ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R M.2528-0

(09/2023)

M Series: Mobile, radiodetermination, amateur and related satellite services

Capabilities of the terrestrial component of IMT-2020 for multimedia communications

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at https://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	**Mobile, radiodetermination, amateur and related satellite services**
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

Note*: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2023

REPORT ITU-R M.2528-0

Capabilities of the terrestrial component of IMT-2020 for multimedia communications

(Question ITU-R 262/5)

(2023)

TABLE OF CONTENTS

Page

1 Scope

This Report addresses the capabilities of IMT-2020 to distribute multimedia content such as video, audio, text and graphics, including support for real-time multimedia interactive applications. This Report also addresses the capabilities of IMT-2020 user devices and base stations to support such multimedia communications with low latency and wide transmission bandwidth.

This new Report complements Report ITU-R M.2373 – Audio-visual capabilities and applications supported by terrestrial IMT systems, which addresses the capabilities of IMT systems for delivering audio-visual services to the consumers and also covers some aspects of production of audio-visual content.

NOTE – The objective of this Report for IMT-2020 for multimedia is to address the aspects of the delivery of multimedia content and to be fully consistent with the preamble of the ITU Constitution.

2 Introduction

Multimedia applications include network video, digital magazine, digital newspaper, digital radio, social media, mobile TV, digital TV, touch media, etc., that are enabled by IMT-2020 technologies. Beyond the traditional media service, the new media application not only supports accurate delivery of content, but also supports real-time interaction and real-time uploading of user-generated content. The users can be both consumers and producers of new media content.

These applications for multimedia content include but are not limited to:

– audio-visual applications,
– network video applications,
– digital online magazine applications,
– digital online newspaper applications,
– internet radio applications,
– social media applications,
– mobile internet TV applications,
– touch media applications,
– online information distribution applications,
– on-demand video applications,
– imaging and audio distribution applications,
– content dissemination applications,
– file delivery application,

– real time uploading of multimedia content,
– electronic classroom presentation technology,
– full motion video conferencing.

This Report covers the utilization of IMT-2020 technology to support the specific applications mentioned above. For details of applications of the Broadcasting service for multimedia, please refer to the list of ITU-R Recommendations and ITU-R Reports in § 3.

3 Relevant ITU-R Recommendations and Reports

Recommendation ITU-R BT.1833 – Broadcasting of multimedia and data applications for mobile reception by handheld receivers

Recommendation ITU-R BT.2016 – Error-correction, data framing, modulation and emission methods for terrestrial multimedia broadcasting for mobile reception using handheld receivers in VHF/UHF bands

Recommendation ITU-R M.2083 – Framework and overall objectives of the future development of IMT for 2020 and beyond

Recommendation ITU-R M.2150 – Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2020 (IMT-2020)

Report ITU-R BT.2049 – Broadcasting of multimedia and data applications for mobile reception

Report ITU-R BT.2295 – Digital terrestrial broadcasting systems

Report ITU-R M.2373 – Audio-visual capabilities and applications supported by terrestrial IMT systems

4 Acronyms

2K resolution	Content having 2000 pixels in horizontal resolution
4K Resolution	3840 wide × 2160 tall pixels = 8.30 megapixels
8K Resolution	7680 wide × 4320 tall pixels = 33.20 megapixels
ACK/NACK	Used in data transmission
AMF	Access and mobility management function
AR	Augmented reality
CA	Carrier aggregation
CLI	Cross-link interference
CLI-RSSI	CLI received signal strength indicator
CN	Core network
CSI	Channel status information
DC	Dual connectivity
DSS	Dynamic spectrum sharing
E1	Interface between gNB-CU-CP and gNB-CU-UP
eMBB	Enhanced mobile broadband
eNB	Evolved Node B – Radio Base Station used in 4G LTE
F1	Interface interface is between gNB-CU and gNB-DU
FDD TDD CA	Carrier aggregation between TDD and FDD bands

FR2	Frequency band above 24.250 GHz
GBR	Guaranteed bit rate
gNB	Next Generation Node B – Radio Base Station used in 5G NR
GPU	Graphic processing unit
HARQ	Hybrid automatic repeat request
HD	1020 tall × 1920 wide pixels = 2 million pixels
HPHT	High power high tower
HSDPA	High speed downlink packet access
HSUPA	High speed uplink packet access
IoT	Internet of Things
IPTV	Internet Protocol Television
L1/L2	Control signalling
LAN-VN	Local area network-virtual network
LTE	Long term evolution
MBMS	Multimedia broadcast multicast service
MCG	Master call group
MIMO	Multiple input and multiple output
mMTC	Massive machine type communications
MPMT	Medium power medium tower
MR	Mixed reality – VR + AR
MR-DC	Multi-radio dual connectivity
NG-RAN	New generation radio access network
NPN	Non-public network (private network)
NR	New radio
O-FDMA	Orthogonal frequency division multiple access
PAPR	Peak to average power ratio
PCF	Policy control function
PDSCH	Physical downlink shared channel
PDU	Protocol data unit
PLMN	Public land mobile network
PRACH	Preamble random access channel
PTM	Point to multipoint
PTP	Point to point
PUCCH	Physical uplink control channel
QFI	QoS Flow ID
QR	Code A type of bar code

RAN	Radio access network
RIM	Remote interference management in 5G
RIT	Radio interface technology
RRC	Radio resource control
SC-FDMA	Single carrier-frequency division multiple access
SFN	Single frequency network
SMF	Session management function
SRIT	Set of radio interface technology
SRS	Sounding reference signal
SRS-RSRP	SRS-Reference signal received power
TRP	Transmission and reception points
UE	User equipment
UHD 5K	Resolution 5000 pixels wide
UMTS	Universal Mobile Telecommunications Service
UPF	Use plane function
URLLC	Ultra-reliable low latency communications
UTRAN	UMTS terrestrial radio access network
V2X	Vehicle to everything
VR	Virtual reality
XRM	External resource management

5 Trends and demands of applications for multimedia content supported by IMT-2020 technologies

By the end of 2025, it is estimated that video traffic will represent 69 percent of all mobile data traffic, a share that is forecast to increase to 79 percent by 2027[1]. Multimedia is an effective method of communicating information because it enriches presentations, retains the audience’s attention, and allows multiple and flexible interaction. And it[2] is also estimated that the global media market over cellular networks will go up to USD 420 billion in 2028 (USD 124 billion), a CAGR of 9.8 percent over ten years. Consumer[2] spending on video, music, and mobile games will nearly double by 2028 to reach nearly USD 150 billion globally (USD 29 billion). Many Multimedia applications are already in use and many more such applications are being developed quite rapidly. The trend to develop latest

[1] *Source*: Ericsson Mobility Report, June 2022. https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf

For additional information refer to the interactive graph at Ericsson Mobility Visualizer - Mobility Report - Ericsson

[2] *Source*: OVUM “How 5G will transform the business of media and entertainment”, October 2018 https://newsroom.intel.com/wp-content/uploads/sites/11/2018/10/ovum%E2%80%93intel%E2%80%935g%E2%80%93ebook.pdf

multimedia application towards eXtended Reality (XR), high-definition video, real time multimedia interaction and real-time uploading of self-produced content.

Virtual reality (VR) being capable of developing immersive and interactive video content, now becoming a key technology to upgrade or even subvert traditional media industries. Its application fields include video games, event live streaming, video entertainment, healthcare, real estate, retail, education, engineering and public safety.

The augmented reality (AR) which is enhanced version of VR is also being developed very rapidly and to create a new way for people to connect with media through virtual items, virtual characters, and augmented contextual information. The hardware and the ecosystem are gradually maturing.

Mixed reality (MR)also referred as hybrid reality is the merging of AR and VR worlds to produce new environments and visualizations where physical and digital objects can co-exist and interact in real time.

XR is a newly added term and defined as an umbrella for all the three realities namely VR, AR and MR. XR refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearable devices.

With the rapid development of IMT-2020 and AR, new businesses including B2B and B2C based on mobile applications have become the important exploration areas. The main serving enterprise-level users are automobiles, machinery manufacturing, real estate, TV stations, publishing houses and exhibition halls. From the perspective of application scenarios, commodity advertising and early childhood education may be the first popular scenarios at present.

Video distribution supported by IMT-2020 is evolving to higher resolution e.g. 4K resolution, whether it is on-demand, live streaming, or video surveillance, thereby improving the user experience on information transmission and image recognition.

The interactive user experience – selection of loading videos, switching channels and other viewing operations through the user device while receiving multimedia content – is another developing trend. Users expect operations to complete virtually instantaneously. In order to meet user expectations, initial video loading should be completed within 1s and the channel switching within 500 ms[3].

Considering the complex visual object perception and processing of VR, it is a trend to move the processing functions to the cloud, then the video streams and action commands are delivered via IMT-2020 wireless technologies to VR terminal. Cloud VR utilizes the powerful processing power of the cloud to reduce the weight and size of the VR terminal and improve user experience, which helps to reduce the complexity and cost of the terminal and greatly help to promote scale development. The cloud VR application scenarios require a large user experienced data rate of around 100 Mbit/s to ensure a high-definition video experience (> 2K), and a network delay of 5~10 ms (MTP delay of 20 ms) to eliminate user motion sickness effects.

