

ETSI TR 126 998 V17.1.0 (2022-10)



TECHNICAL REPORT

LTE;
5G;

**Support of 5G glass-type Augmented Reality / Mixed Reality
(AR/MR) devices**

(3GPP TR 26.998 version 17.1.0 Release 17)

<https://standards.iteh.ai/catalog/standards/sist/94d969f7-f7e0-481d-95cb-26ccdb7990b8/etsi-tr-126-998-v17-1-0-2022-10>



Reference

RTR/TSGS-0426998vh10

Keywords

5G,LTE

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:

<http://www.etsi.org/standards-search>

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status.

Information on the current status of this and other ETSI documents is available at

<https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx>

If you find errors in the present document, please send your comment to one of the following services:

<https://standards.etsi.org/People/CommitteeSupportStaff.aspx>

If you find a security vulnerability in the present document, please report it through our

Coordinated Vulnerability Disclosure Program:

<https://www.etsi.org/standards/coordinated-vulnerability-disclosure>

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.

No recommendation as to products and services or vendors is made or should be implied.

No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.

In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.

The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2022.
All rights reserved.

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<https://ipr.etsi.org/>).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM®** and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under <http://webapp.etsi.org/key/queryform.asp>.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the [ETSI Drafting Rules](#) (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Intellectual Property Rights	2
Legal Notice	2
Modal verbs terminology.....	2
Foreword.....	6
Introduction	6
1 Scope	7
2 References	7
3 Definitions, symbols and abbreviations	10
3.1 Definitions	10
3.2 Symbols.....	11
3.3 Abbreviations	11
4 Introduction to Glass-type AR/MR Devices	13
4.1 General	13
4.2 Device Functional Architecture.....	14
4.2.1 Device Functions	14
4.2.2 Generic reference device functional structure device types.....	16
4.2.2.1 Overview.....	16
4.2.2.2 Type 1: 5G STandalone AR (STAR) UE.....	17
4.2.2.3 Type 2: 5G EDGe-Dependent AR (EDGAR) UE.....	18
4.2.2.4 Type 3: 5G WireLess Tethered AR UE	19
4.2.3 AR Runtime.....	21
4.2.4 Scene Manager.....	22
4.2.5 XR Spatial Computing.....	22
4.2.6 5G Media Access Function.....	24
4.3 Basic Processes for delivering an AR experience	25
4.3.1 Introduction.....	25
4.3.2 AR Media Delivery Pipeline.....	26
4.3.3 XR Spatial Compute Pipeline	28
4.4 AR content formats and codecs.....	29
4.4.1 Overview	29
4.4.2 Scene Graph and Scene Description	30
4.4.3 Metadata	30
4.4.3.1 User pose information.....	30
4.4.3.2 Camera Parameters.....	31
4.4.4 Media Formats/Primitives in AR Scenes	31
4.4.5 Compression Formats	32
4.4.5.1 Elementary stream.....	32
4.4.5.2 Storage and Delivery Formats.....	33
4.4.6 Multiple Media Decoders management and coordination	33
4.4.7 XR Spatial Description	34
4.4.7.1 Overview	34
4.4.7.2 Camera and sensor information.....	35
4.4.7.3 Visual features, Keyframes and Spatial Maps.....	35
4.4.7.4 Spatial Anchors and Trackables.....	37
4.5 Key Performance Indicators and Metrics for AR	37
4.5.1 Summary of TR 26.928	37
4.5.2 Updated KPIs for AR.....	39
4.5.3 Typical Latencies in networked AR Services	41
4.6 Related Work.....	42
4.6.1 3GPP.....	42
4.6.2 MPEG	43
4.6.3 ETSI Industry Specification Group	44
4.6.4 Work related to AR Runtime	46

