**3GPP TSG-SA1 Meeting #96e *S1-21xxxx***

**Electronic Meeting, 8 – 18 November 2021** *(revision of S1-21xxxx)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.1* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | **22.847** | **CR** | **<CR#>** | **rev** | **-** | **Current version:** | **18.0.0** |  |
|  | | | | | | | | |
| *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:*** | Align the use of multi-modal | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Huawei, HiSilicon | | | | | | | | | |
| ***Source to TSG:*** | SA1 | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | FS\_TACMM | | | | |  | ***Date:*** | | | 2021-10-13 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | D |  | | | | | ***Release:*** | | | Rel-18 |
|  | *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:*** | | Both multi-modal and multi-modality are used in the current TR, which need to be aligned to avoid any confusion. In general modal is used as an adjective while modality a noun. In the context of describing a (communication) service or an application or a system, multi-modal should be used instead of multi-modality. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | To replace multi-modality with multi-modal when it is intended to be used as an adjective:   * tactile and multi-modal communication services * multi-modal service * multi-modal interactive system * multi-modal application * tactile and multi-modal interactions * multiple modal representations * multi-modal data and single-modal data * multi-modal outputs | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | Misuse of "multi-modality" and “modality” in the specification. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 1, 3.1, 4.1, 4.2, 5.1.2, 5.1.3, 5.1.6, 5.2.6, 5.3.1, 5.3.2, 5.3.6, 5.4, 5.4.1, 5.4.4, 5.4.6, 5.5.1, 5.5.2, 5.5.4, 5.5.6, 5.6.1, 5.6.3, 5.8.6 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **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:*** | | The name of the specification needs to be updated to “Study on supporting tactile and multi-modal communication services”. | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

**--------- The 1st change ---------**

1 Scope

This present document provides stage 1 use cases and potential 5G requirements on supporting tactile and multi-modal communication services. In the context of the present document, the aspects addressed include:

1. Study new scenarios and identify use cases and potential requirements for immersive real time experience involving tactile and multi-modal interactions, including:
2. Network assistance for coordinated transmission of multiple modal representations associated with the same session,
3. aspects of charging, security and privacy, and
4. KPIs (including network reliability and availability).
5. Gap analysis with existing requirements and functionalities on supporting tactile and multi-modal communication services.

**--------- The 2nd change ---------**

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP 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 3GPP TR 21.905 [1].

**end-to-end latency:** the time that takes to transfer a given piece of information from a source to a destination, measured at the communication interface, from the moment it is transmitted by the source to the moment it is successfully received at the destination.

NOTE 1: This definition was taken from TS 22.261 [6].

**Multi-modal Data:** Multi-modal Data is defined to describe the input data from different kinds of devices/sensors or the output data to different kinds of destinations (e.g. one or more UEs) required for the same task or application. Multi-modal Data consists of more than one Single-modal Data, and there is strong dependency among each Single-modal Data. Single-modal Data can be seen as one type of data.

**reliability**: in the context of network layer packet transmissions, percentage value of the amount of sent network layer packets successfully delivered to a given system entity within the time constraint required by the targeted service, divided by the total number of sent network layer packets.

NOTE 2: This definition was taken from TS 22.261 [6].

**service area:** geographic region where a 3GPP communication service is accessible.

NOTE 3: The service area can be indoors.

NOTE 4: For some deployments, e.g., in process industry, the vertical dimension of the service area can be considerable.

NOTE 5: This definition was taken from TS 22.261 [6].

**synchronisation threshold:** A multi-modal synchronisation threshold can be defined as the maximum tolerable temporal separation of the onset of two stimuli, one of which is presented to one sense and the other to another sense, such that the accompanying sensory objects are perceived as being synchronous.

NOTE 6: This definition is based on [22].

**Tactile Internet**: A network (or network of networks) for remotely accessing, perceiving, manipulating, or controlling real or virtual objects or processes in perceived real time by humans or machines.

NOTE 7: This definition is based on IEEE P1918.1 [3].

**user experienced data rate:** the minimum data rate required to achieve a sufficient quality experience, with the exception of scenario for broadcast like services where the given value is the maximum that is needed.

NOTE 8: This definition was taken from TS 22.261 [6].

**--------- The 3rd change ---------**

4.1 Multi-modal service

Tactile and multi-modal communication services enable multi-modal interactions, combining ultra-low latency with extremely high availability, reliability and security. Tactile Internet can be applied in multiple fields, including: industry, robotics and telepresence, virtual reality, augmented reality, healthcare, road traffic, serious gaming, education and culture, smart grid, etc. [2]. Multiple modalities can be used in combination in a service to provide complementary methods that may convey redundant information but can convey information more effectively. With the benefit of combining input from more than one source and/or output to more than one destination, interpretation in communication services will be more accurate and faster, response can also be quicker, and the communication service will be smoother and more natural.

For a typical tactile and multi-modal communication service/application, there can be different modalities affecting the user experience, e.g.:

* Video/Audio media;
* Information perceived 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).

