International Telecommunication Union

ITU-T Technical Report

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(17 June 2022)

GSTR-5GQoE

Quality of experience (QoE) requirements for real-time multimedia services over 5G networks

Technical Report ITU-T GSTR-5GQoE

Quality of experience (QoE) requirements for real-time multimedia services over 5G networks

Summary

This Technical Report defines a scope for the analysis of QoE in 5G services and several use cases where this scope is applicable. Such use cases are: tele-operated driving, wireless content production, mixed reality offloading and first responder networks.

Addressing this set of use cases is challenging for three different reasons:

– Their requirements and quality of experience (QoE) expectations may be different from the ones present in most QoE-related research and Recommendations, which typically address communication services for consumer-type users (e.g., telephony, videoconference, video delivery or streaming, gaming).

– The experience and expectations of the use case owners may not be applicable to cellular wireless networks, even when quality of service policies are applied. For example, a wireless content production studio will not have the same channel capacity as a wired network, neither from the point of view of bandwidth nor of reliability. Therefore, totally new impairments or artefacts may appear when moving the use case from wired links to 5G.

– Professional and vertical applications typically have fewer users than video consumer applications (there are fewer content producers than content consumers), or video transmission is just one of the pieces of a much more complex ecosystem (as in the automotive industry).

For each of the services, the Technical Report describes:

– Its main characteristics and reference architecture;

– The relevant QoE indicators to be considered on the service;

– A reference implementation, including the order-of-magnitude values of the service key performance indicators; and

– An analysis of the key factors to evaluate the QoE of the service.

Keywords

5G, content production, first responders, mixed reality, tele-operated driving.

Change Log

This document contains Version 1 of the ITU-T Technical Report on "QoE requirements for real-time multimedia services over 5G networks" approved at the ITU-T Study Group 12 meeting held in Geneva, 7-17 June 2022.

Editor:	Pablo Pérez	Tel: +34 91 330 4000
	Nokia	Fax: +34 91 330 5000
	Spain	Email: pablo.perez@nokia-bell-labs.com

Table of Contents

© ITU 2022

Technical Report ITU-T GSTR-5GQoE

Quality of experience (QoE) requirements for real-time multimedia services over 5G networks

1   Scope

The scope of this Technical Report is addressing live multimedia services which are provided over wireless IP networks (5G and beyond).

The proposed scope for this Technical Report is addressing services which are required for 5G and beyond, either because they are related to vertical use cases or because they rely on 5G-specific capacities. They should have the following properties:

– Being based on IP wireless communication.

– Involving multimedia (mostly video) transmission.

– Requiring (or, at least, expecting) specific support from network features so that they are not expected to be provided over-the-top.

– Having specific requirements for latency or expectation on timely delivery.

– In most cases, being vertical or niche applications, which makes it challenging for them to be covered by a dedicated ITU-T quality of service (QoS)/QoE Recommendation.

For all the targeted use cases, the work item will study the specific QoE requirements, as well as the required performance and features from the network. By addressing them in parallel, it will be possible to find synergies between them and, more relevantly, extract the common information that can be used to also analyse other use cases that may arise outside the scope of this work item.

2   References

[ITU-T G.1035]	Recommendation ITU-T G.1035 (2021), *Influencing factors on quality of experience for virtual reality services*.
[ITU-T P.10]	Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience*.
[ITU-T P.805]	Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*.
[ITU-T P.809]	Recommendation ITU-T P.809 (2018), *Subjective evaluation methods for gaming quality*.
[ITU-T P.912]	Recommendation ITU-T P.912 (2016), *Subjective video quality assessment methods for recognition tasks*.
[ITU-T P.913]	Recommendation ITU-T P.913 (2021), *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*.
[ITU-T P.920]	Recommendation ITU-T P.920 (2005), *Interactive test methods for audiovisual communications*.
[ITU-T P.1312]	Recommendation ITU-T P.1312 (2016), *Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance*.
[ITU-R BT.2137-0]	Recommendation ITU-R BT.2137-0 (2020), *Technologies applicable to Internet Protocol interfaces for programme production*.

[3GPP TS 22.261]	3rd Generation Partnership Project, Technical Specification Group Services and System Aspects (2021), *Service requirements for the 5G system*, Stage 1 Release 17, V18.8.0.
[3GPP TR 22.827]	3rd Generation Partnership Project, Technical Specification Group Services and System Aspects (2019), *Study on audio-visual service production,* Release 17, Stage 1.
[3GPP TS 23.287]	3rd Generation Partnership Project, Technical Specification Group Services and System Aspects (2021), *Architecture enhancements for 5G system (5GS) to support vehicle-to-everything (V2X) services*, Release 17, V17.1.0.
[3GPP TS 23.501]	3rd Generation Partnership Project, Technical Specification Group Services and System Aspects (2021), *System architecture for the 5G System (5GS)*, Release 17, Stage 2.
[3GPP TR 26.928]	3rd Generation Partnership Project (2020), *Extended reality (XR) in 5G,* Release 16, V16.0.0.
[3GPP TR 26.998]	3rd Generation Partnership Project (2020), *Support of 5G Glass-type Augmented Reality / Mixed Reality (AR/MR) devices*, Release 17.
[5GAA XWG5-200029]	5GAA Automotive Association (2020), *Tele-operated driving (ToD) – Use cases and technical Requirements*.
[5GAA S-200137]	5GAA Automotive Association (2020), *Study of spectrum needs for safety related intelligent transportation systems – day 1 and advanced use cases*.
[5GAA T-200xxx]	5GAA Automotive Association (2020), *C-V2X Use Cases Volume II: Examples and Service Level Requirements*.
[EBU R 118]	EBU Operating Eurovision and Euroradio R 118 (2017), *Tiering of cameras for use in television production*.
[EBU TR 056]	EBU Operating Eurovision and Euroradio Technical Report 059 (2020), *5G for professional media production and contribution*.
[RDG V4]	Rail Delivery Group's Policing & Security Implementation Group (2015), *National Rail and Underground Closed Circuit Television (CCTV) Guidance document*.
[Dima]	Dima, E., Brunnström, K., Sjöström, M., Andersson, M., Edlund, J., Johanson, M., and Qureshi, T. (2020), *Joint effects of depth-aiding augmentations and viewing positions on the quality of experience in augmented telepresence*, Quality and User Experience, Vol. 5, No. 1, 1–17.
[Henry]	Henry, M., Villemain, O. and Mertens, L. (2021), *Ultrafast ultrasound imaging*. In Multimodality Imaging Innovations In Adult Congenital Heart Disease, Springer, Cham pp. 27–47.
[Janowski]	L. Janowski; L. Malfait; M.H. Pinson (2019), *Evaluating experiment design with unrepeated scenes for video quality subjective assessment*, Quality and User Experience, Vol. 4, Art. 2.
[Lévêque]	Lévêque, L., Outtas, M., Liu, H., and Zhang, L. (2021), *Comparative study of the methodologies used for subjective medical image quality assessment*, Physics in Medicine & Biology, Vol. 66. No. 15, 15TR02.

[Martini]	Martini, M.G., Iacobelli, L., Bergeron, C., Hewage, C.T., Panza, G., Piri, E., Vehkaperä, J., Amon, P., Mazzotti, M., Savino, K. and Bokor, L. (2015), *Real-time multimedia communications in medical emergency-the CONCERTO project solution*. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7324–7327.
[Pérez]	Pérez, Pablo, et al. (2021), *A parametric quality model to evaluate the performance of tele-operated driving services over 5G networks*, Multimedia Tools and Applications, Vol. 81, 1–17.
[Pueo]	B. Pueo (2016), *High speed cameras for motion analysis in sports science*. Journal of Human Sport and Exercise Vol. 11 No. 1, pp. 53-73.
[Razaak]	Razaak, M., Martini, M.G. and Savino, K. (2014), *A study on quality assessment for medical ultrasound video compressed via HEVC*. IEEE Journal of biomedical and health informatics, Vol. 18, No. 5, pp. 1552–1559.
[Villegas]	Villegas, A., Pérez, P., and González-Sosa, E. (2019). *Towards a distributed reality: a multi-video approach to XR*. In Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems, pp. 34–36.

3   Definitions

3.1   Terms defined elsewhere

This Technical Report uses the following terms defined elsewhere:

3.1.1   quality of experience (QoE) [ITU-T P.10]: The degree of delight or annoyance of the user of an application or service.

3.1.2   QoE influencing factors [ITU-T P.10]: Include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service and their fulfilment, the user's cultural background, socio-economic issues, psychological profiles, emotional state of the user, and other factors whose number will likely expand with further research.

3.1.3   QoE assessment [ITU-T P.10]: The process of measuring or estimating the QoE for a set of users of an application or a service with a dedicated procedure, and considering the influencing factors (possibly controlled, measured, or simply collected and reported). The output of the process may be a scalar value, multi-dimensional representation of the results, and/or verbal descriptors. All assessments of QoE should be accompanied by the description of the influencing factors that are included. The assessment of QoE can be described as comprehensive when it includes many of the specific factors, for example a majority of the known factors. Therefore, a limited QoE assessment would include only one or a small number of factors.

