ETSI TR 126 928 V16.1.0 (2021-01)

5G;
Extended Reality (XR) in 5G
(3GPP TR 26.928 version 16.1.0 Release 16)

Reference
RTR/TSGS-0426928vg10

Keywords
5G

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

Copyright Notification

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 IPR Policy, no investigation, 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.

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

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

This Technical Report collects information on eXtended Reality (XR) in the context of 5G radio and network services. Extended reality (XR) refers to all real-and-virtual combined environments and associated human-machine interactions generated by computer technology and wearables. It includes representative forms such as augmented reality (AR), mixed reality (MR), and virtual reality (VR) and the areas interpolated among them. In this Technical Report, baseline technologies for XR type of services and applications are introduced outlining the QoE/QoS issues of XR-based services, the delivery of XR in the 5G system, and an architectural model of 5G media streaming defined in TS 26.501. In addition to the conventional service category, interactive, streaming, download, and split compute/rendering are identified as new delivery categories. A survey of 3D, XR visual and audio formats is also provided.

Use cases and device types are classified, and processing and media centric architectures are introduced. This includes viewport independent and dependent streaming, as well as different distributed computing architectures for XR. Core use cases of XR include those unique to AR and MR in addition to those of VR discussed in 3GPP TR 26.918, ranging from offline sharing of 3D objects, real-time sharing, multimedia streaming, online gaming, mission critical applications, and multi-party call/conferences. Based on the details in the report, proposals for potential standardisation areas are documented.

1 Scope

The present document collects information on eXtended Reality (XR) in the context of 5G radio and network services. The primary scope of the present document is the documentation of the following aspects:

- Introducing Extended Reality by providing definitions, core technology enablers, a summary of devices and form factors, as well as ongoing related work in 3GPP and elsewhere,
- Collecting and documenting core use cases in the context of Extended Reality,
- Identifying relevant client and network architectures, APIs and media processing functions that support XR use cases,
- Analysing and identifying the media formats (including audio and video), metadata, accessibility features, interfaces and delivery procedures between client and network required to offer such an experience,
- Collecting key performance indicators and Quality-of-Experience metrics for relevant XR services and the applied technology components,
- Drawing conclusions on the potential needs for standardisation in 3GPP.

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 26.918: "Virtual Reality (VR) media services over 3GPP".

[3] 3GPP TS 26.118: "3GPP Virtual reality profiles for streaming applications".

[4] ARCore, https://developers.google.com/ar/

[5] ARKit, https://developer.apple.com/arkit/

[6] 3GPP TR 22.842: "Study on Network Controlled Interactive Service in 5GS".

[7] 3GPP TR 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[8] 3GPP TS 23.501: "System Architecture for the 5G System".

[9] Schuemie, Martijn J., Peter Van Der Straaten, Merel Krijn, and Charles A.P.G. Van Der Mast. "Research on Presence in Virtual Reality: A Survey." CyberPsychology & Behavior, Vol. 4, No. 2. April 2001.

[10] Ching, Teo Choong. "The Concept of Presence in Virtual Reality." Medium. 27 August 2016.

[11] Sparks, Matt. "Don't Break the Spell: Creating Presence in Virtual Reality" Learning Solutions Magazine, 17 July 2017.

[12] 3GPP TS 26.501: "5G Media Streaming Architecture".

[13] 3GPP TS 22.173: "IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1".

[14] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction".

[15] 3GPP TR 22.891 "Feasibility Study on New Services and Markets Technology".

[16] Khronos, "The OpenXR Specification", Jan 25, 2020, https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html#introduction

[17] W3C, "WebXR Device API", https://www.w3.org/TR/webxr/

[18] Rolland, Jannick & Holloway, Richard & Fuchs, Henry. (1994). Comparison of optical and video see-through, head-mounted displays. Proceedings of SPIE - The International Society for Optical Engineering. 10.1117/12.197322.

[19] "Cloud Gaming: Architecture and Performance", Ryan Shea and Jiangchuan Liu, Simon Fraser University; Edith C.-H. Ngai, Uppsala University; Yong Cui, Tsinghua University; IEEE Network-July/August 2013.

[20] M. Claypool and K. Claypool. Latency and player actions in online games. Communications of the ACM, 49(11):40–45, 2006.

[21] Quax, P., Monsieurs, P., Lamotte, W., De Vleeschauwer, D., and Degrande, N. Objective and subjective evaluation of the influence of small amounts of delay and jitter on a recent first person shooter game. In Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games (New York, NY, USA, 2004), NetGames '04, ACM, pp. 152–156.

