**3GPP TSG-SA WG4 Meeting** **#131 *S4-250100r01***

**Geneva, Switzerland, 17th – 21th Feb, 2025**

**Source: Huawei, Nokia, Tencent, Qualcomm**

**Title: [FS\_AVATAR] pCR on IMS call flow**

**Spec: 3GPP TR 26.813 v1.0.0**

**Agenda item: 9.7**

**Document for: Discussion and agreement**

**1. Introduction**

During the previous 130 meeting discussion, it was asked to suspend the study of IMS mapping for avatar communication until the conclusion from SA2 is available. Since the conclusion of the architecture and call flow from SA2 NG\_RTC\_Ph2 has been reached in clause AC.11 of TS 23.228 v19.1.0 agreed in last SA2#166 meeting, this contribution proposes to:

1. Add the avatar animation negotiation procedure to TR 26.813, to address the Editor note in step 4 of AC.11.3.2.2 and AC.11.3.3 of TS 23.228 v19.1.0.

2. Move clause 6.3.2.2 of PD to TR 26.813.

**2. Reason for Change**

Update call flow for avatar communication to align with the SA2 conclusion.

**3. Proposal**

It is proposed to agree the following changes to TR 26.813 v1.0.0.

\* \* \* 1st Change (ALL TEXT IS NEW) \* \* \* \*

### 8.6.2 Call Flow

#### 8.6.2.0 Base Avatar generation

Prior to a call session, an Avatar generation session may be required to ensure the availability of the requested Avatar in the BAR. This may include the generation of an Avatar in multiple formats as required by the end devices capabilities. The base avatar generation is out of scope of this specification and the procedure for base avatar generation in this clause is purely illustrative.

Une image contenant diagramme, croquis, Plan, Dessin technique

Le contenu généré par l’IA peut être incorrect.

Figure 8.6.2.0-1: Avatar generation call Flow

The generation of the base avatar representation may happen in different entities depending on the device capabilities (in the UE), the service functionalities offered by the IMS network (in the MF or the BAR), or the availability of a dedicated function under the MNO trusted environment.

**IMS network-centric Avatar generation**

D.1a.1: UE sends captured data needed to generate the base avatar to the MF.

D.1a.2: The MF uses the captured data sent by UE to generate the base avatar for the user.

D.1a.3: MF may store the generated base avatar to MF for future loading.

**External network-centric Avatar generation**

D.2a.1: UE sends captured data needed to generate the base avatar to an Avatar Generator in the MNO trusted domain.

NOTE: The communication between the UE and the external Avatar generator is out of scope of this study.

D.2a.2: The Avatar Generator uses the captured data sent by UE to generate the base avatar for the user.

D.2a.3: MF may store the generated base avatar to MF for future loading.

**UE-centric Avatar generation**

D.3a.1: UE uses the captured data needed to generate the base avatar and generates locally the base avatar for the user.

D.3a.2: MF may store the generated base avatar to MF for future loading.

#### 8.6.2.1 Call Setup and Capability Negotiation

The parameters of the session are negotiated if UE centric mode or network centric mode is needed. This includes exchanging capability information, media and metadata descriptions and formats. The involved entities agree on assignment of avatar generation, animation tasks and media requirements.

##### 8.6.2.1.1 Network Centric



**Figure 8.6.2.1.1-1: Network Centric Call Setup and Capability Negotiation Flow**

For network centric mode, the capability negotiation procedure is based on the avatar type (2D or 3D) and the capability information of UE and MF. The capability information includes the animation data type(s) (e.g., text, expression data and motion signals for joints) supported by UE or MF. After capability negotiation, the IMS AS instructs MF to download UE1’s base avatar from BAR, generate animation data by the source data received from UE1, and animate UE1’s base avatar by the animation data received from UE1 or generated by MF itself.

A.1: An audio/video session is established between UE1 and UE2.

A.2: The bootstrap and application data channels are established between UE1 and IMS.

A.3: The UE1 sends a capability negotiation request using the application data channel through MF to the DC AS. The message carries parameters including an avatar id chosen by UE1 and animation data types (e.g., text, expression data and motion signals for joints) supported by UE1.

A.4: The DC AS sends an avatar capability request to MF.

A.5: The MF responses its avatar capability information to the DC AS.

NOTE1: The step A.4 and A.5 are optional. The DC AS can decide MF’s avatar capability based on its local configuration.

NOTE2: The service of avatar capability provided by MF will be further defined in CT1/CT4 if needed.

A.6: The DC AS gets the avatar type (2D or 3D, from base avatar retrieved from BAR or to be generated by the MF) by avatar id, and confirms the capability negotiation result based on the avatar type and the capability supported by UE1 and MF. The capability negotiation result includes the animation method (e.g., by audio, text or expression data and motion signals for joints).

A.7: The DC AS sends the capability negotiation response to UE1 through MF. The message carries the capability negotiation result.

A.8: The subsequent procedure continues.

##### 8.6.2.1.2 UE Centric



**Figure 8.6.2.1.2-1: UE Centric Call Setup and Capability Negotiation Flow**

For UE centric mode, if UE1 centric procedure is used, there is no capability negotiation initialized by UE1, so only UE2 centric call flow is introduced. In UE2 centric procedure, the base avatar of UE1 is sent from MF to UE2 using data channel, and the animation data is sent from UE1 or MF to UE2 using RTP or data channel.