3.1.4   quality of service (QoS) [ITU-T P.10]: The totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.

3.1.5   tele-operated driving (ToD) [5GAA XWG5-200029]: A Remote Vehicle (RV) user or operator engaging in the act of driving by taking any of the following roles: dispatcher, indirect controller or direct controller.

3.2 Terms defined in this Technical Report

None.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

3D	Three-Dimensional
5G	Fifth Generation (of cellular telephony)
AR	Augmented Reality
AMBR	Aggregate Maximum Bit Rate
AT	Augmented Telepresence
CCAM	Cooperative Connected and Automated Mobility
CCTV	Close Circuit Television camera
DDT	Dynamic Driving Task
DL	DownLink
DR	Distributed Reality
ECG	Electrocardiogram
eMBB	enhanced Mobile Broadband
FPS	Frames Per Second
GFBR	Guaranteed Flow Bit Rate
gNB	g-NodeB
GOP	Group Of Pictures
HEVC	High-Efficiency Video Coding
HMD	Head-Mounted Display
IEM	In-Ear Monitor
IP	Internet Protocol
KPI	Key Performance Indicator
LAiP	Live Audio in Production
LAVP	Live Audiovisual Production
LTE	Long Term Evolution
MCPTT	Mission-Critical Push-To-Talk
MEC	Multiaccess Edge Computing
MFBR	Maximum Flow Bit Rate
mMTC	massive-Machine-Type Communication
MR	Mixed Reality
NR	New Radio
OBU	On-Board Unit
OEDR	Object and Event Detection and Response

PDB	Packet Delay Budget
PER	Packet Error Rate
PDRs	Packet Detection Rules
PTZ	Pan-Tilt-Zoom
QoE	Quality of Experience
QoS	Quality of Service
RGB	Red-Green-Blue
ToD	Tele-operated Driving
UL	Uplink
URLLC	Ultra-Reliable Low Latency Communication
VR	Virtual Reality
WCP	Wireless Content Production
XR	extended Reality

5 Introduction

The fourth generation of mobile networks (LTE) has made it possible to generalize the consumption of video content in mobile. These services are mostly provided over-the-top, using the same internet connection as other applications in the mobile device, since they target end users (individual consumers) with a global outreach.

With the fifth generation of mobile technology (5G), wireless networks are increasing their capabilities in terms of achieved throughput (enhanced Mobile Broadband (eMBB)), latency (ultra-reliable low latency communication (URLLC)), and device density (massive machine type communications (mMTC)). Aside from these aspects, 5G provides improved flexibility to apply these new capabilities to a specific subset of devices, using approaches such as network slicing or non-public networks (NPNs). This opens a new set of video-related services for professional users and vertical industries which, in previous technological generations, were only possible using wired networks, legacy non-IP based wireless links or were even completely impossible to provide. Examples of those services are tele-operated driving (ToD), wireless content production (WCP) and cloud-enabled augmented and/or virtual reality (more generally, mixed reality (MR)).

Addressing this set of use cases is challenging for three different reasons:

– Their requirements and quality of experience (QoE) [ITU-T P.10] expectations may be different from those typically present in most QoE-related research and Recommendations, which typically address communication services for consumer-type users (e.g., telephony, videoconference, video delivery / streaming, gaming).

– The experience and expectations of the use case owners may not be applicable to cellular wireless networks, even when QoS policies are applied. For example, a WCP studio will not have the same channel capacity as a wired network, neither in terms of bandwidth nor from a reliability point of view. Therefore, totally new impairments or artefacts may appear when moving a use case from wired links to 5G.

– Professional and vertical applications typically have fewer users than video consumer applications (there are fewer content producers than content consumers), or video transmission is just one of the pieces of a much more complex ecosystem (as in the automotive industry).

Notwithstanding the characteristics of the professional and vertical use cases, 5G networks also provide the possibility to enhance consumer-oriented services (streaming, gaming, etc.) not only by having higher bandwidth than previous wireless generations, but also by providing added-value functionalities such as traffic prioritization, network slicing or edge computing.

With these premises, this Technical Report addresses services which are required on 5G and beyond, either because they are related to vertical use cases or because they rely on 5G-specific capacities. The addressed services have the following properties:

–   Being based on IP wireless communication.

–   Involving multimedia (mostly video) transmission.

–   Requiring (or, at least, expecting) specific support from network features so that they are not expected to be provided over-the-top.

–   Having specific requirements for latency or expectation on timely delivery.

–   In most cases, being vertical or niche applications, which makes it challenging for them to be covered by a dedicated ITU-T QoS/QoE Recommendation.

–   In most cases, they are task-based applications, performed by a professional or dedicated type of user, with *functional* requirements, rather than purely aesthetic or perceptual.

For all the targeted use cases, the work item will study the specific QoE requirements as well as the required performance and features from the network. By addressing them in parallel, it will be possible to find synergies between them and, more relevantly, extract the common information that can be used to also analyse other use cases that may arise outside the scope of this work item.

6   General considerations

This section provides background information which apply to all the use cases covered by the Report.

6.1   5G QoS model

The system architecture for the 5G system [3GPP TS 23.501] defines the QoS model used by 5G. The model is summarized in Figure 1.
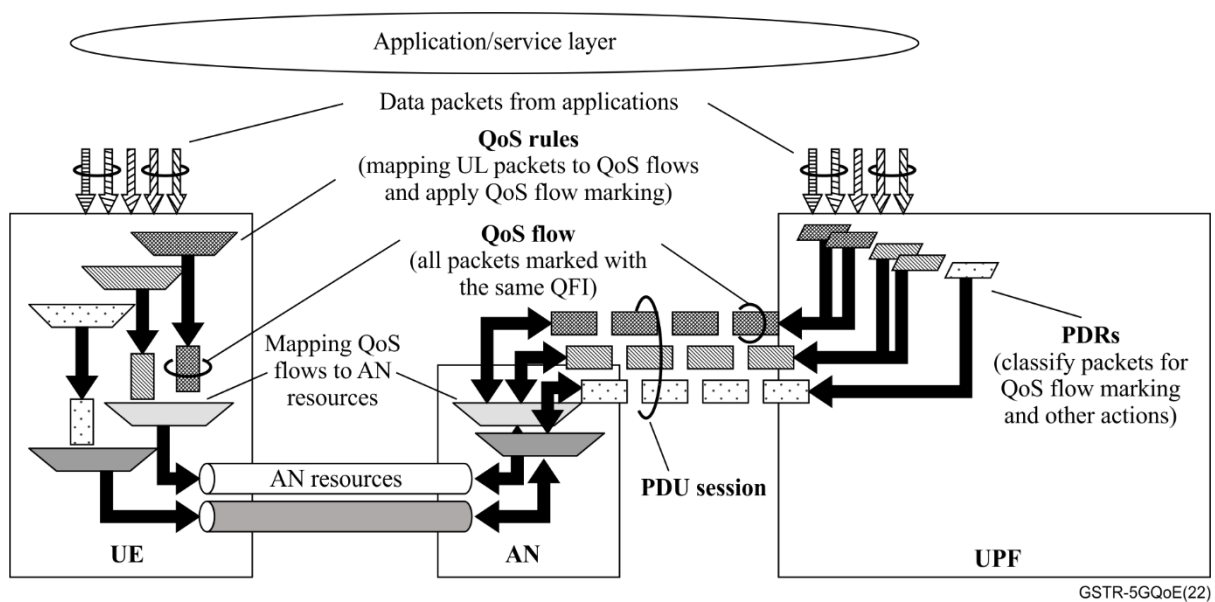

Figure 1 – QoS flows within the 5G QoS model [3GPP TS 23.501]

In the figure, the user equipment (UE) is the user terminal, AN is the access network (the antenna and base station) and UPF is the part of the 5G core which resolves user data. When a UE connects to the

network, it establishes one or several data flows ("PDU sessions") with one or several UPF entities. "Data packets from applications" are typically IP packets[1] that are exchanged between the UE network stack and the internet (or a private IP network), connected directly to the UPF. The traffic coming from the UE is tunnelled across the whole 5G network and it "emerges" to the public internet always at the same point: the UPF.

Each packet of this PDU session belongs to one "QoS Flow", which is tagged with a QoS flow indicator (QFI). There must be at least one default QoS flow in each session, but there could be many. The assignment of IP packets to individual QoS flows is as follows:

– By "packet detection rules" (PDRs) in the downlink (network to UE), which are decided by the core.

– By "QoS rules" in the uplink (UE to network), which must be either explicitly negotiated between the UE and the network or derived by the UE based on received downlink traffic ("reflective QoS").

Both PDRs and QoS rules (explicit or reflective) use packet filter sets to assign IP packets to QoS flows based on IP header fields: source/destination address, port number, protocol ID, type of service, etc.

Each QFI is assigned a QoS management policy or "QoS profile", summarized in Table 1.