With the continuous development of mobile augmented reality, users have increasingly higher requirements for augmented reality application experience: smooth presentation, real-time interaction, and persistent operation, which pose challenges to the computing power and media processing capabilities of the mobile terminal. Efficient invoking the hardware capabilities of mobile terminals, quick identification and capture of augmented reality targets in different business execution environments, and superimposing the smooth display of the augmented reality content of various media types in real time, all greatly affect the user experience. Cloud AR can balance the computing power demand and cost. In the future, with the help of the Cloud AR platform, greater computing

---

[3] https://www-file.huawei.com/-/media/CORPORATE/PDF/white%20paper/White_Paper_on_the_Experience_driven_4K_Bearer_Network.pdf, Tables 2-3 and 2-4.

power required by rendering videos or pictures can be provided with lower cost. It is predicted that the Cloud AR is the future direction of AR development, which provides increased computing power while reducing terminal costs. At the same time, the flexible deployment method of the Cloud AR platform will help to achieve low network delay which is very important for AR applications.

Further, co-working with edge cloud services, IMT-2020 can connect the user to a high-definition virtual world on their mobile device. Edge Cloud is an architecture that is used to decentralise (processing) power to the edges (clients/devices) of the networks. Traditionally, the computing power of servers are used to perform tasks such as data minimisation or to create advanced distributed systems. Within the cloud model, such intelligent tasks can be performed by servers, so that a normal mobile with less or almost no computing power can be used.

Live events with High Definition (HD) and Ultra-High Definition (UHD) can be streamed via an IMT-2020 radio network with higher throughput. HD and UHD TV content can be accessed on mobile devices without any interruptions through IMT-2020 higher user experienced data rate. The entertainment industry will hugely benefit from IMT-2020 wireless networks, which are expected to provide high resolution, and high dynamic range video streaming without interruption. Cloud AR and Cloud VR with HD or UHD video can be supported with higher user experienced data rate and low latency. HD virtual reality games are becoming popular while IMT-2020 network can offer a better real-time interactive gaming experience. It is expected that with the support of IMT-2020 technologies, an amazing virtual experience will bring to the users and the above mentioned multimedia services will become the basic services in the future mobile Internet.

6 Overview of the technical characteristics of IMT-2020 technologies for multimedia communications

6.1 3GPP 5G NR

Three primary 5G NR use cases defined by 3GPP are:

- Enhanced Mobile Broadband (eMBB): data-driven use cases requiring high data rates across a wide coverage area.
- Ultra-reliable Low Latency Communications (URLLC): strict requirements on latency and reliability for mission critical communications, such as remote surgery, autonomous vehicles or the Tactile Internet.
- Massive Machine Type Communication (mMTC): need to support a very large number of devices in a small area, which may only send data sporadically, such as Internet of Things (IoT) use cases.

In 2017-2018, 3GPP, via Release-15, for the first time introduced 5G specification (5G Phase-I) in three phases. The first phase focused primarily on mobile broadband for non-standalone (NSA) 5G architecture. The second phase introduced for standalone (SA) 5G architecture. The third phase introduced an architecture for migration from 4G to 5G. Its primary focus was enabling enhanced Broadband (eMBB), as offer very high uplink throughput, lower latency and/or higher capacity to support real-time multi-media information/interaction and real-time uploading of self-media content.

In 2019-2020, 3GPP, via Release-16 (5G Phase-II), further improved the features of 5G (Release 15) and incorporated Dynamic Spectrum Sharing (DSS), Network Slicing and other features designed for private 5G Network. These developments marked a new focus on enterprise and business centric capabilities. Other capabilities included New Radio (NR) based access to unlicensed spectrum (NR-U) and satellite access. The completion of Release 16 occurred in two stages, Stage 1in late 2019 for the physical layer aspects and Stage 2 in late 2020 for the higher layer aspects.

In 2019-2020, 3GPP Release 17 further improved the features of the 5G network including DSS and private 5G network capabilities and introduced significant enhancements in the Next-Generation Radio Access Network (NG-RAN) and 5G Core Network (5GCN).

With the above development of the IMT-2020 technologies, real-time multimedia content interaction and real-time uploading of user-generated content are supported.

Interference coordination is one of key characteristics of IMT-2020 technologies. If non-synchronized frame structures are deployed for public and non-public networks, interference between gNB and gNB or UE and UE need to be avoided as it would cause big impacts on the overall system performance. 3GPP Release-16 extends interference mitigation techniques to certain operational environments, such as those in non-synchronized TDD networks.

Carrier Aggregation (CA) and Dual Connectivity (DC) are the straightforward and effective ways to boost uplink throughput by directly increasing the uplink frequency resources and possibly cell capacity. With CA technology, a UE can receive or transmit on one or multiple contiguous/non-contiguous component carriers. The CA was firstly introduced in LTE Release 10. CA in IMT-2020 was specified from 3GPP Release 15 and support maximum 16 carriers with 400 MHz each, thus it can support up to 6.4 GHz bandwidth. It is continuously developed in 3GPP Release 16, Release 17 and so on to support more flexible schedule and interference mitigation. The Multi-Radio Dual connectivity (MR-DC) was introduced in 3GPP Release 15. It is comprised of LTE+NR DC, NR+NR DC, NR+LTE DC etc. The RATs utilized in the first node and second node can be different. It is also continuously enhanced in 3GPP Release 15, Release 17 and so on to improve radio resource efficiency and reduce latency.

UL MIMO is another technology to improve up link data rate. It was specified for IMT-2020 from 3GPP Release 15. It introduces CSI feedback and reference signalling design which are more flexible than LTE. With high resolution codebook and beam forming characteristics, the data throughput and capacity of IMT-2020 system are expected to be met. In 3GPP Release 16 and Release 17, it continuously developed CSI feedback, Coordinated Multiple Points (CoMP), beam management, power control, PAPR reference, etc. to offer higher radio resource efficiency and communication performance for new media e.g. XR applications.

There are different types of methods for transporting content data, e.g. broadcast, multicast, and unicast.

Besides the primary 5G NR capabilities, the Multicast Broadcast Services (MBS), including "LTE-based 5G Broadcast"[4] and NR MBS, capabilities of 3GPP 5G-SRIT and 5G RIT MBS enables high-quality multimedia applications within a wide range of services and platforms providing a significant improvement in performance, quality of service and user experience. IMT-2020 specification supporting MBS has been developed step-by-step.

IMT-2020 technologies support group transmission for MBS delivery[5]. Resource efficient delivery of multicast/broadcast services is introduced for services of Point-To-Multipoint nature, such as for example public safety and mission critical services, V2X applications, IPTV, live video, software delivery over wireless and IoT applications.

For broadcast communication service, the same specific content data are provided simultaneously to all UEs in a geographical area. A broadcast communication service is delivered to the UEs using a

---

[4] "LTE-based 5G Broadcast" is the term utilized by 3GPP during the specification development phase. It has been included one of the three radio interfaces of IMT-2020, 3GPP 5G-SRIT, and it is adopted in Recommendation ITU-R M.2150-1 (02/2022) – Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2020 (IMT-2020).

[5] Sourced from 3GPP TR 21.917.

broadcast mode. In the broadcast mode, the UE can receive the content data without any requirement on the RRC state.

For multicast communication service, the same service and the same specific content data are provided simultaneously to a dedicated set of UEs. A multicast communication service is delivered to the UEs using a multicast session. A UE can receive a multicast communication service in connected state with mechanisms such as Point-To-Point (PTP) and/or Point-To-Multipoint (PTM) delivery. HARQ feedback/retransmission can be applied to both PTP and PTM transmission.

The "LTE-based 5G Broadcast" finished in 3GPP Release 16 is based on legacy LTE eMBMS that can be traced back to 3GPP Releases 8 and 9 which lay down the foundation for cellular broadcast. "LTE-based 5G Broadcast" is able to apply 100% of radio resources of one or more radio carriers for the delivery of broadcast/multicast content-with inter-site distances of up to 200 km.

The 3GPP-developed LTE based 5G Broadcast technology is also used as the radio part of the "LTE-based 5G terrestrial broadcast system" described in ETSI TS 103 720 V1.1.1, and included in Recommendation ITU-R BT.2016-3 as "System L".

The NR MBS in 3GPP Release 17[6] enables such a service over a specific geographic area which further enables a more efficient and effective delivery system for real-time and streaming multicast/broadcast content. NR MBS is able to carry IP multicast as well as Ethernet multicast packets with better QoS support using dynamic delivery mode switching.

6.2 DECT-2020 NR component RIT

DECT-2020 NR (Digital Enhanced Cordless Telecommunications 2020 New Radio), a radio interface technology, is fully specified by ETSI in 2021 (ETSI TS 103 636, Part 1 to Part 5). DECT-2020 NR fulfils requirements for IMT-2020 including URLLC and mMTC services. DECT-2020 NR technical performance capabilities are listed in Table 1.

