**3GPP TSG SA WG4#116e S4-211370**

**E-meeting, 10th – 19th November 2021**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.1* | | | | | | | | |
| **PSEUDO CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | 26.998 | **CR** |  | **rev** |  | **Current version:** | 1.0.3 |  |
|  | | | | | | | | |
| *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:*** | [FS\_5GSTAR] Networked AR Latencies | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Qualcomm Incorporated | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | FS\_5GSTAR | | | | |  | ***Date:*** | | | 2021-11-02 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | C |  | | | | | ***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 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:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 2, 4.5.2, 4.5.3 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  |  | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  |  | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  |  | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

**===== 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 TR 26.928: “Extended Reality (XR) in 5G”

[3] Wireless Broadband Alliance, “5G and Wi-Fi RAN Convergence”, April 2021.

[4] Khronos Group, The OpenXR Specification, 1.0, <https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html>

[5] W3C, WebXR Device API, W3C Working Group Draft, <https://www.w3.org/TR/webxr/>

[6] ISO/IEC 23090-13:2021 CD: “Information technology — Coded representation of immersive media — Part 13: Video Decoding Interface for Immersive Media”

[7] Miscrosoft Azure KinectTM DK documentation, <https://docs.microsoft.com/en-us/azure/kinect-dk/>

[8] Google ARCoreTM: Use Depth in your Android app, <https://developers.google.com/ar/develop/java/depth/developer-guide>

[9] Microsoft Azure KinectTM DK documentation: Use Azure Kinect Sensor SDK image transformations, <https://docs.microsoft.com/en-us/azure/kinect-dk/use-image-transformation#overview>

[10] Daniel Wagner, Louahab Noui, Adrian Stannard, "Why is making good AR displays so hard?", LinkedIn Blog, August 7, 2019, <https://www.linkedin.com/pulse/why-making-good-ar-displays-so-hard-daniel-wagner/>

[11] Daniel Wagner, "MOTION TO PHOTON LATENCY IN MOBILE AR AND VR", Medium Blog, August 20, 2018, https://medium.com/@DAQRI/motion-to-photon-latency-in-mobile-ar-and-vr-99f82c480926

[12] Yodayoda, "Why loop closure is so important for global mapping", Medium Blog, December 24, 2020, <https://medium.com/yodayoda/why-loop-closure-is-so-important-for-global-mapping-34ff136be08f>

[13] 3GPP TS 22.261: “Service requirements for the 5G system”

[14] 3GPP TR 22.873: “Study on evolution of the IP Multimedia Subsystem (IMS) multimedia telephony service”

[15] 3GPP TS 26.114: “IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction”

[16] 3GPP RP-193241: “New SID on XR Evaluations for NR”

[17] ISO/IEC 23090-2:2021: “Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format”

[18] ISO/IEC 23090-3:2021: “Information technology — Coded representation of immersive media — Part 3: Versatile video coding”

[19] ISO/IEC 23090-5:2021: “Information technology — Coded representation of immersive media — Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)”

[20] ISO/IEC 23090-8:2020: “Information technology — Coded representation of immersive media — Part 8: Network based media processing”

[21] ETSI GS ISG ARF 003 v1.1.1 (2020-03): “Augmented Reality Framework (ARF) AR framework architecture”

[22] Khronos Group, The GL Transmission Format (glTF) 2.0 Specification, <https://github.com/KhronosGroup/glTF/tree/master/specification/2.0/>

[23] ISO/IEC 23090-14:2021 DIS: “Information technology — Coded representation of immersive media — Part 14: Scene Description for MPEG-I Media”

[24] ISO/IEC 23090-10:2021 FDIS: “Information technology — Coded representation of immersive media — Part 10: Carriage of Visual Volumetric Video-based Coding Data”

[25] ISO/IEC 23090-18:2021 DIS: “Information technology — Coded representation of immersive media — Part 18: Carriage of Geometry-based Point Cloud Compression Data”

[26] 3GPP TS 26.501: “5G Media Streaming (5GMS); General description and architecture”

[27] H. Chen, Y. Dai, H. Meng, Y. Chen and T. Li, "Understanding the Characteristics of Mobile Augmented Reality Applications," 2018 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2018, pp. 128-138.

[28] S. Kang, H. Choi, “Fire in Your Hands: Understanding Thermal Behavior of Smartphones”, The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19)

[29] T. Chihara, A. Seo, “Evaluation of physical workload affected by mass and center of mass of head-mounted display”, Applied Ergonomics, Volume 68, pp. 204-212, 2018

[30] Google Draco: <https://google.github.io/draco/>

[31] T.Ebner, O.Schreer, I. Feldmann, P.Kauff, T.v.Unger, “m42921 HHI Point cloud dataset of boxing trainer”, MPEG 123rd meeting, Ljubljana, Slovenia

[32] Scene understanding, <https://docs.microsoft.com/en-us/windows/mixed-reality/scene-understanding>