[22] Chen, K.-t., Huang, P., Wang, G.-s., Huang, C.-y., and Lei, C.-l. On the Sensitivity of Online Game Playing Time to Network QoS. Proceedings of IEEE INFOCOM 2006 00, c (2006).

[23] 3GPP TS 23.501: "System architecture for the 5G System (5GS)".

[24] 3GPP TS 38.300: "NR; Overall description; Stage-2".

[25] 3GPP TS 22.173: "IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1".

[26] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[27] 3GPP TR 23.758: "Study on application architecture for enabling Edge Applications"

[28] 3GPP TR 23.748: "Study on enhancement of support for Edge Computing in 5G Core network (5GC)".

[29] 3GPP TS 23.558: "Architecture for enabling Edge Applications (EA)".

[30] Recommendation ITU-T H.264 (04/2017): "Advanced video coding for generic audiovisual services" | ISO/IEC 14496-10:2014: "Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding".

[31] Recommendation ITU-T H.265 (12/2016): "High efficiency video coding" | ISO/IEC 23008-2:2015: "High Efficiency Coding and Media Delivery in Heterogeneous Environments – Part 2: High Efficiency Video Coding".

[32] 3GPP TS 26.116: "Television (TV) over 3GPP services; Video profiles". [33] Jens-Rainer Ohm, Gary J. Sullivan, Heiko Schwarz, Thiow Keng Tan, and Thomas Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 22, NO. 12, DECEMBER 2012.

[34] T.K. Tan, M. Mrak, R. Weerakkody, N. Ramzan, V. Baroncini, G.J. Sullivan, J.-R. Ohm, K.D. McCann, "HEVC subjective video quality test results", IBC2014 Conference, 2014.

[35] Thiow Keng Tan ; Rajitha Weerakkody ; Marta Mrak ; Naeem Ramzan ; Vittorio Baroncini, Jens-Rainer Ohm, Gary J. Sullivan, "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance" IEEE Transactions on Circuits and Systems for Video Technology, Volume: 26 , Issue: 1 , Jan. 2016.

[37] ISO/IEC 23090-2: "Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format"

[38] S. Schwarz et al., "Emerging MPEG Standards for Point Cloud Compression," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133-148, March 2019.

[39] Khronos, "The GL Transmission Format (glTF)", Jun 9, 2017, https://github.com/KhronosGroup/glTF/blob/master/specification/2.0/README.md

[40] Long Qian, Alexander Barthel, Alex Johnson, Greg Osgood, Peter Kazanzides, Nassir Navab, and Bernhard Fuerst, "Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display", Int J Comput Assist Radiol Surg. 2017 Jun; 12(6): 901–910, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891507/.

[41] 3GPP TS 22.261: "Service requirements for the 5G system".

[42] 3GPP, VR-IF and the Advanced Imaging Society's 2ND VR ECOSYSTEMS & STANDARDS WORKSHOP, Culver City, CA, US, https://www.vr-if.org/events/3gpp-vrif-ais-workshop/.

3 Definitions of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms 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].

3.2 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].

3DoF	Three Degrees of Freedom
5QI	5G QoS Identifier
6DoF	Three Degrees of Freedom
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
ARP	Allocation and Retention Priority
ASIC	Application-Specific Integrated Circuit
ASTC	Adaptive Scalable Texture Compression
ATW	Asynchronous TimeWarp
AVC	Advanced Video Coding
BC1	Block Compression for RGB
CAD	Computer-Aided Design
CBR	Constant BitRate
CDN	Content Delivery Network
CPU	Compute Processing Unit
CTC	Call for TeChnologies
DASH	Dynamic Adaptive Streaming over HTTP
DL	DownLink
DNS	Domain Name System
DoF	Degrees of Freedom
EAC	Ericsson Alpha Compression