4.6.4.1	OpenXR	46
4.6.4.2	WebXR	48
4.6.5	MPEG Scene Description	49
4.6.6	MPEG-I Video Decoding Interface for Immersive Media.....	50
4.6.7	MPEG-I Carriage of Point Cloud Compression Data	50
4.6.8	Web Real-Time Communication (WebRTC)	51
4.6.8.1	WebRTC as an OTT application	51
4.6.8.2	WebRTC framework for RTC Media Service Enablers.....	52
4.6.9	Joint workshop with Khronos and MPEG on "Streamed Media in Immersive Scene Descriptions"	52
5	Core Use Cases.....	53
6	Mapping to 5G System Architecture.....	54
6.1	General	54
6.2	Immersive media downlink streaming	55
6.2.1	Introduction.....	55
6.2.2	Relevant use cases	55
6.2.3	Architectures	56
6.2.3.1	STAR-based	56
6.2.3.2	EDGAR-based	56
6.2.4	Procedures and call flows	57
6.2.4.1	STAR-based media streaming.....	57
6.2.4.2	EDGAR-based media streaming	59
6.2.5	Content formats and codecs	60
6.2.6	KPIs and QoS	61
6.2.7	Standardization areas	62
6.3	Interactive immersive services	62
6.3.1	Introduction.....	62
6.3.2	Relevant use cases	62
6.3.3	Architectures	63
6.3.3.1	STAR-based	63
6.3.3.2	EDGAR-based	63
6.3.4	Procedures and call flows	63
6.3.4.1	STAR-based interactive immersive service	63
6.3.4.2	EDGAR-based interactive immersive service	66
6.3.5	Content formats and codecs	67
6.3.6	KPIs and QoS	67
6.3.7	Standardization areas	69
6.4	5G cognitive immersive service	69
6.4.1	Introduction.....	69
6.4.2	Relevant use cases	69
6.4.3	Architectures	70
6.4.3.1	STAR-based	70
6.4.3.2	EDGAR-based	70
6.4.4	Procedures and call flows	70
6.4.6	KPIs and QoS	73
6.5	AR conversational services	74
6.5.1	Introduction.....	74
6.5.2	Relevant use cases	75
6.5.3	Basic architecture and call flows	75
6.5.3.1	Symmetrical case	77
6.5.3.2	Asymmetrical case	77
6.5.4	Instantiation #1: MTSI-based architecture extension.....	78
6.5.5	Instantiation #2: DCMTSI-based architecture extension with immersive media processing	81
6.5.6	Content formats and codecs	84
6.5.7	Summary of AR conversational instantiations.....	84
6.5.8	Standardization areas	85
6.6	Shared AR conversational experience	85
6.6.1	Introduction.....	85
6.6.2	Relevant use cases	86
6.6.3	Basic architecture.....	86
6.6.4	Generic Call flow.....	87

6.6.5	Various instantiations	89
6.6.6	Content formats and codecs	90
6.6.7	Standardization areas	90
7	Considerations on Devices Form-factor	90
7.1	General	90
7.2	Battery/Power consumption	91
7.3	Camera	91
7.4	Display	91
7.5	Heat dissipation	92
7.6	Weight	92
8	Potential New Work and Study Area	93
8.1	General	93
8.2	5G Generic Architectures for Real-Time Media Delivery	93
8.3	5G-Media Service Enablers.....	94
8.4	5G Real-time Communication.....	95
8.5	Media Capabilities for Augmented Reality Glasses	95
8.6	Split Rendering Media Service Enabler with AR profile	96
8.7	Tethering AR Glasses.....	97
8.8	IMS-based AR conversational services	97
8.9	Audio Media Pipelines for AR Experiences.....	98
9	Conclusions	98
Annex A:	Collection of Glass-type AR/MR Use Cases	100
A.1	Use Case 16: AR remote cooperation	100
A.2	Use Case 17: AR remote advertising	102
A.3	Use Case 18: Streaming of volumetric video for glass-type MR devices	104
A.3.1	Use case description	104
A.3.2	Call flow for STAR UE.....	107
A.3.3	Call flow for EDGAR UE	109
A.4	Use Case 19: AR Conferencing.....	111
3.2.1	AR Conferencing (1:1)	111
3.2.1	AR Conferencing (1:many).....	111
A.5	Use Case 20: AR IoT control	112
A.6	Use Case 21: AR gaming	114
A.7	Use Case 22: Shared AR Conferencing Experience.....	116
Annex B:	Change history	118
History	119