TABLE 1

Technical performance capabilities of DECT- 2020 NR

Parameter	Direction	Value	Unit
User plane latency	Uplink and Downlink	0.11-0.96	ms
Control plane latency	Not applicable	2.1-16.83	ms
Reliability	Uplink or Downlink	> 99.999	%
Maximum bandwidth	Not applicable	221.184	MHz
Scalability	Not applicable	1.728-221.184	MHz

DECT-2020 NR is designed to provide a technology foundation for wireless applications such as wireless microphones which can be used in multimedia use cases. This radio technology is based on OFDM (Orthogonal Frequency Division Multiplexing) and targets high flexibility and scalability to support among others cordless telephony, audio streaming applications, professional audio applications and in general solutions for local area deployments for URLLC applications.

[6] Refer the 3GPP TR 21.917 clause 6.3.4.

7 Use cases and requirements

This section provides examples of recently emerged multimedia use cases and reflects the current trend of moving towards a more interactive and dynamic environments.

7.1 Ultra-high-definition multimedia content

The visual and audio applications are being developed towards Ultra-high-definition (UHD) multimedia content. For instance, in the concert, the 4K-based multi-screen and multi-camera can capture extremely vivid details such as subtle "micro-expression" changes, bringing a strong visual impact to the audience. Meanwhile, the HDR and wide colour gamut technology can perfectly present the effect of live lighting and stage design to the audience. In band performances, multi-camera allows the audience to follow their favourite players. For large concerts (such as orchestras), multi-channel audio transmission with Hi-Fi music quality with inaudible noise and distortion, and a flat frequency response within the human hearing range, can allow listeners to experience a truly immersive experience.

Huge user experienced data rate and the low transmission latency are typical transmission objectives from above scenarios. With the communication capability of IMT-2020, the above scenarios can be supported.

7.2 Virtual reality (VR) panoramic video

Virtual Reality panoramic video is developed and extended on the technology of 720-degree or 360-degree panorama. It converts static panoramic pictures into dynamic video images, and can be viewed at any angle of 360 degrees from left to right, up and down, so that we have a truly immersive feeling. VR Panoramic video is not a single static panoramic picture, but has depth of field, dynamic image and sound, etc., and also has sound and picture alignment, sound and picture synchronization. It can be applied to multiple multimedia applications, e.g. network video applications, digital online magazine applications, digital online newspaper applications, social media applications, on-demand video applications, mobile internet TV applications etc. Compared with the traditional 720-degree panorama, VR panoramic video has a huge leap in quality, quantity, form and content. But it requires very higher user experienced data rate, lower latency and higher availability and multiple information sources synchronization capabilities. With the support of IMT-2020 capable of high data rate and ultra-reliability and low latency, the VR panoramic video can be delivered to the user.

7.3 Augmented reality (AR)

Augmented reality (AR) allows the virtual world on the screen to be combined and interacted with the real-world scene through the calculation of the position and angle of the camera image and the addition of image analysis technology. AR can be utilized in multiple scenarios. For example, AR can help visualize construction projects. Computer-generated structural images can be superimposed onto a real partial view of the property prior to constructing a physical building on top of it. AR systems have been used as collaborative tools for design and planning in a build environment. For example, AR can be used to create augmented reality maps, buildings and data sources projected onto a desktop for collaborative viewing by built environment professionals. In educational settings, AR has been used to supplement standard curriculum. Text, graphics, video and audio can be overlaid into the student's real-time environment. Textbooks, flashcards, and other educational reading materials may contain embedded "markers" or triggers that, when scanned by AR devices, provide supplemental information to students presented in multimedia formats.

The combination display of virtual and real objects, motion tracking, network communication, fusion rendering and human-computer interaction are the key elements of AR which require very high use experienced data rate, higher accuracy positioning and tracing, and real-time interaction between human and computer. With the capabilities of IMT-2020 especially for uplink higher throughput capability, the AR is expected to be developed step by step.

7.4 Extended Reality (XR)

eXtended Reality (XR) is a term for different types of realities (VR, AR and MR) and refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables. XR could be seen as the "container" for the representative forms such as AR and VR and the various interpolated in-between areas. The boundary between the real and the simulated worlds becomes vague as users dive into an extended reality visually, acoustically or even haptically.

Practical applications of XR include retail (trying out something, e.g. a wearable or a cloth, before buying), remote working and teaching (the capabilities of connection through XR with team members, clients or students will fundamentally change the way of interaction and could be the only means of maintaining activities during pandemics), training and coaching (XR can provide training facilities of a principally new level, especially for the scenarios associated with challenges which cannot be easily modelled in a real world), real estate (renting or buying after a virtual visit), museums and exhibitions (virtual tours on-the-coach), and other, not yet addressed categories of human activity.

XR traffics have the characteristics of high throughput, low latency, and high reliability requirement, and the UE battery level may impact user's experience since the high throughput require the high power consumption in terminal side. So considering the limited radio resource and end-to-end QoS policy control from system perspective, the 5G System should be enhanced to support the trade-off among throughput, latency and reliability and device battery life.

7.5 Tactile and multi-modal communication services

Some advanced multimedia services may include more modalities besides video and audio stream, such as information from different sensors and tactile or emotion data for more immersing experience e.g. haptic data or sensor data.

The tactile and multi-modal communication service can be applied in multiple fields, e.g. industry, robotics and telepresence, virtual, augmented and extended reality, healthcare, road traffic, serious gaming, education, culture and smart grid. These services support applications enabling input from more than one sources and/or output to more than one destination to convey information more effectively.

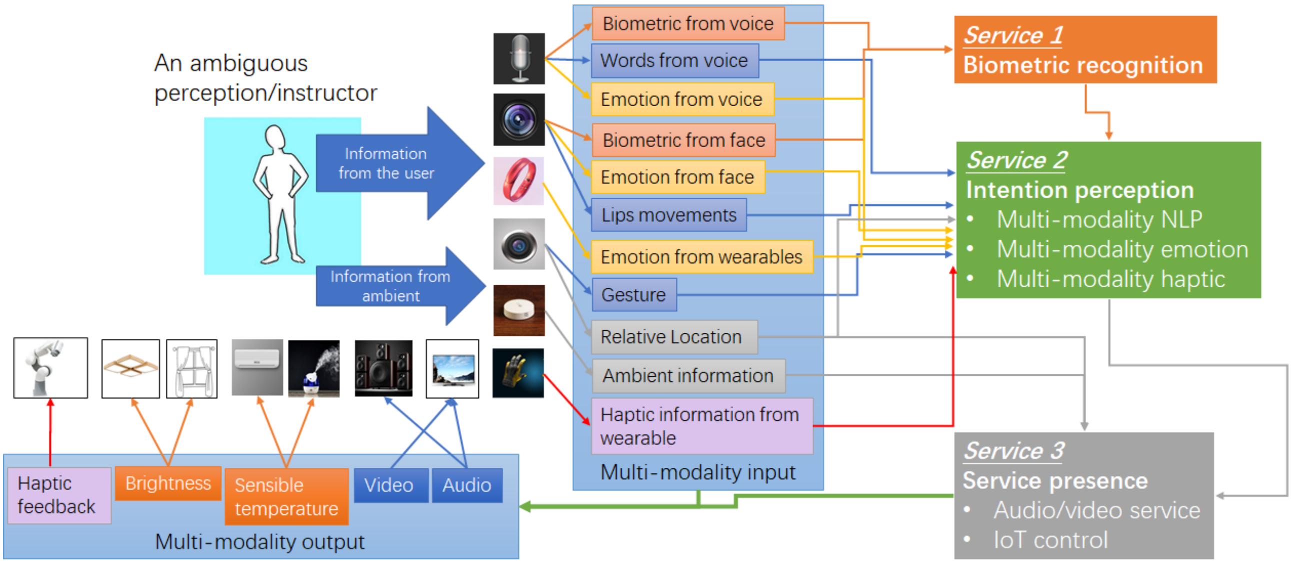
As Fig. 1[7] below illustrates, the input and output can be different modalities including:

– video/audio media;
– information received by sensors about the environment, e.g. brightness, temperature, humidity;
– 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).

[7] Sourced from 3GPP TS22.261 Section 6.4.3.

FIGURE 1

Multi-modal interactive system



To support such tactile and multi-modality communication services, the 5G system may need to address service requirement of different types of traffic steams with coordinated QoS selection and packet processing, guaranteed latency and reliability, time synchronization of these parallel information, in order to ensure best service experience.

7.6 Entertainment live streaming

Live streaming which is an important information distribution application has gradually become a very popular form of entertainment. The popularity of live streaming is no less than the media that traditional entertainment platforms rely on (such as TV or Radio). For instance, top tier live streamers can attract millions of viewers when they are streaming. However, unlike those produced programs, there is no fixed schedule for live streaming, which means the live streamers could start their streaming anytime, anywhere. This characteristic will trigger an explosive surge in online viewing within a geographic area or within a period of time, which brings great technical challenges to the network transmission of the streaming. Moreover, the way of live streaming is evolving as well. For instance, the streaming video is evolving to ultra-high definition and immersive, which requires the higher video bit rate and stricter latency tolerance.

IMT-2020 network empowered streaming can be of benefit to solve above issues. By integrating communication and information distribution storage and processing supported by IMT-2020, the mass access and the explosive traffic peak challenges caused by super-influencer across the country or region can be effectively addressed.