The ambient information may be further processed to generate IoT control instructions as the feedback. The haptic data, according to the physiological perception, has specific characteristics, e.g. frequency and latency, and may require adequate periodic, deterministic and reliable communication path. For example, the sampling rate of the haptic device for teleoperation systems may reach 1000 times per second and samples are typically transmitted individually hence 1000 packets per second, while the video is 60/90 frames per second. The high frequency transmission of small packets over a long distance would be a great challenge for 5G system.

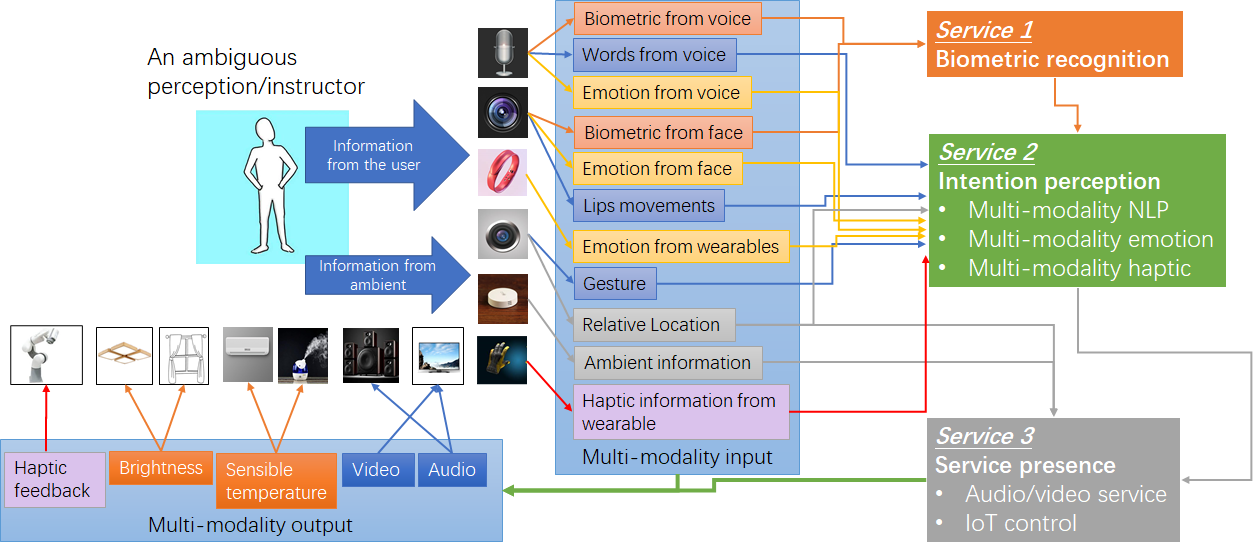
Multiple modalities can be transmitted at the same time to multiple application servers for further processing in a coordinated manner, in terms of QoS coordination, traffic synchronization, power saving, etc.

* Multiple outcomes may be generated as the feedback. In the scenario of real time remote virtual reality service, a VR user may use a plurality of independent devices to separately collect video, audio, ambient and haptic data from the person and to receive video, audio, ambient and haptic feedback from one or multiple application servers for a same VR application. In this case, an end user could wear VR glasses to receive images and sounds, and a touch glove to receive a touch sensation, a camera to collect video inputs, a microphone to collect audio inputs, multiple wearable sensors to provide haptic information and environmental information associated to the user. The real time remote virtual reality service can also be conducted between two users.
* Multiple outcomes may need to reach the distributed UEs at the very same time. In the scenario of sound field reappearing, different channels of sounds are sent to the distributed sound boxes to simulate the sound from a particular direction. A small time difference may cause big direction error to impact user experience. In some cases, time difference of 1ms may cause more than 30° angle error.
* Multi-modal applications may involve a big number of UEs at a long distance. In the scenario of multi-modal telepresence, tens of UEs may need synchronization for time, control signal and visual signal.

In another scenario, the devices associated to the same tactile and multi-modal communication service may be triggered to wake up by the discovery of a tactile and multi-modality capable user/UE in proximity. And a different group of tactile and multi-modality capable devices can serve the user as he moves.

Other scenarios that can be investigated are industrial manufacturing and drones real-time applications, which require synchronous control of visual-haptic feedback.

4.2 Multi-modal interactive system

****

**Figure 4.2-1. Multi-modal interactive system**

As shown in figure 4.2-1, multi-modal outputs are generated based on the inputs from multiple sources. In the multi-modal interactive system, modality is a type or representation of information in a specific interactive system. Multi-modal interaction is the process during which information of multiple modalities are exchanged. Modal types consists of motion, sentiment, gesture, etc. Modal representations consists of video, audio, tactition (vibrations or other movements which provide haptic or tactile feelings to a person), etc.