Table 1 – 5G QoS profile definition

	Resource type	Non-GBR*	GBR*	Delay-critical GBR*
5G QoS Identifier	Priority level	X	X	X
	Packet delay budget	X	X	X
	Packet error rate	X	X	X
	Average window		X	X
	Maximum data burst volume			X
GBR* specific	Guaranteed flow bit rate		X	X
	Maximum flow bit rate		X	X
	Maximum packet loss rate		X	X
	Notification control		X	X
Aggregate Maximum bit rate	Per session	X		
	Per UE	X		
	Per slice	X	X	X
Allocation and retention priority		X	X	X
Reflective QoS attribute		X		
* GBR – guaranteed bit rate				

The definition of the QoS profile is mostly carried out by the assignment of a 5G QoS identifier (5QI). A 5QI defines the QoS policies which will be applied to the specified QoS flow. It comprises the following components:

– Resource type. Three resource types are identified in 5G: Guaranteed bit rate (GBR), delay-critical GBR, and non-guaranteed bit rate (non-GBR). As seen in the table, some QoS parameters apply only to specific resource types.

_____

[1] PDU sessions can also be established at layer 2 using Ethernet packets, with similar functionality.

–   The total packet delay budget (PDB) and packet error rate (PER) allowed for the flow in the whole 5G network.

–   The average time window used to compute bit rate statistics (for GBR flows).

–   The maximum volume of data that is expected (and allowed) in each burst.

5QIs can be defined explicitly for each flow. However, the [3GPP TS 23.501] defines some standard values for different types of services. In this Report, the standard 5QIs which apply to each of the use cases under study will be presented.

GBR flows require additional definition beyond a 5QI: the service QoS guaranteed flow bit rate (GFBR) and packet loss rate, as well as the maximum flow bitrate (MFBR) allowed. Additionally, the notification control specifies how to handle a situation when the 5G network is not able to fulfil the guaranteed QoS conditions.

Finally, some additional QoS parameters are defined:

–   Aggregate maximum bit rate (AMBR), to limit the maximum bitrate that a session, UE or slice can use (to prevent individual users overusing the shared resources of the network).

–   Allocation and retention policy (ARP), which specifies the priority to decide whether a new QoS flow may be accepted or needs to be rejected, as well as which existing QoS flow to pre-empt, in the case of resource limitations.

–   Reflective QoS attribute (RQA), to indicate whether reflective QoS is enabled for a (non-GBR) QoS flow.

6.2   Identifying relevant QoE indicators

Most Recommendations involving auditory, visual or audiovisual quality aim at modelling or predicting the quality of experience perceived by the user. This model may involve analysing the audiovisual signal involved, either as perceived by the user or in its encoded form (bitstream), or solely analysing the interaction of the audiovisual communication with the network quality of service (QoS). For the purpose of this Report, two sets of relevant Recommendations can be considered:

–   **Planning models**. These models allow the dimensioning of the network, by estimating the expected QoE that can be achieved given generic characteristics of the content (e.g., resolution, codec, compression bit rate) and the network (e.g., throughput, loss rate). Examples are ITU-T P.107 and the P.1070–1072 series.

–   **Monitoring models**. These models allow the prediction of the subjective quality as perceived by a user from the actual bitstream that would be received by the user device, either in compressed or in decoded format. Examples are the ITU-T P.1201–1204 series.

Those models are built on:

–   A target **QoE indicator** representing the subjective opinion of the users about the system. This QoE indicator is typically evaluated in terms of the mean opinion score (MOS) of a panel of users evaluating the system under test.

–   A set of **technical factors** which influence the QoE indicator. Those factors are related to:

•   **Implementation** of the system (e.g., image resolution, codec).

•   Restrictions on the **communication channel** (i.e., communication QoS indicators: throughput, loss rate, latency).

–   A **mathematical model** which relates variations in the technical factors (causes) with variations in the QoE indicator (effect).

When analysing other types of services, especially the ones related to professional usage and vertical use cases, additional considerations are needed.

First, the basic **QoE indicator** is normally not a MOS value rating aesthetic quality, but an indicator of the **effectiveness** of the system for its designed task. Each use case will have a different key quality indicator (KQI). Identifying this QoE indicator is key to being able to properly build a QoE analysis of the system.

Understanding the system **implementation** is also key for a proper analysis of the QoE. The systems covered by many Recommendations comprise a very limited number of potential systems: voice calls on telephones, video consumption on screens, and the like. However, vertical use cases may have very different setups, involving non-conventional streams (e.g., control signals) and devices (e.g., vehicles).

Restrictions on the **communication channel** still apply. A relevant difference with respect to over-the-top communication is that the 5G network **may guarantee certain levels of QoS**, according to the model described in the previous section. In this sense, network planning models may be more relevant than network monitoring ones, unlike in over-the-top scenarios.

Additionally, there is an additional set of relevant technical factors involving **energy consumption**, from two different perspectives. On the one hand, the energy efficiency on mobile devices (bits per joule) affects battery life, which might be more relevant than other performance indicators for mobility use cases. On the other hand, there is a trade-off between the complexity of the algorithms used to implement each use case (for compression, scene understanding, video processing, etc.) and the performance they offer. The complexity is also related to the energy consumption of those algorithms, either on mobile or on infrastructure computing power. Therefore, there is an energy-distortion trade-off that needs to be considered in the evaluation of the use cases.

Finally, it is worth noting that the development of mathematical models for QoE, either for planning or monitoring purposes, is out of the scope of this Report. Some QoE models have been proposed by the research community for some of the use cases, and they will be referenced where appropriate. In addition, this Report should provide guidelines to develop network planning and monitoring strategies for the mentioned use cases.

6.3 QoE assessment for verticals

The objective for QoE assessment is identifying how the variation of different conditions, either technical (e.g., video resolution or bitrate) or non-technical (e.g., the type of task) can affect the overall QoE. For this, subjective tests are performed to assess the response of a panel of subjects to the variation of such conditions.

QoE assessment should be done by the type of subjects that are typically going to use the system. From the perspective of subjective evaluation, subjects can be classified into (based on [ITU-T P.805]):

– **Untrained subjects (naïve)**. They are neither experienced in subjective testing methodology, nor are they experts in technical implementations of the system under test. They should have not participated in any subjective test in the previous 6 months.

– **Application experts**. They are experienced in subjective testing, and they are able to describe their subjective impressions in detail. However, they have neither a background in technical implementations of the system under test, nor detailed knowledge of the influence of these implementations on subjective quality.

– **System experts**. They have a background in the technical implementations of the equipment under test and detailed knowledge of the influence of particular implementations on subjective quality.

In most scenarios described in this Report, users will have deep knowledge of the use case, and therefore they will be system experts (if they also have knowledge of the underlying technical implementation) or application experts (otherwise).

In most scenarios, the relevant QoE indicator will not be a perceptual property, but a functional one: driving safety, medical situation evaluation, etc. This requires that the evaluation be carried out by the proper type of users (application or system experts), and the evaluation methodology reflect the actual use case.

Regarding the evaluation methodologies, there are two possible approaches:

–   **Active tests**: create an evaluation scenario where the actual task is performed.

–   **Passive tests**: evaluate pre-recorded videos of content relevant to the use case.

Active tests must be specifically designed for each use case by subject matter experts. The exact definition of the task is outside the scope of this Technical Report. However, several guidelines can be taken from existing ITU-T Recommendations, for different types of tasks:

–   **Recognition tasks** [ITU-T P.912]. This is appropriate methodology when there is a question whose answer can be objectively considered as correct or incorrect. The recommendation provides guidelines to several types of tasks, including how to perform statistical analysis of the results:

  •   Multiple choice. The video is shown above a list of verbal labels representing the possible answers.

  •   Single answer. There is an unambiguous answer to an identification question, e.g., alphanumeric character recognition.

  •   Timed task method. A viewer may be asked to watch for a particular action or object to be recognized in the video clip. When the viewer perceives that the target has occurred, a timer button can be pushed.

–   **Conversational tasks** are covered by several Recommendations: [ITU-T P.805] for voice conversations, [ITU-T P.920] for audiovisual conversations and [ITU-T P.1312] for multiparty voice telemeetings.

–   **Interactive tasks** can be based on [ITU-T P.809], design for the evaluation of interactive game scenes. Two types of test stimuli are considered:

  •   Short interactive (90-120 seconds). It is possible to assess the interaction quality (e.g., the impact of delay on the control), but the assessment of more complex player experience features highly depends on the player and the game content.

  •   Long interactive. It is reasonable to use a duration of 10-15 minutes to ensure that players get emotionally attached to a game scenario while aiming to measure emotions and other QoE aspects such as flow.

It is important to mention that these active test Recommendations will probably not completely cover the requirements of the use case that need to be tested. They can be used as guidelines to develop a subjective evaluation methodology for the use case, but other important aspects may need to be evaluated which exceed the scope of such Recommendations. Once more, a deep knowledge of the use case is required.

Passive tests should be based on conventional multimedia evaluation methodologies such as [ITU-T P.913]. The selection of source content and added impairments should be representative of the actual use case (e.g., a ToD test should use content captured from a vehicle under a representative set of driving conditions).

