TELECOM INFRA PROJECT®

MRN PG Report

QoE/QoS Measurement Framework

Approach to QoE Engineering

V1.1

Table of Contents

This technical report results from a collaborative effort within TIP's Metaverse Ready Network Project Group.

The main objective of this project group is to create and develop a systematic approach and corresponding guiding principles for designing and optimizing multimedia networks and emerging Metaverse immersive applications to deliver an enhanced Quality of Experience (QoE) to ensure customer and end-user satisfaction.

Authors and Contributors

François Blouin
QoE Engineering, Meta Platforms Inc,
francoisb@meta.com

Xinli Hou
Connectivity Technologies and Ecosystems Manager, Meta Platforms Inc
xinlihou@meta.com

Chris Murphy
Regional CTO, EMEA, VIAVI Solutions
chris.murphy@viavisolutions.com

Kafi Hassan
Technology Development Strategist, T-Mobile USA
kafi.hassan@t-mobile.com

Javed Rahman
Technology Development Strategist, T-Mobile USA
Javed.2.rahman@t-mobile.com

Minqi Wang
Research engineer in Orange Innovation (France)
minqi.wang@orange.com

Ricardo Serrano Gutierrez
Core & service platforms engineer, Telefónica
ricardo.serranogutierrez@telefonica.com

Kevin Smith
Distinguished Engineer, Vodafone Networks
kevin.smith@vodafone.com

Gavin Young
Head Of Fixed Access Centre Of Excellence, Vodafone Networks
gavin.young2@vodafone.com

Mayur Channegowda
Broadband Architect, Vodafone Networks
mayur.channegowda@vodafone.com

Introduction and Background

The Quality of Experience (QoE) engineering framework consists of five steps. First, we propose a method for determining QoE metrics for different applications and services (Step 1). Next, we identify the relevant Quality of Service (QoS) impairments that can affect and degrade overall application performance QoE (Step 2). In Step 3, we correlate QoE (end-user level) to QoS (network infrastructure level). Step 4 focuses on measurement and telemetry approaches for monitoring QoS and QoE. Finally, in Step 5, we implement the previous four steps for specific use cases.

This QoE engineering framework's main contribution is the novel, end-user-centric, top-down approach that can be used by application and network architects, network planners, and content application providers delivering multimedia services to achieve the QoE requirements associated with multimedia and immersive services and ensure commercial success.

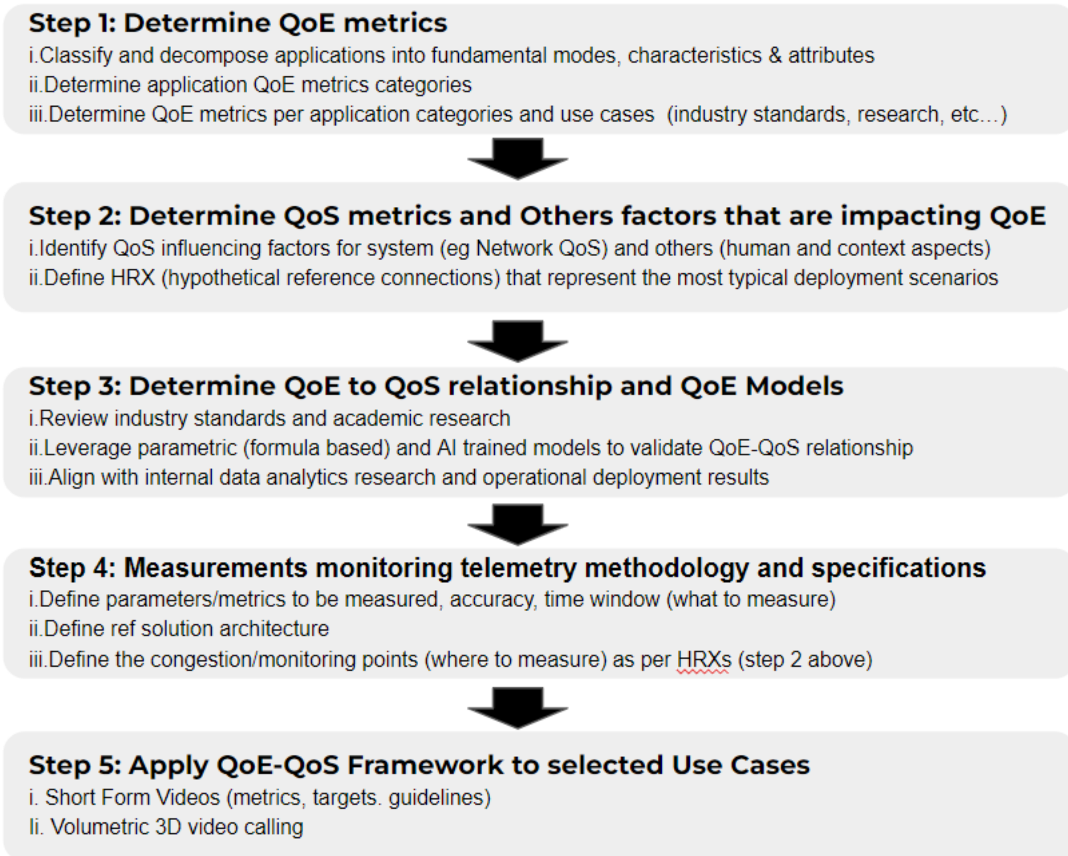

Figure 1: Top-Down QoE Engineering Framework

We propose a five-step top-down model that moves away from the traditional single-service, network-centric design cycle techniques commonly used in today's networks.

The key objectives of this framework are:

- Adopting a user-centric approach rather than the commonly followed network-centric approach, where information about the user needs, expectations, and services are considered input into the product requirements and definition stage.
- From a provider’s perspective, the framework provides a comprehensive view of factors that should be considered during the network's design, planning, and operational management.

Quality of Experience (QoE) refers to a system's overall performance from the user's perspective when using a service or application. It reflects how effectively the system enables users to achieve their goals. QoE focuses on end-user satisfaction, which includes sound quality, video quality, and interaction speed, with a strong emphasis on the application layer. However, QoE does not address Quality of Service (QoS) metrics such as throughput, latency, and packet loss.

Quality of Service (QoS) focuses on the infrastructure—such as networks, servers, cloud services, and end devices—that supports product applications. This includes aspects like delay, bandwidth, and capacity. Understanding infrastructure performance is crucial, as it directly affects the Quality of Experience (QoE) layer above. The challenge is defining the relationship and dependencies between QoE and QoS. Various performance metrics are established at different layers, catering to the needs of different stakeholders.

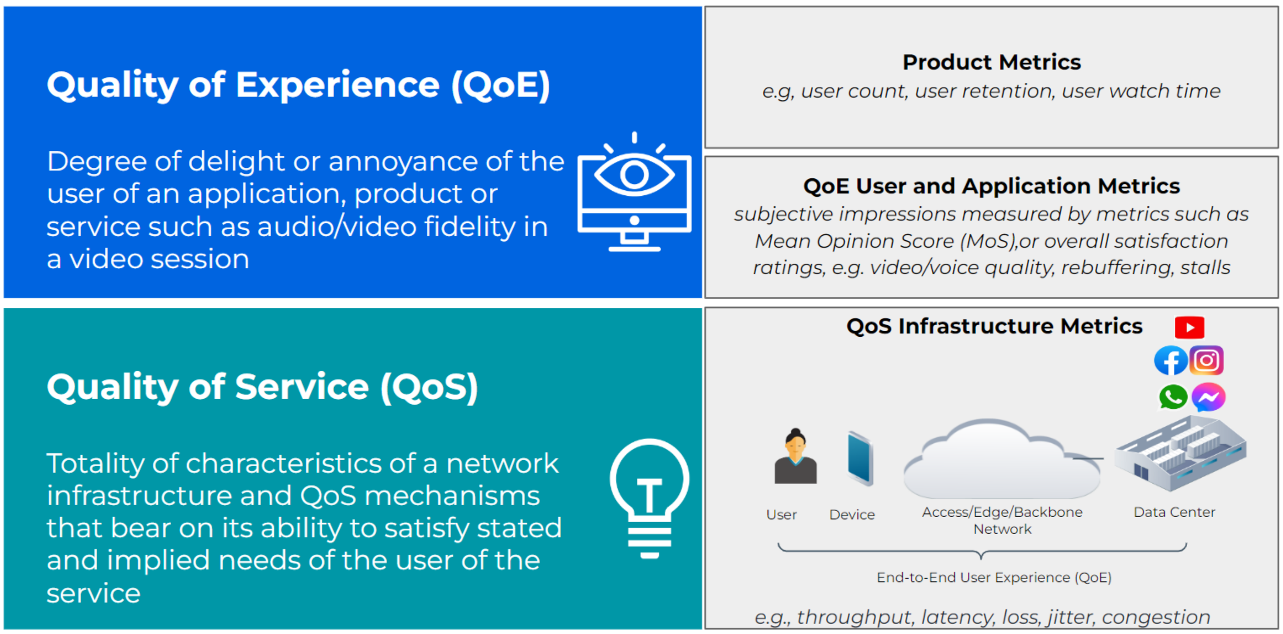

Figure 2: Definition of QoE and QoS

For many applications, efforts have been made to translate subjective measures of user experience into objective metrics (e.g., ITU-G.107). These translations help define objective requirements, including the metrics and targets that networks should meet.

Using the QoE requirements to guide network engineering and design has two significant benefits:

1. Network design targets are grounded in user needs and experience for the services carried, making them as attractive as possible to potential users,
2. We avoid over-engineering or under-engineering the network to ensure the provider can deliver high-quality content without wasting resources.

Step 1: Determining QoE metrics

I. Classify and decompose applications into fundamental modes, characteristics & attributes
II. Determine application QoE metrics categories
III. Determine QoE metrics per application categories (industry standards, internal research, etc…)

Process to Derive QoE metrics

Understanding and defining a service's characteristics and impact on user expectations is complex and challenging. Quality of Experience requirements is dispersed across international and industry standards, research reports, academia, and practical industry experience.

Defining Quality of Experience (QoE) requires a multidisciplinary approach. The influencing factors can be categorized into three distinct yet interconnected perspectives: Human, System, and Context. As we develop metrics, we must consider elements from these three perspectives and optimize them to meet user experience expectations, including delight-related and annoyance-related metrics.

We must first break down applications into their modalities and attributes to define QoE metrics. Then, we categorize these metrics based on their relevance to specific applications. Let's explore the first important step: the QoE metrics, which we divided into three subsections.

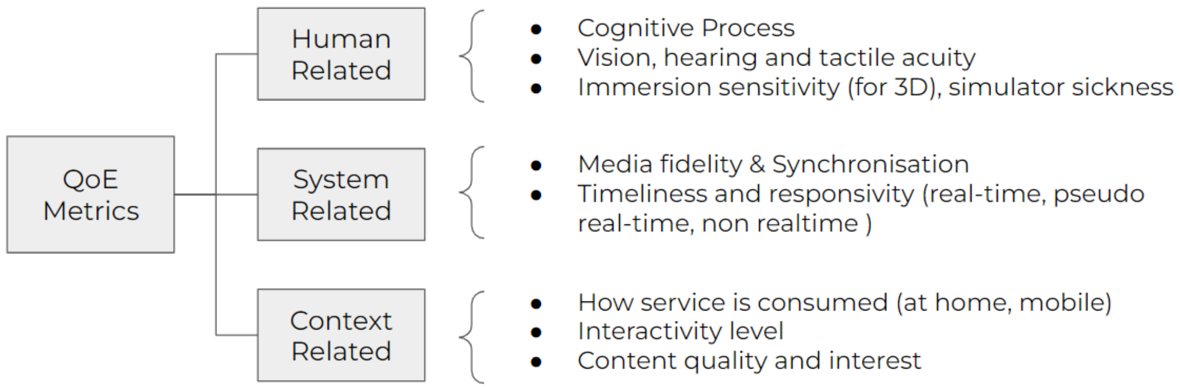

Figure 3 QoE metrics breakdown into three fundamental categories

To define QoE metrics, we need a perceptual model that addresses the key factors influencing end-user appreciation and satisfaction with a service.