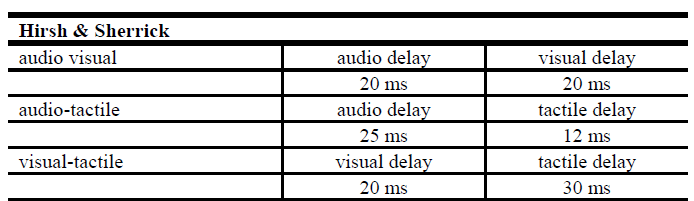
**--------- The 4th change ---------**

5.1 Immersive multi-modal Virtual Reality (VR) application

5.1.1 Description

Immersive multi-modal VR 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 VR applications. Therefore, a high-field virtual environment (high-resolution images and 3-D stereo audio) is essential to achieve an ultimately immersive experience.

One of the major objectives of VR designers and researchers is to obtain more realistic and compelling virtual environments. As the asynchrony between different modalities increases, users’ sense of presence and realism will decrease. There have been efforts (since 1960s or even earlier) in multi-modal-interaction research regarding the detection of synchronisation thresholds. The obtained results vary, depending on the kind of stimuli and the psychometric methods employed. Hirsh and Sherrick measured the synchronisation thresholds regarding visual, auditory and tactile modalities [23].



M.E. Altinsoy and co. believe the audio-tactile synchronization has to be at least within an accuracy of ±40 ms [22]. More results have been reported based on extensive theoretical and experimental efforts. [24] further indicated the perceptual threshold values were 50 ms for the audio lag and the 25 ms for audio lead.

As to the visual-tactile synchronisation threshold, Massimiliano Di Luca and Arash Mahnan provided test results in [25] that indicate that none of the participants could reliably detect the asynchrony if haptic feedback was presented less than 50ms after the view of the contact with an object. The asynchrony tolerated for haptic before visual feedback was instead only 15ms.

5.1.2 Pre-conditions

The devices for immersive multi-modal VR application may include multiple types of devices such as VR glass type device, the gloves and other potential devices that support haptic and/or kinaesthetic modal. These devices which are 5G UEs are connected to the immersive multi-modal VR application server via the 5G network without any UE relays, 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 VR application operator, the application operator may in advance provide the 5G network with 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.



5G UE (gloves)

Application

Server



5G network

Service data flows

5G UE (VR glasses)

**Figure 5.1.2-1. Immersive multi-modal VR application with multiple 5G UEs directly connected to 5G network**

5.1.3 Service Flows

1. The application user utilizes the devices to experience immersive multi-modal VR application. The user powers on the devices to connect to the application server, then the user starts the gaming application.

2. During the gaming running period, the devices periodically send 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 glasses.

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 necessary process operations 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, with different time periods respectively, 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, respectively, receive the data from the application server and present the related sensing including video, audio and haptic to the user.

NOTE 3: To obtain more realistic and compelling virtual environments, network assistance may be needed to ensure synchronisation thresholds between different modal data thus improve the end users’ sense of presence and realism.

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.

Support of audio-video synchronisation thresholds has been captured in TS 22.261:

*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.*

5.1.6 Potential New Requirements needed to support the use case

[PR 5.1.6-1] The 5G System shall provide the network connection to address the KPIs for immersive multi-modal VR applications, see table 5.1.6-1.

**Table 5.1.6-1 – Potential key performance requirements for immersive multi-modal VR applications**

| **Use Cases** | **Characteristic parameter (KPI)** | | | **Influence quantity** | | | | **Remarks** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Max allowed end-to-end latency** | **Service bit rate: user-experienced data rate** | **Reliability** | **Message size (byte)** | **# of UEs** | **UE Speed** | **Service Area** |
| Immersive multi-modal VR (UL: device 🡪 application sever) | 5 ms  (note 2) | 16 kbit/s -2 Mbit/s  (without haptic compression encoding);  0.8 - 200 kbit/s  (with haptic compression encoding) | [99.9%] (without haptic compression encoding)  [99.999%] (with haptic compression encoding) | 1 DoF: 2-8  3 DoFs: 6-24  6 DoFs: 12-48  More DoFs may supported by the haptic device | - | Stationary or Pedestrian | several km2  (note 3) | Haptic feedback |
| 5 ms | < 1Mbit/s | [99.99%] | MTU | - | Stationary or Pedestrian | several km2  (note 3) | Sensing information e.g. user poisoning and view |
| Immersive multi-modal VR (DL: application sever 🡪 device) | 10 ms  (note1) | 1-100 Mbit/s | [99.9%] | 1500 | - | Stationary or Pedestrian | several km2  (note 3) | Video |
| 10 ms | 5-512 kbit/s | [99.9%] | 50 | - | Stationary or Pedestrian | several km2  (note 3) | Audio |
| 5 ms  (note 2) | 16 kbit/s -2 Mbit/s  (without haptic compression encoding);  0.8 - 200 kbit/s  (with haptic compression encoding) | [99.9%] (without haptic compression encoding)  [99.999%] (with haptic compression encoding) | 1 DoF: 2-8  3 DoFs: 6-24  6 DoFs: 12-48 | - | Stationary or Pedestrian | several km2  (note 3) | Haptic feedback |
| 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.  NOTE 3: In practice, the service area can vary depends on the actual deployment. In some cases a local approach (e.g. the application servers are hosted at the network edge) is preferred in order to satisfy the requirements of low latency and high reliability. | | | | | | | | |

Editor's Note: Service Area is FFS.