7.7 Live commerce

The term "Live commerce" refers to the use of live webcast technology to carry out new sales methods such as online display of merchandise, customer Q&A, and shopping guide sales through the Internet platform. Manufacturers usually use professional platforms to build their live streaming booth to sell a variety of products by streamers. Unlike traditional home shopping channel, live commerce could interact real-time with customers. Considering the characteristics of live commerce, the following interactions are needed:

– The live streamers can receive the feedback from the viewer and response further with supplemental product introduction, which help enhance the viewer's confidence in the quality of the product.

– The live streamers can use the VR/AR to provide product trials and related functions, allowing customers to “experience” the product.
– The viewer can order and pay for the products online in real-time through the payment methods of live streamers (Payment link/QR code, etc.).
– The live streamers can conduct online prize draws or mini-games, to improve viewer stickiness.

Therefore, for live commerce application, distributing low-latency and high-quality live video to a large number of viewers, and real-time interaction between the live streamer and the viewers are the key objectives for IMT-2020.

7.8 Smart venue

With the construction of IMT-2020 networks, the concept of “Smart Venue” that empower large venues with IMT-2020 networks and provide a new viewing experience to live events or activities has emerged. In top events, new viewing methods such as multi-view point and free-view point have been introduced, bringing a different experience to the website viewer and spectator. Taking “multi-view point” as an example, by deploying multiple cameras on stadium, content service providers can allow the viewer watch the game with different viewing angles from the regular streaming.

In addition to providing a diverse experience for viewers, another core concept of “Smart Venue” is to provide spectator with live services. For instance, during a game or event, the spectator can choose a different perspective from the current seat to watch the game via the client. During the intermission, the spectator could watch the replay of highlights or behind-the-scenes.

Generally, a large event often has tens or even hundreds of thousands of spectators, and millions of viewers. Such a dense population leads to huge challenges for communication network coverage, communication latency and connectivity capacity.

IMT-2020 empowered streaming could provide multimedia distribution service to users in or out of the certain area e.g. the venue, which will effectively overcome the wired network transmission bandwidth limitation and greatly reduce the cost of venue network construction. Meanwhile, the venue owners or event organizers will be able to leverage 5G empowered streaming to bring viewers (or spectators) an unprecedented experience.

7.9 Live multimedia streaming and distribution

The production of traditional large sports events is mainly implemented by sportscast production systems and equipment such as sportscast vans with wired communication. However, ultra-high-definition sports video has gradually become the mainstream, and the content of live events is becoming more and more complex with a large number of cameras, microphones and sensors working at the same time, a large amount of data needs to be transmitted back to the sportscast production control centre. In addition, as events become more globalized, sportscast resources need to be quickly collected across the world in a very short period of time. The event live streaming production is gradually developing towards centralized, remote and lightweight.

IMT-2020 has the characteristics of low latency, high bandwidth and wide coverage, which could improve the sportscast production capability, optimize resource allocation, and efficiently distribute real-time content from multi-viewing point, leaderboard and other content generated on the event to viewers. In particular, the high bandwidth and low latency transmission of IMT-2020 enables the data of some independent camera (such as Video Adjudication System) to be collected and delivered on time. For cameras with special mobility (such as Spidercam), IMT-2020 will expand its moving range, help it working more flexible and efficient.

Specifically, the following aspects could be enhanced:

– Short-distance transmission from multi-camera to live streaming system: multiple cameras (such as Steadicam, panoramic VR, AR hardware, etc.) are simultaneously transmitted to the sportscast van through IMT-2020 for live streaming production. The Single-channel high-definition video (mainly 1920 × 1080i) is encoded with a bit rate of 15 Mbit/s, and the uplink bit rate of the transmission channel will not be lower than this value. The relative delay between multiple cameras (same signal types) is less than 40 ms to meet synchronization requirements.
– Lightweight deployment of live streaming equipment: IMT-2020 has extremely low delay for both DL&UL transmission, multiple signals (such as video/audio/data) can be transmitted to the remote production centre through wireless transmission. Therefore, the transmission of control, tally, and voice signals can ensure extremely low delay in both uplink and downlink. These devices only need to be connected to IMT-2020 network, whereas in the past they required thousands of cables to connect to each other, which avoids the complicated deployment process.
– Multi-content streaming services: Through high speed streaming, various content (such as multi-camera, VR, Box score) can be distributed to a large number of spectators in real time, allowing them to fully grasp the details of the event and provide a better viewing experience.

7.10 Interactive Broadcast to Connected Vehicle

The Connected Vehicle ecosystem has evolved from early emergency services and basic telematics, to 5G-enabled autonomous driving applications and Cellular Vehicle to Everything (C-V2X) advanced safety use cases. In ETSI/3GPP standards[8], Connected Vehicles are also referred to as "Intelligent Transport Systems", and two connectivity modes are defined: Vehicle-to-Vehicle (V2V) over the PC5 interface and Vehicle-to-Network (V2N) over the Uu interface. The V2N transmission over the Uu interface is designed for multimedia usage[9], while the V2V over the PC5 interface is not. In summary, V2X communication enables a vehicle to interact with other vehicles (V2V), pedestrians (V2P), infrastructure (V2I), and general networked services (V2N) with cellular connection. By providing broadcast capability, new ecosystem opportunities can be enabled.

This use case would benefit from close coordination between mobile network and the 5G NR based broadcast network. Broadcast network can be used to push information that will enhance features such as improved safety, traffic congestion awareness, less environmental impacts, and lower capital expenditure. The broadcasted information can also be combined with information exchanges among vehicles to improve situational awareness thus avoiding accidents. A return path via the mobile network can be used to provide real-time traffic, sensor, and high-resolution mapping data. The combination of broadcast data and returned data can be useful not only for today's drivers but can also potentially help future autonomous vehicles. In addition, potential services, such as infotainment, software/firmware update, and advanced emergency information (AEI), should be considered.

7.11 Live audio production

Programme Making and Special Events (PMSE) includes wireless applications that are used for live professional audio/video productions such as concerts, musicals, or other staging of entertainment,

---

[8] 3GPP TS23.287 version 17.6.0, "Architecture enhancements for 5G System (5GS) to support Vehicle-to-Everything (V2X) services", March 2023.

[9] 3GPP TS23.247 version 17.6.0, "Architectural enhancements for 5G multicast-broadcast services", March 2023.

meetings, conferences, cultural and education activities, trade fairs, sport, religious, and other public or private events. PMSE is defined as follows:

– Programme Making: the creation of content for broadcast, the production of films, presentations, advertisements, audio or video recordings; and the staging or performance of an entertainment, sporting, social or other public/private event.

– Special Events: occurrences of limited duration, typically from one day to several weeks or longer, which take place in specifically defined locations.

Audio PMSE applications require low latency audio transmission in an indoor or outdoor service area. During a live event, audio is captured through a wireless microphone and processed. Live performers rely on receiving a personalized audio mix of the event streamed back to their in-ear monitoring device. An in-ear monitoring system transmits and receives an audio stream ranging between 200 kbit/s and 5 Mbit/s. Professional performers tolerate a maximum round-trip latency of 4 ms[10]. This round-trip latency is the time between capturing the audio by their wireless microphones (audio input) and replay by their in-ear monitoring devices (audio output).

Currently, the audio PMSE industry utilize analogue and digital technologies employing narrowband modulation techniques with each link typically occupying a bandwidth of e.g. 200 kHz[11]. IMT-2020 capabilities for reliability, latency, synchronicity and spectral efficiency need to support the requirements of professional audio transmission to cater for increased demand for capacity in audio PMSE applications.

8 Capabilities of multimedia communications supported by IMT-2020 technologies

8.1 Capabilities of multimedia communications

Capabilities of multimedia communications are introduced based on the 3GPP TS22.261[12], 3GPP TS22.263[13] and 3GPP TR38.913[14]. They include not only broadcast, multicast communication methods, but also unicast to flexible and efficient support UHD, AR/VR, live streaming production and distribution, live commerce, and smart venue, etc.

8.1.1 IMT-2020 capabilities for real-time multimedia interaction and media content uploading

Interactive multimedia allows the user to control, combine and manipulate a variety of media types, such as text, computer graphics audio and video materials, animation and virtual reality.

To enable interactive task completion during voice conversation, IMT-2020 is capable of supporting low-delay speech coding for interactive conversational services (refer 3GPP TS22.261, 100 ms, one-way mouth-to-ear).

---

[10] 3GPP TR22.827 – Study on Audio-Visual Service Production

[11] ETSI EN 300 422-1, V2.2.1 – Wireless Microphones; Audio PMSE up to 3 GHz, Part 1: Audio PMSE Equipment up to 3 GHz; Harmonised Standard for access to radio spectrum

[12] 3GPP TS22.261 V18.6.1 (2022-06) – Service requirements for the 5G system

[13] 3GPP TS22.263 V17.4.0 (2021-06) – Service requirements for video, imaging and audio for professional applications (VIAPA)

[14] 3GPP TR38.913 V17.0.0 (2022-04) – Study on scenarios and requirements for next generation access technologies

Table 2 gives a capability example of IMT-2020 to support high data rate and traffic density scenario for an interactive audio and video application in indoor hotspot.