Human-related influencing factors can be physiological (e.g., sim sickness, hearing acuity, visual acuity) or psychological (e.g., perception, cognition), including audio/visual perception, user engagement, and the user's emotional state. Immersion is an essential factor for 3D applications.

System-related influencing factors are the technical aspects that engineers and architects are generally more familiar with. They are associated with media capture, transmission, networking, coding, storage, rendering, reproduction/display, and the communication of information from content production to the user. Ease of use, often referred to as User Experience (UX), is also part of this category but has human-related aspects.

Context-related influencing factors describe the user's surrounding environment in terms of physical aspects (location and space, mobile vs. fixed access), temporal factors, social elements (people present or involved in the experience), economic aspects (costs, subscription type, or brand of the service/system), task characteristics, and technical characteristics. For example, mobile subscribers have different expectations than fixed broadband users, but how significant are these differences? Audio and video (A/V) modalities are both critical, but they vary according to their context of use. In high-motion sports, video quality is more crucial than audio quality. Meanwhile, audio quality and synchronization are paramount in talking heads and music videos. The annoyance caused by audio/video drop frames or interruptions varies depending on the type of content.

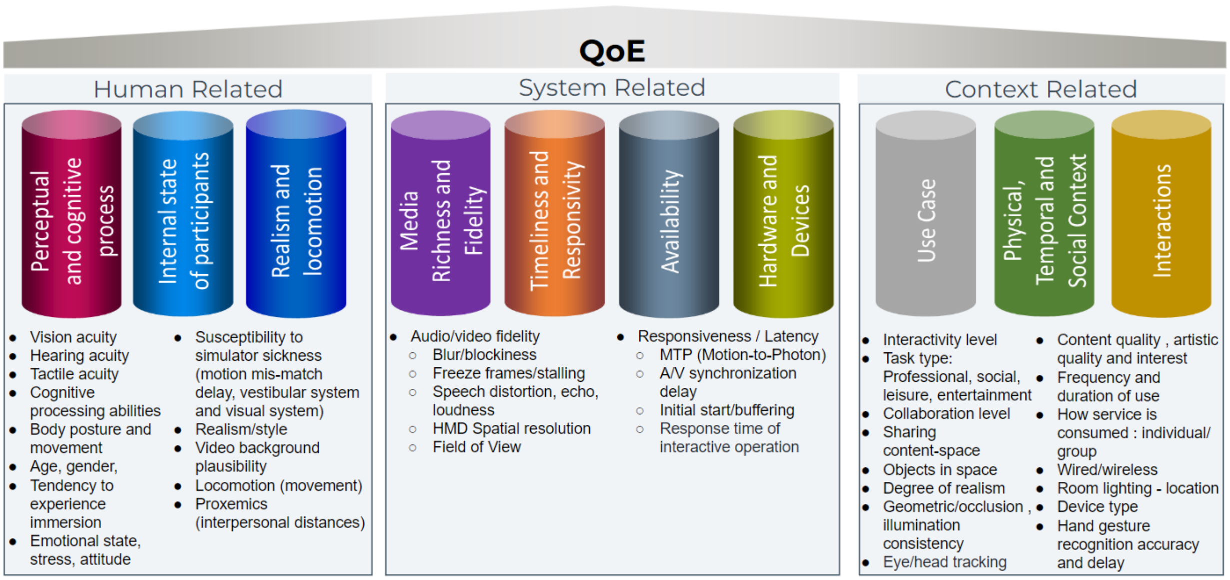

Figure 4 We have organized and identified the most important QoE factors that affect a wide variety of applications. In the ideal world, we would have a model that accounts for all of them; however, this is not practical, and we need to select a subset relevant to our use case.

As we define product use cases, it would be beneficial to examine the various trade-offs and constraints related to the human, system, and context influencing factors.

We have conducted an extensive review of industry standards and academic research. The importance of specific Quality of Experience (QoE) and Quality of Service (QoS) metrics depends on the application. The more immersive the application, the more factors influence user QoE and satisfaction. The QoE metrics shown in Figure 5 below emphasize the impact on the end-user rather than the networking level (QoS) across various application categories.

To understand the properties of services, we classify applications by their main characteristics and relevant attributes, such as *timeliness* (also called responsivity) and *modality,* which are independent of the transport network's features. *Timeliness indicates* whether interactions happen in real time or not. *Modalities* include audio, video, text, still graphics, data, and other formats used in the service or application.

Application	QoE Metrics Examples
Conversational VoIP Call	**Voice QoE = f (Responsivity, Media Fidelity)** Responsivity: One way delay; Media Fidelity: speech distortion, Synchro echo and sound level Example of well accepted QoE metrics/measurement methods: MOS, PESQ (ITU-T P.862), POLQA(ITU-T P.863), E-model R-factor (ITU-T G.107)
Video	**Video QoE = f (Temporal video quality, Spatial video quality, Timeliness / Responsivity Context)** Temporal video quality: fluidness/jerkiness; Spatial video quality: media fidelity, blur/blockiness,; Timeliness / Responsivity: initial loading, Stalls (#, duration, timing), one-way delay, A/V Synch (conversational); Context: content type Example of well accepted QoE metrics/measurement methods: PSNR, SSIM, VMAF, ITU-T G.1070/71, ITU-P.1204
Gaming	**Gaming QoE = f (Temporal video quality, Spatial video quality, Responsivity, Context, Human)** Temporal video quality: fluidness/jerkiness; Spatial video quality: media fidelity, blur/blockiness,; Responsivity: action motion response Stalls (#, duration, timing); Context: game classification; Human: frame Loss Sensitivity Delay sensitivity Example QoE metrics/measurement methods: R-factor for cloud gaming (ITU-T G.1072/G.1032)
3D VR//SR/XR	**3D Virtual Reality QoE = f (Temporal video quality, Spatial video quality, Responsivity, Context Human)** Temporal video quality: fluidness/jerkiness; Spatial video quality: media fidelity, blur/blockiness,; Responsivity: motion- to-photon Stalls (#, duration, timing); Context: content interest, task type: collab or individual; Human: Locomotion immersion, motion sensitivity degree of realism Emerging/under development QoE metrics/measurement 2023 ITU-T PSTR-OQMXR "Objective quality modelling for XR services"

Figure 5: QoE Metrics per application categories with relevant industry standards references

For instance, in video applications, video fidelity is a key aspect of the Quality of Experience (QoE), which includes both temporal (motion artifacts) and spatial (compression artifacts) factors. In newer services like short-form video, where content typically lasts less than 45 seconds, timeliness and interactivity have become more important than video fidelity, according to recent research [1].

Video conferences are essential for maintaining effective communication and audio/visual synchronization among participants. The nature of the collaboration, whether a formal business meeting or an informal leisure call, influences Quality of Experience (QoE) expectations and should be considered when setting QoE targets.

In gaming applications, trade-offs are often made between temporal factors (video frame rate), spatial factors (video fidelity), and interactivity (response to commands). The type of game (context) also influences the encoding complexity and sensitivity to latency and video distortion. Players of fast-paced games prefer to accept a reduction in video or animation quality rather than tolerate high delays, as gameplay and player success heavily depend on their ability to react quickly.

An important consideration for determining QoE metrics is the context, including the scenario and environment where the application is used. This context influences end-user expectations and is often overlooked. Factors like video types, game genres and complexity, and the gaming platform (mobile smartphones versus dedicated terminals) impact expectations and, consequently, the acceptability of QoE targets. Additionally, AI-based inference engines are now widely used to provide better recommendations aligned with user interests, particularly in short-form videos, which makes measuring QoE objectively even more challenging.

In step 2, we will describe the method for identifying how QoE is affected by the system level and, particularly, the most influential QoS factors.

Step 2: Determine QoS metrics and other factors that impact QoE

I. Identify factors that influence Quality of Service (QoS) for the system, such as network QoS, as well as human and contextual aspects.

II. Define HRX (Hypothetical Reference Connections) that represent the most common deployment scenarios.

In step 2, we provide a procedure for identifying a service's QoS impairments from an end-to-end network perspective. We consider the network topology and crucial intermediate network elements involved in realizing the service.

Overall, Quality of Experience (QoE) is mainly influenced by three key metrics: (1) Responsivity, (2) Media Fidelity, and (3) Availability. These metrics help derive Quality of Service (QoS) factors for each major segment of the end-to-end path: client, network, and server-side, as illustrated in Figure 6.

Client

Hardware
- Processing power
- Display resolution
- Camera sensor noise

Firmware stack
- Sensor delay
- Transmission queuing delay

Application stack
- Packet Loss Concealment
- Bandwidth estimator
- ABR lane switching
- Processing (encoding / decoding delay
- Asynch/Synch rendering delay

User stack
- De-jitter/playback buffer
- media flow synch delay
- Frame rate, bitrate /resolution
- Refresh rate
- Codec compression artefacts
- Call setup delay
- Request delay

Network

Capacity
- Loading
- Congestion
- Packet loss
- queuing/jitter

Bandwidth
- Link Speed & variation
- Throttling

Routing
- switching/forwarding
- propagation delay

Server

Edge
- GPU/CPU Processing power
- Workloads
- Cache queuing delay
- Codec Encoding parameters

Origine
- GPU/CPU Processing power
- Workloads
- Cache-miss delay
- Propagation delay
- Server queuing delay
- Server processing delay

Figure 6: End-to-end QoS impairment factors divided by core system elements, including the client device, network, and server side where applications and services are typically hosted.

Quality of Experience (QoE) should be measured from an end-to-end perspective, considering the effects of all three segments: Client, Network (typically Quality of

Service), and Server on QoE and overall user satisfaction. Each segment introduces specific Quality of Service (QoS) impairments. The management of these impairments depends on the ownership of each segment, reinforcing the need for standardized end-to-end guidelines provided by TIP and this working group.

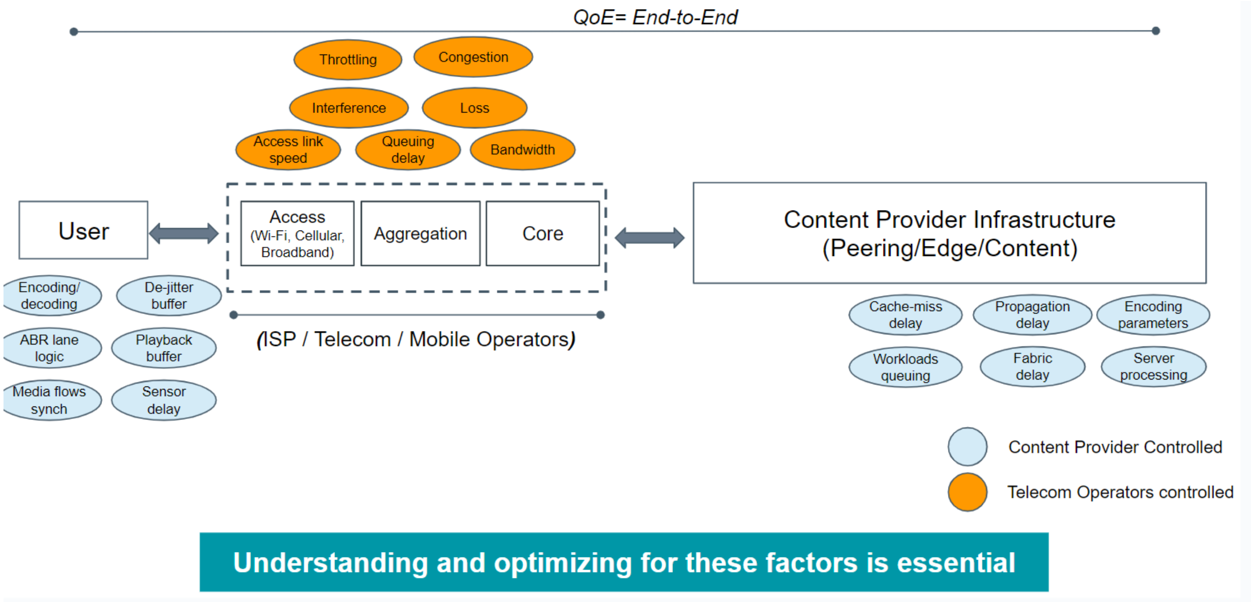

Figure 7 illustrates a typical end-to-end reference connection for an end user accessing a service on the Internet. It highlights the relevant QoS factors that impact QoE, which should be considered in planning and optimizing to ensure service delivery quality.