ERP	Equi-Rectangular Projection
ETC2	Ericsson Texture Compression version 2
EVC	Essential Video Coding
FFS	For Further Study
FLUS	Framework for Live Uplink Streaming
FOV	Field-Of-View
FPS	Frames Per Second
GBR	Guaranteed BitRate
GFBR	Guaranteed Flow Bit Rate
GNSS	Global Navigation Satellite System
G-PCC	Geometry-based Point Cloud Compression
GPS	Global Positioning System
GPU	Graphics Processing Unit
HEVC	High-Efficiency Video Coding
HMD	Head-Mounted Display
HRTF	Head-Related Transfer Function
HTTP	Hyper-Text Transfer Protocol
HUD	Heads-Up Display
IDMS	Inter-destination Multimedia Synchronization
IMU	Inertial Measurement Unit
IOD	Inter-aural Output Difference
IVAS	Immersive Voice and Audio Services
JPEG	Joint Photographic Experts Group
JVET	Joint Video Exploration Team
LIDAR	Light Detection and Ranging
MCPTT	Mission Critical Push To Talk
MCU	Multipoint Control Unit
MEC	Multi-access Edge Computing
MFBR	Maximum Flow Bit Rate
MMS	Multimedia Messaging Service
MOBA	Multiplayer Online Battle Arena
MPEG	Moving Pictures Expert Group
MR	Mixed Reality
NBMP	Network-Based Media Processing
NCIS	Network Controlled Interactive Service
NEF	Network Exposure Function
PBR	Physically-Based Rendering
PCC	Point Cloud Compression
PCF	Policy Control Function
PDB	Packet Delay Budget
PDU	Packet Data Unit
PER	Packet Error Rate
PLY	PoLYgon
PNG	Portable Network Graphics
PPI	Pixels Per Inch
PQI	PC5 QoS Identifier
PSS	Packet-Switched Streaming
PTT	Push To Talk
PVRTC	PowerVR Texture Compression
QCI	QoS Class Identifier
QFI	QoS Flow ID
QoE	Quality of EXperience
QoS	Quality of Service
RCS	Rich Communication Service
RGB	Red-Green-Blue colour space
RGBD	Red-Green-Blue-Depth
RPG	Role Playing Game
RQA	Reflective QoS Attribute
RTP	Real-Time Protocol
RTS	Real-time Strategy
RTT	Round Trip Time
SCS	Spatial Compute Server

SDP	Session Description Protocol
SIP	Session Initiation Protocol
SLAM	Simultaneous Localization and Mapping
SWB	Super WideBand
TCP	Transmission Control Protocol
ToF	Time of Flight
TPU	Tensor Processing Unit
UL	UpLink
USB	Universal Serial Bus
VCL	Video Coding Layer
V-PCC	Video-based Point Cloud Compression
VPS	Visual Positioning System
VR	Virtual Reality
VVC	Versatile Video Coding
XR	Extended reality
YUV	Luminance-Bandwidth-Chrominance

4 Introduction to Extended Reality

4.1 XR Terms and Definitions

4.1.1 Different Types of Realities

The scope of this clause is the introduction of eXtended Reality (XR) to 3GPP services and networks. eXtended Reality (XR) is an umbrella term for different types of realities as shown in Figure 4.1-1. The figure also shows different application domains of XR such as entertainment, healthcare, education, etc. The different terms are defined in the following, reusing and extending some definitions from 3GPP TR26.918 [2].

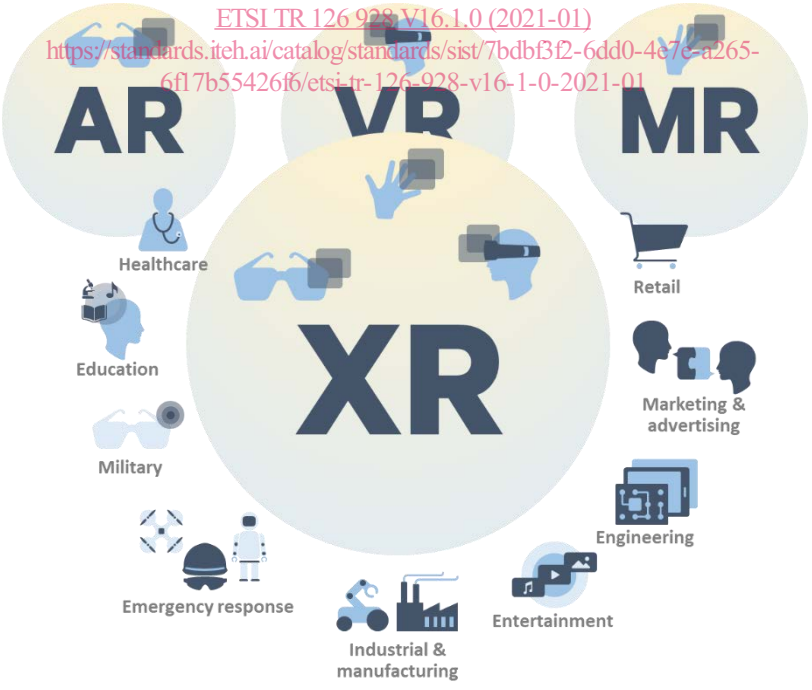

Figure 4.1-1: Different Types of Realities and some applications

Virtual reality (VR) is a rendered version of a delivered visual and audio scene. The rendering is designed to mimic the visual and audio sensory stimuli of the real world as naturally as possible to an observer or user as they move within the limits defined by the application. Virtual reality usually, but not necessarily, requires a user to wear a head mounted display (HMD), to completely replace the user's field of view with a simulated visual component, and to wear headphones, to provide the user with the accompanying audio. Some form of head and motion tracking of the user in

VR is usually also necessary to allow the simulated visual and audio components to be updated in order to ensure that, from the user's perspective, items and sound sources remain consistent with the user's movements. Additional means to interact with the virtual reality simulation may be provided but are not strictly necessary.