TABLE 2[15]

High data rate and traffic density scenario of IMT-2020 for an interactive audio and video application

Scenario	Experienced data rate (DL)	Experienced data rate (UL)	Area traffic capacity (DL)	Area traffic capacity (UL)	Overall user density	Activity factor	UE speed	Coverage
Indoor hotspot	1 Gbit/s	500 Mbit/s	15 Tbit/s/km^2	2 Tbit/s/km^2	250 000/km^2	(1)	Pedestrians	Office and residential (1)

(1) A certain traffic mix is assumed; only some users use services that require the highest data rates.

Table 3 provides capabilities of IMT-2020 to support AR/VR high data rate and low latency usages.

TABLE 3[16]

Capabilities of IMT-2020 to support AR/VR high data rate and low latency service

Use Cases	Characteristic parameter (KPI)			Influence quantity		
	Max allowed end-to-end latency	Service bit rate: user-experienced data rate	Reliability	# of UEs	UE speed	Service area (2)
Cloud/Edge/Split Rendering (1)	5 ms (i.e. UL+DL between UE and the interface to data network) (4)	0.1 to [1] Gbit/s supporting visual content (e.g. VR based or high definition video) with 4K, 8K resolution and up to 120 frames per second content.	99.99 % in uplink and 99.9 % in downlink (4)	–	Stationary or Pedestrian	Countrywide
Gaming or interactive data exchanging (3)	10 ms (4)	0.1 to [1] Gbit/s supporting visual content (e.g. VR based or high definition video) with 4K, 8K resolution and up to 120 frames per second content.	99.99% (4)	≤ [10]	Stationary or Pedestrian	20 m × 10 m; in one vehicle (up to 120 km/h) and in one train (up to 500 km/h)
Consumption of VR content via tethered VR headset (6)	[5 to 10] ms (5)	0.1 to [10] Gbit/s (5)	[99.99%]	–	Stationary or Pedestrian	–

15 Sourced from 3GPP TS22.261 Table 7.1-1.

16 Sourced from 3GPP TS22.261 Table 7.6.1-1

Notes relatives to Table 3:

(1) Unless otherwise specified, all communication via wireless link is between UEs and network node (UE to network node and/or network node to UE) rather than direct wireless links (UE to UE).

(2) Length × width (× height).

(3) Communication includes direct wireless links (UE to UE).

(4) Latency and reliability KPIs can vary based on specific use case/architecture, e.g. for cloud/edge/split rendering, and can be represented by a range of values.

(5) The decoding capability in the VR headset and the encoding/decoding complexity/time of the stream will set the required bit rate and latency over the direct wireless link between the tethered VR headset and its connected UE, bit rate from 100 Mbit/s to [10] Gbit/s and latency from 5 ms to 10 ms.

(6) The performance requirement is valid for the direct wireless link between the tethered VR headset and its connected UE.

NOTE – The [] means that it is better to be supported but not has to.

The capabilities of IMT-2020 to support professional low-latency periodic deterministic audio transport service which are used for Video and audio production applications are listed in Table 4.

TABLE 4[17]

Capabilities of IMT-2020 to support professional low-latency periodic deterministic audio transport service

Profile	**No. of active UEs**	**UE speed (km/h)**	**Service area (m)**	**E2E latency (1) (µs)**	**Transfer interval(1) (µs)**	**Packet error rate (2), (3)**	**Data rate UL**	**Data rate DL (Mbit/s)**
Music Festival	200	10	500 × 500	750	250	10^{-6}	500 kbit/s	-
	100	10	500 × 500	750	250	10^{-6}	-	1 Mbit/s
Musical	30	50	50 × 50	750	250	10^{-6}	500 kbit/s	-
	20	50	50 × 50	750	250	10^{-6}	-	1 Mbit/s
	10	-	50 × 50	750	250	10^{-6}	-	500 kbit/s
Semi-professional	10	5	5 × 5	750	250	10^{-6}	100 kbit/s	-
	10	5	5 × 5	750	250	10^{-6}	-	200 kbit/s
	2	-	5 × 5	750	250	10^{-6}	-	100 kbit/s

[17] Sourced from 3GPP TS22.263 Table 6.2.1-1.

TABLE 4 (*end*)

Profile	No. of active UEs	UE speed (km/h)	Service area (m)	E2E latency (1) (µs)	Transfer interval(1) (µs)	Packet error rate (2), (3)	Data rate UL	Data rate DL (Mbit/s)
AV production	20	5	30 × 30	750	250	10^{-6}	1.5 Mbit/s	-
	10	5	30 × 30	750	250	10^{-6}	-	3 Mbit/s
Audio Studio	30	-	10 × 10	750	250	10^{-6}	5 Mbit/s	-
	10	5	10 × 10	750	250	10^{-6}	-	1 Mbit/s

(1) Transfer interval refers to periodicity of the packet transfers. It has to be constant during the whole operation. The value given in the table is a typical one, however other transfer intervals are possible as long as the end-to-end latency is ≤ (1 ms – Transfer interval).

(2) Packet error rate is related to a packet size of (transfer interval × data rate). Packets that do not conform with the end-to-end latency are also accounted as error.

(3) The given requirement for a packet error rate assumes a uniform error distribution. The requirement for packet error rate is stricter if packet errors occur in bursts.

The capabilities of IMT-2020 to support low latency Video applications are listed in Table 5.

TABLE 5[18]

Capabilities of IMT-2020 to support low latency video

Profile	# of active UEs	UE speed (km/h)	Service area	E2E latency	Packet error rate (1)	Data rate UL	Data rate DL (Mbit/s)
Uncompressed UHD video	1	0	1 km^2	400 ms	10^{-10} UL 10^{-7} DL	12 Gbit/s	20
Uncompressed HD video	1	0	1 km^2	400 ms	10^{-9} UL 10^{-7} DL	3 .2 Gbit/s	20
Mezzanine compression UHD video	5	0	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	3 Gbit/s	20
Mezzanine compression HD video	5	0	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	1 Gbit/s	20
Tier one events UHD	5	0	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	500 Mbit/s	20
Tier one events HD	5	0	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	200 Mbit/s	20
Tier two events UHD	5	7	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	100 Mbit/s	20

18 Sourced from 3GPP TS22.263 Table 6.2.1-3.

TABLE 5 (*end*)

Profile	# of active UEs	UE speed (km/h)	Service area	E2E latency	Packet error rate (1)	Data rate UL	Data rate DL (Mbit/s)
Tier two events HD	5	7	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	80 Mbit/s	2
Tier three events UHD (2)	5	200	1 000 m^2	1 s	10^{-7} UL 10^{-7} DL	20 Mbit/s	10
Tier three events HD (2)	5	200	1 000 m^2	1 s	10^{-7} UL 10^{-7} DL	10 Mbit/s	10
Remote OB	5	7	1 000 m^2	6 ms	10^{-8} UL 10^{-7} DL	200 Mbit/s	20

(1) Packets that do not conform with the end-to-end latency are also accounted as error. The packet error rate requirement is calculated considering 1500 B packets, and 1 packet error per hour is $10^{-5}/(3*x)$, where x is the data rate in Mbit/s.

(2) Could use either professional equipment or mobile phone equipped with dedicated news gathering app.

The capabilities of IMT-2020 to support UHD streaming via airborne base stations are listed in Table 6.

TABLE 6[19]

Capabilities of IMT-2020 to support UHD via airborne base stations

Profile	No. of active UEs	UE speed (km/h)	Service area	E2E latency (ms)	Packet error rate (1)	Data rate UL (Mbit/s)	Data rate DL (Mbit/s)
NPN ground to air UHD up Link	10	500	700 km^2 × 6 000 m (2)	40	10^{-8} UL 10^{-7} DL	100	20
NPN ground to air HD up link	10	500	700 km^2 × 6 000 m (2)	40	10^{-8} UL 10^{-7} DL	80	20
NPN air to ground UHD down Link	2	500	700 km^2 × 6 000 m (2)	40	10^{-7} UL 10^{-8} DL	20	100
NPN air to ground HD down link	2	500	700 km^2 × 6 000 m (2)	40	10^{-7} UL 10^{-8} DL	20	80
NPN radio Camera UHD	10	200	1 km^2	3	10^{-8} UL 10^{-7} DL	100	20

19 Sourced from TS22.261 Table 6.2.1-4.

TABLE 6 (*end*)

Profile	No. of active UEs	UE speed (km/h)	Service area	E2E latency (ms)	Packet error rate (1)	Data rate UL (Mbit/s)	Data rate DL (Mbit/s)
NPN radio camera HD	10	200	1 km^2	3	10^{-8} UL 10^{-7} DL	80	20

(1) Packets that do not conform with the end-to-end latency are also accounted as error. The packet error rate requirement is calculated considering 1 500 B packets, and 1 packet error per hour is $10^{-5}/(3*x)$, where x is the data rate in Mbit/s.

(2) 6 000 m = height but in a cone formation (i.e. ground coverage with a circle of diameter 30 km).

Further, with development of immersive multiple modal integrated with AR/VR, IMT-2020 are capable of supporting the communication KPIs listed in Table 7, which will greatly enrich and enhance user experience of the immersive multimedia service.

