**3GPP TSG-SA WG1 Meeting #98e S1-221035r3**

**Electronic Meeting, 9 – 19 May 2022** *(revision of S1-22xxxx)*

**Source: Charter Communications**

**pCR Title: Pseudo-CR on Identification of a User and object**

**Draft Spec: 3GPP TS / TR 22.856**

**Agenda item: 7.4**

**Document for: Approval**

**Contact: Deh-Min Richard Wu <deh-minRichard.wu@charter.com>**

*Abstract: to add a new use case in the TR 22.856 to enable virtual representation of a user and an object*

**1. Introduction**

This PCR proposes that a VR/AR session supports a visual object representing a user and an object.

**2. Reason for Change**

Without the CR, a VR/AR session establishment is displaying MSISDN or NAI. A visual representation is a non-3GPP solution.

**3. Conclusions**

It is a service enabler that a visual representation of a user and an object with the support of 3GPP solutions.

**4. Proposal**

It is proposed to agree the following changes to 3GPP TR 22.856.

\* \* \* First Change \* \* \* \*

## 5.X Collaboration Service in Audio/Visual Media with No Real Physical Touch

### 5.X.1 Description

An avatar can be generated by different software libraries and applications external to the 5G system. A user avatar can be used in his social networks such as Facebook, Twitter and mail providers. A user needs to pass a username to retrieve a user’s social network profile to obtain her visual representation. The avatar service is currently not enabled by 5GS to use a visual representation combined with IMSI for session establishment and during an active session.

Some popular OTTs integrate an avatar service to collaborate with 5GS as a communication platform. An avatar used identifies solely within a specific service. There is no avatar shared across services, no integration with 5GS. We are looking for an avatar service enabled within or outside the 5GS, which can utilize the user visual representation at the session establishment as well as during an active session. The VR/AR session supported by 5GS can therefore support a better latency via an avatar.

Joe and Jerry are business partners and living in different communities. User Joe and Jerry wearing a VR mobile headset and both hands controllers are making an immersive experience. Joe and Jerry are situated in fully virtual environments and interact with virtual objects through user interaction techniques. They both are feeling in the virtual environment to paint collaboratively a picture which is a virtual projection. They feel commonly in a physical space together. They feel that they can collaborate a painting in real-time and align with their painting with the virtual objects. They are able to share the sense of presence, interaction and communication.



Figure 5.x-1: Collaborative Painting

Joe and Jerry both have picked an avatar from a picture or in a more advanced format, such as a cartoonish avatar, a photorealistic user representation, and stored them in their mobile subscriber profile.

### 5.X.2 Pre-conditions

Mobile headsets of Joe and Jerry are both registered as UEs in different operators’ CPNs.

An application supported by a 3rd party service provider for drawing a virtually projected picture, for example VR painting and modelling tools, is running as one or many virtual objects in the 5GC by the network service provider. The users can use the tool to draw freehand in 3D space using smooth curves, grab and move points to adjust splines, and use both controllers to create a surface, assign a color and group objects.

3rd party service provider has an e-agreement to provide a VR service via 5GC operators to Joe and Jerry.

### 5.X.3 Service Flows

1. Joe makes a VR connection invite from his mobile headset to Jerry’s mobile headset connected to his home 5G-RG CPN. Joe’s VR stereo depth cameras are connected with the 3rd party service via 5G-RG CPN.
2. Jerry’s mobile headset alerts him about this invite via Joe's visual representation, e.g. avatar, with a gesture of invite and Jerry accepts the session.

Note: Joe’s identification can be represented alternatively as MSISDN or NAI which is linked to the avatar.

1. Jerry and Joe are able to have the sense of presence in the same room though they are not physically next to each other. Joe shakes his hand with Jerry with visual sense of touch, though physically they are separated by few streets away.
2. Joe is talking about a painting together. Jerry can feel Joe’s audio sense from his visual presence via his visual representation. Both users experience symmetrically the other’s presence in his own environment with the “canvas” supported by the service.
3. Joe sets up paint, paper and colors from the 3rd party application which represents an object in a visual representation. Both Joe and Jerry have the same visual and audio experience on the happening though they have not physically touched these objects.
4. Jerry and Joe can collaborate the drawing together and exchange their opinions via conversation, view of angles and gestures in 360° video.
5. They terminate the session after an hour.
6. Jerry and Joe are able to retrieve their painting from 3rd party service provider via 5GS and store it locally at their own desktop.

### 5.X.4 Post-conditions

A VR session is charged by the service provider and network operators. The users and objects are viewable with a visual representation during the session establishment and connectivity.

### 5.X.5 Existing features partly or fully covering the use case functionality

TS 22.261 v18.6.0 chapter 7.6.1 AR/VR:

Audio-visual interaction is characterised by a human being interacting with the environment or people, or controlling a UE, and relying on audio-visual feedback. In the use cases like VR and interactive conversation the latency requirements include the latencies at the application layer (e.g. codecs), which could be specified outside of 3GPP.