When characterizing a network for a particular video application, realism is critical and camera impairments are a major constraint. The subjective test must include sufficient variety of subject matter and environmental conditions. This can be done by using unrepeated scene experiment design as described in [Janowski]. For instance, an unrepeated scene experiment design can characterize the full range of camera impairments; a conventional matrix experiment design, where all combinations of source sequence and impairments are tested, cannot.

The result of passive tests must be taken with proper precaution. Application and system experts tend to provide severe judgement in the evaluation of systems with passive tests. However, in practice, they can adapt to bad quality conditions and still perform the desired task.

Nonetheless, passive tests are useful to understand the effect of the different technical factors into the perceived QoE, as well as to prototype active tests.

7   Tele-operated driving (ToD)

The availability of 5G networks is becoming a reality, and it brings the opportunity to support the deployment of new cooperative connected and automated mobility (CCAM) services. One of the most challenging CCAM uses is ToD. ToD can be seen as a side effect of tackling potential issues that autonomous driving cannot solve by itself. In this case human intervention might be required to drive the car. Such intervention can be feasible by making use of faster 5G infrastructure, a dedicated protocol, video data channels and cockpit setup under the supervision of a control centre.

Depending on the level of implication of the remote operator in the act of driving, different ToD types may be defined (Table 2) [5GAA XWG5-200029]. For the purpose of this Report, the types under consideration are those where part of or all of the dynamic driving tasks are performed by a remote driver on a sustained basis (ToD Type 2 or 3).

Table 2 – ToD taxonomy. Source: [5GAA XWG5-200029]

ToD type (role of ToD operator when engaging in the act of driving)	Act of driving		
	Strategic operation (travel planning, route and itinerary selection)	Dynamic driving task (DDT)	
		Tactical operation (object and event detection and response (OEDR))	Operational operation (sustained lateral and longitudinal vehicle motion control)
0 (No role)	In-vehicle user or system	In-vehicle user or system	In-vehicle user or system
1 (Dispatcher)	ToD operator	In-vehicle user or system	In-vehicle user or system
2 (Indirect controller)	ToD operator	ToD operator	In-vehicle user or system
3 (Direct controller)	ToD operator	ToD operator	ToD operator

Although several implementations of the ToD service are possible, a typical reference architecture assumes that there are four cameras boarded on the vehicle, one at each side, as well as other sensors (positioning system, radar, lidar, etc.). The information coming from those sources is processed by an on-board unit (OBU) and sent through a 5G network to the remote driver cockpit. The cockpit has several screens, or maybe a virtual environment using VR, where the video from the cameras is displayed, as well as the information from the sensors. Additionally, the cockpit contains a remote control (driving wheel, throttle) whose information is also sent back to the car.
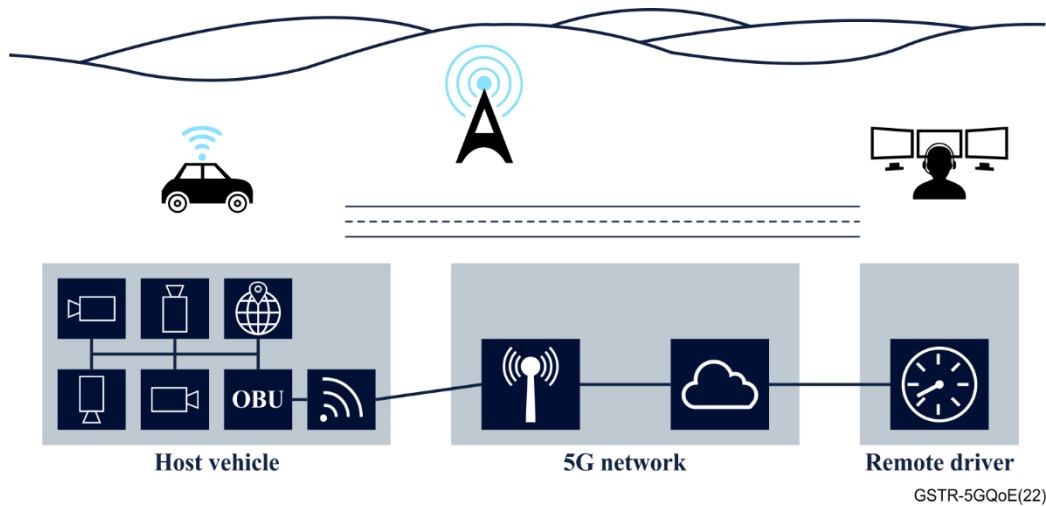
Figure 2 – Example architecture of a ToD system

7.1 Service implementation

Typical communication performance requirements, for vehicle speed < 50 km/h, are described in Table 3 [5GAA T-200xxx], [Pérez 2021]. Service latency is critical, and it covers the video and data stream (UL) and the messages from the ToD operator to the host vehicle (DL). The uplink data rate accounts for 1 to 4 cameras, plus additional telemetry data. Note that UL latency includes video encoding plus decoding delay, which might be a significant part of the end-to-end latency budget.

Table 3 – Service KPIs for tele-operated driving

ToD type	Latency (UL)	Latency (DL)	Data rate (UL)	Data rate (DL)
2	100–200 ms	200 ms	4–36 Mbit/s	25 kbit/s
3	100–200 ms	20–100 ms	4–36 Mbit/s	400 kbit/s

It is assumed that these services are going to be provided on a dedicated spectrum. Due to the requirements for throughput, density and coverage, they will probably need to be deployed on the midbands, allocating a minimum of 30 MHz for urban areas and up to 100 MHz for rural areas [5GAA S-200137]. Specific enhancements of 3GPP standards are being proposed to support this, as well as other V2X scenarios [3GPP TS 23.287].

Table 4 shows the relevant 5QIs for ToD services, from those provided in [3GPP TS 23.501]. The UL, whose most relevant information is video, should use the 5QIs proposed for real-time video transmission: 2, 3 or 7. The DL, where the information is mostly related to control commands, should use V2X messaging 5QIs: 79 or 85.

Table 4 – Relevant 5QIs for tele-operated driving

5QI Value	Resource type	Default priority level	Packet delay budget	Packet error rate	Default maximum data burst volume	Default averaging window	Example services
2	GBR	40	150 ms	10^{-3}	N/A	2 000 ms	Conversational video (live streaming)
3		30	50 ms	10^{-3}	N/A	2 000 ms	Real-time gaming, V2X messages
7	Non-GBR	70	100 ms	10^{-3}	N/A	N/A	Video (live streaming) interactive gaming
79		65	50 ms	10^{-2}	N/A	N/A	V2X messages
85	Delay-critical GBR	21	5 ms	10^{-5}	255 bytes	2 000 ms	V2X messages (remote driving)

7.2 Relevant QoE indicators

Tele-operated driving is a task-based media experience, whose quality is mostly evaluated through the ability of the remote driver to safely control the vehicle. Safety is therefore the most relevant driver for ToD QoE.

In more detail:

– Video quality and latency have safety impacts on ToD with human remote operators.

– The relation between video quality and ToD performance is complex and should be studied.

– Video quality and latency must be addressed as part of the functional safety standardization.

– Different ToD types may have different video quality requirements.

– There are strong influences from the environments and the vehicle speed. For instance, it is not the same to deploy tele-operated valet parking as being able to perform ToD on a main road.

The relationship between network throughput, latency and losses can be used to build a parametric planning model for ToD, such as the one proposed in [Pérez 2021].

7.3 Key factors to evaluate user QoE

Two kinds of factors are critical to evaluate the QoE for ToD services: technical factors, regarding the specificity of the data streams that are used for the tele-operation, and human factors.

From the technical perspective, the main components of the final QoE are:

– The end-to-end latency, including coding and network delays;

– The quality of the video signal from the cameras, including the coding quality and the effect of transmission errors;

– The quality of the sensors and control signals, especially the potential effect of transmission errors.

The end-to-end latency is critical to a safe ToD execution. The tolerance to latency is directly related to the driving speed and the safety distance between the vehicle and any possible obstacle or hazard. The higher the latency, the lower the maximum speed and the higher the required safety distance.

Three factors are relevant for the evaluation of the video quality for a given camera: the speed of driving, the camera location and the current manoeuvre in which the car is involved. The speed of driving affects the video quality because differences between one frame and the next increase with speed, so the compression algorithms require more bandwidth to maintain the same video quality. Camera location is also critical for quality as lateral cameras are less relevant for some type of driving and also present more differences between one frame to the next compared with the front camera, so we need to consider the relevance of the camera information and the quality that can be provided for a given bitrate. Finally, manoeuvres sometimes imply that laterals or rear cameras become more relevant in some scenarios.

The quality of the rest of the data streams (sensors and actuators) will be strongly dependent on the implementation of each ToD system.

Regarding the subjective assessment of QoE (human factors), there are two elements to be considered:

–   Evaluation of ToD systems must be based on a task-oriented methodology, i.e., the relevant quality is not related to aesthetics or perceptual fidelity, but to the ability of the remote driver to execute the ToD task in a safe manner.

–   The subjective experience of the driver will be influenced by the operation cockpit. Different options (e.g., multiscreen vs VR implementations) may show different QoE results for a similar task and signal quality.