TABLE 7[20]

Capabilities of IMT-2020 to support immersive multimedia communication service

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)	UE speed	Service area	
Immersive multi-modal VR (UL: device → application sever)	5 ms (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 can be supported by the haptic device	Stationary or Pedestrian	typically < 100 km^2 (4)	Haptic feedback (3)
	5 ms	< 1 Mbit/s	99.99%	1500	Stationary or Pedestrian	typically < 100 km^2 (4)	Sensing information e.g. position and view information generated by the VR glasses

[20] Sourced from 3GPP TS22.261 Table 7.11-1.

TABLE 7 (*continued*)

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)**	**UE speed**	**Service area**	
Immersive multi-modal VR (DL: application sever → device)	10 ms (1)	1-100 Mbit/s	99.9%	1500	Stationary or Pedestrian	typically < 100 km^2 (4)	Video
	10 ms	5-512 kbit/s	99.9%	50	Stationary or Pedestrian	typically < 100 km^2 (4)	Audio
	5 ms (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	typically < 100 km^2 (4)	Haptic feedback (3)
Immersive multi-modal navigation applications Remote Site → Local Site (DL)	50 ms	16 kbit/s - 2 Mbit/s (without haptic compression encoding); 0.8-200 kbit/s (with haptic compression encoding)	99.999%	1 DoF: 2 to 8 10 DoF: 20 to 80 100 DoF: 200 to 800	Stationary or Pedestrian	≤ 100 km^2 (4)	Haptic feedback (3)
	< 400 ms	1-100 Mbit/s	99.999%	1500	Stationary/ or Pedestrian	≤ 100 km^2 (4)	Video
	< 150 ms	5-512 kbit/s	99.9%	50	Stationary or Pedestrian	≤ 100 km^2 (4)	Audio
	< 300 ms	600 Mbit/s	99.9%	1500	Stationary or Pedestrian	≤ 100 km^2 (4)	VR

TABLE 7 (*end*)

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)**	**UE speed**	**Service area**	
Immersive multi-modal navigation applications Local Site → Remote Site (UL)	< 300 ms	12 kbit/s	99.999%	1500	Stationary or Pedestrian	≤ 100 km² (4)	Biometric / Affective
	< 400 ms	1-100 Mbit/s	99.999%	1500	Workers: Stationary/ or Pedestrian, UAV: [30-300 mph]	≤ 100 km² (4)	Video
	< 150 ms	5-512 kbit/s	99.9 %	50	Stationary or Pedestrian	≤ 100 km² (4)	Audio
	< 300 ms	600 Mbit/s	99.9 %	1500	Stationary or Pedestrian	≤ 100 km² (4)	VR

(1) Motion-to-photon delay (the time difference between the user's motion and corresponding change of the video image on display) is less than 20 ms, and the communication latency for transferring the packets of one audio-visual media is less than 10 ms, e.g. the packets corresponding to one video/audio frame are transferred to the devices within 10 ms.

(2) According to IEEE 1918.1 [40] as for haptic feedback, the latency is less than 25 ms for accurately completing haptic operations. As rendering and hardware introduce some delay, the communication delay for haptic modality can be reasonably less than 5 ms, i.e. the packets related to one haptic feedback are transferred to the devices within 10 ms.

(3) Haptic feedback is typically haptic signal, such as force level, torque level, vibration and texture.

(4) In practice, the service area 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.

NOTE – The latency requirements are expected to be satisfied even when multimodal communication for skillset sharing is via indirect network connection (i.e. relayed by one UE to network relay).

Beyond these above capabilities of IMT-2020, refer 3GPP TS22.263, following capabilities are developed:

1 IMT-2020 is capable of enabling an NPN for video, imaging and audio for professional applications.

2 IMT-2020 network is capable of providing a time reference information to a third party application acting as a master clock with an accuracy of 1 microsecond.

3 IMT-2020 is capable of securely reconnecting within a short period of time (< 1 s) from UE starting first network reconnection attempt after the UE has detected a UE network connection loss.

4 IMT-2020 is capable of supporting uplink and downlink service continuity maintaining acceptable performance requirements while switching between co-located PLMN and NPN (e.g. due to mobility).

5 IMT-2020 is capable of supporting service continuity maintaining acceptable performance requirements: for an uplink stream while performing traffic steering, switching, and splitting

among co-located PLMN(s) and NPN(s); for downlink while switching between co-located PLMN and NPN.

8.1.2 IMT-2020 capabilities for broadcast/multicast in multimedia communication

The capability examples of multimedia broadcast/multicast supported by IMT-2020 are:

1 to support operation of downlink only broadcast/multicast over a specific geographic area (e.g. a cell sector, a cell or a group of cells). In addition, UE may have the ability to perform MBS Interest Indication (MII) in the uplink direction, as specified in 3GPP TS38.331.

2 To support operation of a downlink only broadcast/multicast system over a wide geographic area in a spectrally efficient manner for stationary and mobile UEs. In addition, UE may have the ability to perform MBS Interest Indication (MII) in the uplink direction, as specified in 3GPP TS38.331.

3 To enable the operator to reserve 0% to 100% of radio resources of one or more radio carriers for the delivery of broadcast/multicast content.

4 To allow the UE to receive content via a broadcast/multicast radio carrier while a concurrent data session is ongoing over another radio carrier.

5 To support broadcast/multicast of UHD streaming video (e.g. 4K/8K UHD).

6 To allow the operator to configure and broadcast multiple quality levels (i.e. video resolutions) of broadcast/multicast content for the same user service in a stand-alone IMT-2020 based broadcast/multicast system.

7 To support parallel transfer of multiple quality levels (i.e. video resolutions) of broadcast/multicast content for the same user service to the same UE taking into account e.g. UE capability, radio characteristics, application information.

8 To support parallel transfer of multiple multicast/broadcast user services to a UE.

9 To be able to setup or modify a broadcast/multicast service area within [1 s][21].

10 To be able to apply QoS, priority and pre-emption to a broadcast/multicast service area.

11 To support downlink parallel transfer of the same content, via broadcast/multicast and/or unicast, such that all receiver group members in a given area receive the media at the same time according to user perception.

12 To support a mechanism to inform a media source of relevant changes in conditions in the system (e.g. capacity, failures).

13 To provide means for a media source to provide QoS requirement requests to the broadcast/multicast service.

14 To provide means for the broadcast/multicast service to inform the media source of the available QoS, including modification of available QoS characteristics and availability of the broadcast/multicast service.

15 To support broadcast/multicast of voice, data and video group communication, allowing at least 800 concurrently operating groups per geographic area.

16 To support delivery of the same UE-originated data in a resource-efficient manner in terms of service bit rate to UEs distributed over a large geographical area.

17 To allow a UE to request a communication service to simultaneously send data to different groups of UEs at the same time.

18 To allow different QoS policy for each group the UE communicates with.

---

[21] The [] means it is better to support this KPI, but not has to.

Refer to 3GPP TR38.913, following capabilities for broadcast/multicast that have been described:

1 To support existing Multicast/Broadcast services (e.g. download, streaming, group communication, TV).

2 To support dynamic adjustment of the Multicast/Broadcast area based on e.g. the user distribution or service requirements.

3 To support concurrent delivery of both unicast and Multicast/Broadcast services to the users.

4 To support efficient multiplexing with unicast transmissions in at least frequency domain and time domain.

5 To support static and dynamic resource allocation between Multicast/Broadcast and unicast; the new RAT is needed to in particular allow support of up to 100% of DL resources for Multicast/Broadcast (100% meaning a dedicated MBMS carrier).

6 To cover large geographical areas up to the size of an entire country in SFN mode with network synchronization. It is needed to also support local, regional and national broadcast areas.

7 To support Multicast/Broadcast services for fixed, portable and mobile UEs.

8 To leverage usage of RAN equipment (hard- and software) including e.g. multi-antenna capabilities (e.g. MIMO) to improve Multicast/Broadcast capacity and reliability.

9 To support on-demand establishment of UE to UE, multicast and broadcast private communication between member UEs of the same network.

10 To allow member UEs of local area virtual network to join an authorized multicast session.

11 To support for real time and non-real time multimedia services and applications with advanced Quality of Experience (QoE)

8.2 Multimedia communications supported by IMT-2020 technologies

The capabilities can be supported by different IMT-2020 technical components e.g. radio interface, architecture enhancements. This section describes examples of these technical components.

8.2.1 Radio Access Network

The radio technical components of IMT-2020 to support multimedia communication can be found in Recommendation ITU-R M.2083. Follow here examples of them.

8.2.1.1 UL enhancement for unicast multimedia communication

A) UL MIMO

While UL MIMO offers the capability for reduction in overhead and/or latency, high-speed vehicular scenarios (e.g. a UE traveling at high speed on highways) at FR2 require more aggressive reduction in latency and overhead – not only for intra-cell, but also for L1/L2 centric inter-cell mobility. This also includes reducing the occurrence of beam failure events. Besides of it, enhancements for enabling panel-specific UL beam selection were investigated and specified continuously. This offers some potential for increasing UL coverage including. Then, channels other than PDSCH can benefit from multi-TRP transmission (as well as multi-panel reception) which also includes multi-TRP for inter-cell operations. This includes some new use cases such as UL dense deployment within a macro-cell and/or heterogeneous-network-type deployment scenarios. The SRS can be further enhanced for capacity and coverage.