Content application providers (CAPs) control certain factors, illustrated in the blue bubbles in Figure 7, including application, network portion, and user software stack. Communication service providers (CSPs) also manage factors, shown in the orange bubbles, such as throttling or shaping video traffic on cellular networks. Achieving complete end-to-end (E2E) visibility of the ecosystem is very challenging for any one provider. Therefore, there is a strong desire to share information among the various participants in the end-to-end ecosystem. Refer to reference SADCDN IETF reference RFC [22].

If we examine the middlebox (ISP / Telecom / Mobile Operators) further, we'll understand that it is a multi-segment network with some complexity (e.g., home, Wi-Fi, cellular, broadband, etc.). Figures 6 and 7 show various network factors that must be accounted for, such as delay and jitter.

Defining HRX (hypothetical reference connections)

We begin by abstracting the complete topology for a specific service configuration and then apply a set of predetermined use cases known as Hypothetical Reference Connections. The Hypothetical Reference Connection modeling technique is a well-established telecom practice used by Communications Service Providers (CSPs) to understand budget allocation standards within their networks. These connections are hypothetical because they consider variables like distances and the type and number of equipment used, which involves abstracting domains rather than modeling the exact conditions of a real network connection. In the context of packet networking, HRX can be used to analyze Quality of Service (QoS) requirements for specific applications.

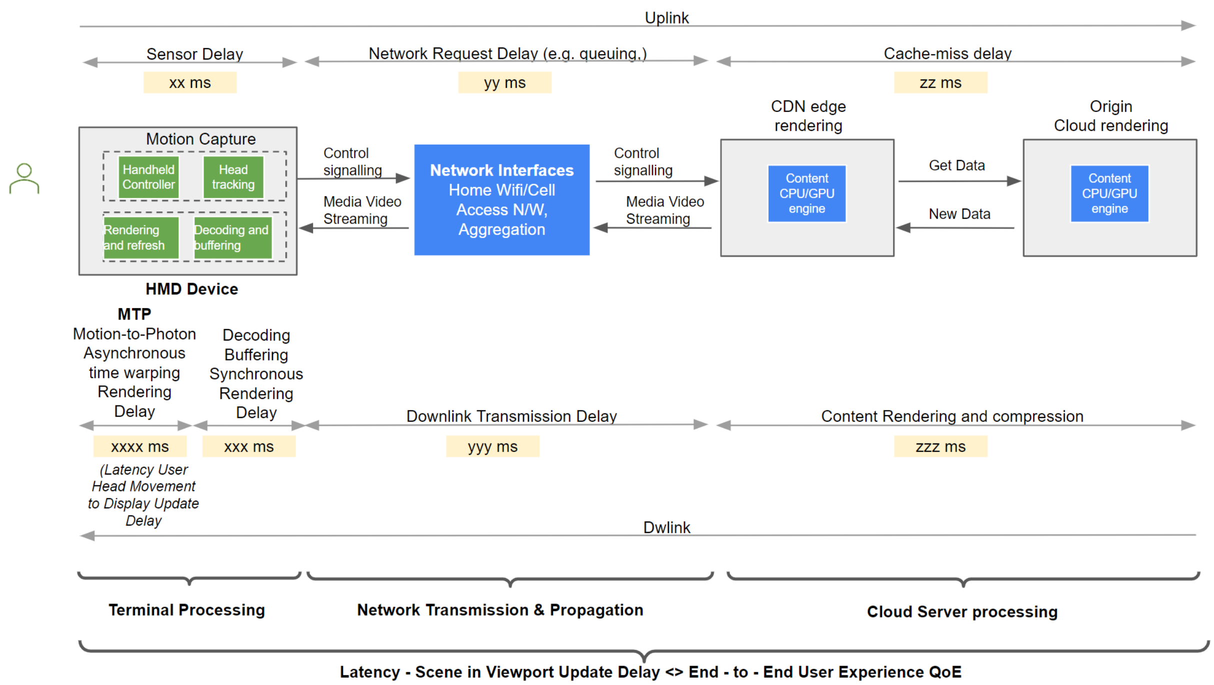

Figure 8 HRX shows a high-level overview of the HRX approach applied in planning the virtual reality (VR) service, focusing on the motion-to-photon latency budget allocation. To ensure the overall network meets the QoE targets, an impairment budget planning exercise is used to determine allowances for each network segment and the nodes within those segments. Note: Time warping is a reprojection technique that maps the previously rendered frame to the correct position based on the latest head orientation information and runs in parallel with the synchronous rendering process. It translates the image by a certain number of pixels based on changes in head position between the start of rendering and the initiation of the time warp operation.

Figure 8 illustrates an example of how to apply HRX in the product planning phase, where QoE engineering must be conducted, such as the example delay budget highlighted in yellow boxes. This process defines the end-to-end (E2E) tolerable delays, allocates per network segment where QoE could be significantly impacted, and outlines the network requirements necessary to achieve specific QoE targets. Additionally, it aids in understanding the trade-offs between delay and QoE impacts.

Step 3: Determine QoE to QoS relationship and QoE Models

I. Review industry standards and academic research
II. Utilize parametric (formula-based) and AI-trained models to validate the relationship between QoE and QoS.
III. Align with internal data analytics research and operational deployment results

This phase's objective is to understand the relationship between a given service’s impairments and the parameters in the application (QoE) and network layer (QoS) that affect an application’s QoE.

The goal is to translate the QoE requirements into lower-layer definitions (application and network layer) to define our QoS requirements. These include metrics and targets, as well as guidelines and rules for traffic engineering, allowable operating ranges, and so forth. The QoS requirements are used to engineer the network so that the services carried out will meet their QoE targets.

Packet-switched networks have probabilistic (non-deterministic) aspects, and the related Quality of Service (QoS) requirements reflect this, often expressed as the proportion of time that a requirement will be met. From this understanding, we can derive QoS requirements that align with specific desired Quality of Experience (QoE) outcomes, or even balance QoE against other factors such as cost. This translation is complex because developing a universal equation or algorithm to compute or derive an individual QoS requirement from a corresponding QoE target is difficult.

The industry and academic research have been actively pursuing the topic of Quality of Experience (QoE) for several decades. In the early 2000s, there was a particular focus on services such as VoIP, IPTV, and various over-the-top (OTT) services. QoE research remains a vibrant field, originating from the telecommunication and multimedia engineering domains, that aims to understand, measure, and design the quality of experience for multimedia technologies. A summary of the numerous standardization and implementation guidelines from this work is presented in Table 1-5.

Overview of the QoE standardization activities

To establish a framework for Quality of Experience (QoE) engineering for next-generation services, this section provides an overview of current standardization activities related to QoE as perceived by users. Below are examples of standardization efforts in QoE for Virtual Reality (VR) / Extended Reality (XR), AI-based video quality analysis, and multimedia. Readers are encouraged to review Tables 1-5, which offer a more comprehensive list underpinning much of the discussed work and include standards relevant to other services. There is significant research in academia on Augmented Reality (AR) / Virtual Reality (VR), including examples of proposed QoE models [14] [15].

1. IEEE HFVE_WG - Human Factors for Visual Experiences Working Group

The IEEE has a working group standard that focuses on human factors for AR/VR and Metaverse-related applications. The IEEE Standard 3333.1.3 [2], titled "Deep Learning-Based Assessment of Visual Experience Based on Human Factors," identifies factors contributing to a user's perceptual experience, including human, system, and context factors. The standard specifically investigates how to estimate the mechanisms of human visual perception. The assessment of human visual perception is divided into two subgroups: perceptual quality and VR cybersickness. To measure Quality of Experience (QoE), the standard uses two evaluation methods: deep learning models that consider human factors for various QoE assessments and a subjective test methodology with a content database. For the subjective test methodology, the standard developed an immersive VR content database to evaluate cybersickness and the sense of presence. This VR content database is available for free download and use in scientific research [2].

In addition, another relevant IEEE working group focusing on Metaverse AR/VR interoperability is the IEEE 2048 VR/AR Working Group (VRARWG).

2. ITU-T Study Group 12

ITU-T SG12 focuses on performance, quality of service (QoS), and quality of experience (QoE). It is organized into three working parties (WPs) and a series of questions addressing the field of QoE.

1. WP1 Testing methodology

   a. Q7/Q10 Subjective test methods for VR/XR, crowdsourcing
2. WP2: QoE objective models
   a. Q14 Audio/Video QoE modeling, streaming, cloud gaming, QoE vs user retention
   b. Q17 testbed framework
   c. Q19 No reference QoE models, collaboration with VQEG,
3. WP3 multimedia QoS and QoE , operational aspect, requirements for new services
   a. Q13, QoE influencing factors, QoE requirements for 5G services

Moreover, a focus group on metaverse (FG-MV) was established in 2022 to analyze technical requirements and enabling technologies in this domain.

3. Immersive Media Group (IMG)- Video Quality Experts Group (VQEG)

The Immersive Media Group (IMG) within the Video Quality Experts Group (VQEG) conducts studies on the quality assessment of immersive media, including virtual reality, augmented reality, stereoscopic content, and Multiview. Its goal is to establish guidelines for quality experience. The group performs baseline assessments using both traditional and updated virtual reality content, incorporating 360-degree and light field cameras. The VQEG Immersive Media Group has published a Phase 1 baseline test plan for assessing the quality of 360-degree video, focusing specifically on short video sequences.

4. 3GPP Standardization for extended reality (XR) in 5G and beyond

The 3rd Generation Partnership Project (3GPP) issued several relevant QoE and QoS service requirements for LTE and 5G systems, covering media streaming interoperability for emerging AR/VR/XR technologies. Some reference standards are listed below.

1. 3GPP TR 26.918 version 15.2.0 Release 15) Universal Mobile Telecommunications System (UMTS); LTE; Virtual Reality (VR) media services over 3GPP
2. 3GPP TS 22.261 (2021) Version 16.14.0, Release 16: Service Requirements for the 5G System
3. 3GPP TR 26.928 (2020) is a standard document that contains the Technical Specification Group on Services and System Aspects of Extended Reality (XR) in 5G (Release 18). This Release 18 standard provides a baseline for XR services and applications technologies.

4. 3GPP TR 26.929 (2020) QoE parameters and metrics relevant to the Virtual Reality (VR) user experience.
5. 3GPP TR 26.909 version 17.0.0 Release 17 (2022) Study on improved streaming Quality of Experience (QoE) reporting in 3GPP services and networks.
6. Also, it introduces an outline for QoE/QoS issues of XR-based services, the delivery of XR in the 5G system, and an architectural model of 5G media streaming defined in 3GPP TS 26.501 (2020)
7. 3GPP TR 26.998 version 17.0.0 Release 17 (2022) LTE; 5G; Support of 5G glass-type Augmented Reality / Mixed Reality (AR/MR) devices

5. European Network on Quality of Experience in Multimedia Systems and Services (Qualinet)

The European Network on Quality of Experience in Multimedia Systems and Services (Qualinet) published a white paper defining Quality of Experience (QoE) for immersive media experiences [3].