8   Wireless content production (WCP)

Wireless content production covers several uses cases and broadly falls under the following categories: newsgathering, live broadcast of events, live audio in production and non-live production (media file transfer). Non-live production will not be covered by this Technical Report.

Newsgathering is vital for broadcast companies worldwide, offering news coverage of any kind of event that may interest the public. This includes planned events such as those related to politics and the economy or disasters that occur without prior notice and therefore cannot be planned for, and reporters may have to capture audio and video data to be sent back to production facilities as an event unfolds. Typically, a single camera or multiple cameras simultaneously feedback to a curator in a central production facility, depending on the circumstance [3GPP TR 22.827], [EBU TR 056].

Live audiovisual production (LAVP) involves capturing and distributing both video and audio data to an audience in real-time. There are different scales of live video production. These range from using mobile handsets on social media to live contribution using a single camera up to the typical multicamera broadcast of a high-profile event (including outside broadcast). 5G is one of the enablers that can add value to a production budget. For example, in remote live TV production, wireless cameras and microphones could be deployed in a political chamber to provide an automated feed of a meeting or a small cultural venue to cover an arts event. In large sporting and cultural events such as football matches or musical concerts, where many cameras are needed, you can have a mix of both wireless and wired cameras, wireless microphones, telemetry and remote control. Currently most of these sources will be mixed locally on-site, but in both cases it is desirable that captured data is sent to a data centre or studio where live content selection can occur, and 5G may offer a reliable connection [3GPP TR 22.827], [EBU TR 056].
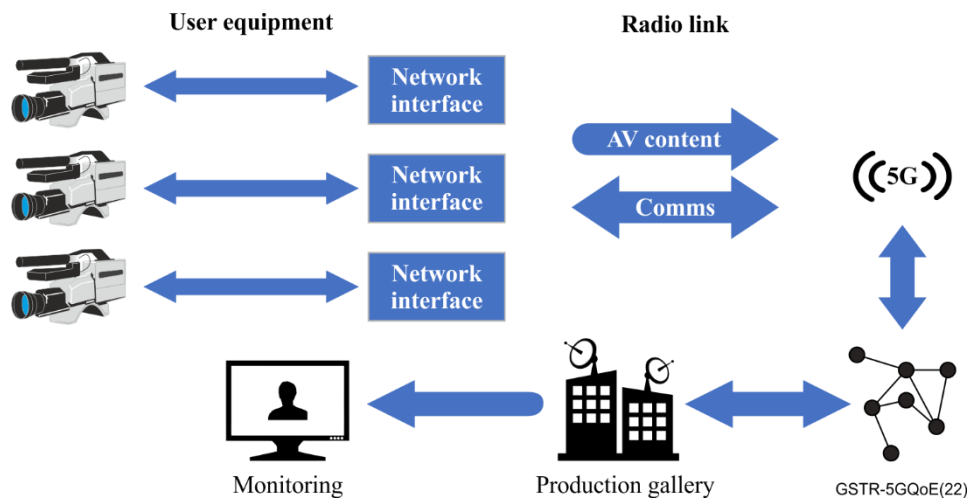
Figure 3 – Example of a multicamera source use case for live video production [EBU TR 056]

Live audio in production (LAiP) is relevant for events such as concerts or theatrical events where one or more artists perform before an audience. Producing and capturing a live event for live programme audio or subsequent use of the cultural and creative content involve many wireless audio streams. The technical crew, the production team and the security staff are normally backstage ensuring the successful realization of the live event and are normally connected via an intercom system using microphones, loudspeakers and in-ear monitoring. For the live audio content, reception of the wireless communication service is only required locally, limited to the event area, and all audio processing such as audio mixing is done in real-time during operation. A number between 5 and 300 simultaneously active wireless audio links can be expected. Each wireless audio source is streamed to or from a central audio mixing console. Therefore, the setup and local network must be fast to optimize the usage of resources (equipment, radio spectrum, working time of the artists and crew) [3GPP TR 22.827], [EBU TR 056].

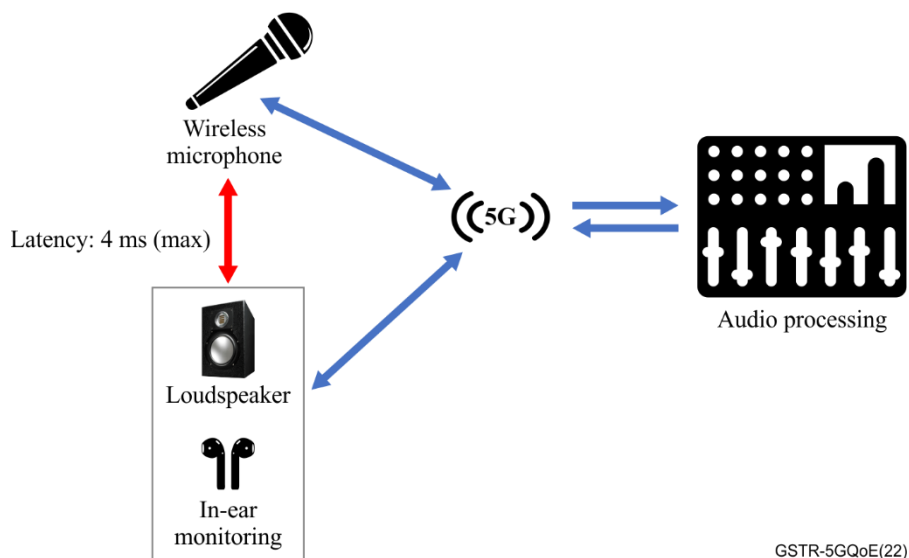**Figure 4 – Live audio in production use case [EBU TR 056]**

8.1 Service implementation

WCP services are provided by connecting different media production elements (microphones, IEMs, cameras, etc.) to the 5G network. A wide range of implementations is possible, starting from connecting the system to the commercial mobile network (a typical case for newsgathering) to using dedicated slices or NPNs.

Table 5 shows a set of KPIs for different use cases. When applicable, it is based on a reference configuration using NR midband (3.8 GHz) with 100 MHz of bandwidth, which would result in around 120-200 Mbit/s of UL [3GPP TR 22.827]. This limits the maximum UL throughput that can be considered for the service.

In this respect, for LAVP, codec decision is critical. I-frame-only could exceed the target data rate, IBBP is too much latency, so a form of IPP is probably required. The impact of network errors on the video stream is dependent on codec decision and group of pictures (GOP) length, since packet sizes are constant [ITU-R BT.2137-0].

For LAiP, the quality of the audio is dependent on the different user data rates per audio stream that need to be supported for different audio demands (e.g., compressed vs. uncompressed audio). For example, 5 Mb/s is a target maximum desired for studio usage (24 bit @ 192 kHz).

A relevant KPI for all services is end-to-end latency, even though requirements vary strongly between services. For newsgathering, it is enough to have an audio return channel from the news curator allowing two-way conversation. For LAVP, low and consistent latency is critical, so that a constant frame rate can be achieved, as well as frame-level synchronization among all the cameras. In LAiP, latency requirements are even more strict to guarantee minimum mouth-to-ear delay for the performers.

Table 5 – Service KPIs for wireless content production

Use case	Number of UEs	Latency (RTT)	Synchro/ jitter	Data rate per UE (UL)	Data rate per UE (DL)	PER
Newsgathering	1	1 000 ms	100 ms	15-50 Mbit/s	0.1-0.3 Mbit/s	10^{-6}
LAVP	1–10	100 ms	6 ms	80-200 Mbit/s	20 Mbit/s	10^{-8}
LAiP	5–300	4 ms	0.01 ms	0.1-5 Mbit/s	0.1-5 Mbit/s	10^{-6}

Table 6 shows the relevant 5QIs for WCP services, from the ones provided in [3GPP TS 23.501]. Depending on the specific use case, different 5QIs will apply. Newsgathering will use less restrictive QoS requirements for the UL, e.g., the ones defined for non-conversational video (4, 8, 9), while requiring voice only for the DL (e.g., 1). LAVP has more demanding QoS requirements, mostly due to the latency limits. 5QIs designed for conversational video or gaming could be applied. Finally, LAiP requirements are so far not supported by any standardized 5QI.

Table 6 – Relevant 5QIs for wireless content production

5QI Value	Resource type	Default priority level	Packet delay budget	Packet error rate	Default maximum data burst volume	Default averaging window	Example services
1	GBR	20	100 ms	10^{-2}	N/A	2 000 ms	Conversational voice
2		40	150 ms	10^{-3}	N/A	2 000 ms	Conversational video (live streaming)
3		30	50 ms	10^{-3}	N/A	2 000 ms	Real-time gaming
4		50	300 ms	10^{-6}	N/A	2 000 ms	Non-conversational video (buffered streaming)

Table 6 – Relevant 5QIs for wireless content production

5QI Value	Resource type	Default priority level	Packet delay budget	Packet error rate	Default maximum data burst volume	Default averaging window	Example services
6		60	300 ms	10^{-6}	N/A	N/A	Video (buffered streaming)
7	Non-GBR	70	100 ms	10^{-3}	N/A	N/A	Voice, video (live streaming) interactive gaming
8		80	300 ms	10^{-6}	N/A	N/A	Video (buffered streaming)
9		90					

8.2 Relevant QoE indicators

WCP covers a set of use cases where connectivity between production devices and the production centre is provided via a 5G network. In most cases, it replaces (or complements) previous solutions, based either on wired connectivity or on ad hoc wireless technologies. Using 5G mobile technology increases the flexibility of the system, making it possible to deploy production pipelines in locations where it was not possible before, or with a more cost-effective solution. In such cases, this may be at the expense of the audiovisual quality. Balancing these factors is quite critical for the QoE evaluation of these services.

