**3GPP TSG-SA WG1 Meeting #94bis-e S1-212079**

**Electronic Meeting, 5 – 12 July 2021** *(revision of S1-21xxxx)*

Title: Updating 5.1 use case of Immersive multi-modality VR application

Agenda Item: 2.6.1 FS\_TACMM

Source: China Mobile

Contact: Xiaonan Shi (shixiaonan@chinamobile.com)

*Abstract: This paper proposes a update about the immersive multi-modal VR application with introducing the supporting of one UE transmission multi-modal signals.*

Reason for change: The use case of 5.1 describes a virtual reality application case, and there is one application server in this case to process the virtual reality video and haptic information. For the terminal sides, there can be one UE which collects all the packets, and another case that there are multiple UEs which can transmit different dimensional information to application server (e.g. access in different frequency).

Revision: to clarify that in this scenario multiple UEs transmit different dimensional information to application server.

5.1 Immersive Multi-modal Virtual Reality application

5.1.1 Description

Immersive Multi-modal Virtual Reality application describes the case of a human interacting with virtual entities in a remote environment such that the perception of interaction with a real physical world is achieved. Users are supposed to perceive multiple senses (vision, sound, touch) for full immersion in the virtual environment. The degree of immersion achieved indicates how real the created virtual environment is. Even a tiny error in the preparation of the remote environment might be noticed, as humans are quite sensitive when using Immersive Multi-modal Virtual Reality applications. Therefore, a high-field virtual environment (high-resolution images and 3-D stereo audio) is essential to achieve an ultimately immersive experience.

5.1.2 Pre-conditions

The devices for Immersive Multi-modal Virtual Reality application may include multiple type of devices such as VR glass type device, the gloves and other potential devices that support haptic and/or kinaesthetic modal. These devices are connected to the Immersive Multi-modal Virtual Reality application via the 5G network, see Figure 5.1.2-1.

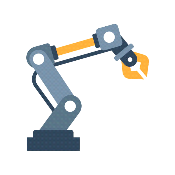
NOTE: The devices that are connected to VR application via the 5G network are assumed to be 3GPP UEs.

Based on the service agreement between MNO and Immersive Multi-modal Reality application operator, the application operator may in advance provide the 5G network the application information including the application traffic characteristics and the service requirement for network connection. For example, the packet size for haptic data is related to the Degrees Of Freedom (DOF) that the haptic devices supports, and packet size for one DoF is 2-8 Bytes [3] and the haptic device generates and sends 500 haptic packets within one second.



Application

Servers



Service data flows

5G network

**Figure 5.1.2-1: Immersive Multi-modality Virtual Reality application**

5.1.3 Service Flows

1. The application user utilizes the devices to experience Immersive Multi-modal Virtual Reality application. The user powers on the devices to connect to the application server, then the user start the gaming application.

2. During the gaming running period, the devices periodically sends the sensing information to the application server, including: haptic and/or kinesthetic feedback signal information which is generated by haptic device, and the sensing information such as positioning and view information which is generated by the VR glass.

NOTE 1: The devices may send the haptic data and the sensing data with different periodic time. As an example, the device may send one packet containing haptic information to the application server every 2ms, and send the packets related to sensing information to application server every 4ms. Thus the haptic data and sensing data may be transferred in 5G network via two separate flows. The amount of haptic packets that are generated and transferred within one second may be 1K - 4K packets (without haptic compression encoding), or 100-500 packets (with haptic compression encoding). As indicated in IEEE 1918.1 [3], the size of each haptic packet is related to the DoF capacity that haptic device supports, the data size for one DoF is 2-8 Byte.

3. According to the uplink data from the devices, the application server performs the process operation on immersive game reality including rendering and coding the video, the audio and haptic model data, then application server periodically sends the downlink data to the devices via 5G network.

NOTE 2: The application server may also send the haptic data and the video/audio data with different periodic time. For example, the application server sends one packet containing haptic information to the device every 2ms, and it sends the packets related to one video/audio frame to the device every 16.7ms in case 60 Frame Per Second which forms one burst traffic that goes on 3ms. Thus the haptic data and audio/video data may be transferred via two separate service data flows of a single session.

4. The devices receive the data from the application server and present the related sensing including video, audio and haptic to the user.

5.1.4 Post-conditions

The user experiences the immersive game reality application enabled by 5G network, and the 5G system address the service requirements of the application.

5.1.5 Existing features partly or fully covering the use case functionality

3GPP TS 22.261 [6] specifies KPIs for high data rate and low latency interactive services including Cloud/Edge/Split Rendering, Gaming or Interactive Data Exchanging, Consumption of VR content via tethered VR headset, and audio-video synchronization thresholds.

5.1.6 Potential New Requirements needed to support the use case

5.1.6.1 KPIs for Immersive Multi-modal Virtual Reality application

The 5G System shall provide the network connection to address the KPIs for Immersive Multi-modal Virtual Reality application, see Table 5.1.6.1-1.

**Table 5.1.6.1-1 – Potential key performance requirements for Immersive Multi-modality Virtual Reality application**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Traffic direction** | **Traffic types** | **Packet Size** | **Reliability (%)** | **Latency (ms)** | **Average data rate** |
| Device 🡪 Application Server | Haptic feedback | 1 DoF: 2-8 B  3 DoFs: 6-24 B  6 DoFs: 12-48 B  More DoFs may supported by the haptic device | [99.9] (without haptic compression encoding)  [99.999] (with haptic compression encoding) | <5 for one packet [Note 2] | 1k-4k packets/s (without haptic compression encoding);  100-500 packets/s, (with haptic compression encoding) |
| Sensing information i.e. User poisoning and view |  | [99.99] | <5 | < 1Mbps |
| Application Server 🡪 Device | Video |  | [99.9] | <10 [Note 1] for one video frame | 1-100 Mbps |
| Audio |  | [99.9] | <10 for one audio frame | 5-512 kbps |
| Haptic feedback | 1 DoF: 2-8 B  3 DoFs: 6-24 B  6 DoFs: 12-48 B | [99.9] (without haptic compression encoding)  [99.999] (with haptic compression encoding) | <5 for one packet [Note 2] | 1k-4k packets/s (without haptic compression encoding);  100-500 packets/s, (with haptic compression encoding) |
| NOTE 1: Motion-to-photon delay (the time difference between the user’s motion and corresponding change of the video image on display) should be less than 20ms, the communication latency for transfer the packets of one audio-visual media is less than 10ms, e.g. the packets corresponding to one video/audio frame are transferred to the devices within 10ms.  NOTE 2: Refer to IEEE 1918.1 [3] , as for haptic feedback, the latency should be less than 25ms for accurately completing haptic operations. As rendering and hardware introduce some delay, the communication delay for haptic modality should be reasonably less than 5ms, i.e. the packets related to one haptic feedback are transferred to the devices within 10ms. | | | | | |

5.1.6.2 Service requirements for Immersive Multi-modal Virtual Reality application

Editor’s Note: The service requirements for Immersive Multi-modal VR application needs FFS.

Editor’s Note: KPIs for haptic- audio-video synchronisation needs FFS.