6. Others

Additionally, there are other Quality of Experience (QoE) related work performed by other industry groups such as:

- Virtual Reality Industry Forum (VRIF)
- Moving Picture Experts Group (MPEG)
- Khronos Group the OpenXR™ Specification (khronos.org)
- World Wide Web Consortium (W3C)
- ITU-T SG16/Q8 Immersive Live Experiences
- ETSI Technical Committee (TC) Human Factors (HF)
- WiFi Alliance XR

Tables 1-5 below present an initial list of various standards organizations that focus on Quality of Experience (QoE) in real-time applications, including the metaverse, voice and video, gaming, planning aspects, and telemetry. This list is not exhaustive, and we welcome feedback on other industry standards, areas of interest, or any anticipated issues, such as licensing terms and royalties. The list also includes ongoing work related to QoE in Metaverse AR/VR/XR.

Metaverse AR/VR/XR
Human and system factors, metrics, affecting the user perceived experience of virtual reality (VR) and augmented reality (AR) services. Service quality monitoring requirements. Latency and synchronization aspects including motion-to-photon latency, motion-to-sound latency, A/V synchronization. Subjective test methodologies to evaluate aspects of QoE for 360 videos viewed in head-mounted display. Measurement methods to spatial audio telemeeting systems. QoS networking level performance requirements. Model for multimedia Quality of Service (QoS) categories from an end-user viewpoint Metaverse QoE requirements development.
1. ITU-T Y.3109 (2021) – QoS assurance related requirements for VR 2. ITU-T P.1320 (2022) – QoE assessment of XR meetings 3. ITU-T G.1035 (2021) – Influencing factors of QoE for VR services 4. ETSI TR 126 918 v17 (2022) – VR medias services over 3GPP 5. ITU-T G.1036 (2022) – QoE influencing factors for AR services 6. ITU-T P.1310 (2017) – Spacial audio meetings quality evaluation 7. Metaverse Standard Forum (2022) – Exploratory group on Networking 8. IEEE Std 3333.1.3 (2022) Assessment of Visual Experience Based on Human Factors 9. NASA Research (2004) – Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity 10. ITU-T SG16, (2023) – Correspondence Group on Metaverse 11. ITU-T H.EMG, (2023) – A cooperative architecture for enhanced multimedia QoS/QoE 12. IEEE P2048, (2023) – Standards for Virtual Reality and Augmented Reality

Table 1: Metaverse AR/VR/XR Industry standards

Video
Objective parametric quality assessment model to predict the impact of audio and video media encodings and observed IP network impairments on QoE in multimedia streaming applications. Measurement approaches, diagnostic analysis and KPIs/KQIs for video-based services, including video, audio quality estimation and quality integration. Methodology to conduct subjective quality assessment of multi-party telemeeting

systems at remote locations Methodology to conduct subjective quality assessment for multimedia applications
1. ITU-T P.1301 (2017) – Subjective quality evaluation of audio and audiovisual multiparty telemeetings 2. ITU-T P.1305 (2016) Effect of delays on telemeeting quality 3. ITU-T P.1305 (2016) Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance 4. ITU-T P.1203 (2019) – Parametric bitstream-based quality assessment 5. ITU-T P.1204 (2020) – Video quality assessment of streaming services up to 4K 6. ITU-T G.1070 (2018) – Opinion model for video-telephony applications 7. ITU-T G.1071 (2016) – Opinion model for network planning of video and audio streaming applications 8. ITU-T P.NATS – Quality integration module for adaptive video streaming QoE 9. VQEG (Video Quality Expert Group) – Industry and academia collaboration to advance subjective methodology and objective tool development/verification for video quality (Meta is a member) 10. ITU-T P.91X (2007-2022) – Subjective video quality assessment methods for multimedia applications including 3D video quality 11. ITU-R BT.500 (2019) – Methodologies for the subjective assessment of the quality of television images 12. IEEE P2048.3 – Immersive video file and stream formats 13. 3GPP TR 26.909 version 17.0.0 Release 17 (2022) Study on improved streaming Quality of Experience (QoE) reporting in 3GPP services and networks.

Table 2: Video QoE Industry standards

Audio
The E-model ITU G.107 offers a standard method for prediction and planning of telecom networks. An analytical tool for estimating End-to-End VoIP conversation quality across networks, considers a wide range of impairments including coded type, packet loss, delay, echo etc. Useful for transmission planning tools, to assess VoIP audio performance, establish benchmark networks for comparison, and compare design alternatives.
1. ITU-T G.107 (2016) – The E-model: a computational model for use in transmission planning 2. ITU-T G.109 (1999) – Definition of categories of speech transmission quality 3. ITU-T P.1305 (2016) – Effect of delays on telemeeting quality 4. ITU-T P.1310 (2017) - Spatial audio meetings quality evaluation 5. ITU-T G.114 (2003) – General Recommendations on the transmission quality for an entire international telephone connection

Table 3: Audio QoE Industry standards

Gaming (Cloud and terminal based)
Defines a list of factors which may influence the quality of experience (QoE) of cloud gaming and online gaming. Also provides a parametric model to predict gaming quality of experience (QoE) by considering relevant factors. The model is a network planning tool which can be used by various stakeholders for purposes such as resource allocation and configuration of IP-network transmission settings such as the selection of resolution and bitrates, under the assumption that the network is prone to packet loss, throughput and latency.
1. ITU-T G.1072 (2020) – Opinion model predicting gaming quality of experience for cloud gaming services 2. ITU-T G.1032 (2017) – Influence factors on gaming quality of experience 3. IEEE P2948/P2949 – Recommended practice for the evaluation of cloud gaming user experiences

Table 4: Gaming QoE Industry standards

Telemetry & QoE-QoS Planning
Measurement approaches, diagnostic analysis and KPIs/KQIs. Proactive analysis of network performance and support for customer service troubleshooting. New on-path per-packet telemetry information (piggybacking metadata on packet) to be collected and extracted from the network, and techniques being developed with real-time notification to complement ping/traceroute.
1. ITU-T P.DiAQoSE – Diagnostic assessment of QoS and QoE for adaptive video streaming sessions 2. Broadband Forum PEAT – Performance, Experience and Application Testing 3. Broadband Forum QED – Quality Experience Delivered 4. BBF TR-452.2 – Quality Attenuation Measurements using Active Test Protocols 5. IETF IOAM – In-Situ flow and on-path telemetry 6. MEF 23.2 (2016) – Carrier Ethernet Class of Service 7. ITU-T Y.1541 (2011) Network performance objectives for IP-based services 8. ITU-T GSTR-5G QoE (2022) – Quality of experience (QoE) requirements for real-time multimedia services over 5G networks 9. ITU-T G.QoE-5G (2024) – QoE factors for new services in 5G network 10. ITU-T G.CMVTQS – Computational model used as a QoE/QoS monitor to assess video telephony services 11. ITU-T J.1631 Functional requirements of E2E network platforms to enhance the delivery of cloud-VR services

Table 5: Telemetry & QoE-QoS Planning Industry standards

To define QoE requirements for services, user perception models of quality are often used. Several approaches to QoE modeling exist, and these are typically divided into three broad categories:

- subjective
- objective
- hybrid

The subjective approach relies on user testing, opinions, past experiences, expectations, user perception, judgment, and descriptive capabilities. It primarily evaluates the effectiveness, efficiency, and overall satisfaction of using a service. This method is considered the most reliable only if the subjective tests are designed carefully and the user pool reflects current expectations and demographics. Although direct user testing is the most time-consuming and costly method, making it impractical for continuous monitoring, parametric models based on user subjective testing can still be useful in certain circumstances, as we will explore in the following section.

On the other hand, objective models can be digitally automated and deployed at scale. We will focus the next part of the discussion on these models.

To develop a mapping from Quality of Experience (QoE) to Quality of Service (QoS), several parametric models for QoE have been created and standardized as viable options. These models use specific formulas for estimating QoE based on its correlation with QoS. The parametric models incorporate inputs from network parameters and measurements, as well as human factors such as subjective testing (like audio-visual perception) and context, including the types of devices and applications used. These inputs are then translated into an estimated QoE score calculated using the chosen parametric model.

Parametric QoE models are generally derived by conducting subjective experiments (lab or crowdsourcing) and then performing statistical analysis (e.g., regression analysis) on the acquired evaluation results. This creates a mathematical model with a set of parameters.

Parametric QoE models are commonly used for product design, live network quality monitoring, and product development planning. They help assess the impact of interventions on Quality of Experience (QoE). Telecom operators and vendors are major users in the planning phase of deploying network services for multimedia options such as VoIP and video streaming.

Video QoE Models

Objective video QoE models are generally categorized into three families: media layer, packet layer, and parametric planning.

- Media layer (pixel-based): This examines the pixel level of a decoded stream, also known as bitstream models. It may use both packet headers and payload data.
- Packet layer: Models extract information from packet headers, including frame rate, motion vector (scene complexity), QP, and packet loss from the encoded stream.
- Parametric planning: These models are typically used with network planning parameters and inputs measured from specific QoS network performance metrics to compute an estimated QoE score.

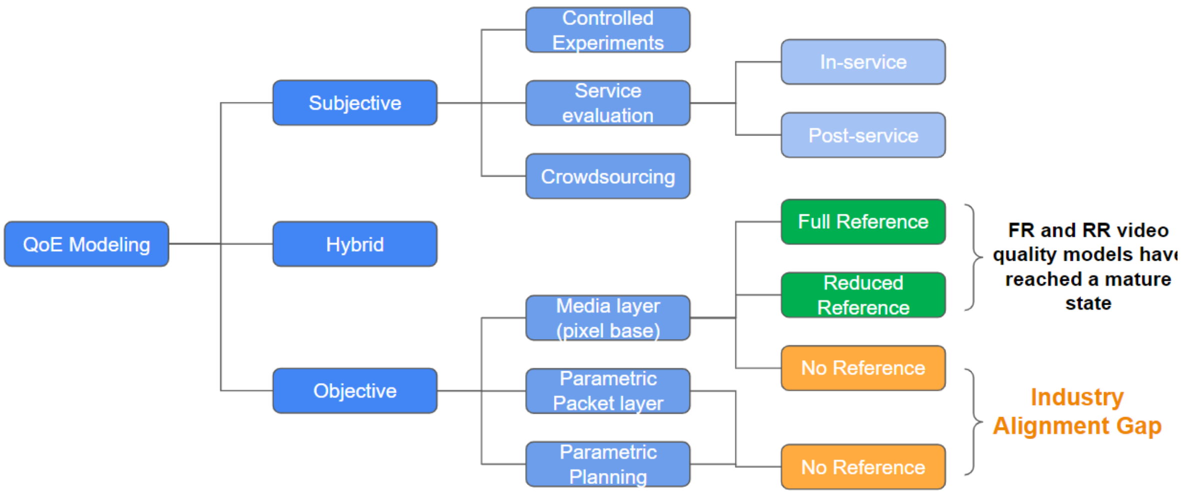

Figure 9 Overview of QoE models classifications for video services

Full Reference Video QoE Models

Full reference models provide the most accuracy but are not always practical, as telecom operators typically lack access to both the original source material (the reference) and the final delivered video (possibly impaired stream). This limitation makes it difficult to generate a QoE score. On the other hand, content access providers usually have access to both the original source material and the delivered stream. Additionally, these models can be computationally intensive, so scaling needs to be considered in the QoE model selection process. While full reference video quality models have achieved a mature state with superior accuracy, they focus solely on video fidelity and may not apply to emerging short-form video applications (like YouTube Shorts, Facebook/Instagram Reels, TikTok, etc.), which are affected by the overall timeliness and responsivity due to users rapidly swiping between videos. To address this, ongoing research is working to develop a composite QoE metric that incorporates both video fidelity and timeliness aspects, but these models have yet to be standardized in the industry.