For instance, connectivity is the most important factor for newsgathering QoE. Another factor is the quality of the captured audio and video data. However, connectivity takes priority over the visual aesthetics, and therefore, quality compromises can be made. Latency is also a critical factor with two-way communications needed between studio and location services. Increasingly there is a need to control remote contribution facilities and this requires lower latency connectivity. Ideally latency should be fixed and not drift over time and be able to be synchronized across multiple sources.

In LAVP, quality is evaluated through ability of the production team to deliver a visual experience for broadcast, based on the programme's running order. This can be impacted by the following:

Video quality and latency from capture sources.

– Multicamera feeds synchronization. Camera feeds have to be synchronized to ensure switching and/or composing mismatches.

– Different LAVP implementations may have different video quality requirements. Typical wired solutions use uncompressed video but where wireless links are used they need to meet minimal contractual quality as set out by broadcast regulators.

– Camera accessories such as the lens also have an impact on video quality.

– Often there is a contractual or legal obligation to provide a level of quality. Broadcasters can be penalized for breaking these obligations and may lose future rights if the quality of the broadcast is not seen to be high enough by a rights holder.

In LAiP, the two most critical requirements are reliability of connection and mouth-to-ear latency between microphone and the in-ear monitoring (IEM) equipment used by the artists or presenters. The tolerance for disturbance to QoE for this application is extremely low as there is no opportunity of recovery (e.g., no possibility to ask the singer to repeat).

8.3 Key factors to evaluate user QoE

The two main factors to evaluate for the QoE of WCP services are technical factors and human factors. From the technical perspective, the main components of the final QoE are:

– Reliable connectivity;

– Audio and video signal quality;

– For LAVP: low and, more importantly, predictable end-to-end latency and multicamera synchronization;

– For LAiP: extremely low mouth-to-ear latency.

Since WCP covers live content for professional use, reliable connectivity is the most critical element. Fast and efficient deployment of communication links are crucial for newsgathering. This is directly related to the location of the audio and video capture and the connectivity within the area. So reduced bandwidth links should be taken into consideration, to emulate areas with poor reception capacity. For LAVP and LAiP, reliability is particularly crucial for the successful coordination of activities of different members of the production.

Video and audio quality are also critical, particularly for LAVP and LAiP. In the case of newsgathering, since connectivity is crucial, quality compromises can be made using compression algorithms to ensure the video can be easily transmitted using the bandwidth available for upload at the location of capture.

As we move to IP based production solutions there are also requirements for return path AV at lower bandwidth and control signals such as tally or pan-tilt-zoom (PTZ) control.

To evaluate video quality for a given camera, the following factors should be considered: image sensor size, sensor resolution, video raster resolution, frame rate, digital signal processing and optical path. While the image sensor size on any given camera is fixed, the other factors can be easily changed and thus affect the perceptual quality of the resulting video. Objective metrics are desirable to avoid the costs of subjective assessments and/or expert viewing. These are defined for Europe in [EBU R 118].

In a multicamera feed setup, capture-to-production latency will have a direct impact on audiovisual and camera synchronization and this could lead to audio and video mismatch, multiple feed mismatch and compositing artefacts.

In a live production, the end-to-end latency that can be tolerated by a live performer between their microphone and IEM typically ranges from 0.125 to 2 ms and cannot exceed 4 ms.

Regarding the subjective assessment of QoE (human factors), there are two elements to be considered:

– The QoE of the system is provided to content production professionals: news curators, video producers, performers, etc. Therefore, QoE evaluation should be carried out by such professionals.

– Although conventional quality evaluation methodologies can be of use (e.g., ITU-T P.800, P.910, P.913, etc.), a task-oriented methodology could provide better ecological validity, e.g.:

  • The ability of the remote curator to weave a storyline from audio and video data received from new sources to fit in with other parts of the editorial workflow.

  • A live vision engineer confirming the quality of the incoming video feed before passing on to be made available for selection in a live broadcast by editorial production staff.

  • The ability of the members of the production team and live performers to be able to communicate effectively.

In addition, objective metrics would be desirable to avoid the costs of subjective assessments and/or expert viewing.

9 Mixed reality offloading

The fast-paced development of mixed reality (MR), composed of augmented reality (AR) and virtual reality (VR) technologies, have raised the computation demands of MR applications. High rendering resolution is not enough to provide the users with fully immersive experiences; other novel algorithms such as hand tracking, occlusion handling or object recognition are becoming fundamental for both AR and VR. However, robust examples of such algorithms require high-end GPU-enabled computing platforms which hinder the goal of reducing the size and weight of current state-of-the-art AR/VR devices. The global deployment of 5G networks enables the chance of offloading some or all of these heavy-duty algorithms to a multiaccess edge computing (MEC) platform. Successful MEC-delivered AR/VR can lead to lighter, more comfortable and more affordable devices.

Mixed reality offloading is, then, a solution which, using 5G-network capabilities, enables running state-of-the-art MR algorithms while wearing a lightweight head-mounted display (HMD) –an item of edge-dependent UE [3GPP TR 26.998]. This approach is also called XR (extended reality) distributed computing [3GPP TR 26.928].

MR offloading assumes that the immersive environment where the application takes place is either the real world around the user (AR) or a completely virtual environment (VR). There is an alternative scenario where the immersive environment is a real-time capture of a remote location, sent via a 360-degree videostream or similar means, on to which virtual elements are added. This scenario, called augmented telepresence (AT) [Dima] or distributed reality (DR) [Villegas], is not covered by this Technical Report.

9.1 Service implementation

AR/VR applications run at rates of 60 Hz and above. These rates set severe latency constraints that must be fulfilled to avoid visual artefacts or user simulator sickness that can completely ruin the experience. Processing latencies have different effects in AR and VR:

– In AR, processing latencies lead to virtual objects drifting, having incorrect positions and dynamics. Besides, AR demands synchronized coherence between the real and virtual worlds.

– In VR these latencies lead to very severe simulator sickness which might force the user to leave the experience. In some VR applications in which the real world is incorporated (e.g., users' hands are segmented and rendered in the virtual scene) these latencies also produce incorrect drifts between the real and virtual content.

Ideally, the maximum delay between the sensor capture process and the rendered frame should be bounded to the frame update period (16.6 ms for 60 Hz). Besides, recent MR devices incorporate several high-definition cameras, including both RGB and depth feeds, which generate a wide stream of sensor data. Consequently, any successful MEC-delivered AR/VR will require a careful network design to fulfil with the latency and data rates. We propose a set of KPIs from a thorough mathematical analysis using two different scenarios for both AR and VR.

A) Full offloading scenario: the sensor feed (RGB and depth) is streamed to the MEC and processed. The new ultra-high-definition frame is rendered, sent and displayed on the device.

B) Algorithm offloading scenario: a machine-learning-based heavy-duty algorithm is offloaded from the device. In this scenario, a part or the whole of the sensor feed is streamed to the MEC and processed, and the result is sent back to the device. The result is used to render the new frame. For simplification, we consider this result to be a low-definition segmentation mask, common in scene recognition or hand segmentation algorithms.

Table 7 shows the KPIs for both scenarios. More detail on the derivation of such KPIs is presented in Annex A. Note that KPIs are aligned with those described in [3GPP TS 22.261].

Table 7 – Service KPIs for mixed reality offloading

Scenario	Latency (RTT)	Average data rate (UL)	Average data rate (DL)	Peak data rate (UL)	Peak data rate (DL)
A	2-3 ms	180 Mbit/s	1 000 Mbit/s	1 250-1 700 Mbit/s	2 800-3 000 Mbit/s
B	3.5-6 ms	180 Mbit/s	18 Mbit/s	750-900 Mbit/s	250-300 Mbit/s

The estimated KPIs impose tight demands which require a well-designed configuration. The latency requirements impose a need for the MEC to be placed as close to the gNB as possible, avoiding any backhauling connectivity. User prioritization is also a key component in a successful AR/VR offloading architecture: packets should be allocated as soon as they become available to provide high-throughput and low latency transmissions. Besides, the gNB should be configured with at least a subcarrier spacing of 120 KHz and a bandwidth of 400 MHz. Consequently, high-frequency bands are the most suitable for the proposed offloading scenarios: mmWave is a key enabler in MEC-enabled AR/VR. Figure 5 shows a possible network architecture for MR offloading.