[PR 5.1.6-2] The 5G system shall enable means to meet a synchronization threshold for flows of multiple UEs associated with an application based on input received from an authorized 3rd party.

Editor's Note: this requirement is FFS.

[PR 5.1.6-3] To support immersive multi-modal VR applications, the 5G system shall support the following synchronisation thresholds in order to avoid having a negative impact on the user experience (i.e. viewers detecting lack of synchronisation).

**Table 5.1.6-2: Potential Key performance requirements for synchronization thresholds for immersive multi-modal VR applications**

|  |  |  |
| --- | --- | --- |
|  | **synchronisation threshold** | |
| **audio-tactile** | audio delay:  [50 ms] | tactile delay:  [25 ms] |
| **visual-tactile** | visual delay:  [15 ms] | tactile delay:  [50 ms] |

Editor's Note: this requirement is FFS.

**--------- The 5th change ---------**

5.2.6 Potential New Requirements needed to support the use case

[PR 5.2.6-1] The 5G system shall be able to support tactile and multi-modal communication service with following KPIs.

**--------- The 6th change ---------**

5.3 Immersive VR games

5.3.1 Description

This use case is about supporting immersive VR games with tactile and multi-modal communication services. VR games have provided a better experience comparing to traditional games. As customers ask for more immersive game experience, haptic information has been taken into account including force and DOF. Traditional VR games provide video and audio information for players to create the real game scenarios. For better immersive VR games, haptic feedback is introduced and which provides the reality of touching things in games as well as the interaction of team players.

To play the VR game, players buy VR games and related equipment like VR glasses, hand shank, analogue steering wheel and haptic gloves, which may be produced by different manufactures. In multi-player VR games, they act as different UEs and need to corporate to complete the mission. On application level, the VR gaming application will be able to distinguish these UEs and share the information with network that these UEs and data flows are grouping under this VR gaming service, and network need to provide corresponding QoS accordingly.

For smooth experience in VR, the motion-to-photon latency should be less than 20ms [7], which indicates that the latency between player do one movement and the corresponding new pixels show in VR sights should be less than 20ms. The uplink dataflow in this loop is motion or haptic information, while the downlink in this loop is the video data. The QoS requirements for these two types of data flows are different, which includes latency requirements. But for the service level, it requires the joint latency consist both uplink and downlink. [5]

5.3.2 Pre-conditions

Alice and Bob are playing a VR game together. They need to cover each other’s back, find weapons and fight with zombies. The VR gaming equipment they wear are all connected to 5G network. The VR game application interacted with 5G network about the UE and dataflow information, and network provides the pre-agreed policy between application and operator on QoS requirements of each kind of modal data flow.

5.3.3 Service Flows

1. Alice and Bob both joined this VR games, and they can see each other’s character in the view.
2. Alice found two stones on the ground, she picks up both the stones. The delay between the real pick-up action and the virtual hand pick-up video is ignorable.
3. Bob has nothing to arm himself. So he asks Alice to throw him a stone.
4. Alice heard Bob and throws one of the stone to Bob.
5. Bob catches the stone can feel the weight of the stone.

5.3.4 Post-conditions

Alice and Bob can feel the things in the game as they are in the real world. The experience of this VR games is very realistic and smooth.

5.3.5 Existing features partly or fully covering the use case functionality

In TS 22.261 [6], there are performance requirements for supporting VR services.

5.3.6 Potential New Requirements needed to support the use case

[PR 5.3.6-1] The 5G system shall be able to support tactile and multi-modal communication service with following KPIs.

**--------- The 7th change ---------**

5.4 Support of skillset sharing for cooperative perception and manoeuvring of robots

5.4.1 Description

Today, most automated driving vehicles rely on a single controller, which is the vehicle itself: sensing and controlling features of its own [3]. Since 3GPP Rel-14, LTE-based support for V2V features have been developed and tested through collaborative participation from automotive and communication industry. However, it is still challenging to use automated driving functionalities in general unstructured settings, if the controlling features are based on a single controller, having no idea on how neighbouring vehicles will behave.

This consequently requires that the automated driving system (ADS) should allocate an extra safety margin into the planned trajectory which in turn causes traffic flow to be reduced and causes inefficiency to happen in a large scale network of vehicle networks where non-V2X vehicles and V2X-enabled vehicles possibly coexist. This is not even a problem for such an automated driving of road vehicles – the same applies to the operations of “automated manoeuvring robots” in unstructured settings. Without cooperation, the field of perception of a vehicle/robots is limited to the local coverage of the onboard sensors – not only for the relative distance, relative angle.