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
 - 1 presented to TSG for information;
 - 2 presented to TSG for approval;
 - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Augmented Reality (AR) and Mixed Reality (MR) promise to provide new experiences for immersive media services. The form factors of the devices for these services are typically not expected to deviate significantly from those of typical glasses, resulting in less physical space for the various required components such as sensors, circuit boards, antennas, cameras, and batteries, when compared with typical smartphones. Such physical limitations also reduce the media processing and communication capabilities that may be supported by AR/MR devices, in some cases requiring the devices to offload certain processing functions to a tethered device and/or a server.

This report addresses the integration of such new devices into 5G system networks and identifies potential needs for specifications to support AR glasses and AR/MR experiences in 5G.

The focus of this document is on general system aspects, especially targeting visual rendering on glasses, and may not be equally balanced or equally precise on all media types (e.g. on haptics, GPUs, audio). For example, extrapolations on architectural aspects derived for primarily visual media pipelines to audio pipelines may require confirmation based on further study.

1 Scope

The present document collects information on glass-type AR/MR devices in the context of 5G radio and network services. The primary scope of this Technical Report is the documentation of the following aspects:

- Providing formal definitions for the functional structures of AR glasses, including their capabilities and constraints,
- Documenting core use cases for AR services over 5G and defining relevant processing functions and reference architectures,
- Identifying media exchange formats and profiles relevant to the core use cases,
- Identifying necessary content delivery transport protocols and capability exchange mechanisms, as well as suitable 5G system functionalities (including device, edge, and network) and required QoS (including radio access and core network technologies),
- Identifying key performance indicators and quality of experience factors,
- Identifying relevant radio and system parameters (required bitrates, latencies, loss rates, range, etc.) to support the identified AR use cases and the required QoE,
- Providing a detailed overall power analysis for media AR related processing and communication.

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 23090-13:2022 DIS: "Information technology — Coded representation of immersive media — Part 13: Video Decoding Interface for Immersive Media"
- [7] Microsoft Azure Kinect™ DK documentation, <https://docs.microsoft.com/en-us/azure/kinect-dk/>
- [8] Google ARCore™: Use Depth in your Android app, <https://developers.google.com/ar/develop/java/depth/developer-guide>
- [9] Microsoft Azure Kinect™ 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 TR 38.838: "Study on XR (Extended Reality) 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 23090-14:2022: "Information technology — Coded representation of immersive media — Part 14: Scene Description for MPEG-I Media"
- [24] ISO/IEC 23090-10:2021: "Information technology — Coded representation of immersive media — Part 10: Carriage of Visual Volumetric Video-based Coding Data"
- [25] ISO/IEC 23090-18:2021: "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 23.706: “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, 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
- [55] 3GPP TS 23.501: "System architecture for the 5G System (5GS)"
- [56] 3GPP TS 26.260: "Objective test methodologies for the evaluation of immersive audio systems"
- [57] 3GPP TS 26.261: "Terminal audio quality performance requirements for immersive audio services"
- [58] Younes, Georges, et al. "Keyframe-based monocular SLAM: design, survey, and future directions." *Robotics and Autonomous Systems* 98 (2017): 67-88. <https://arxiv.org/abs/1607.00470>
- [59] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". <https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>
- [60] Herber Bay et. al., "SURF: Speeded Up Robust Features". <https://people.ee.ethz.ch/~surf/eccv06.pdf>
- [61] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011 International Conference on Computer Vision, 2011, pp. 2564-2571, doi: 10.1109/ICCV.2011.6126544.
- [62] ETSI GS ISG ARF 004 v1.1.1 (2020-08): "Augmented Reality Framework (ARF) Interoperability Requirements for AR components, systems and services"
- [63] Microsoft Flight Simulator, <https://www.xboxachievements.com/news/news-34054-microsoft-flight-simulators-world-is-two-million-gigabytes-in-size.html>

3 Definitions, symbols and abbreviations

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].