GSTR-5GQoE(22)
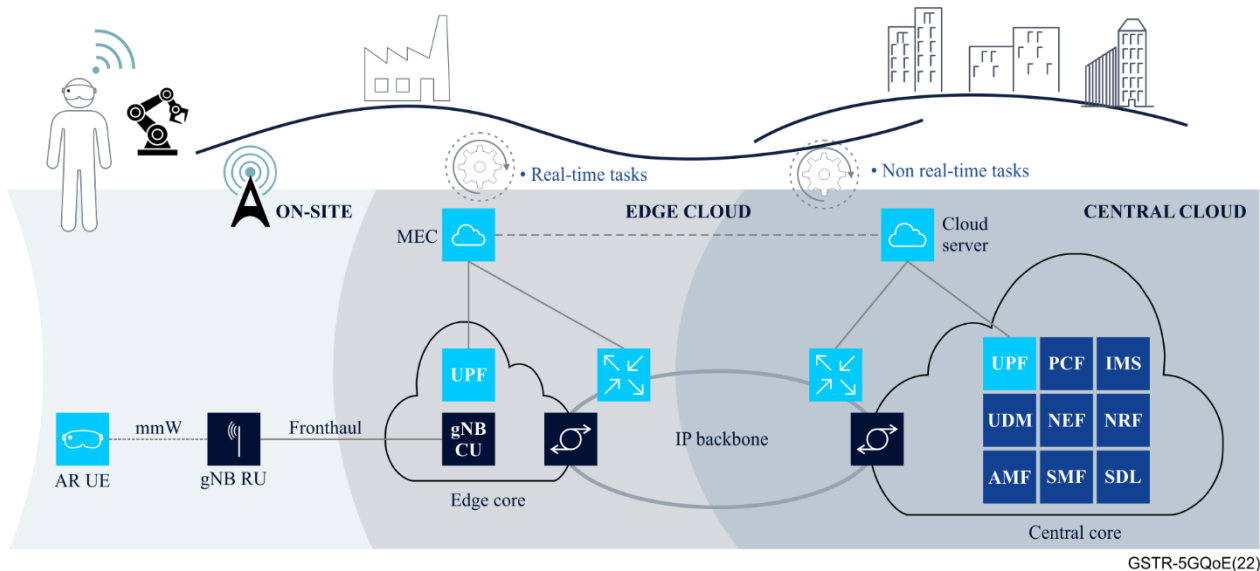
Legend: AMF – Access and Management mobility Function; IMS – IP Multimedia Subsystem; NEF – Network Exposure Function; NRF – Network Repository Function; PCF – Policy Control Function; SDL – Shared Data Layer; SMF – Session Management Function; UDM – United Data Function.

Figure 5 – Proposed 5G architecture for MR offloading

Table 8 shows the relevant 5QIs for MR offloading of those provided in [3GPP TS 23.501]. As MR is one the critical drivers for 5G, specific 5QIs have been defined (80, 87–90). It is worth noting that, even if they are the most restrictive ones so far defined in terms of PDB, they might be insufficient to fulfil the requirements of the scenarios as we have defined them.

Table 8 – Relevant 5QIs for mixed reality offloading

5QI value	Resource type	Default priority level	Packet delay budget	Packet error rate	Default maximum data burst volume	Default averaging window	Example services
80	Non-GBR	68	10 ms	10^{-6}	N/A	N/A	Low latency eMBB applications AR
87	Delay-critical GBR	25	5 ms	10^{-3}	500 bytes	2 000 ms	Interactive service – motion tracking data
88		25	10 ms	10^{-3}	1 125 bytes	2 000 ms	Interactive service – motion tracking data
89		25	15 ms	10^{-4}	17 000 bytes	2 000 ms	Visual content for cloud/edge/split rendering
90		25	20 ms	10^{-4}	63 000 bytes	2 000 ms	Visual content for cloud/edge/split rendering

9.2 Relevant QoE indicators

The QoE in both AR and VR is measured in terms of the level of immersion into the virtual experience [ITU-T G.1035]: how the augmented objects are perceived as part of the real-world experience (AR) or, similarly, whether the user feels fully immersed in the virtual environment (VR).

However, the immersion is related to several aspects of the overall AR/VR experience:

– The quality and resolution of the rendered frames or augmented elements directly affect the immersion.

– The responsiveness during the interaction with virtual objects and 3D UIs is closely related to the QoEs. This indicator is crucial in the applications in which the user has to accomplish a specific task.

– The correct alignment and relative positions of the virtual and real objects, including the correct resolution of occlusions, is a key factor of AR applications.

– The sense of embodiment, which measures how the user perceives himself within the virtual experience, is a key factor of any VR application.

– Motion-to-photon latency is another key indicator for the quality of the experience.

– The manner each of these metrics affects the experience is tightly related to the specific use case or application.

9.3 Key factors to evaluate user QoE

The correlation between the network KPIs and the user QoE can be roughly achieved by directly measuring the performance of the network. Network KPIs can serve as proxy to estimate certain factors of the end user's QoE.

Two-way end-to-end latency is the most critical technical factor influencing MR offloading QoE. Latencies above a modelled threshold indicate:

– High motion-to-photon latency, which causes simulator sickness and considerably degrades the experience in both AR and VR.

–   Low responsiveness in human–virtual interaction.

–   Decreased feeling of embodiment in the cases in which the hand/body tracking or segmentation algorithms are offloaded.

End-to-end latency is influenced by network round-trip-time, peak UL and DL throughputs, and coding and processing delay.

Additionally, mean network throughput directly affects the number and quality of visual elements that are rendered in the MEC and streamed to the end device.

Besides, the level of QoE degradation is tightly correlated to the type of MR application. To estimate a highly detailed QoE-KPI correlation, is necessary to evaluate the QoE indicators in a subjective manner. The relation between the QoE indicators and the network KPIs is application-dependent, as is is the evaluation methodology. As an example, the QoE indicators might not be the same in an industrial training task as in an immersive video game.

10   First responder networks

A relevant use case of 5G networks is the ability to provide connectivity to first responders attending emergency situations. Whenever a medical or security emergency occurs, an effective connectivity and coordination of the different teams involved may make the difference in saving lives. To support the work of the emergency services, this communication should be provided by dedicated network resources, either specific frequency bands or prioritized network slices.

Figure 6 shows an example of a first responder network. In general, an emergency situation requires coordination between on-site teams (e.g., first responders in an ambulance), infrastructure elements (e.g., closed circuit television (CCTV)) and central stations (e.g., a hospital). Different kinds of first responders and public safety officers may have different requirements, depending on the emergency scenario. For instance, attending a traffic accident, a natural disaster or a criminal threat will involve different personnel and systems.

For the purpose of this Report, we will use two specific examples:

–   On-site medical emergencies: communication between an ambulance and the hospital.

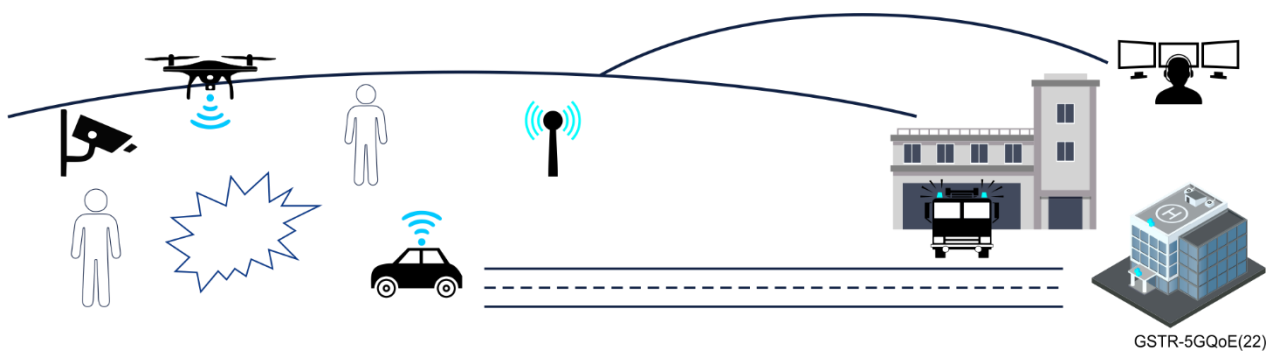–   Infrastructure elements: deployment of CCTV for live monitoring of an area (e.g., railway infrastructure).

GSTR-5GQoE(22)

Figure 6 – Elements of a first-responder network

In a medical emergency multiple data flows are transmitted from the emergency site or ambulance to the hospital (UL). Typically, voice from emergency room staff or medical specialists in the hospital guiding operations can be transmitted in the DL. Video can be used for the DL to visually show paramedics how a specific operation should be performed. Access to medical records in the hospital or national health system may be required (also DL).

Figure 7 (from a demonstration that was part of the CONCERTO European project) illustrates the different sources of data flows in a connected ambulance.

Figure 7 – Demonstration in the CONCERTO EU project, emergency services at the Hospital of Perugia, Italy [Martini]

CCTV systems are used on various parts of the rail infrastructure including intercity rail or metro rail stations, trains and at car parks [RDG V4]. They are usually deployed for a variety of reasons including:

– Crime prevention and prosecution;

– Investigation of rail related accidents and incidents;

– To reassure staff and members of the public;

– To aid decisions on train movements;

– To meet statutory requirements and obligations.