No Reference QoE Models

No reference models are becoming an attractive option for satisfying all actors in the ecosystem and aligning on a unified metric or model, but their maturity and accuracy are not as high as those of comprehensive reference models.

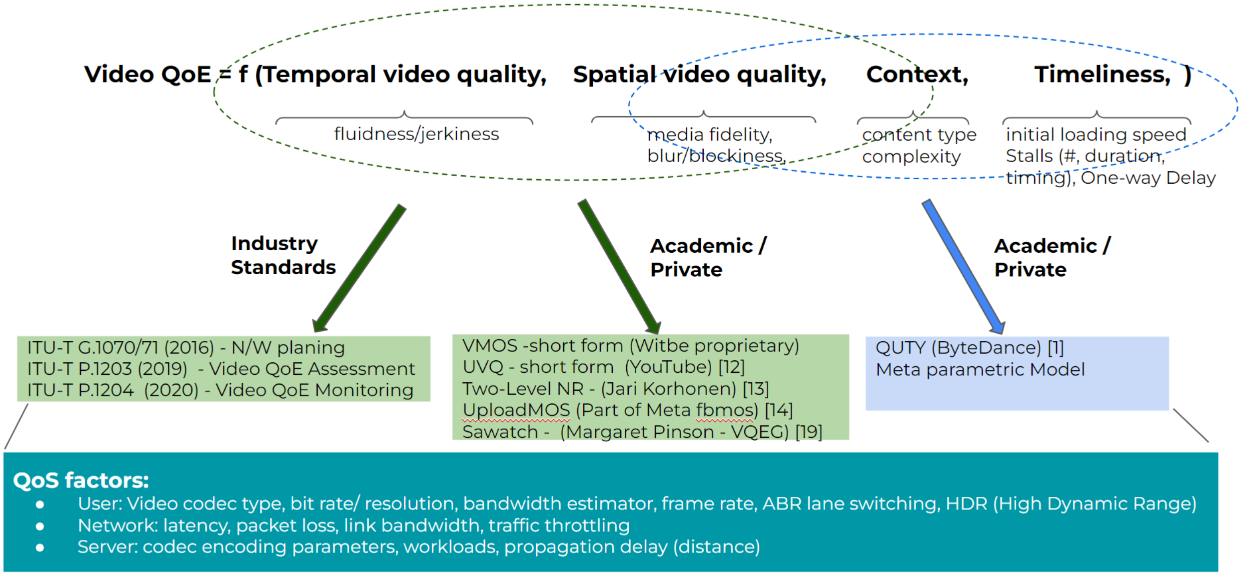

Figure 10a Compilation of No reference video QoE models from academia, private and industry standards

So far, the industry and academic research has focused on either video fidelity (left side) or timeliness and interactivity (right side) of Figure 10. An ideal composite QoE metric would incorporate both aspects, but it is not currently widely available. The QoS factors listed in Figure 10 are an example of what influences these QoE models.

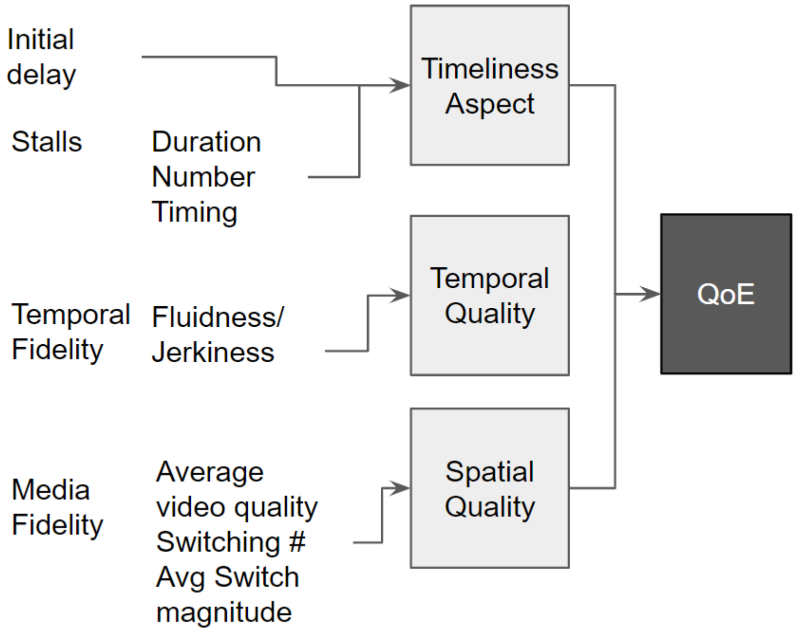

Figure 10b Composite short form video QoE model incorporating both video fidelity aspect as well as timeliness for loading/rebuffering

There are a few challenges regarding the lack of reference video metrics. These issues are being addressed but have not yet been resolved, namely:

- There are no sufficiently accepted consolidated composite metrics in the industry that include both the fidelity aspect and timeliness for short-form video services [18].
- There are no recommendations or standards that define common testing methodologies.
- they are still limited in their accuracy and not yet universally adopted.

Active research is ongoing to understand and model the impact of loading time [1] [17] and rebuffering in short-form video applications. The big challenge is in the No no-reference models [18] [19] where more research is needed, and industry standardization and adoption of commonly accepted QoE models are required.

Interactive Video Conferencing QoE Models

The QoE study of interactive video services has a different approach to traditional video streaming services and has attracted the attention of standard bodies since 2016 [4-7] [16].

For example, in video conferences (like Zoom or Microsoft Teams), the overall perception of audio and video quality is susceptible to unstable network conditions due to limited buffer sizes (or even no buffer). This instability affects participants' interactivity and conversational dynamics, making jitter noticeable to end users. ITU-T G.107 recommends a one-way delay of less than 150 ms to maintain good conversation dynamics. Chrome for WebRTC provides a minimum audio jitter buffer of 0 ms, with a default maximum number of packets in the audio jitter buffer set at 50. In contrast, the buffer for video streaming, even for live broadcast TV, can extend to several minutes.

Moreover, in telepresence/video conferences, the audio and video are generated by end-user terminals and impacted by the performance of their local/access networks. Thus, the source could be degraded any time before being sent to another end-user. The non-optimal and variable quality of the audio and video source should be considered for subjective evaluations such as the Mean Opinion Score (MOS) commonly used for audio QoE measurement.

MOS standards for interactive audio-video applications have yet to be defined, but researchers are beginning to show interest in this area. Technically, the research paper [8] reveals the differing impacts of various protocols used for video streaming (DASH) compared to protocols for real-time services (WebRTC). Subjectively, the study [9] uses algorithms to analyze facial and speech features to assess the MOS of audio-visual conversations. Some conferencing application providers have researched user needs to prioritize the performance of specific modalities, such as audio and screen sharing, over others, like video quality [10]. We believe that by using the MOS indicator, a more optimized strategy can be applied across all network segments to enhance user experience.

Audio QoE Models

One of the most popular voice quality QoE models is the E-model, standardized as ITU-T G.107. This subtractive model begins with a maximum transmission rating factor (R) and subtracts various impairments to estimate user perception of quality.

The user's perception of conversation quality depends on the following parameters:

Conv. Voice QoE = f (Conversation dynamics, Distortion)

= f (Delay, Speech Distortion, Sound & Echo levels)

The E-model calculates a Transmission Rating Factor, R, given by

$$R = Ro - Is - Id - Ie + A$$

R_o = the basic signal-to-noise ratio based on send, receive loudness, electrical, and background noise

I_s = real-time impairment factor, e.g., loudness, sidetone, and quantizing distortion

I_d = impairment from delay factors: e.g., talker echo, listener echo, and all delay (packetization, de-jitter, etc...)

I_e = the equipment impairment factor for special equipment: e.g., codecs, loss concealment algorithm, loss distribution, burst (determined subjectively for each codec, for each % packet loss)

A = the Advantage factor, an adjustment for the advantage of access, e.g., mobile devices

Proper control of these four parameters ensures satisfactory end-user voice quality and therefore provides good QoE.

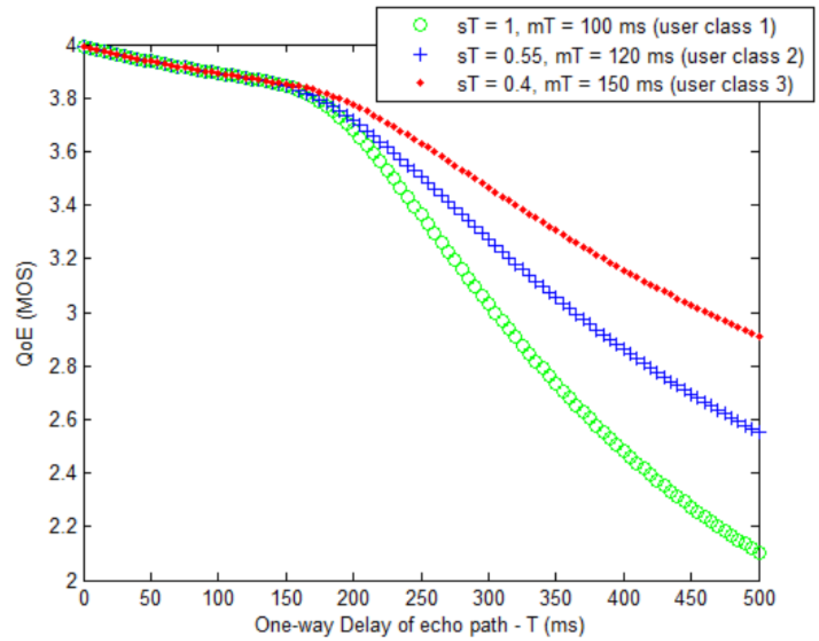

Figure 11 ITU-T G.107 QoE Vs QoS (User Sensitivity Latency). The parameters used in the model are shown here, the model has an interesting features to account for the user profile in terms of sensitivity to delay, eg a business conversation vs casual

Gaming QoE Model

One of the leading industry-standard cloud gaming models is the QoE Parametric Model ITU-T G.1072/G.1032. An interesting feature of this model is that you can choose the game's complexity or classification to adjust the output requirements. The game type is determined by encoding complexity, as well as sensitivity to delay and frame losses. Encoding complexity is influenced by the movements of a virtual camera, texture details, and the frequency of movements of game objects, as shown in Figure 12.

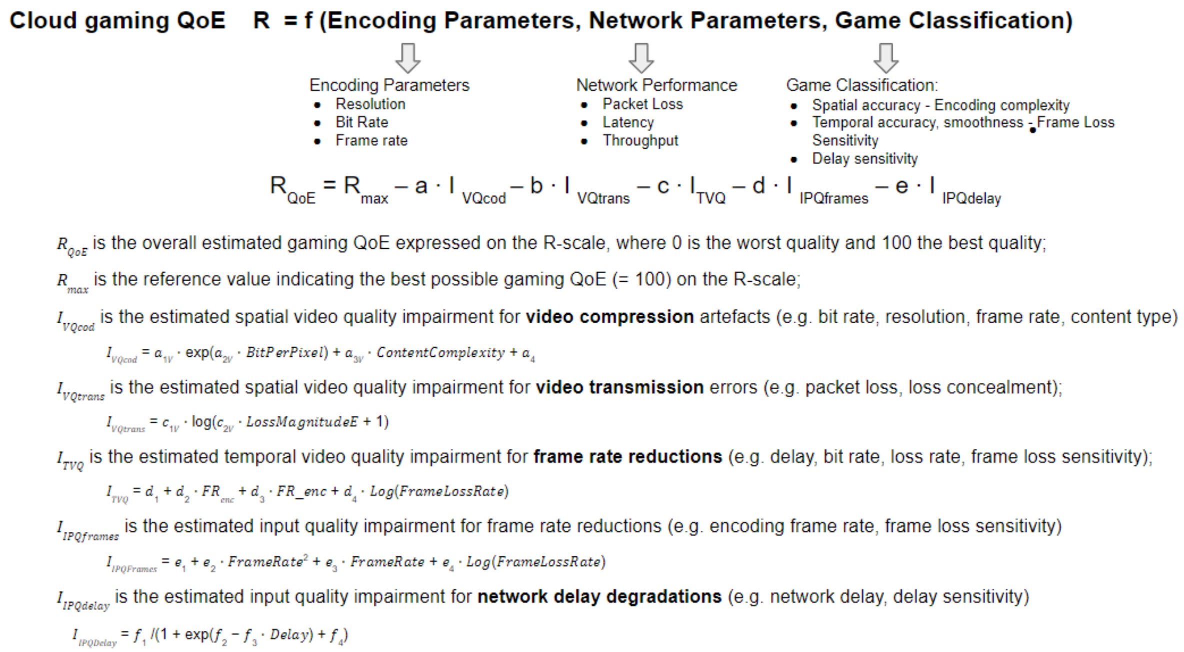

Figure 12 ITU-T G.1072/G.1032 Cloud Gaming QoE Parametric Model

For gaming applications, cloud gaming requires more stringent overall network performance, namely bandwidth, latency (critical), and packet loss control, because rendering is partially done in the cloud. The game type impacts encoding complexity and sensitivity to delay and frame losses.