To support VR environments with low motion-to-photon capabilities, the 5G system shall support:

- motion-to-photon latency in the range of 7 ms to 15ms while maintaining the required resolution of up to 8k giving user data rate of up to [1Gbit/s] and

- motion-to-sound delay of [< 20 ms].

To support interactive task completion during voice conversation, the 5G system shall support low-delay speech coding for interactive conversational services (100 ms, one-way mouth-to-ear).

Due to the separate handling of the audio and video component, the 5G system will have to cater for the VR audio-video synchronisation in order to avoid having a negative impact on the user experience (i.e. viewers detecting lack of synchronization). To support VR environments the 5G system shall support audio-video synchronisation thresholds:

- in the range of [125 ms to 5 ms] for audio delayed and

- in the range of [45 ms to 5 ms] for audio advanced.

When it comes to implementation of applications containing AR/VR components, the requirements on the 5G network could depend on architectural choices implementing these services. Note 3 in table 7.1-1 above gives an example on such dependences for a VR application in a 5G system. Table 7.6.1-1 below illustrates additional use cases and provides more corresponding requirements on the 5G system.

- Cloud/Edge/Split Rendering – Cloud/Edge/Split Rendering is characterised by the transition and exchange of the rendering data between the rendering server and device.

- Gaming or Training Data Exchanging – This use case is characterised by the exchange of the gaming or training service data between two 5G connected AR/VR devices.

- Consume VR content via tethered VR headset – This use case involves a tethered VR headset receiving VR content via a connected UE; this approach alleviates some of the computation complexity required at the VR headset, by allowing some or all decoding functionality to run locally at the connected UE. The requirements in the table below refer to the direct wireless link between the tethered VR headset and the corresponding connected UE.

TS 22.261 v18.6.0 chapter 6.43 Tactile and multi-modal communication service:

The tactile and multi-modal communication service can be applied in multiple fields, e.g. industry, robotics and telepresence, virtual reality, augmented reality, healthcare, road traffic, serious gaming, education, culture and smart grid [38]. These services support applications enabling input from more than one sources and/or output to more than one destinations to convey information more effectively. As figure 6.43.1-1 illustrates, the input and output can be different modalities including:

* Video/Audio media;
* Information received by sensors about the environment, e.g. brightness, temperature, humidity, etc.;
* Haptic data: can be feelings when touching a surface (e.g., pressure, texture, vibration, temperature), or kinaesthetic senses (e.g. gravity, pull forces, sense of position awareness).

For immersive multi-modal VR applications, synchronization between different media components is critical in order to avoid having a negative impact on the user experience (i.e. viewers detecting lack of synchronization), particularly when the synchronization threshold between two or more modalities is less than the latency KPI for the application. Example synchronization thresholds [41] [42] [43] [44] are summarised in table 6.43.1-1.

TS 22.261 v18.6.0 chapter 7.11 KPIs for tactile and multi-modal communication service.

TS 22.263 v17.4.0:

AVProd workflows also require accurate timing protocols for 2 reasons

1. To enable multiple cameras and microphones to be synchronized thus avoiding the capture of mis-matched audio and video.
2. To provide IEEE-1588-2008 PTP [6] with an SMPTE 2059-2 [5] profile which is used for the accurate time stamping of IP packets

It is anticipated that the 5G system will act as a master clock and media clocks will be generated by UE applications. Requirements for this are in line with those in 22.104. If suitable sources are available, then each device my operate from its own master clock.

The 5G network shall be able to provide a time reference information to a 3rd party application acting as a master clock with an accuracy of 1 microsecond.

The 5G system shall support mechanisms to allow 3rd party application to update information associated to UE configuration (e.g. media compression, resolution, frame rate) for a UE or group of UEs using the application.

The 3GPP system shall be able to enable a UE to receive low-latency downlink multicast traffic from one network (e.g. NPN), and paging as well as data services from another network (e.g. PLMN) simultaneously.

NOTE: Depending on the capabilities and configurations of the UE, limitations of data-rate and latency may be acceptable.

Due to the separate handling of the audio and video component, the 5G system will have to cater for the VR audio-video synchronisation in order to avoid having a negative impact on the user experience (i.e. viewers detecting lack of synchronization). To support VR environments the 5G system shall support audio-video synchronisation thresholds:

- in the range of [125 ms to 5 ms] for audio delay and

- in the range of [45 ms to 5 ms] for audio advanced.

### 5.X.6 Potential New Requirements needed to support the use case

Editor Note: KPI requirements are FFS

[PR 5.x.6-1] 5GS shall enable a mechanism for users to control (e.g. create, delete, access rights to an avatar) their own related data which are associated with the subscription and managed by the network service provider.

[PR 5.x.6-1] 5GS shall enable a mechanism for users to manage their visual rendering relationships between a device and a user (e.g. which device associated with which user).

\*\*\* End of Change \*\*\*