5G is one of the enablers that can add value to CCTV systems and workflows. For example, wireless CCTV cameras can be easily deployed in various locations of the rail infrastructure. In addition, it can provide both the bandwidth and reliable connectivity needed for real-time very-high-quality video remote monitoring of rail and non-rail assets. Currently, most rail CCTV systems are deployed with local storage and then transferred later when required bandwidth is available.

10.1 Service implementation

In a medical emergency scenario, there are several uplink flows that need to be transmitted from the incident scenario or the ambulance (UL):

– Ambient video, from camera on ambulance and from body worn/head mounted cameras from paramedics. High resolution video is required for specific tasks, such as the observation of eyes.

– Medical data, such as ECG or ultrasound video. The specific case of ultrasound video is important because it requires high data rate. Frame rate is typically 25-40 fps, but higher temporal resolution is required to identify specific cardiac mechanical events: specific applications of ultrafast cardiac imaging [Henry] would need 500-600 fps. The typical space resolution is 640×480 pixels, but higher resolutions are also being considered (e.g., 1280×768, 1400×1050). This results in an UL data rate from 92 Mbit/s (uncompressed ultrasound

video), which can be reduced with compression, e.g., starting from 658 kbit/s (HEVC medium quality) [Razaak].

In the DL direction, fast access to patients' medical records from a hospital or via the national health system is also required.

There are different sub-scenarios:

–   Data transmission from the emergency site.

–   Data transmission from the ambulance, while in motion (high speed possible).

–   Guided intervention from the hospital. In this scenario, an external expert (e.g., a specialist in the hospital) needs to support a paramedic in real-time. Besides the mentioned flows, it requires low-latency bidirectional communication.

A typical deployment of CCTV for railway infrastructure monitoring will contain a dense number of IoT cameras. Each camera will produce an upstream video flow, starting from a low resolution and low frame rate (e.g., 640×480 pixels, 1 fps) to a high or ultra-high-definition at 30 fps. This translates into a data rate range from about 200 kbit/s up to 20 Mbit/s. The DL rate is negligible.

Table 9 – Service KPIs for first responder networks

Use case	Number of items of UE per cell	Latency (RTT)	Data rate per item of UE (UL)	Data rate per item of UE (DL)	PER
Medical emergency	1–5	100–1 000 ms	10–100 Mbit/s	0.1–0.3 Mbit/s	10^{-8}
Wireless CCTV	5–50	1 000 ms	0.1–20 Mbit/s	–	10^{-6}

Different network configurations may be applied to the described scenarios. However, first responder networks require coverage over wide areas, either permanently (CCTV monitoring of a large infrastructure) or temporarily (medical emergency). For such reason, low to mid frequency ranges are proposed (900 MHz to 2.8 GHz).

Some first responder use cases include specifically regulated frequencies and even dedicated network services. An example of this is mission-critical push-to-talk (MCPTT), which provides "walkie-talkie" videoconference functionality on top of a cellular network. MCPTT is not specifically covered in this Report, although it presents similarities with the described use cases.

Since the first responder network use case is particularly complex, a wide range of 5QIs are applicable to it (see Table 10). They can be divided into three groups:

–   5QIs regarding video communication, for the UL of the medical emergency scenario: 2, 7; and voice communication for the DL: 1, 7.

–   Specific 5QIs for push-to-talk communication and mission-critical video (65, 66, 67, 69, 70), specifically designed for first responder communications.

–   5QIs for buffered video, to connect wireless CCTV cameras: 8, 9. If the real-time access of this video became critical during an emergency scenario, then the traffic should be moved to conversational video (or even mission critical) 5QIs.

Table 10 – Relevant 5QIs for first responder networks

5QI Value	Resource type	Default priority level	Packet delay budget	Packet error rate	Default maximum data burst volume	Default averaging window	Example services
1	GBR	20	100 ms	10^{-2}	N/A	2 000 ms	Conversational voice
2		40	150 ms	10^{-3}	N/A	2 000 ms	Conversational video (live streaming)
65		7	75 ms	10^{-2}	N/A	2 000 ms	Mission-critical user plane push-to-talk voice (e.g., MCPTT)
66		20	100 ms	10^{-2}	N/A	2 000 ms	Non-mission-critical user plane push-To-talk voice
67		15	100 ms	10^{-3}	N/A	2 000 ms	Mission-critical video user plane
7	Non-GBR	70	100 ms	10^{-3}	N/A	N/A	Voice, video (live streaming)
8		80	300 ms	10^{-6}	N/A	N/A	Video (buffered streaming)
9		90					
69		5	60 ms	10^{-6}	N/A	N/A	Mission-critical delay sensitive signalling (e.g., MC-PTT signalling)
70		55	200 ms	10^{-6}	N/A	N/A	Mission-critical data (e.g., example services are the same as 5QI 6/8/9)

10.2 Relevant QoE indicators

First-responder use cases are strongly focused on the required task. For those described in this Report, this task is either the coordination of the attendance of patients or the surveillance of the railway infrastructure. The quality of the system is evaluated based on ability of the first responders to make swift decisions, based on video and data provided.

This can be impacted by the following:

– Video resolution and compression quality.
– Video Frame Rate is key in two scenarios:
 • Ultrasound video, as described above.

- CCTV monitoring of fast-moving trains. Lower frame rates provide slow sampling rates especially for fast-moving objects. For example, a 15 fps forward facing camera running on a 11 m/s train can only sample objects within 0.7 m [Basilio].

– Camera capture conditions may be challenging, especially on moving scenarios (trains, ambulances) and outdoors: camera stability, weather and light conditions, etc.

– Streaming from vehicles at high speed has impact on several parameters of the access network quality (cell changes, beamforming, coherence time, etc.).

10.3 Key factors to evaluate user QoE

From the technical perspective, the main components of the final QoE are:

– Capture-to-viewing latency, including coding and network delays.

– The quality of the video signal from the cameras and sensors, including spatial resolution, frame rate, coding quality and the effects of transmission errors.

– For the specific use case of guided intervention from a hospital, the two-way latency of the whole interaction loop.

In a multicamera CCTV feed setup, these components are crucial for an effective execution. Capture-to-viewing latency will have a direct impact on camera synchronization and this could lead to multiple feed mismatch and delay. To evaluate the video quality for a given camera, the following factors should be considered: image sensor size, sensor resolution, frame rate, digital signal processing and optical path. While the image sensor size is fixed on any given camera, the other factors can be easily changed and thus affect the perceptual quality of the resulting video.

Objective metrics are desirable to avoid the costs of subjective assessments and/or expert viewing.

Regarding the subjective assessment of QoE: Evaluation of first responder systems must be based on a task-oriented methodology. Due to the high level of specialization of the users of these systems, experts are required in all the phases of the subjective evaluation: design, execution and evaluation. A comparative study of the methodologies used for subjective medical image quality assessment can be found in [Lévêque].

Annex A

Derivation of KPIs for mixed reality offloading

Assuming that a single RGB and single depth feeds are compressed and transmitted, after compression, each pixel is estimated to weigh one bit. These assumptions are used to estimate the uplink and downlink frame sizes summarized in Table A.1.

Table A.1 – Uplink and downlink streams weights for scenarios A and B

Scenario A & B			
Uplink			
Data feed	*Frame size*	*Frequency*	*Mean throughput*
RGB 1080p	2 Mbit		120 Mbit/s
Depth 1024 × 1024	1 Mbit	60 Hz	60 Mbit/s
Total	3 Mbit		180 Mbit/s
Scenario A			
Downlink			
Data feed	*Frame size*	*Frequency*	*Mean throughput*
Rendered left eye 4K	8.3 Mbit		498 Mbit/s
Rendered right eye 4K	8.3 Mbit	60 Hz	498 Mbit/s
Total	16.6 Mbit		996 Mbit/s
Scenario B			
Downlink			
Data feed	*Frame size*	*Frequency*	*Mean throughput*
Low resolution mask	0.3 Mbit	60 Hz	18 Mbit/s

Using the UL and DL stream sizes shown in Table A.1, a simple optimization algorithm is applied to estimate an initial set of network latency and peak throughput KPIs for both scenarios. The estimated KPIs are included in Table A.2.

**Table A.2 – Set of peak throughput and round trip latencies KPIs
for the offloading scenarios A and B**

Scenario A					
Size	*Data rate*	*Processing – 5 ms*		*Processing – 10 ms*	
		Peak throughput	*Latency*	*Peak throughput*	*Latency*
3 Mbit	180 Mbit/s	~1.25 Gbit/s	~3 ms	~1.7 Gbit/s	~ 2 ms
16.6 Mbit	996 Mbit/s	~2.8 Gbit/s		> 3 Gbit/s	
Scenario B					
Size	*Data rate*	*Processing – 5 ms*		*Processing – 10 ms*	
		Peak throughput	*Latency*	*Peak throughput*	*Latency*
3 Mbit	180 Mbit/s	~ 0.75 Gbit/s	~6 ms	~0.9 Gbit/s	~3.5 ms
18 Mbit	18 Mbit/s	~ 0.25 Gbit/s		~0.3 Gbit/s	

————————