[33] Serhan Gül, Dimitri Podborski, Jangwoo Son, Gurdeep Singh Bhullar, Thomas Buchholz, Thomas Schierl, Cornelius Hellge, “Cloud Rendering-based Volumetric Video Streaming System for Mixed Reality Services”, Proceedings of the 11th ACM Multimedia Systems Conference (MMSys'20), June 2020

[34] Scene lighting: <https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/lights>

[35] PBR material: <https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/pbr-materials>

[36] Colour Material: <https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/color-materials>

[37] S. N. B. Gunkel, H. M. Stokking, M. J. Prins, N. van der Stap, F.B.T. Haar, and O.A. Niamut, 2018, June. Virtual Reality Conferencing: Multi-user immersive VR experiences on the web. In Proceedings of the 9th ACM Multimedia Systems Conference (pp. 498-501). ACM.

[38] Dijkstra-Soudarissanane, Sylvie, et al. "Multi-sensor capture and network processing for virtual reality conferencing." Proceedings of the 10th ACM Multimedia Systems Conference. 2019.

[39] VRTogether, a media project funded by the European Commission as part of the H2020 program, <https://vrtogether.eu/>, November 2020.

[40] MPEG131 Press Release: Point Cloud Compression – WG11 (MPEG) promotes a Video-based Point Cloud Compression Technology to the FDIS stage: <https://multimediacommunication.blogspot.com/2020/07/mpeg131-press-release-point-cloud.html>

[41] 3GPP TR 23.701: “Study on Web Real Time Communication (WebRTC) access to IP Multimedia Subsystem (IMS); Stage 2”

[42] 3GPP TR 23.701: “Study on enhancements to Web Real Time Communication (WebRTC) access to IP Multimedia Subsystem (IMS); Stage 2”

[43] 3GPP TS 24.371: “Web Real-Time Communications (WebRTC) access to the IP Multimedia (IM) Core Network (CN) subsystem (IMS); Stage 3; Protocol specification”

[44] IETF RFC 8831: WebRTC Data Channels

[45] IETF RFC 8864: Negotiation Data Channels Using the Session Description Protocol (SDP)

[46] IETF RFC 8827: WebRTC Security Architecture

[47] ISO/IEC 23090-6:2021: “Information technology — Coded representation of immersive media — Part 6: Immersive media metrics”

[48] 3GPP TR 26.926: “Traffic Models and Quality Evaluation Methods for Media and XR Services in 5G Systems”

[49] Oscar Falmer: “AR Headsets Landscape”, 10th September 2021, https://docs.google.com/spreadsheets/d/1aUO8nuXWCnL1xYnpe9tvSgCphifhMRLrFj1enayv0X8/edit#gid=0

[50] Oscar Falmer: “Mobile AR Features Landscape”, 20th September 2021, <https://docs.google.com/spreadsheets/d/1S1qEyDRCqH_UkcSS4xVQLgcMSEpIu_mPtfHjsN02GNw/edit#gid=0>

[51] Microsoft, Coordinate Systems, <https://docs.microsoft.com/en-us/windows/mixed-reality/design/coordinate-systems>

[52] Google WebRTC project update & Stadia review, see <https://www.youtube.com/watch?v=avtlQeaxd_I&t=438s>

[53] Joint MPEG/Khronos/3GPP Workshop on "Streamed Media in Immersive Scene Descriptions", September 29/30, 2021, <http://mpeg-sd.org/workshop.html>

[54] MPEG Systems Output WG3 N21042 "Report of Joint Workshop on Streamed Media in Immersive Scene Descriptions", MPEG#136, October 2021, https://www.mpegstandards.org/wp-content/uploads/mpeg\_meetings/136\_OnLine/w21042.zip

**===== CHANGE =====**

### 4.5.2 Updated KPIs for AR

In TR 26.928 [2], some high-level statements on experience KPIs for AR are provided. To achieve Presence in Augmented Reality, seamless integration of virtual content and physical environment is required. Like in VR, the virtual content has to align with user's expectations. For truly immersive AR and in particular MR, it is expected that users cannot discern virtual objects from real objects.

Also relevant for VR and AR, but in particular AR, is not only the awareness for the user for the environment. This includes, safe zone discovery, dynamic obstacle warning, geometric and semantic environment parsing, environmental lighting and world mapping.

Based on updated information, Table 4.6.2-1 provides new KPIs with focus on AR and in particular glasses. For some background and additional details refer for example to [10],[11], [49], [50] and [51].