As a technology enabler solution against such problems of guaranteeing safety and traffic efficiency, it is being studied to share the sensor information [9] and manoeuver sharing [8] in SAE. Tactile Internet for V2N (potentially with assistance from edge cloud instead of general cloud servers) or V2V can enable an ultra-fast and reliable exchange of highly detailed sensor data sets between nearby vehicles, along with haptic information on trajectory [3]. Also, it would be one of the key factors for so-called “cooperative perception and manoeuvring” functionalities [10]: planning cooperative manoeuvers among multiple automated driving vehicle (or robots), such as plan creation, target point (TP) generation and target point risk assessment. It is by the TI connectivity that vehicles can perform a cooperative perception of the driving environment based on fast fusion of high definition local and remote maps collected by the onboard sensors of the surrounding vehicles (e.g., video streaming from camera, radar, or lidar). This allows to augment the sensing range of each vehicle and to extend the time horizon for situation prediction, with huge benefits for safety [3]. The onboard sensors in today automated driving vehicles generate data flows up to 8 Gb/s [3]. All these requirements call for new network architectures interconnecting vehicles and infrastructure utilizing ultralow-latency networks based on the Tactile Internet for cooperative driving services [3].

This use case is related to the support of (1) cooperative perception and manoeuvring and (2) extension of sensing range for cooperative automated driving scenarios using Tactile Internet, with some examples of moving robots (e.g., local delivery robots). Manoeuvring and perception obtained via haptic and multi-modal communications (also known as skillset sharing) are very timely shared between the controller and controlee.

5.4.2 Pre-conditions

Four robots S1, S2, C1 and C2 are working on delivery tasks from a geographic point to another, respectively.

Robots UEs S1 and S2 are automated driving robots with a standalone steer/control, manoeuvring in a crowded village.

Robots UEs C1 and C2 are automated driving robots with steer/control and manoeuver/skillset sharing functionalities, manoeuvring in another crowded village.

The roads that robots are using in these villages are not structured road (i.e., no lane separator, no lane marks, etc.) and they are under the same conditions for robots to move.

5.4.3 Service Flows

1. S1 is moving at speed of X and is getting close to S2.
   1. S1 does not know exactly what trajectory S2 is planning to move along. Therefore, S1 reduces the current moving speed to (0.5 \* X).
   2. During the operation at this reduced speed, S1 still does not know the detailed trajectory of S2. Therefore, S1 reserves a distance of Y1 meters relative to S2 and gets ready to further reduce the speed or to make a full stop, monitoring the movement of S2.
   3. S1 and S2 pass each other and continue their trip to the destinations, respectively.
2. C1 is moving at speed of X and is getting close to C2.
   1. C1 knows exactly what trajectory C2 is planning to move along as they share the manoeuvers. Therefore, C1 reduces the current moving speed to (0.9 \* X).
   2. During the operation at this reduced speed, C1 still knows the detailed trajectory of C2 through manoeuver sharing. Therefore, C1 reserves a distance of Y2 (Y2 << Y1) meters relative to C2 and gets ready to further reduce the speed to avoid any change of collision but there is very little chance that both should make a full stop as both are sharing the steer/control: one can yield the space for the other only when necessary.
   3. C1 and C2 pass each other and continue their trip to the destinations, respectively.

5.4.4 Post-conditions

The total time that S1 and S2 should spend is much greater than the total time that C1 and C2 should spend.

The total energy consumption (e.g., to accelerate from a low speed level to X) for S1 and S2 is greater than that for C1 and C2.

Fig. 5.4.4-1 provides a simplified explanation on the behaviours of speed changes for examples without (left section) and with (right section) real-time multi-modal communication for interactive haptic control and feedback (skillset sharing).



**Fig. 5.4.4-1. Simplified examples on the stochastic behaviours of the speed change (and the minimum margin to set between robots (or vehicle-styled robots)) without (left section) and with (right section) real-time multi-modal communication for interactive haptic control and feedback (skillset sharing). The road is assumed to be in general unstructured setting, e.g., no lane separator or marks.**

5.4.5 Existing features partly or fully covering the use case functionality

V2X performance requirements found in TS 22.185, TS 22.186. (e)CAV requirements in TS 22.104 [21]. VIAPA requirements in TS 22.263 [4]

5.4.6 Potential New Requirements needed to support the use case

[PR 5.4.6-1] 5G system shall be able to support real-time multi-modal communication for interactive haptic control and feedback with KPIs as summarized in Table 5.4.6-1.