Asymmetric network conditions can severely impact Quality of Experience (QoE), making the ability to model both uplink and downlink highly desirable. When packets are lost in the uplink, the gaming experience suffers due to the lack of user actions sent to the server, resulting in delays in game responsiveness. On the other hand, when packets are lost in the downlink, video quality degrades due to frame losses, leading to video distortion, unnaturalness, or discontinuity. Players of fast-paced games prefer to tolerate higher packet loss rates rather than high delays, as gameplay and success depend significantly on their ability to react quickly.

Quantitative Timeliness Agreement (QTA)

Beyond the "Quantity of Bandwidth" required to support an application, the "Quality of Bandwidth" required also needs to be understood and specified. Essentially, this is the latency and loss of the packets transporting the application. A cumulative distribution function (CDF) can capture these in a unified way. Expressing application requirements (of the network) in this way is known as a "Quantitative Timeliness Agreement" or QTA.

Thresholds on the CDF can be useful for expressing network capability (end-to-end and per link), application requirements, and even Service Level Agreements (SLAs). However, the exact "threshold" (e.g., 99%, 99.5%, 99.9% ..., etc.) varies by application, for example, control plane vs. user/data plane traffic.

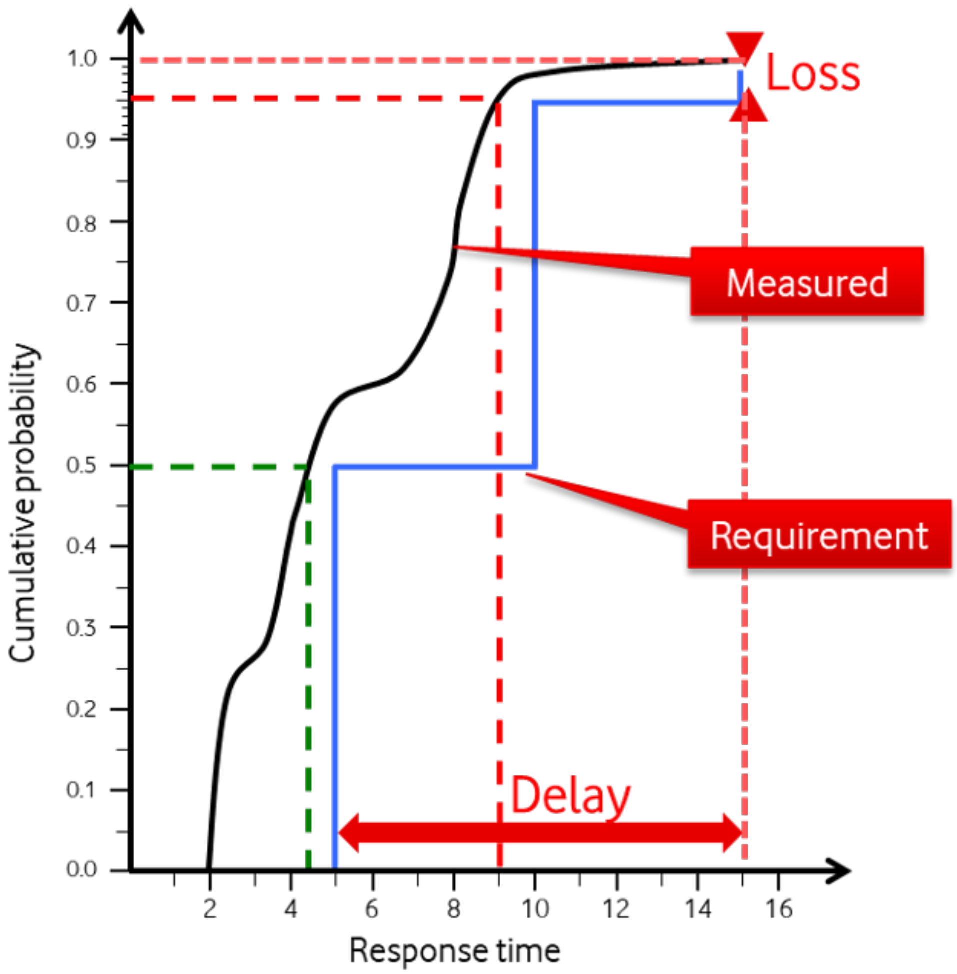

Figure 13 Quantitative Timeliness Agreement" or QTA example. Source Broadband Forum MR-452.2

In the QTA example in Figure 13, the blue line indicates that 50% of packets should arrive within 5 ms and 95% within 10 ms, with a packet loss rate of 0.5%. The black line depicts the measured network performance as a cumulative distribution function (CDF). This means the timeliness requirement is satisfied since it is to the left (i.e., better than) the specified requirements CDF. If it were not happy, the application's outcome would be at risk of not meeting the quality of experience (QoE) expectation. Table 6 presents a set of QoS metrics from developers working with Vodafone for online gaming.

Online game developer (source Vodafone)	
Metric	Tolerance for acceptable QoE
Packet round-trip delay ('latency')	< 75ms
Packet delay variation ('jitter')	< 15ms
Packet loss	< 0.5%
Throughput to sustain 720p	> 4 Mbps
HD Video Conferencing developer	
Packet round-trip delay	< 100ms
Throughput	> 1.2 Mbps

Table 6: Example QoS metrics from developers working with Vodafone

Table 7 below illustrates how developers' requirements are represented as QTAs for Video on Demand.

Scenario: Video on demand	
Metric	QTA (minimum centiles for values in CDF)
Time to first frame (round trip)	75% < 50ms 99.5% < 85ms
First frame loss	<0.5%
Packet delay after first frame (round trip)	50% < 100ms 95% < 200ms
Packet loss after first frame	<5%

Table 7: Illustrative examples of QTAs for video on demand

QTAs provide a common language for comparing application requirements and network measurements. Every application has a certain level of Quality Attenuation (packet latency/loss) that, if exceeded, will lead to a poor Quality of Experience (QoE) for the customer.

Quality of Outcome (QoO) quantifies the gap between application requirements and actual measured network performance within the QTA. This allows application developers to understand the quality users can expect during a network session. If needed, they can adapt application behavior to optimize user experience based on network constraints. Rather than using calculus to calculate the area between required and measured performance, QoO approximates this by analyzing key percentiles in the CDF. It measures how close the performance is to a threshold that ensures a great application outcome versus one that results in a poor outcome. Ultimately, it simplifies this into a percentage that quantifies the probability of a successful application outcome and, therefore, the user experience (QoE).

QTA and QoO are detailed in the following Broadband Forum (BBF) and Internet Engineering Task Force (IETF) references:

BBF: https://www.broadband-forum.org/marketing/download/MR-452.2.pdf

IETF: https://datatracker.ietf.org/doc/draft-olden-ippm-qoo/
https://datatracker.ietf.org/doc/draft-teigen-ippm-app-quality-metric-reqs/

Step 4: Measurements monitoring telemetry methodology & specifications

Once we define the requirements of the endpoints—service source and client device—for delivering next-generation services like the Metaverse, we must understand how these services are offered, the key performance indicators (KPIs) for quality of service (QoS), and whether they meet the quality of experience (QoE) needs. Endpoints that deliver the service, such as capture and replay devices, can measure QoS from end to end, including delay and jitter. When considering customer satisfaction, churn avoidance, and service level agreements (SLAs), it is essential to ensure that the network provides the necessary QoS for Metaverse services. This directly impacts service success and revenue.

Determining if a network meets the requirements for a specific Metaverse service is just one application of QoE measurements. There are additional applications for measurements and monitoring:

1. What is the capacity of a network to deliver a specific proposed Metaverse service before its deployment?
2. Is the network delivering the required QoS for specific Metaverse services?
3. When a network is not delivering the QoS required for a specific or Metaverse service, what are the root causes, and where are they located?
4. What vulnerabilities exist in a network that could jeopardize the reliable delivery of a specific Metaverse service?

Networks are becoming more complex. Disaggregation in the Radio Access Network means that more components can suffer impairments that impact QoE. The networks are also more dynamic, with network functions and service delivery components spinning up in different locations depending on demand and conditions. New radio technologies mean that radio resource control can fail in new ways. These technological evolutions have many benefits, but they also come with the challenge of detecting problems, pinpointing them when they occur, and knowing how to fix them.

Another trend is the increased richness and immersion of many Metaverse services.

These services are becoming more susceptible to temporary issues; even a slight delay can greatly impact the Quality of Service (QoS) and Quality of Experience (QoE) of specific Metaverse applications. Traditional network monitoring collects Key Performance Indicators (KPIs) over time, which can mask temporary impairments that affect QoE during these collection intervals. More detailed measurement aggregation can help in detecting these transient issues. Furthermore, impairments must be quickly identified to ensure fast resolutions and minimal impact on Service Level Agreements (SLAs). Consequently, lower latency in collecting and analyzing QoS measurements is likely necessary. The increased data needed for swiftly identifying issues can lead to high costs associated with generating, managing, and analyzing this data. Possible solutions include distributed QoS analysis to identify key service metrics, along with anomaly detection to trigger detailed QoS measurements only when necessary for diagnostics.

Distributed QoS Measurement Points

To deliver the measurement and monitoring use cases described above, granular visibility into the service endpoints and the communication network between them is required. The network is complex and may contain mobile and fixed aspects. Endpoints may be in the cloud or with the service consumer/user. Since impairments to the delivery of immersive services with target quality can arise from a variety of pathologies, including congestion, link failure, node failure, and service failure, for example, we can consider the various ways that networks can be measured.

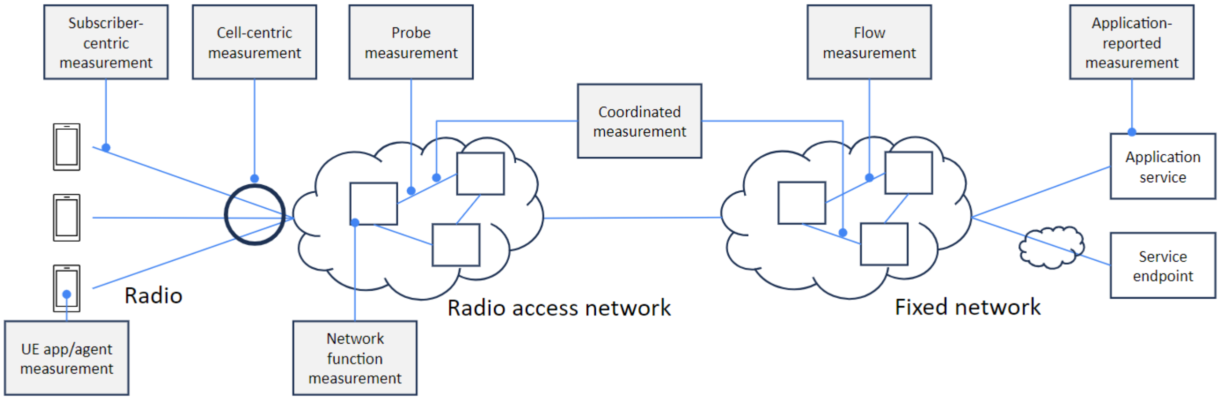

Figure 14: types of measurements and measurement points mapped onto a mobile network architecture.