5G AR/MR media service enabler: A 5G AR/MR media service enabler is supporting an AR/MR application to provide AR/MR experience using at least partially 5G System tools.

5G System (Uu): Modem and system functionalities to support 5G-based delivery using the Uu radio interface.

AR/MR Application: a software application that integrates audio-visual content into the user's real-world environment.

AR/MR content: AR/MR content consists of a scene with typically one or more AR objects and is agnostic to a specific service.

AR Data: Data generated by the AR Runtime that is accessible through API by an AR/MR application such as pose information, sensors outputs, and camera outputs.

AR Media Delivery Pipeline: pipeline for accessing AR scenes and related media over the network.

AR/MR object: An AR/MR object provides a component of an AR scene agnostic to a renderer capability.

AR Runtime: a set of functions that interface with a platform to perform commonly required operations such as accessing controller/peripheral state, getting current and/or predicted tracking positions, general spatial computing, and submitting rendered frames to the display processing unit.

Lightweight Scene Manager: A scene manager that is capable to handle a limited set of 3D media and typically requires some form of pre-rendering in a network element such as the edge or cloud.

Media Access Function: A set of functions that enables access to media and other AR related data that is needed in the scene manager or AR Runtime in order to provide an AR experience.

NOTE: In the context of this report, the Media Access function typically uses 5G system functionalities to access media.

Peripherals: The collection of sensors, cameras, displays and other functionalities on the device that provide a physical connection to the environment.

Scene Manager: a set of functions that supports the application in arranging the logical and spatial representation of a multisensorial scene with support of the AR Runtime.

Simplified Entry Point: An entry point that is generated by 5G cloud/edge processes to support offloading processing workloads from UE by lowering the complexity of the AR/MR content.

Spatial Computing: AR functions which process sensor data to generate information about the world 3D space surrounding the AR user.

XR Spatial Description: a data structure describing the spatial organisation of the real world using anchors, trackables, camera parameters and visual features.

XR Spatial Compute Pipeline: pipeline that uses sensor data to provide an understanding of the physical space surrounding the device and uses XR Spatial Description information from the network.

XR Spatial Compute server: an edge or cloud server that provides spatial computing AR functions.

XR Spatial Description server: a cloud server for storing, updating and retrieving XR Spatial Description.

3.2 Symbols

Void

[ETSI TR 126 998 V17.1.0 \(2022-10\)](https://standards.iteh.ai/catalog/standards/sist/94d969f7-f7e0-481d-95cb-26ccdb7990b8/etsi-tr-126-998-v17-1-0-2022-10)

<https://standards.iteh.ai/catalog/standards/sist/94d969f7-f7e0-481d-95cb-26ccdb7990b8/etsi-tr-126-998-v17-1-0-2022-10>

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

5GMS	5G Media Streaming
AAC	Advanced Audio Coding
AF	Application Function
AGW	Access GateWay
API	Application Programming Interface
AR	Augmented Reality
ARF	Augmented Reality Framework
AS	Application Server
ATIAS	Terminal Audio quality performance and Test methods for Immersive Audio Services
ATW	Asynchronous Time Warp
AVC	Advanced Video Coding
BLE	Bluetooth Low Energy
BMFF	Based Media File Format
BoW	Bag-Of-visual-Words
CAD	Computer Aided Design
CGI	Computer Generated Imagery
CMAF	Common Media Application Format
CoM	Centre of Mass
CPU	Central Processing Unit
CSCF	Call Session Control Function