**Table 5.4.6-1: Multi-modal communication service performance requirements.**

| **Use Cases** | **Characteristic parameter (KPI)** | | | **Influence quantity** | | | | **Remarks**  **(**NOTE 1**)** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Max allowed end-to-end latency**  **(**NOTE 2**)** | **Service bit rate: user-experienced data rate** | **Reliability** | **Message size (byte)** | **# of UEs** | **UE Speed** | **Service Area** |
| Skillset sharing low- dynamic robotics  (including teleoperation) Controller to controlee | [5-10ms] | 0.8 - 200 kbit/s (with compression) | [99,999%] | n DoFs: (2n)-(8n)  (n=1,3,6) | **-** | Stationary or Pedestrian | TBD | Haptic  (position, velocity) |
| Skillset sharing low- dynamic robotics  (including teleoperation)  Controlee to controller | [5-10ms] | 0.8 - 200 kbit/s (with compression) | [99,999%] | n DoFs: (2n)-(8n)  (n=1,10,100) | **-** | Stationary or Pedestrian | TBD | Haptic feedback |
| 10ms | 1-100 Mbit/s | [99,999%] | 1500 | **-** | Stationary or Pedestrian | TBD | Video |
| 10ms | 5-512 kbit/s | [99,9%] | 50 | **-** | Stationary or Pedestrian | TBD | Audio |
| Highly dynamic/ mobile robotics  Controller to controlee | [1-5ms] | 16 kbit/s -2 Mbit/s  (without haptic compression encoding);  0.8 - 200 kbit/s  (with haptic compression encoding) | [99,999%] (with compression)  [99,9%] (w/o compression) | n DoFs: (2n)-(8n)  (n=1,3,6) | **-** | high-dynamic | TBD | Haptic  (position, velocity) |
| Highly dynamic/ mobile robotics  Controlee to controller | [1-5ms] | 0.8 - 200 kbit/s | [99,999%] (with compression)  [99,9%] (w/o compression) | n DoFs: (2n)-(8n)  (n=1,10,100) | **-** | high-dynamic | TBD | Haptic feedback |
| [1-10ms] | 1-10 Mbit/s | [99,999%] | [2-4] | **-** | high-dynamic | TBD | Video |
| [1-10ms] | 100-500 kbit/s | [99,9%] | 100 | **-** | high-dynamic | TBD | Audio |
| NOTE 1: Haptic feedback is typically haptic signal, such as force level, torque level, vibration and texture.  NOTE 2: The latency requirements are expected to be satisfied even when multi-modal communication for skillset sharing is via indirect network connection (i.e., relayed by one UE to network relay). | | | | | | | | |

5.5 Haptic feedback for a personal exclusion zone in dangerous remote environments

5.5.1 Description

With the assistance of 5G networks, many industries including mining, operate unmanned and automated. In mining scenarios, drilling safety and precise control for automated rigs and sending effective alarms when needed to the onsite crew is vital for their safety. When crew move heavy lifting equipment wearing personal protection equipment (PPE) for their safety or work in a noisy/poor visibility environment, audio/light alarm systems coverage may not be detected for immediate reaction therefore use of wearables (belts, shoe sole, arm/shoulder tactile equipment) can improve the reliability of alarm system.

Furthermore, the nature of human brain response time to light and audio makes the use of haptic feedback to alert faster and more reliable. Human brain response to the sense of touch in range of 1 ms where the response to audio and video is in 100s of milliseconds. Therefore, the alarm system can be enriched with additional haptic information and multi-modal session can be relayed to the on-site crew over to accelerate human response time and improve the system reliability.

While haptic alarm can be actuated on an extended PPE haptic device (e.g., belt) to improve alerting the on-site crew, the audio siren and light evacuation signals (where possible) need to be sent, as an alert to ambient actuators requiring a multi-modal information transfer via separate service data flows, to the proximity devices during an emergency for directing them to a safe location. Furthermore, affective wearables feedback and affective or biometric data collected from workers can adjust the navigation process to the workers’ temperament in order to effectively guide them away from the hazard locations.

As an exmple, in a large mining environment, a hazard scenario is detected by the remote control unit to notif and navigate on-site workers by a multi-modal alarm system to avoid exclusion zones. The multi-modal alarm system monitors the environment using surveillance cameras fixed in the local site, on drones, or on workers helmets as well as haptic information relayed by the workers. Ambient sensory information (smoke, temperature, audio/video) can be relayed to a remote central monitoring and control unit to predict/detect hazard scenarios.

The personal exclusion zone needs to be defined to prohibit the entry of the on-site workers which can change over time. The detected exclusion zone information from remote control unit will actuate devices (e.g., siren and light) fixed in the local site or in drones as well PPE haptic belt actuator to navigate the workers.

5.5.2 Pre-conditions

Multi-modal UEs, with sensor, haptic and audio/visual capabilities, are interconnected, through the 5G Network provided by the MNO, both locally, or remotly over the MNO 5G Network.

The MNO provides service data flows for multi-modal information delivered over one common session to different UEs. E.g., a single user may receive haptic information be sent to her/his haptic glove and belt and audio to be sent to her/his headset to help navigate the worker through the exclusion zone.

****

**Figure 5.5.2-1. Example of a personal exclusion zone**

5.5.3 Service Flows

1. A Worker using PPE connected to a monitoring and control unit walks through a mining site where multiple exclusion zones are defined to protect her/him from hazardous occurrences. As the worker gets closer to the exclusion zone, her/his PPE establishes a session to enable transmission of alarms to the monitoring and control unit, e.g., noise levels, visibility levels and user location.

