**3GPP TSG-SA WG4 Meeting #132 S4-250975r1**

**Fukuoka, Japan, 19th – 23rd May 2025** *revision of S4aR250099*

|  |
| --- |
| *CR-Form-v12.2* |
| **CHANGE REQUEST** |
|  |
|  |  | **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 | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | [AvCall-MED] Update on Avatar Formats |
|  |  |
| ***Source to WG:*** | , Qualcomm Inc. |
| ***Source to TSG:*** |  |
|  |  |
| ***Work item code:*** |  |  | ***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 canbe 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-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** | AR Calls have been enhanced with Avatar communication capaibilities. The stage 3 aspects related to formats and signaling are missing. |
|  |  |
| ***Summary of change:*** | This CR adds the formats, protocols, and signaling to add support for Avatar communication to AR calls. |
|  |  |
| ***Consequences if not approved:*** | Rel-19 will be missing stage 3 support for Avatar communication. |
|  |  |
| ***Clauses affected:*** |  |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

First change

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[3] 3GPP TS 26.119: "Media Capabilities for Augmented Reality".

[4] 3GPP TS 23.228: "IP Multimedia Subsystem (IMS); Stage 2".

[5] 3GPP TS 24.229: "IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3".

[6] 3GPP TS 26.565: "Split Rendering Media Service Enabler".

[7] ISO/IEC 23090-14 AMD 2, Information technology — Coded representation of immersive media — Part 14: Scene description — Amendment 2: Support for haptics, augmented reality, avatars, Interactivity, MPEG-I audio, and lighting

[8] 3GPP TS 26.522: "5G Real-time Media Transport Protocol Configurations".

[9] ISO/IEC 23090-39, Information technology — Coded representation of immersive media — Part 39: Avatar Representation Format

Next change

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

**AR data:** Collection of information to be exchanged among participants in a call with AR experience. It includes AR media and AR metadata.

**AR media:** Media (e.g., audio, video, text or image) that will be rendered by the AR-MTSI client as an overlay over the user’s real perception. This includes traditional 2D media (e.g., a 2D audio stream rendered to be perceived by the user to originate from their left side) and 3D media (e.g., spatial audio and volumetric video).

**AR metadata:** Data that provides information on AR media and its rendering. This includes pose, spatial descriptions and scene descriptions.

**AR-MTSI client:** A DCMTSI client supporting AR capabilities as defined by this specification.

**AR MRF:** An AR-MTSI client implemented by functionality included in the MRFC and the MRFP.

**AR MF:** An AR-MTSI client implemented by functionality included in the MF.

**AR-MTSI client in terminal:** An AR-MTSI client that is implemented in a terminal or UE. The term "AR-MTSI client in terminal" is used in this document when entities such as AR MF/MRF is excluded.

**Split rendering**: The procedure in which a UE offloads some of the media processing related to rendering tasks to a media function as considered for network centric AR IMS session procedures in TS 23.228 [4].

**Asset:** An independently accessible component of an avatar.

**Avatar:** A digital representation of a user.

**Animation data:** Skeletal, blend shape set, and other animation-related information.

**Animation stream:** Timed animation data sequence used to animate the base avatar.

**Base avatar model:** Personalized and animatable model of the user.

Next change

## 5.6 Avatars

### 5.6.1 General

An AR-MTSI Tx client in terminal offering an avatar in an AR Call shall support the Avatar Representation Format (ARF) as specified in [9] and shall have at least one base avatar stored in the Base Avatar Repository (BAR) in one of the two supported container formats defined in [9].

An AR-MTSI Rx client that supports 3D avatars shall support both the ARF ISOBMFF and Zip container formats.

Editor’s Note: The ARF specification is still under development. Some details may change.

The base avatar shall comply with the ARF specification [9]. In addition, the ARF document shall include the following information:

- A list of the supported animations, which includes at least one animation type (e.g., face or landmark animation),

- At least one asset with at least one level of detail, and

- All data of relevant assets shall be contained in the ARF container of the base avatar.

### 5.6.2 3D Avatar Format

An AR-MTSI client that supports 3D avatars shall support the following data formats for the avatar assets:

- 3D meshes that conform to the binary glTF (GLB) format,

- Texture data components that conform to still image formats as defined in section 5.5, and

- Sparse and dense tensor data formats (e.g., used for skinning weights) as described in the ARF specification [9].

Data items of the base avatar should signal no compression or protection schemes.

### 5.6.3 2D Avatar Format

TBD.

Next change

# 6 AR Metadata

## 6.1 General

Real-time scene creation for an AR conference with two or more participants may be done by the MF to create a symmetric experience for all participants. For an MF to create a scene, it may request the following information from the UEs:

- spatial description of the space surrounding the UE e.g., the occlusion-free space around the user in which the AR media will be rendered.

- media properties indicating the AR media that the UE will be sending, and thus have to be incorporated in the scene.

- receiving media capabilities of the UEs, which may include

- UE media decoding capabilities

- UE hardware capabilities (e.g., the display resolution)

- information based on detecting the location, orientation, and capabilities of physical world devices, eligible for usage in an audio-visual communications session- information on whether each user should be represented by their avatar, and if so, the Avatar ID of the user, and the avatar capabilities of the UEsBased on this information the MF creates a scene which includes:

- defining the placement of the user and the AR media in that scene, including e.g., the position, size, depth from the user, anchor type, and recommended resolution (or quality)

- specific rendering properties for the AR media, e.g., for a 2D object to be rendered with a billboarding effect

The MF can then share the scene with the participant UEs using a supported scene description format. This scene description may be different for different UEs.

NOTE: The scene as sent by the MF allows the UE to 1) select and request any related media (for example, in a quality and bitrate based on the rendering characteristics or network connection), 2) render the complete scene on a (virtual) display device, and 3) update the rendering and requested media dynamically (e.g., according to the movement and view orientation of the user).

An AR-MTSI terminal may request the MF to generate the animation streams based on the UE’s supplied media streams.

Next change

### 6.3.2 Avatar Animation Stream Format

An AR-MTSI client or MF that supports avatars shall support the exchange of avatar animation data over the data channel according to the sample formats described in clause 8 of ISO/IEC 23090-39 [9].

NOTE: Support for other means to transport animation streams (e.g., over the media channel) may be added in the future.

When the data channel is used to send animation data, the metadata data channel message format defined in clause 6.2 shall be used and the avatar animation messages shall have the type “urn:3gpp:ar:v2:avatar:animation” and the format shown in Table 6.6-1.

Table 6.6-1: Message format for avatar animation messages

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Cardinality | Description |
| id | string | 1..1 | A unique identifier of the message in the scope of the data channel session. |
| type | string | 1..1 | urn:3gpp:ar:v2:avatar:animation |
| message | Object | 1..1 | Message content  |
|  subtype | number | 1..1 | An identifier of the subtype of the animation message. Value 1 indicates a facial animation, value 2 indicates a joint animation, and value 3 indicates a landmark animation. Other values are reserved for future use. |
|  payload | Object | 1..1 | The avatar animation sample format corresponding to the message subtype. |

No compression scheme is defined for the animation samples.

End of changes