Table 4.6.2-1 KPIs from TR 26.928 with focus on AR glasses

|  |  |
| --- | --- |
| **Feature** | **KPIs for AR glasses** |
| **Tracking** | |
| Freedom Tracking | 6DoF |
| Translational Tracking Accuracy | Sub-centimeter accuracy - tracking accuracy of less than a centimeter |
| Rotational Tracking Accuracy | Quarter-degree-accurate rotation tracking is desired |
| AR tracking space | In AR, the tracking space is theoretically unlimited. However, when moving, tracking accuracy may not be assured beyond a certain level of space or trajectory distance. SLAM based methods quickly introduce a large drift in large scale mapping. To correct the scaling issues, a loop closure technique [12] needs to be applied in order to continuously harmonize the local coordinate systems with global ones. |
| World-scale experience | World-scale experiences that let users wander beyond   * orientation-only or seated-scale experiences * standing-scale or room-scale experiences   To build a world-scale experience, techniques beyond those used for room-scale experiences, namely creating an absolute room-scale coordinate system that is continuously registered with the world coordinate system, typically requiring dynamic sensor-driven understanding of the world, continuously adjusting its knowledge over time of the user's surroundings. |
| Tracking frequency | At least 1000 Hz |
| **Latency (for more details refer to clause 4.5.3)** | |
| motion-to-photon latency | Less than 20 ms, and preferably even sub 10ms for AR as you may observe movement against the real world. |
| pose-to-render-to-photon latency | 50-60ms for render to photon is desired in order to avoid wrongly rendered content with late warping applied. |
| **Video Rendering and Display** | |
| Persistence – Duty time | Turn pixels on and off every 2 - 4 ms to avoid smearing / motion blur |
| Display refresh rate | 60 Hz minimum  90 Hz acceptable  120 Hz and beyond desired  240 Hz would allow always on display at 4ms |
| Colour | RGB colours  Accurate colours independent of viewpoint. |
| Spatial Resolution per eye | for 30 x 20 degrees  - 1.5K by 1K per eye is required  - 1.8K by 1.2K per eye is desired  for 40 x 40 degrees  - 2K by 2K required  - 2.5 K by 2.5 K desired  ultimate goal for display resolution is reaching or going slightly beyond the human vision limit of roughly one arcmin (1/60°) |
| Content frame rates | Preferably matching the display refresh rate for lowest latency  Lower frame rates for example 60 fps or 90 fps may be used but add to overall end to end delay. |
| Brightness | 200-500 nits for indoor  Up to 2K for state-of-the-art devices in 2021 [49]  10K to 100K nits for full outdoor experiences |
| **Optics** | |
| Field of View | Augmentable FoV   * typically, 30 by 20 degrees FoV acceptable * 40 by 40 degrees desired   maximize the non-obscured field of view |
| Eye Relief | the minimum and maximum eye-lens distance wherein a comfortable image can be viewed through the lenses.  at least 10mm, ideally rather 20mm |
| Calibration | correction for distortion and chromatic aberration that exactly matches the lens characteristics |
| Depth Perception | Avoid vergence and accommodation conflict (VAC) for accommodation being different for the real and virtual object |
| **Physics** | |
| Maximum Available Power | AR Glass: below 1 W, typically 500mW  For less design-oriented devices, additional power may be available. |
| Maximum Weight | AR Glass: around 70g. However, if the weight is well distributed, several hundred grams may be acceptable. |

**===== CHANGE =====**

### 4.5.3 Typical Latencies in networked AR Services

Building on top of the architectures introduced in clause 4.2 in this document as well as the latency considerations in TR 26.928 [2], Figure 4.5.3-1 provides a summary of different latencies involved networked AR services. Based on TR 26.928 as well as Table 4.5.2-1, two relevant latency requirements for adequate user experience matter:

1. motion-to-photon latency being less 20ms, but preferably even single digit latency below 10ms.
2. pose-to-render-to-photon latency: as small as 50-60ms

It is important to note that the motion-to-photon latency is primarily a function of the device implementation as it is basically covered within the AR run time. What matters and is relevant is the time used to provide the pose information from the AR run time to the renderer and the renderer using this pose to generate the displayed media. Final pose correction to the latest pose can always be done in the AR run time.

Figure 4.5.3-1 provides different latency critical uplink and downlink operations, depending on where the rendering is done, locally, in the edge or in the cloud. If done in the edge or cloud, rendered data needs to be delivered in low-latency and high-quality over the network. The typical operations in this case include:

* pose detection in the UE
* sending the pose through a 5G uplink network to the edge of cloud.
* Rendering the scene in the edge or cloud
* Compressing and encrypting the rendered scene and delivering to the UE
* Decrypting and decompressing the rendered scene
* Composition of the scene
* Applying the latest pose in the pose correction and display the immersive media.



Figure 4.5.3-1: Typical Latencies in networked AR services

It is ultimately relevant that in case of networking the rendering loop, the processes in the loop are executed such that the end-to-end latency requirements for the pose-to-render-to-photon latency can be met. Obviously the “closer” the rendering happens at the AR UE, the easier it is to meet latency requirements. However, with proper support of 5G system and media functionalities, these networked AR challenges can be solved. This is subject of the remaining discussion of this report.

With reference to TR 26.928 [2], other types of latencies impact the user experience, for example when used for cloud gaming, interactive scenes or in case of real-time network-based processing of sensor data. These aspects are not specific to AR but are also relevant. Some more details are provided in clause 6 for the different scenarios.