B) UL carrier aggregation

Carrier aggregation was developed the unaligned frame boundary capability to provide more flexible beginning Tx frame structure configurations among different carriers to offer higher uplink throughput and lower latency. Tx switching is to specify the dynamic switching mechanisms among two uplink carriers. Unaligned frame boundary and Tx switching can be implemented together to achieve larger uplink throughput for TDD CA operation. In one cell group, it supports PUCCH carrier switching semi-statically or dynamically. This could reduce the latency of PUCCH transmissions significantly for CA operation.

C) UL dual connectivity (DC)

Dual connectivity is capable of Uplink power controlling i.e. limiting UE's transmission power to assure edge user's communication QoS, migration interference among the users, and reducing energy consumption of the UE. Uplink DC is also capable of earlier measurement and fast recovery from MCG failure to reduce latency. Beyond these, it is capable of indicating UE entering the third state to maintain context of the UE which will reduce configuration overhead.

D) Interference coordination

Interference coordination includes two aspects, i.e. cross-link interference (CLI and remote interference management (RIM). For CLI, it is left to network implementation for gNB-gNB interference. The UE-UE interference coordination is capable of such as SRS-RSRP/CLI-RSSI-based layer-3 CLI measurement and reporting, and network coordination mechanism for CLI with inter-gNB exchange of intended UL/DL configuration. RIM targets to migrate the interference occurring in specific weather conditions with the distance between aggressor gNB and victim gNB hundreds of kilometres. The RIM reference signal (RS) based on PRACH preamble-like RS is introduced for better interference measurement while the detailed remote interference mitigation mechanisms are left to implementation.

E) MBS Interest Indication (MII) in the uplink direction

As specified in 3GPP TS38.331, the purpose of this procedure is to inform the network that the UE in RRC_CONNECTED is receiving or is interested to receive MBS broadcast service(s) and to inform the network about the priority of MBS broadcast versus unicast and multicast MRB reception. MBS Interest Indication can only be sent after AS security activation.

8.2.1.2 NR-based 5G multicast and broadcast

IMT-2020 can utilize flexible and dynamic resources allocation to enable both multicast and broadcast in use cases of public safety and mission critical, V2X applications, transparent IPv4/IPv6 multicast delivery, IPTV, software delivery over wireless network, group communications and IoT applications, instead of broader area broadcast like services.

With this technology, RAN basic functions of broadcast/multicast for UEs in RRC_CONNECTED state are to be supported with better reliability and better service continuity. For multicast communication service, the same service and the same specific content data are provided simultaneously to a dedicated set of UEs (i.e. not all UEs in the Multicast service area are authorized to receive the data). A multicast communication service is delivered to the UEs using a multicast session. A UE can receive a multicast communication service in RRC_CONNECTED state with mechanisms such as point to point (PTP) and/or point to multi-point (PTM) delivery, with a balance between network efficiency. A UE can also receive a multicast communication service in RRC_INACTIVE, e.g. to fulfil the requirements of, e.g. safety related applications, especially for cells with a large number of UEs.

HARQ feedback/retransmission can be applied to both PTP and PTM transmissions. For PTM transmission there can be two HARQ feedback schemes: UE specific feedback and NACK only feedback, depending on network decision based on level of reliability requirement or network resources. Network is able to dynamically change Multicast service delivery between PTM and PTP with service continuity for a given UE.

Basic mobility with service continuity is supported. Unicast like mobility mechanism is designed to offer the basic mobility, e.g. the Multicast session resources are to be established along with UE's mobility in the target RAN node which supports the MBS service.

For broadcast communication service, the same service and the same specific content data are provided simultaneously to all UEs in a geographical area (i.e. all UEs in the Broadcast service area are authorized to receive the data). A broadcast communication service is delivered to the UEs using a broadcast session. A UE can receive a broadcast communication service in RRC_IDLE, RRC_INACTIVE and RRC_CONNECTED state.

There are other enhancement including, UE are able to receive MBS with simultaneous operation in unicast reception; enhancement on RAN network interface like Xn, F1, and E1 interfaces; support on dynamic broadcast area; a UE may be able to use shared processing for MBS broadcast and unicast reception from the same or different operators; resource efficiency for MBS reception in RAN sharing scenarios (if the same Broadcast service is provided by two or more operators separately, a single PTM radio resources in the same cell for transmission of the same content will be used), etc.

8.2.1.3 LTE-based 5G Broadcast

IMT-2020 can deliver audio-visual services (including free-to-air services) in single frequency network (SFN) on stand-alone infrastructure, usually in High-Power High-Tower Single Frequency Networks to support larger inter-site distance (e.g. to allow cell radii of up to 100 km).

To allow cell radii of up to 100 km, numerology enhancement with a new 0.37 kHz subcarrier spacing and CP duration ~300 µs was introduced to support broadcast in medium power medium tower (MPMT) and HPHT. To allow reception with UE mobility up to 250 km/h, numerology enhancement with a new 2.5 kHz wider sub-carrier spacing, CP duration ~100 µs with better Doppler resiliency. There are other related enhancements to support broader coverage for the control information like dedicated reference signals (RS) accompany each numerology, enhanced subframe structure, control channel, and less dense RS pattern with reducing overheads.

8.2.2 Architecture enhancements

The unicast architecture enhancement of IMT-2020 to support real-time multimedia interaction and uploading follows IMT-2020 architectural principles defined in 3GPP TS23.501 and 3GPP TS26.501[22]. Following architecture enhancements are supported in 5GS.

5G-XR Application Provider provides a 5G-XR Aware Application on the UE to make use of a 3GPP 5G-XR client and network functions via 5G media streaming model which is introduced in 3GPP TS26.501.

The 5G QoS model supports both:

– QoS Flows that require guaranteed flow bit rate (GBR QoS Flows).

– and QoS Flows that do not require guaranteed flow bit rate (Non-GBR QoS Flows).

The QoS model also supports Reflective QoS. A QoS Flow ID (QFI) is used to identify a QoS Flow in IMT-2020. User Plane traffic assigned to the same QoS Flow within a Protocol Data Unit (PDU) Session receives the same traffic forwarding treatment (e.g. scheduling, admission threshold). The QFI may be dynamically assigned or may be equal to the 3GPP 5G QoS Identifier (5QI). A QoS Flow may either be 'GBR', 'Non-GBR' or "Delay Tolerant GBR" depending on its QoS profile and it contains QoS parameters as follows:

– For each QoS Flow, the QoS profile includes the QoS parameters:

---

[22] 3GPP TS26.501 V17.3.0 (2022-09) – 5G Media Streaming (5GMS); General description and architecture.

- 3GPP 5G QoS Identifier (5QI); and
- Allocation and Retention Priority (ARP).

– For each Non-GBR QoS Flow only, the QoS profile can also include the QoS parameter:

- Reflective QoS Attribute (RQA).

– For each GBR QoS Flow only, the QoS profile also include the QoS parameters:

- Guaranteed Flow Bit Rate (GFBR) - uplink (UL) and downlink (DL); and
- Maximum Flow Bit Rate (MFBR) - UL and DL; and

– In the case of a GBR QoS Flow only, the QoS profile can also include one or more of the QoS parameters:

- Notification control;
- Maximum Packet Loss Rate - UL and DL.

PDU Set is introduced to optimize the delivery of XRM service in 5GS. A PDU Set is composed of one or more PDUs carrying the payload of one unit of information generated at the application level (e.g. a frame or video slice for XRM Services), which are of same importance at application layer. All PDUs in a PDU Set are needed by the application layer to use the corresponding unit of information. In some cases, the application layer can still recover parts of the information unit, when some PDUs are missing.

Another approach of NR multimedia communications quality assurance in 3GPP specifications is called NR QoE[23] (Quality of Experience). The QoE Measurement Collection function enables the collection of application layer measurements from the UE. The measurements are supported for Streaming services, Multimedia Telephony Services for IMS (MTSI), and VR services. Both signalling based and management based QoE measurement collection are supported. In 3GPP Release 17, basic application layer measurement collection mechanism is supported, and RAN visible QoE measurement mechanism and metrics are supported.

The following key issues are currently under investigation:

WT#1: Enhancements for supporting multi-modality service: Study whether and how to enable delivery of related tactile and multi-modal data (e.g. audio, video and haptic data related to a specific time) with an application to the user at a similar time, focusing on the need for policy control enhancements (e.g. QoS policy coordination).

WT#2: Enhancements of network exposure to support interaction between 5GS and application:

– WT#2.1: Study whether and how interaction between AF and 5GS is needed for application synchronization and QoS policy coordination among multiple UEs or between multiple QoS flows per UE.

– WT#2.2: Study exposure of 5GS QoS information (e.g. QoS capabilities) and network conditions to the Application to enable quick codec/rate adaptation helps to provide desired QoE (e.g. such as assist in alleviating 5GS congestion).

WT#3: Study whether and how the following QoS and policy enhancements for XR service and media service transmission are performed:

– WT#3.1: Study the traffic characteristics of media service enabling improved network resources usage and QoE.

[23] 3GPP TR21.917 Section 12.2.

– WT#3.2: Enhance QoS framework to support media units granularity (e.g. video/audio frame/tile, Application Data Unit, control information), where media units consist of PDUs that have the same QoS requirements.