Figure 14 illustrates the types of measurements and endpoints mapped onto a mobile architecture. The measurements may relate to a specific service user, a group of service users (such as users of a network slice), a flow between two endpoints, or aggregated data on a connection. Measurements can be taken at specific points in the network, such as the mobile endpoint device, the radio link, the RAN network functions, the transport layer, the core network, or the application server. They may focus on a specific point in the network or involve coordinated measurements between two or more locations (for instance, using two-way active management protocol (TWAMP)).

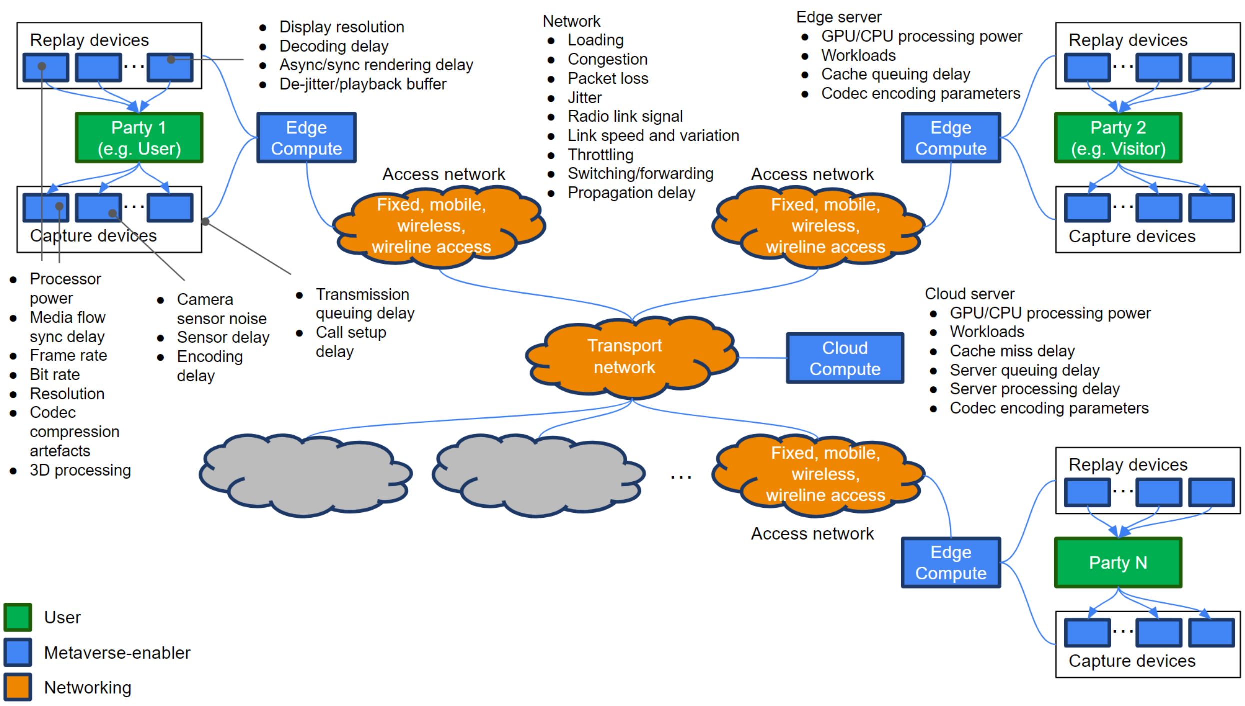

Figure 15: QoS impairment factors mapped to a Metaverse architecture highlight the main contributors, including end-user devices (capture and replay), network, and edge/cloud computing. Telemetry at both the application and network levels is becoming a strategic and essential metric for traffic, customer retention, operations, troubleshooting, and capacity planning. Transport protocols like QUIC and BBR impact traffic volume efficiency.

In Step 2, we introduced specific QoS impairment factors envisaged for metaverse-type immersive services. In Figure 15, we move from the generic architecture of a mobile network to an architecture specific to metaverse services and map those impairments identified in Step 2 to this metaverse-specific architecture.

Industry Initiatives Towards Network-as-a-Service

The Network as a Platform (NaaP) concept is emerging. It provides network service APIs that allow service and application developers to interact with networks without needing to understand or directly integrate with the underlying network technology. This leads to programmable service platforms encapsulated by NaaP, which enable functions to extract information from the network and configure it for specific services.

As a result, rich immersive services can be built on top of NaaP, ensuring that the QoS characteristics defined by the target QoE for that service are met. This area is evolving with the GSMA Open Gateway initiative, which develops and publishes APIs through the Linux Foundation CAMARA open-source project. This effort produces a set of APIs that network operators can use for network interaction. For instance, the CAMARA Connectivity Insights subproject enables developers to request performance-related information about a network's capability to meet specific SLAs through a standardized API.

The GSMA OpenGW project establishes interoperability between operators to utilize CAMARA APIs. While it is not yet widely adopted, the IETF Quality of Outcome (QoO) framework suggests mechanisms for sampling network quality, defining service quality requirements, and evaluating the likelihood of meeting them based on sampled quality measurements. Some commercial solutions are available; for example, Domos provides solutions that combine Quality Attenuation (QA) with customer experience according to application tolerance. Questions such as whether a network can support HD calls can be addressed beforehand.

The CableLabs Network as a Platform (NaaP) initiative is developing a set of APIs allowing developers to access fixed-access network features. This project works with other standardization efforts, such as CAMARA and TMForum, to ensure consistency between mobile and fixed networks, thus providing a similar experience for application developers in all areas.

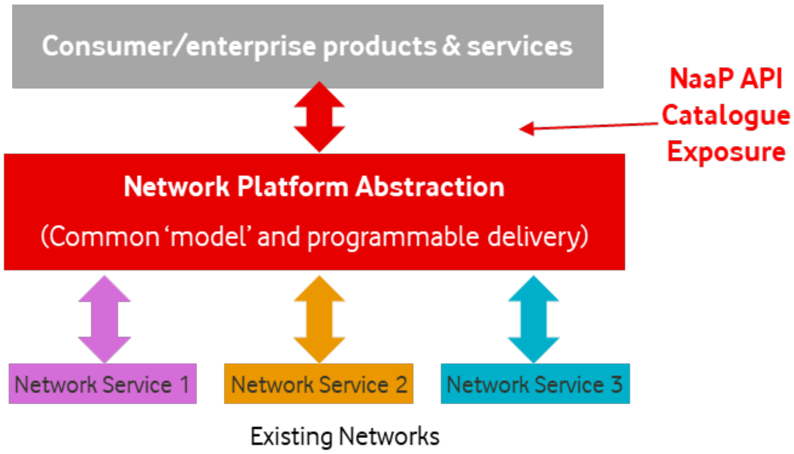

Figure 16: CABLELABS NaaP. Open Gateway

As the industry converges on NaaP APIs, shaping the resulting APIs can lead to more successful delivery of Metaverse services. For example, a network API that exposes the QoO for a particular application requirement would quantify the probability of a successful application outcome in terms of the target QoE for that network connection. Based on this, a decision could be made on whether to offer that service in general or on a specific occasion.

Survey of QoS Measurements

Passive and active network measurements are classified based on their assessment methods. They play an essential role in evaluating application Quality of Experience (QoE).

The active measurement method tracks the behavior of applications and end-users in real-time to determine network quality. This measurement involves injecting test traffic at various network points to monitor user or application traffic and measure its performance. Because test traffic mimics service traffic, active testing is ideal for providing a real-time view of end-to-end performance concerning latency (delay), jitter (delay variation), and packet loss. It helps segment the network, providing an end-to-end view, and validating and reporting on varying network path characteristics. Examples of active probing include Ping, Traceroute, TWAMP Light, STAMP, IRTT, varying latency under load tests, and simulating real traffic.

Passive measurements involve capturing and observing live traffic between hosts and applications at specific points in the network, such as aggregation routers or home gateways. This type of monitoring allows for offline analysis of signaling protocols, application usage, or the top bandwidth consumers. Passive tracking is best suited for in-depth traffic and protocol analysis and can provide visibility into the actual quality of experience for users and applications. Examples of passive monitoring techniques include DPIs, NetFlow, IPFIX, and others.

ITU-T Y.1540, 3GPP TS 26.234, ITU-T P.1203, and IETF RFC 6703 are some notable application QoE measurement standards that provide guidelines for active and passive monitoring.

The industry has various approaches for measuring QoS (latency).

- Broadband Forum Quality Attenuation: As described in Step 2, the Quality Attenuation measurement method standardized by the BBF assesses the network's impact on application outcomes. Rather than relying on a single round-trip time (RTT) parameter, it provides a statistical distribution of six separate latency components—three for upstream and three for downstream, as shown in Figure 17. Understanding the contribution of each element to overall network quality attenuation helps identify the most effective improvement strategies. Quality attenuation enables the breakdown of network quality into components that represent different root causes. This method allows us to add and subtract Quality Attenuation, which helps measure the performance of the WiFi link, the link to the ISP, and the link from the ISP to the server, allowing us to combine these measurements for end-to-end performance. In contrast, other metrics, such as the 99th percentile of latency or average values, do not provide this level of detail.

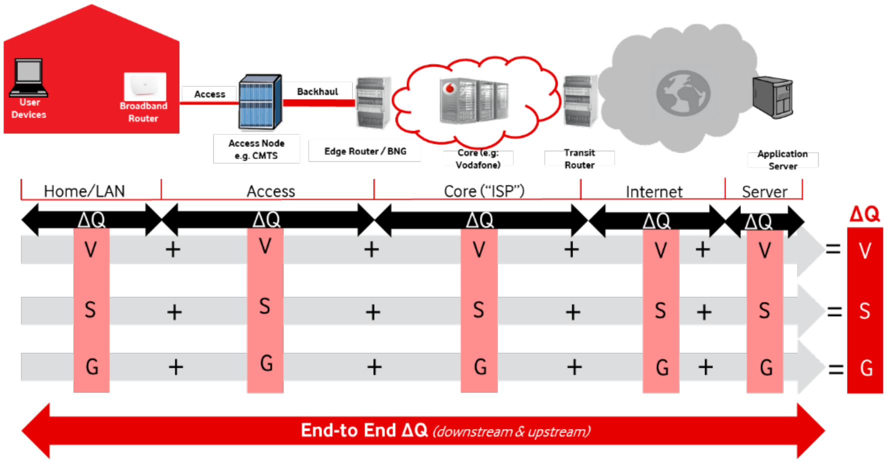

Figure 17: BBF TR 452 Quality Attenuation: The G (Geographic) component is related to propagation delay, which is determined by physical distance and the speed of light. The S (Serialization) component arises from clocking packets in and out of network nodes. The V (Variable) component is due to queuing, buffering, and scheduling, which are impacted by network load.

- CableLabs: The CableLabs "Latency Measurement Metrics and Architecture" report recommends tracking latency metrics and describes an architecture to implement in operator networks. The simple end-to-end latency measurement framework uses IETF's STAMP (RFC 8762, 'Simple Two-way Active Measurement Protocol) and LMAP (RFC 7954, 'A Framework for Large-Scale Measurement of Broadband Performance') technologies. A measurement software prototype from the CableLabs Common Code Community (C3) program is available for use. It contains session reflectors, measurement agents, and controller/collector components. A specification of latency test definitions and STAMP extensions will be published soon.

