 (Release 17) Rel-18 (Release 18)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | This pCR provides guidelines on the usage of scene description for overlay signaling. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  |  | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  |  | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  |  | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

|  |
| --- |
| First Change |

## Scene Description-based Overlay

### Scene Description

#### Overview

A scene graph is a directed acyclic graph, usually just a plain tree structure, that represents an object-based hierarchy of the geometry of a scene. The leaf nodes of the graph represent geometric primitives such as polygons. Each node in the graph holds pointers to its children. The child nodes can among others be a group of other nodes, a geometry element, a transformation matrix, etc.

Spatial transformations are attached to nodes of the graph and represented by a transformation matrix.

#### glTF 2.0

glTF 2.0 is a new standard that was developed by Khronos to enable Physically Based Rendering. glTF 2.0 offers a compact and low-level representation of a scene graph. glTF 2.0 offers a flat hierarchy of the scene graph representation to simplify the processing. glTF 2.0 scene graphs are represented in JSON to ease the integration in web environments. The glTF 2.0 specification [ref] is designed to eliminate redundancy in the representation and to offer efficient indexing of the different objects in the scene graph.

The structure of a glTF 2.0 scene graph document is arranged as follows:



### MPEG-I Scene Description

MPEG has developed an architecture [MPEG-I part 1] for immersive media, where a Presentation Engine plays a central role. The Presentation Engine is responsible for receiving the scene description and its updates, parsing it, and rendering the scene. In order to access the media required to render the scene, the Presentation Engine invokes functionality provided by the Media Access Function API.

The MPEG-I scene description [MPEG-I SD] architecture is depicted by the following figure:



The MPEG solution also specifies a key set of extensions to glTF 2.0 to support timed media such as dynamic objects, video textures, and audio. The MPEG\_media, MPEG\_accessor\_timed, and MPEG\_buffer\_circular extensions make up the core of these extensions and enable integrating timed media in the glTF described scene. The MPEG\_texture\_video allows the integration of video textures that could for example be the sources of an overlay in an ITT4RT conference.

The node structure of a MPEG-I scene description is depicted by the following diagram:

![Graphical user interface

Description automatically generated]()

The MPEG-I scene description design allows for decoupling media access, which would be provided by the Media Access Function, from the scene rendering, which is provided by the Presentation Engine.

### Scene Description for ITT4RT Sessions

The composition of scene description may be performed at an ITT4RT MRF that supports scene description-based overlays. There is at mone one scene description for an ITT4RT session. The participants may contribute nodes to that scene through sending scene updates after a successful negotiation of the scene description-based overlay configuration. In case the ITT4RT session has only 2 participants, the scene is created by the offerer.

An ITT4RT session is considered to be based on one or more scenes. Each scene contains a 360 degree video and some overlays. The ITT4RT MRF is responsible for positioning the overlays in the scene through proper node transformatiosn.

The scene description describes the whole scene, including all audible/visible participants and the main conference room. Each scene node should contain at most one node with 360 degree content. Switching between different 360 conference rooms is then equivalent to switching between scenes in the scene description.

A visual node in the scene graph may describe the 360 degree content or overlay content. The node provides the geometry and the associated texture, which references a 360 or an overlay image/video source.

The 360 degree content is described through a sphere or cube-map geometry with associated video texture coming from the 360 video. The type of geometry depends on the selected projection for the ITT4RT session. The overlay nodes are typically rectangular plane regions with an associated video texture coming from the associated overlay video stream.

The remote participants are required to indicate if they support receiving scene description by accepting an SDP offer that contains the data channel, which indicates “mpeg-sd” as the sub-protocol. A participant that does not support receiving scene description will receive the overlay descriptions in the SDP and may declare its overlay streams using the SDP 3gpp-overlay attribute.

In a scene, node names are expected to be unique to ensure there are no naming conflicts in nodes provided by different parties in a call. Nodes in the scene description may reference external media streams, such as other media streams that are declared in the SDP. A participant may selectively make one or more nodes of the received scene description from certain parties invisible in the rendering process, e.g. based on user input.

### Referencing Media Streams

In order to reference the source video streams for the 360 content and the overlays in a scene description document, the URL format as specified in 23090-14 Annex B is to be used.

The ABNF syntax for the URI format shall be as follows:

url="rtp://" fqdn\_or\_ip [":" port] "/" session\_id "/" ssrc "/" mid

The ssrc corresponds to the SSRC identifier [RFC 3550] of the source. The mid corresponds to the mid attribute of the referenced media stream in the SDP. Only the SSRC and the mid are provided, all other fields are set to a a generic matching pattern “\*”.

If the SSRC of the source is not known at the time of the scene description creation, a random identifier is used instead.

Editor’s Note: a liaison is sent to MPEG to clarify this referencing scheme and to consider addressing its shortcomings.

### Example Scene Description with Overlays

An example for a scene description with a sphere projection 360 video and two overlay streams is TBD.