2. If a hazard situation is detected, on-site workers need to be alerted immediately and defined personal exclusion zone(s) should be communicated with the on-site miners according to their proximity to the exclusion zone. Thus, the monitoring system determines that the user must be alerted, through haptic and audio alerts, since the monitoring and control unit determined that those are the modal types the user can utilized safely in the dangerous environment.

3. The monitoring and control system, navigates workers using both a trained haptic language instruction (single buzz on the right side: move right, single buzz on the left side: move left, etc.) set and audio/visual mechanism to enable the user to safely leave the hazard area.

5.5.4 Post-conditions

The site worker is safely navigated away from the dangerous exclusion zone as the 5G network enables the PPE to alert the monitoring and control unit of dangerous situations which triggers the monitor and control unit to use the 5G network to transmit multi-modal commands to steer away site workers from dangerous exclusion zones.

5.5.5 Existing features partly or fully covering the use case functionality

In 3GPP TS 22.263 [4], there are requirements for supporting video, imaging, and audio for professional applications.

In 3GPP TS 22.261 [6] there are requirements to support the following:

Clause 6.3 Multiple Access Technologies

“The 5G system shall support a set of identities for a single user in order to provide a consistent set of policies and a single set of services across 3GPP and non-3GPP access type”

Clause 6.26.2.5 Traffic Types

“The 5G system shall support traffic scenarios typically found in a home setting (from sensors to video streaming, relatively low amount of UEs per group, many devices are used only occasionally) for 5G LAN-type service”

5.5.6 Potential New Requirements needed to support the use case

[PR. 5.5.6-1] The 5G network shall support a mechanism to allow an authorized 3rd party to provide QoS policy for flows of multiple UEs associated with an application. The policy may contain e.g. the expected 5GS handling and the associated triggering event.

[PR. 5.5.6-2] The 5G system shall support a mechanism to apply QoS policy for flows of multiple UEs associated with an application received from an authorized 3rd party.

[PR 5.1.6-3] The 5G system shall provide a network connection to address the KPIs for immersive multi-modal navigation applications, see Table 5.5.6-1.

**Table 5.5.6-1: Potential Key performance requirements for a personal exclusion zone in dangerous remote environments.**

| **Use Cases** | **Characteristic parameter (KPI)** | | | **Influence quantity** | | | | **Remarks** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Max allowed end-to-end latency (ms)** | **Service bit rate: user-experienced data rate** | **Reliability** | **Message size (byte)** | **# of UEs** | **UE Speed** | **Service Area** |
| Immersive multi-modal navigation applications  Remote Site 🡪 Local Site (DL) | 0-50 [11], [12] | 1-4k pkts/s  (w/o compression)  100-500 pkts/s,  (w/ compression.) | [99.999 %] | 1 DoF: 2 to 8  10 DoF: 20 to 80  100 DoF: 200 to 800 | - | Stationary or Pedestrian | TBD | Haptic feedback |
| <400 [11] | 1-100 Mbit/s | [99.999 %] | 1500 | - | Workers: Stationary/ or Pedestrian, | TBD | Video |
| <150 [11] | 5-512 kbit/s | [99.9 %] | 50 | - | Stationary or Pedestrian | TBD | Audio |
| <300 | 600 Mbit/s | [99.9 %] | MTU | - | Stationary or Pedestrian | TBD | VR |
| Local Site 🡪 Remote Site (UL) | <300 | 12 kbit/s [26] | [99.999 %] | 1 500 |  | Stationary or Pedestrian | TBD | Biometric / Affective |
| <400 [11] | 1-100 Mbit/s | [99.999 %] | 1 500 | - | Workers: Stationary/ or Pedestrian, UAV: [30-300mph] | TBD | Video |
| <150 [11] | 5-512 kbit/s | [99.9 %] | 50 | - | Stationary or Pedestrian | TBD | Audio |
| <300 | 600 Mbit/s | [99.9 %] | MTU | - | Stationary or Pedestrian | TBD | VR |

Editor’s Note: KPIs for personal exclusion zone in dangerous remote environments may need further discussion and verification.

[PR 5.5.6-4] The 5G system shall support the following synchronization thresholds to support immersive multi-modal navigation applications, see Table 5.5.6-2.