- IETF IPPM draft-cpaasch-ippm-responsiveness-03 aims to raise awareness of and encourage solutions to the issue of bufferbloat in networks, a major cause of latency and packet loss. The draft specifies the "Responsiveness Test," expressed as "Round Trips Per Minute," utilizing common protocols and mechanisms to measure user experience, specifically under normal working conditions, to identify degradation in network performance related to bufferbloat.

Step 5: Apply QoE-QoS Framework to selected Use Cases

In step 5, we apply the QoE engineering framework to a specific use case. This section presents a brief overview of the use cases; for more details, the reader is invited to read the full use case white papers.

Short Form Video - use case summary

Short-form video (SFV) is increasingly popular. According to the Ericsson Mobility Report (June 2023 edition), SFV accounts for 20 to 30% of total mobile broadband traffic in North America. According to Statista, typical content lasts 30 to 42 seconds.

Different from traditional video streaming (Youtube™, Netflix™ ...), new features brought by SFV services impact QoE in various degrees:

- Short video contents are pre-loaded according to the recommendation algorithm. An optimized pre-loading strategy should be applied to satisfy the user experience with reasonable loading time while not over-preloading the contents that cause network congestion, thus impacting QoE.
- End-users frequently scroll or swipe their screens in a short period, even if they haven't finished watching the entire video. Therefore, Quality of Experience (QoE) models should consider the length of the video, as users will be more sensitive to initial loading times and buffering events when the video's content is brief.

QoE Aspects	QoE metrics	How to measure
Audio quality	Audio MOS	Codec type, Bit rate, PESQ, POLQA, E-model R factor ITU G.107
Timeliness	Click to play time (CTPT)	Measured on the client, the interval between the time when a user click a video and the time when the video starts to play on the screen
	Play success rate n (PSRn)	Percentage of SFV views which has a CTPT less than n seconds
	Stalls	Measured on the client side per viewing session by some/combination of (1) number of stalls (longer than xxx ms) (2) total time of stalls (milliseconds) (3) meantime between stalls during a session
Temporal quality	Fluidity	Measure on client by number of frames per second
	Synchronicity	Measured on the client side per viewing session by some/combination of (1) numbers of audio/video out-of-synch (2) total time of audio/video out-of-synch (3) meantime between audio/video out-of-synch
Spatial quality	Video fidelity	No Reference: Under Development Full Reference: PSNR, SSIM, or VMAF
Context**	Client Device, Location	Display resolution and audio fidelity, mobile or stationary, network type
Human factors	User rating	Users are asked, during or after their viewing, to rate their satisfaction in scale of 1 - 5 Content interest

Table 8 : QoE metrics under consideration for short form video
*** The context may impact on target values of QOE metrics for what an acceptable/good/excellent QoE is but not as a QoE metric.*

Short-form video requires substantial backend effort, including streaming and the AI inference engine. This setup ensures smooth playback and provides the content that users enjoy, helping to maintain their attention and retention. The network is crucial for these aspects, and monitoring at both the application and network levels has become a strategic metric for traffic analysis, customer retention, operations, troubleshooting, and capacity planning. Furthermore, transport protocols like QUIC and BBR affect traffic volume efficiency and should also be considered when delivering these services.

3D Volumetric Video Telepresence – use case summary

The Real-Time 3D Immersive Telepresence use case is a perfect fit for the evolving world of digital connectivity and realism, known as the metaverse. With media types such as volumetric video data and network performance levels that allow for scalability, these digital spaces are filled with content that can be fully interacted with. In contrast, 2D video conferencing struggles with disconnects, such as a lack of detail or difficulty in achieving proper perspective, making capturing a narrative's essence challenging. This often leads to inadequate productivity and emotional fulfillment.

The following tables present the initial Quality of Experience (QoE) requirements for this implementation. These requirements cover system components and various QoE indicators aimed at positively influencing user experience. The proposed values mainly focus on providing satisfactory experience levels. Advanced values, where proposed, are based on industry trends and reflect highly desirable QoE levels. Testing and data collection will further identify or validate these advanced values.

QoE Indicator	QoE Influenced	Experience Type Response Type QoE Type	Proposed Value
Display Resolution	Visual Quality (Objective)	Cognitive + Perceptual Satisfaction Usability	Initial - 4K 4096 x 2160 Advanced - 8K min. or comparable for PPD = 60 w/ H-FoV = 210 deg.
FoV	Visual Quality (Objective)	*As above.*	Initial – 110 deg. Advanced – 210 deg.
Spatial Pixel Density	Visual Quality (Objective)	*As above.*	3200 PPI min.
Angular Pixel Density (PPD)	Visual Quality (Objective)	*As above.*	30 (Initial) 60 (Advanced)
Brightness	See Through Capability (Objective)	Cognitive Attitude Usability	Variable as Required 200 - 500 nits

Display Refresh Rate	Normal Visual Fatigue (Subjective)	Emotional + Cognitive Attitude / Satisfaction Subjective / Usability	120 Hz min. 144 Hz Preferred 360 Hz (Advanced)
Display Refresh Rate	With Judder or non-smooth motion. To Tackle Blur. With Multiple Imaging. (Subjective)	Emotional + Perceptual Attitude / Satisfaction Subjective / User-Behavior	144 Hz
Viewport Drift	Visual Quality (Subjective)	Emotional + Perceptual Immersion / Satisfaction Usability / User-Behavior	< 0.1m
Viewport Smoothness	Visual Quality (Subjective)	*As above*	< 0.01m
Perturbation (L2 Norm w/ countermeasure)	Visual Quality (Subjective)	*As above*	< 0.05

Table 9: QoE Influencing Requirements for the "Consumption" Stage of Volumetric Video-based Live and Real-Time Telepresence Use Case. Source: T-Mobile Research

Transmitting volumetric video content requires large amounts of data. Therefore, live streaming this content needs a network capable of handling a moderately large bandwidth necessary for multiple streaming objects within a scene. Fast and efficient compression algorithms are essential, and this network should also provide low latency to enable rapid interaction.

Requirement	Proposed Initial Value
Latency - Scene in Viewport Update Delay (Fig. 8 for reference) ● 1- way peer to peer latency / end-to-end ● User to visitor ● Pose to Render to Photon ● Comparable to Glass (Camera) to Glass (HMD) Latency ● Viewport Independent Delivery (Fixed Viewport)	=< 100 ms * (Over Metaverse Ready 5G Network)
Latency User Head Movement to Display Update Delay (Fig. 8 for reference) ● Motion to Photon Delay ● Viewport Dependent ● Sensor to optimized pose prediction /correction to rendering to Display ● Optimized Rendering latency < 10 ms (With pose prediction, correction)	< 20 ms**
Downlink Bandwidth (Consumption End Content Delivery)	50 Mbps ***
Start-up-Delay (Application Metric)[7]	< 2s
Stall Duration / Re-buffering (Application Metric)[7]	< 10ms

Table 10 Essential QoS
* Utilizing appropriate 5G QoS class.
** This can be the time to display a pre-rendered or based upon pre-rendered volumetric content (3D content may be already available or partially available). Update viewport to match visitor's 6DoF shift.
*** Point Cloud Size = 100,000 & FPS = 30. 4K texture resolution with H.265 compression

Conclusions

Delivering an improved product user experience (QoE) will ensure customer satisfaction and market success while enhancing the ecosystem and value proposition for all partners. For any multimedia service to succeed, it is essential to plan from an end-to-end perspective to identify the critical QoE and QoS factors that influence success. A proposed top-down, user-centric design approach includes an end-to-end system that establishes perceptually based QoE targets and maps the corresponding QoS impacts specific to different use cases. Additionally, it features standardized and practical methods for ongoing measurement and monitoring of user impact on quality aspects.

Using QoE requirements to guide network engineering and design offers two key benefits: (1) network design targets focus on user needs and experiences, making services more appealing to potential users; (2) it prevents over- or under-engineering, enabling providers to deliver high-quality content efficiently. QoE is also connected to application and network layer attributes and their associated QoS through a framework utilizing HRX models, which we reviewed in detail within the top-down framework.

Products and services must be designed to achieve quality-of-experience targets. We proposed a top-down approach based on end-user perceptual requirements and network infrastructure to attain this goal and align with commercial impact.

References

[1] QUTY: Towards Better Understanding and Optimization of Short Video Quality. H Zhang, Y Ban, Z Guo, Z Xu, Q Ma, Y Wang, X Zhang, Proceedings of the 14th Conference on ACM Multimedia Systems, 2023

[2] https://standards.ieee.org/project/3333_1_3.html

[3] QUALINET White Paper on Definitions of Immersive Media Experience (IMEx), Perkis, A., Timmerer, C., et al.,European Network on Quality of Experience in Multime This technical report results from a collaborative effort within TIP's Metaverse Ready Network Project Group

[4] ITU-T P.1301 Subjective quality evaluation of audio and audiovisual multiparty telemeetings (2017)

[5] ITU-T P.1305 Effect of delays on telemeeting quality (2016)

[6] ITU-T P.1310 Spatial audio meetings quality evaluation (2017)

[7] ITU-T P1312 Method for the measurement of the communication effectiveness of multiparty telemeetings (2016)

[8] Y. Maehara et T. Nunome, « WebRTC-Based Multi-View Video and Audio Transmission and its QoE », in 2019 International Conference on Information Networking (ICOIN), janv. 2019, p. 181-186. doi

[9] G. Bingöl, S. Porcu, A. Floris, et L. Atzori, « QoE Estimation of WebRTC-based Audio-visual Conversations from Facial and Speech Features », ACM Trans. Multimedia Comput. Commun. Appl., p. 3638251,

[10] Optimizing Performance of Zoom in Low Bandwidth Environments

[11] UVQ - Measuring YouTube's Perceptual Video Quality, 2022 Yilin Wang, Staff Software Engineer, YouTube and Feng Yang, Senior Staff Software Engineer, Google Research

[12] Two-level approach for no-reference consumer video quality assessment, J Korhonen, IEEE Transactions on Image Processing, 2019

[13] Efficient Measurement of Quality at Scale in Facebook Video Ecosystem, SPIE Optics + Photonics, Meta Research 2020

[14] Quantifying the value of 5G and edge cloud on QoE for AR/VR. B Krogfoss, J Duran, P Perez. 2020. ieeexplore.ieee.org

[15] A survey on QoE-oriented VR video streaming: Some research issues and challenges. J Ruan, D Xie - Electronics, 2021 - mdpi.com

[16] Quality of experience in telemeetings and videoconferencing: a comprehensive survey. J Skowronek, A Raake, GH Berndtsson, OS Rummukainen, P Usai, SNB Gunkel. 2022 ieeexplore.ieee.org

[17] On additive and multiplicative QoS-QoE models for multiple QoS parameters. T Hossfelt, L Skorin-Kapov, PE Heegaard. 2016 - ntnuopen.ntnu.no

[18] Why no reference metrics for image and video quality lack accuracy and

reproducibility. MH Pinson - IEEE Transactions on Broadcasting, 2022 - ieeexplore.ieee.org
[19] No-reference image and video quality assessment: a classification and review of recent approaches M Shahid, A Rossholm, B Lövström, HJ Zepernick. EURASIP Journal on image and Video Processing, 2014·Springer
[20] VQEG Immersive Media Group.
[21] Swipe along: a measurement study of short video services S Zhu, T Karagioules, E Halepovic, A Mohammed, AD Striegel Proceedings of the 13th ACM Multimedia Systems Conference, 2022·dl.acm.org
[22] M. Joras, “Securing Ancillary Data for Communicating with Devices in the Network,” Internet Engineering Task Force, Internet Draft draft-joras-sadcdn-01, Jul. 2023. Accessed: Nov. 08, 2024. [Online]. Available: https://datatracker.ietf.org/doc/draft-joras-sadcdn