DASH	Dynamic Adaptive Streaming over HTTP
DC	Data Channel
DCMTSI	Data Channel Multimedia Telephony Service over IMS
DIS	Draft International Standard
DoF	Degree of Freedom
DRX	Discontinuous Reception
DTLS	Datagram Transport Layer Security
EAS	Edge Application Server
EDGAR	EDGE-Dependent AR
EEL	End-to-End Latency
eMMTEL	Evolution of IMS Multimedia Telephony Service
EMSA	Streaming Architecture extensions for Edge processing
ERP	Equirectangular Projection
EVC	Essential Video Coding
FDIS	Final Draft International Standard
FFS	For Future Study
FLUS	Framework for Live Uplink Streaming
FoV	Field of View
FPS	Frame Per Second
G-PCC	Geometry-based Point Cloud Compression
GBR	Guaranteed Bit Rate
gITF	Graphics Library Transmission Format
GPS	Global Positioning System
GPU	Graphics Processing Unit
HDCA	High Density Camera Array
HEVC	High Efficiency Video Coding
HLS	HTTP Live Streaming
HMD	Head-Mounted Display
HOA	Higher-Order Ambisonics
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ICE	Interactive Connectivity Establishment
IMS	IP Multimedia Subsystem
IMU	Inertial Measurement Unit
ISG	Industry Specification Group
ISOBMFF	ISO Based Media File Format
ITT4RT	Immersive Teleconferencing and Telepresence for Remote Terminals
IVAS	Immersive Voice and Audio Services
JPEG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LI	Lawful Interception
LIDAR	LIght Detection And Ranging
LSR	Late Stage Reprojection
MAF	Media Access Function
MBS	Multicast and Broadcast Services
MIoT	Mobile Internet of Things
MMT	MPEG Media Transport
MPD	Media Presentation Description
MPR	Maximum Power Reduction
MR	Mixed Reality
MRF	Multimedia Resource Function
MSH	Media Session Handler
MTSI	Multimedia Telephony Service over IMS
NAT	Network Address Translation
NBMP	Network Based Media Processing
OHMD	Optical Head Mount Display
OMAF	Omnidirectional MediA Format
OpenGL	Open Graphics Library
OTT	Over-The-Top
PBR	Physically-Based Rendering
PCC	Point Cloud Compression

PCF	Policy Control Function
PDU	Protocol Data Unit
PLY	PoLYgon file format
PRACK	Provisional Response Acknowledgement
RAN	Radio Access Network
RFC	Request For Comments
RP	Reference Point
RTC	Real-Time Communication
RTP	Real-time Transport Protocol
SDK	Software Development Kit
SDP	Session Description Protocol
SIFT	Scale Invariant Feature Transform
SID	Study Item Description
SIP	Session Initiation Protocol
SLAM	Simultaneous Localization And Mapping
SRTCP	Secure Real-time Transport Control Protocol
SRTP	Secure Real-time Transport Protocol
SSE	Server-Sent Events
STAR	STandalone AR
STUN	Session Traversal Utilities for NAT
SURF	Speeded Up Robust Features
TCP	Transmission Control Protocol
TLS	Transport Layer Security
ToF	Time of Flight
TPU	Tensor Processing Unit
TURN	Traversal Using Relays around NAT
UE	User Equipment
USB	Universal Serial Bus
V3C	Visual Volumetric Video-based Coding
V-PCC	Video-based Point Cloud Compression
VDE	Video Decoding Engine
VDI	Video Decoding Interface
VPU	Vision Processing Unit
VVC	Versatile Video Coding
WebRTC	Web Real-Time Communication
WLAR	WireLess tethered AR
WTAR	Wired Tethered AR
XHR	XML HTTP Request
XMPP	eXtensible Messaging and Presence Protocol
XR	eXtended Reality

4 Introduction to Glass-type AR/MR Devices

4.1 General

After a series of feasibility studies and normative work on Virtual Reality (VR), the feasibility study on eXtended Reality (XR) in 5G (FS_XR5G), documented in TR 26.928 [2], analysed the frameworks for eXtended Reality (XR) as a conceptual umbrella for representing the virtual, augmented, and mixed realities in a consistent fashion. This study, 5G glass-type Augmented Reality (AR) / Mixed Reality (MR) classified the XR devices into different form factors, including AR glasses, also referred to as optical head-mounted displays (OHMDs) and pointed out their power and tethering issues with the sidelink. A key aspect of this study is to identify the details of AR glasses including the capabilities for communication, processing, media handling, and offloading of power consumption.