– WT#3.3: Support differentiated QoS handling considering different importance of media units. e.g. eligible drop packets belong to less important media units to reduce the resource wasting.

– WT#3.4: Whether and how to support uplink-downlink transmission coordination to meet RTT (Round-Trip Time) latency requirements between UE and N6 termination point at the UPF.

– WT#3.5: Potential policy enhancements to minimize the jitter, focusing on i.e. requirement provisioning from AF, extension of PCC rule.

WT#4: Study potential enhancements of Mobility and power management considering traffic pattern of media services:

– WT 4.1: void

– WT 4.2: Power saving enhancement e.g. support trade-off of throughput/latency/reliability considering device battery life, whether and how to enhance CDRX, considering XR/media traffic pattern.

Addressing solutions to the aforementioned issues are captured in 3GPP TS23.501 as well.

9 Case studies

Followings are intended to provide examples of IMT-2020 capabilities to support multimedia content on demand transmission, real-time multimedia interactive in different scenarios.

9.1 News live report

FIGURE 2

5G based interview with holographic remote location on the same screen[24]

Since 2019, 5G based multimedia technology was adopted to support the live reports of the two sessions. In 2022, Xinhua News Agency launched the first immersive screen fusion interview the With 5G-based multimedia interaction and broadcast technology.

It provided 5G network transmission and cloud-based production of ultra-high-definition videos to media users. Through 5G multimedia services, visitors can not only use the 5G network to return ultra-high-definition video interviews, but also produce program content on the cloud platform. It

---

24 http://www.zgjx.cn/2022-03/09/c_1310506639.htm (in Chinese)

makes full use of the characteristics of 5G large bandwidths and ultra-low delay, thus the news reports can use 4K, VR and other video services to improve the picture quality, to smooth the interaction, to ensure high-quality user experience and to diversify needs in all aspects. Based on 5G transmission, the all-round director can realize the integration of the core functions of the studio, such as the input of multiple SDI camera signals, HDMI signals, and IP signals, real-time director switching, programme production, and graphic packaging. It can experience and display highly integrated and efficient broadcast programme. And through 5G transmission, it can realize multi-camera indoor and outdoor multi-signal access live broadcast, providing real-time experience of virtual studio.

9.2 Smart Venues

FIGURE 3

Smart Venues for 2022 Beijing Winter Olympics

Located in Beijing and Zhangjiakou City, Hebei Province, the 87 venues and the roads connecting the venues for 2022 Beijing Winter Olympics were full coverage of 5G networks. This is also the largest commercial use of 5G networks in the history of the Olympic Games. With the help of 5G network, many Winter Olympics games e.g. alpine skiing, which operated in high mountains with high-speed competitions and difficult-to-capture images are turned into visual feasts in a clear and timely manner. Many of these wonderful videos and pictures are presented through 5G live broadcast technology.

– For the audience, the picture supported by 5G is more high-definition and richer. For example, at the opening ceremony of the Winter Olympics, some Bird's Nest venue shots and athletes entering the arena were captured by the Olympic Broadcasting Service (OBS) through the 5G network that fully covers the Bird's Nest venues. And for the first time it is broadcasted in 4K ultra-high-definition, and some important events are broadcasted in 8K which has higher requirements for transmission bandwidth, e.g. at least 130 Mbit/s or more. Only gigabit-level 5G networks can guarantee image quality. In addition, the combination of the VR camera and the 5G network allows the audience to enjoy the game from a 180-degree free angle of view, e.g. the audience in the stands, the environment of the stadium. A lot of spatial information that would have been omitted in the past can be displayed this time, allowing the audience to see the game from different angles.

– The Winter Olympics have built a dedicated 5G media network to serve media live broadcasts and other special activities. 4K/8K picture quality, free viewing angles, and 'bullet time' and other images are all broadcast through low-latency, high-bandwidth 5G media dedicated networks. Even in venues with a high density of people, the network capacity can still meet media needs to upload the multimedia content quickly. Additional, on the Beijing-Zhangjiakou high-speed railway connecting Beijing and Zhangjiakou, the two major Winter Olympic competition areas, 356 5G base stations were deployed to construct the 3.5G 200M high-speed rail network, thus even at a speed of up to 350 km/h and across mountains and mountains, global media workers can watch the game in real time and use 4K live broadcast throughout the whole process without any lag. During the competition, many media reporters need to work in the mountains, while the 5G backpack works like a mobile signal station can follow the footsteps of the camera reporters to ensure the mobile network signal in sub-zero temperatures. Thus, it allows high-definition video pictures to be uploaded to the cloud in real time, which enable the background staff directly edits, produces, and broadcasts on time to provide the wonderful scenes.

9.3 Sports event live broadcast

FIGURE 4

2019 International Gymnastics Federation Chengdu Parkour World Cup[25]

In the programme "2019 International Gymnastics Federation Chengdu Parkour World Cup" from April 6th to 7th, 2019, China Media Group (CMG) realized the live broadcast of VR sports events based on 5G network transmission. This is the first time that 5G technology and ultra-high-definition VR production technology have been combined and applied to the broadcast of sports events. Considering the characteristics of parkour, the technical team took advantage of 5G's advantages of low latency, higher data rate, and large capacity to design the VR video live broadcast solution, technically shortening the time delay of real-time VR video signals from collection and production to distribution and improving its performance. In addition, the refined production of event-related VR video content covers the scenery of the host city, highlights around the stadium, VR video aerial photography, and the unique first-person perspective of contestants. The game-related content is displayed in all directions and from multiple angles, enabling viewers to experience the game atmosphere at close range.

In the solution, the multi-angle videos were shot in the competition area, collected and stitched in real time via the 5G network. Then, the high-definition VR video signal is transmitted to the 5G media application laboratory in Beijing's Guanghua Road office through the 5G base station and network. After the VR video was produced including switching, subtitle packaging, and special effects in the media production platform, the media content from publishing platform is distributed to CCTV5 APP client via 5G network for live broadcast.

[25] National Radio and Television Administration (PRC) 5G Advanced Video – White Paper on VR Video Technology (2020) http://www.nrta.gov.cn/module/download/downfile.jsp?classid=0&filename=cf359da741ac4bdba25f4b77b9cfc160.pdf

9.4 Interactive cloud gaming

FIGURE 5

ShouGang No. 1 Furnace Paradise[26]

Integrated with 5G, cloud XR, multiple layer's computing network, the traditional sports competitions are merged with modern virtual reality technology, build the century-old industrial site of ShouGang No. 1 blast furnace into the largest XR technology experience park in North China, turn the abandoned industrial site into the entrance of the universe, build the capital's high-tech cultural industry, and form 5G entertainment new benchmark. Through 5G+cloud XR technology, gamers only need to wear a lightweight head-mounted display device to enter the metaverse venue and experience the subversive and innovative immersive experience.

The "wide-area and large-space VR interaction" scenario supported by IMT-2020, further provides differentiated computing power for cloud XR gaming, saving GPU resources by up to 50%, processing delay can be reduced to 35ms, and end-to-end communication delay is less than 20ms, ensuring the high performance of cloud XR gaming service experience. Through the intelligent scheduling of IMT-2020 resources, it greatly reduces the initial CAPX investment and OPAX of network operation in every venue. At the same time, the number of customers has increased at least 30 percent with new interactive games (e.g. Player versus Player, Player versus Environment, skydiving game). And with IMT-2020 supporting the social and interactive capabilities in different locations, the different venues that were originally independent of each other can be upgraded to a "metaverse" entrance, realize the online and offline linkage of experiences, support multiple gamers to interact in the metaverse space.

10 Summary

This Report summarizes various capabilities of terrestrial component of IMT-2020 for multimedia communications.

Multimedia is an immersive technological way of presenting information that combines audio, video, images and animations with textual data. Multimedia applications include network video, digital magazine, digital newspaper, digital radio, social media, touch media, etc., and can be easily enabled using IMT-2020. New emerging technologies such as Virtual Reality (VR) and Augmented Reality (AR) are becoming key technologies to upgrade the traditional multimedia industries.

The IMT-2020 capabilities can support the evolving interactive multimedia communication with the capabilities not only broader bandwidth, higher data rate, but also lower latency and higher reliability.

The typical technologies are flexible and dynamic resources allocation, uplink enhancement e.g. UL MIMO, UL Carrier Aggregation and Dual connectivity, and related architecture improvement, which can connect the user to a high-definition video, real-time multimedia interaction virtual world on their mobile device.

---

[26] http://www.5gaia.org.cn/FourthBloomingCup/News/Detail/360; https://www.laoyaoba.com/n/805626

Live events with HD and UHD content can be streamed via IMT-2020 radio network with higher throughput. HD and UHD content (e.g. news, sport event) can be real-time produced and on demand distributed to mobile devices without any interruptions through IMT-2020 higher user experienced data rate and low latency. The entertainment industry will hugely benefit from IMT-2020 wireless networks, which are expected to enable HD virtual reality games with a better real-time interactive gaming experience, and high dynamic range video streaming without interruption. Cloud AR and Cloud VR with HD or UHD video can be supported with higher user experienced data rate and low latency supported by IMT-2020.

It is foreseeable that with the support of IMT-2020 technology, it will gradually bring consumers more amazing virtual experience.