Augmented reality (AR) is when a user is provided with additional information or artificially generated items or content overlaid upon their current environment. Such additional information or content will usually be visual and/or audible and their observation of their current environment may be direct, with no intermediate sensing, processing and rendering, or indirect, where their perception of their environment is relayed via sensors and may be enhanced or processed.

Mixed reality (MR) is an advanced form of AR where some virtual elements are inserted into the physical scene with the intent to provide the illusion that these elements are part of the real scene.

Extended reality (XR) refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. It includes representative forms such as AR, MR and VR and the areas interpolated among them. The levels of virtuality range from partially sensory inputs to fully immersive VR. A key aspect of XR is the extension of human experiences especially relating to the senses of existence (represented by VR) and the acquisition of cognition (represented by AR).

Other terms used in the context of XR are *Immersion* as the sense of being surrounded by the virtual environment as well as *Presence* providing the feeling of being physically and spatially located in the virtual environment. The sense of presence provides significant minimum performance requirements for different technologies such as tracking, latency, persistency, resolution and optics. For more details, refer to clause 4.2.

Other relevant terms in the context of XR experiences are:

- **Parallax** is the relative movement of objects as a result of a change in point of view. When objects move relative to each other, users tend to estimate their size and distance.
- **Occlusion** is the phenomena when one object in a 3D space is blocking another object from being viewed.

This document uses the acronym XR throughout to refer to equipment, applications and functions used for Virtual Reality, Augmented Reality, and other related technologies. Examples include, but are not limited to:

- Head-mounted displays for Virtual Reality,
- Optical see-through glasses and camera see-through HMDs for Augmented and Mixed Reality,
- Mobile devices with positional tracking and camera.

All in common with them is the ability that they offer some degree of spatial tracking and the spatial tracking results in an interaction to view some form of virtual content. More details on XR devices are provided in clause 4.8.

4.1.2 Degrees of Freedom and XR Spaces

A user acts in and interacts with extended realities as shown in Figure 4.1-2. Actions and interactions involve movements, gestures, body reactions. Thereby, the *Degrees of Freedom (DoF)* describe the number of independent parameters used to define movement of a viewport in the 3D space.

Any consistent interaction for an XR application with XR hardware is assumed to be restricted to an XR session. Once an XR session has been successfully established, it can be used to poll the viewer pose, query information about the user’s environment, and present imagery to the user.

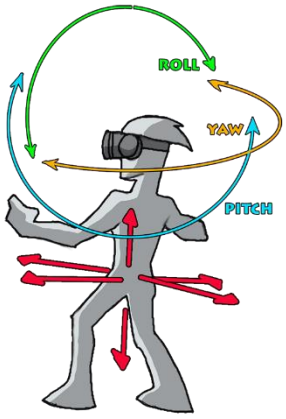

Figure 4.1-2: Different degrees of freedom for a user in extended realities

Typically, the following different types of Degrees-of-Freedom are described (and also shown in Figure 4.1-3).

- **3DoF**: Three rotational and un-limited movements around the X, Y and Z axes (respectively pitch, yaw and roll). A typical use case is a user sitting in a chair looking at 3D 360 VR content on an HMD (see Figure 4.1-3 (a)).
- **3DoF+**: 3DoF with additional limited translational movements (typically, head movements) along X, Y and Z axes. A typical use case is a user sitting in a chair looking at 3D 360 VR content on an HMD with the capability to slightly move his head up/down, left/right and forward/backward (see Figure 4.1-3 (b)).
- **6DoF**: 3DoF with full translational movements along X, Y and Z axes. Beyond the 3DoF experience, it adds (i) moving up and down (elevating/heaving), (ii) moving left and right (strafing/swaying); and (iii) moving forward and backward (walking/surging). A typical use case is a user freely walking through 3D 360 VR content (physically or via dedicated user input means) displayed on an HMD (see Figure 4.1-3 (d)).
- **Constrained 6DoF**: 6DoF with constrained translational movements along X, Y and Z axes (typically, a couple of steps walking distance). A typical use case is a user freely walking through VR content (physically or via dedicated user input means) displayed on an HMD but within a constrained walking area (see Figure 4.1-3 (c)).