Augmented reality composites virtual objects with reality. The compositing is a combination of light from real world and light presented on display to make them visible together to a user's eyes. Challenges in AR include the prediction of the real-world image visible to the human eye, and the accurate positioning and presentation of the virtual image on the display of AR glasses. To display a virtual image in front of a user's eyes, an optical see-through display is installed between the real world and the user's eyes, typically and most conveniently done with AR glasses.

In order to track the real-world space in which to place the virtual objects, sensors and in particular cameras are required to capture a live-action image visible to the human eye. Typically, multiple sensors and cameras are needed to construct a three-dimensional geometry around the user, called the user's geometry. The perception of geometry and the mapping of AR glass in geometry is referred to Simultaneous Localization and Mapping (SLAM), also introduced in some details in TR 26.928.

When AR objects are placed in the user's geometry, these objects are anchored to a part of the real-world geometry. With users moving, maintaining augmentation and consistency between reality and the user's geometry is challenging and requires continuous data flows and processing. In order to support devices with low power consumption, split processing such as split rendering or split perception are technologies to offload processing to powerful network servers. Such split processing approaches are considered beneficial or even essential for satisfying AR experiences, but add new necessary processes and challenges. This encoding, transmission, decoding, correction of rendering data and sensor/camera data over 5G networks, altogether pose challenges on bitrates, reliability requirements and latencies.

Based on the findings in clause 8 of TR 26.928, this clause follows up on some parts of the conclusions and proposed short term actions:

- Develop a flexible XR centric device reference architecture as well as a collection of device requirements and recommendations for XR device classes based on the considerations in clause 7.2 of TR 26.928.
- Study detailed functionalities and requirements for glass-type AR/MR UEs with standalone capabilities according to clause 7.6 of TR 26.928 and address exchange formats for AR centric media, taking into account different processing capabilities of AR devices.

Three different types of device reference architectures are identified in this clause. One major distinction among these types is the device capabilities of whether stand-alone processing of required AR media processing (in clause 4.2.2.2) or having dependencies on an entity in charge of offloading of power consuming processes, which the entity may be a cloud/edge service (in clause 4.2.2.3) and 5G wireless connectivity (in clause 4.2.2.4).

For the detailed functionalities for the device reference architecture of AR glasses, AR runtime (in clause 4.2.3) is identified for AR/MR system capability discovery, AR/MR session management, tracking of surrounding area, and rendering of AR/MR content in scene graph. Scene manager (in clause 4.2.4) is able to process a scene graph and render the corresponding 3D scene. 5G Media Access Function (in clause 4.2.4) is identified to support AR UE and the scene manager to access and stream components of AR content (in clause 4.4).

AR content consists of one or more AR objects in terms of primitives (in clause 4.4.4) and their spatial and temporal compositions described by a scene description (in clause 4.4.2). Processing of AR/MR functions may require additional metadata (in clause 4.4.3) to properly recognize user's pose and surrounding area.

Key performance indicators and metrics for AR/MR based on TR 26.928 are provided (in clause 4.5) and related works (in clause 4.6) on AR/MR in 3GPP, MPEG, and ETSI are identified for the considerations on collaborative work on device function architecture and AR content formats and codecs.

4.2 Device Functional Architecture

4.2.1 Device Functions

AR glasses contain various functions that are used to support a variety of different AR services as highlighted by the different use cases in clause 5. AR devices share some common functionalities to create AR/XR experiences. Figure 4.2.1-1 provides a basic overview of the relevant functions of an AR device.

The primary defined functions are

- **AR/MR Application:** a software application that integrates audio-visual content into the user's real-world environment.
- **AR Runtime:** a set of functions that interface with a platform to perform commonly required operations such as accessing controller/peripheral state, getting current and/or predicted tracking positions, general spatial computing, and submitting rendered frames to the display processing unit.
- **Media Access Function:** A set of functions that enables access to media and other AR related data that is needed in the scene manager or AR Runtime in order to provide an AR experience. In the context of this report, the Media Access function typically uses 5G system functionalities to access media.