A.1: An audio/video session is established between UE1 and UE2.

A.2: The bootstrap and application data channels are established between UE1 and IMS, UE2 and IMS.

A.3: The UE1 sends a capability negotiation request using the application data channel through MF to the DC AS. The message carries an avatar id chosen by UE1 and the animation data types supported by UE1.

A.4: The DC AS check if UE2 centric mode is used, then the DC AS transfers the capability negotiation request to UE2 through MF. The request carries the animation data types supported by MF in addition to the capability negotiation parameters.

A.5: The terminating network/UE2 finishes the capability negotiation.

A.6: The terminating network/UE2 returns the capability negotiation response carrying the negotiation result. The negotiation result includes the animation method (e.g., by audio, text or expression data and motion signals for joints).

A.7: The DC AS transfers the capability negotiation response to UE1 through MF.

A.8: The subsequent procedure continues.

#### 8.6.2.2 Avatar Delivery and Animation



**Figure 8.6.2.2-1: IMS Avatar Delivery and Animation Flow**

**A. Call Setup and Capability Negotiation**

An audio/video session is established between UE1 and UE2 and parameters of the session are negotiated as described in clause 8.6.2.1.

**B. Scene Description Retrieval**

The MF and the participating UEs retrieve scene descriptions, the scene description may be shared by the MF with the UEs, or the UEs may have their own scene descriptions.

**C. Scene Description Update**

A scene update trigger occurs, e.g., if an object is added to or removed from a scene or if spatial information is updated. The update trigger may originate from the MF itself or the UEs. The UEs may update their scene descriptions independently or the MF may generate an updated scene description and share it with the UEs.

NOTE1: The step B and C are not needed for 2D avatar.

**D.1. Avatar Acquisition**

Alternative #1: Network-centric Avatar Generation

D.1a.1: UE1 sends captured data needed to generate the base avatar to the MF.

D.1a.2: The MF uses the captured data sent by UE1 to generate the base avatar for the user.

D.1a.3: MF may store the generated base avatar to MF for future loading.

NOTE2: This network-centric avatar generation may occur before avatar communication starts.

Alternative #2: Network-centric Avatar Loading

D.1b.1: The MF loads the base avatar (for 2D avatar, the base avatar is comprised of a DNN model and a base image/video; for 3D avatar, the base avatar can be a 3D model or an INR model) for UE1 from BAR.

**D.2. Avatar Delivery**

Alternative #1: UE1 centric

D.2a.1: The MF delivers the base avatar to UE1 through data channel.

Alternative #2: UE2 centric

D.2b.1: The MF delivers the base avatar to UE2 through data channel.

**D.3. Animation Data Generation**

Based on the capability negotiation result in step A, the UE or network may generate animation data.

Alternative #1: UE centric animation data generation

D.3a.1: The UE1 generates the animation data based on the source data (e.g., audio, video, text). The animation data may be transformed from the source data (e.g., from audio to text), or the same as the source data.

D.3a.2: UE1 delivers the animation data to the entity actuating avatar animation through RTP or data channel. The animating entity may be the MF or UE2.

Alternative #2: Network centric animation data generation

D.3b.1: UE1 sends source data for animation data generation to the MF over RTP (audio, video, text) or data channel (text).

D.3b.2: The MF processes the received source data to generate animation data during the session. The animation data may be transformed from the source data (e.g., from audio to text, video to motion data), or the same as the source data.

D.3b.3: The MF delivers animation data over RTP or data channel to the UE2 animating the base avatar. If network centric avatar animation is used, this step will be skipped. The animation data may be delivered to UE1 as well.

**D.4. Avatar Animation**

Based on the capability negotiation result in step A, the UE or network may animate the avatar.

Alternative #1: UE centric avatar animation

Alternative #1a: UE1 does avatar animation

D.4a.1: UE1 animates and renders the base avatar using animation data. The animation data is generated by UE1 in step D.3a.1.1.

D.4a.2: UE1 delivers the animated and rendered avatar to UE2. The animated and rendered avatar (e.g., 3D or 2D video) may be delivered through RTP.

Alternative #1b: UE2 does avatar animation

D.4b.1: UE2 animates and renders the base avatar using animation data. The animation data may be generated by the MF, following steps D.3b.1 to D.3b.2 and received by UE2 in step D.3b.3 or it may be generated by UE1 in step D.3a.1 and received by UE2 in step D.3a.2.

Alternative #2: Network centric avatar animation

D.4c.1: The MF animates and renders the UE1’s base avatar using animation data. The animation data may be generated by the MF, following step D.3b.1 and D.3b.2 or it may be received from UE1 following steps D.3a.1 and D.3a.2.

D.4c.2: The MF delivers the animated and rendered avatar to the UEs. In the figure, delivery to UE2 is shown as example. The animated and rendered avatar (e.g., 3D or 2D video) may be delivered through RTP.

NOTE3: Rendering is not needed for 2D avatar.

\* \* \* End of Changes \* \* \* \*