**Table 5.5.6-2: Potential Key performance requirements for synchronization thresholds for a personal exclusion zone in dangerous remote environments.**

|  |  |  |
| --- | --- | --- |
| **synchronisation threshold** | | |
| **audio-tactile** | audio delay:  [50 ms] | tactile delay:  [25 ms] |
| **visual-tactile** | visual delay:  [15 ms] | tactile delay:  [50 ms] |

5.6 Live Event Selective Immersion

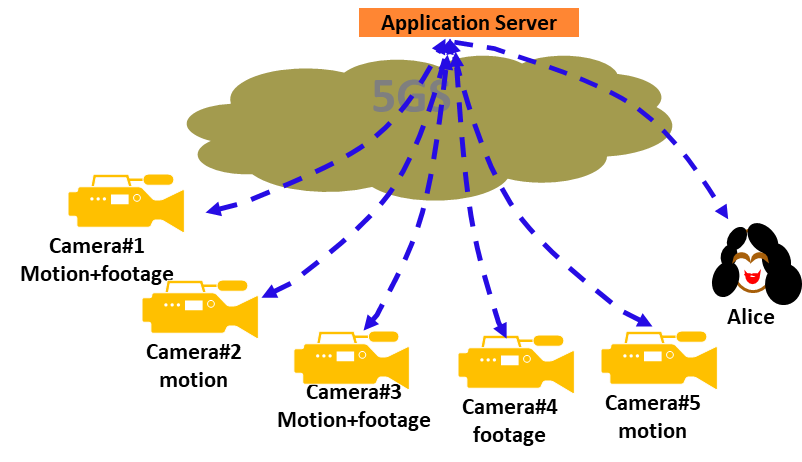
5.6.1 Description

To provide immersive experience of a live football game to the audience not at the scene, multiple AI cameras are deployed in the stadium collecting data of the video and audio from different angles for the generation of footages at the Application Server. The footages are provided to the audience for selection.

The AI camera predictively follows the live action, depending on the objects to follow e.g. ball or player. The Application Server predicts the potential actions of the object based on the data collected from the cameras and instructs the camera on what object to follow and how to follow. The AI camera acts based on the instructor from the Application Server.

This use case shows an example of multi-modal service as defined in clause 4.1 and 4.2, in which the application server requires inputs from multiple UEs to generate the outputs. The service flows and potential new requirements of this use case also apply to other multi-modal interactive system use case scenarios.

5.6.2 Pre-conditions

****

**Figure 5.6.2-1. Live Event Selective Immersion**

Each AI camera interacts with the Application Server over 5GS network as a UE.

Alice is watching the football game, and the footages for selection are transmitted to her UE over 5GS network.

Footages and motion prediction are generated at the Application Server based on the video and audio data collected by the AI camera.

Each AI camera has its own responsibilities:

* Camera#1: data collection for motion prediction and footage generation, the primary camera placed in a location owning the best view
* Camera#2: data collection for motion prediction, placed in the best location for capturing the motion
* Camera#3: data collection for motion prediction and footage generation
* Camera#4: data collection for footage generation
* Camera#5: data collection for motion prediction

5.6.3 Service Flows

1. The football game starts. Camera#1, Camera#2, Camera#3, Camera#4 and Camera#5, as the UEs, are switched on and registered to the 5GS network to collect video and audio data to conduct Live Event Selective Immersion service.
2. The Application Server informs 5GS that UEs corresponding to Camera#1, Camera#2, Camera#3, Camera#4 and Camera#5 are subject to the service application, and provides QoS requirements of these UEs and the coordination policies for this multi-modal service for assistance from 5GS.
3. The QoS requirements and the coordination policy are applied at 5GS. Camera#1, Camera#2, Camera#3, Camera#4 and Camera#5 transmit the collected data over 5GS to the Application Server at the target QoS.
4. The Application Server makes motion prediction based on data received from Camera#1, Camera#2, Camera#3 and Camera#5, and generates footages based on data received from Camera#1, Camera#3 and Camera#4.
5. The Application Server transmits motion prediction of the object(s) over 5GS to Camera#1, Camera#2, Camera#3 and Camera#5 and transmits footages over 5GS to Alice’s UE.
6. Some people are gathering at the gate of the stadium for the celebration of winning scores, then network congestion happens from time to time. Based on the coordination policy of the multi-modal service, when the target QoS of Camera#1 can’t be guaranteed, 5GS reduces QoS of Camera#4 to make sure QoS of Camera#1 is guaranteed; when the congestion is relieved, 5GS increases QoS of Camera#4 while target QoS of Camera#1 is still guaranteed.
7. The network congestion gets more serious, and target QoS of Camera#2 can’t be guaranteed. Since the motion data collected by Camera#2 is mandatory for motion prediction without which the motion prediction can’t be made, 5GS releases resources of Camera#2 and Camera#5 based on the coordination policy of the multi-modal service.

NOTE: When motion prediction doesn’t work, the Application Server may request the 5GS network to only guarantee QoS of Camera#1 or to release resources of all the cameras. This is beyond what 5GS can coordinate.

**--------- The 8th change ---------**

5.8.6 Potential New Requirements needed to support the use case

[PR 5.8.6-1] 5G system shall be able to support the interaction with applications on UEs or data flows grouping information within one tactile and multi-modal communication service.

[PR 5.8.6-2] 5G system shall be able to provide a dynamic mechanism to transfer different data flows with different latency to achieve a certain transmission time difference within one tactile and multi-modal communication service.

Editor’s note: The above requirements are for FFS.