ACCORDION

Adaptive edge/cloud compute and network continuum over a heterogeneous sparse edge infrastructure to support nextgen applications

DRAFT

Deliverable D6.2

Requirements on Quality Assessment I

ACCORDION receives funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 871793

DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	ACCORDION
PROJECT FULL NAME	Adaptive edge/cloud compute and network continuum over a heterogeneous sparse edge infrastructure to support nextgen applications
STARTING DATE	01/01/2020 (36 months)
ENDING DATE	31/12/2022
PROJECT WEBSITE	http://www.accordion-project.eu/
TOPIC	ICT-15-2019-2020 Cloud Computing
GRANT AGREEMENT N.	871793
COORDINATOR	CNR
DELIVERABLE INFORMATION	
WORKPACKAGE N. | TITLE	WP 6 | Integration, Pilot Implementation & Evaluation
WORKPACKAGE LEADER	OVR
DELIVERABLE N. | TITLE	D6.2 Requirements on Quality Assessment I
EDITOR	S. Zadtootaghaj (TUB)
CONTRIBUTOR(S)	S.Zadtootaghaj(TU Berlin), Maria Pateraki (OVR), Zbyszek Ledwoń (ORBK), Tom Loven (PLEX), F. Diego Andilla (TID), Mike Kentros (OVR), Patrizio Dazzi (CNR)
REVIEWER	Ferran Diego Andilla (TID)
CONTRACTUAL DELIVERY DATE	Original: 31 December 2020, Amended: 31 January 2021
ACTUAL DELIVERY DATE	31 January 2021
VERSION	1.0
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	42
KEYWORDS	Quality Assessment, Subjective Test, QoE Modeling

EXECUTIVE SUMMARY

This deliverable reports about the requirements for the development of the Quality of Experience (QoE) model for ACCORDION. The document focuses on a parametric-based monitoring models that can predict the mean opinion scores measuring the impact of impairments introduced by typical networks and compression on the quality experienced by users.

The deliverable focuses on the OVR, PLEX and ORBK applications and its main objective is to plan the necessary steps supporting the development of QoE models. The deliverable includes information on the scope of the model, list of influencing factors, range of parameters per use case, model structure, test structure, and information about test setup, test participants, selection of stimuli, considered quality features, test procedure, condition plan and the tool that will be used for the subjective test. This report also provides the plan for future work.

Details about the assessment tools developed for the ORBK use case are given in Appendix A.

DISCLAIMER

ACCORDION (871793) is a H2020 ICT project funded by the European Commission.

ACCORDION establishes an opportunistic approach in bringing together edge resource/infrastructures (public clouds, on-premise infrastructures, telco resources, even end-devices) in pools defined in terms of latency, that can support NextGen application requirements. To mitigate the expectation that these pools will be "sparse", providing low availability guarantees, ACCORDION will intelligently orchestrate the compute & network continuum formed between edge and public clouds, using the latter as a capacitor. Deployment decisions will be taken also based on privacy, security, cost, time and resource type criteria.

This document contains information on ACCORDION core activities. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the ACCORDION Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 members states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

REVISION HISTORY LOG

VERSION No.	DATE	AUTHOR(S)	SUMMARY OF CHANGES
0.1	10/12/2020	S.Zadtootaghaj(TU Berlin)	First Draft
0.2	21/01/2021	S.Zadtootaghaj(TU Berlin) Maria Pateraki (OVR) Zbyszek Ledwoń (ORBK) Tom Loven (PLEX)	Revision
0.3	27/01/2021	S.Schmidt (TU Berlin) Tom Loven (PLEX)	Revision
0.4	29/01/2021	F. Diego Andilla (TID) S.Zadtootaghaj(TU Berlin)	Internal Review Revision according to comments
1.0	31/12/2021	Patrizio Dazzi (CNR)	Executive Summary and proofreading

GLOSSARY

AR	Augmented Reality
EC	European Commission
FEC	Forward Error Correction
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
HMD	Head-Mounted Displays
IPQ	Input Quality
MOS	Mean Opinion Scores
QoE	Quality of Experience
RTT	Round-Trip Time
UDP	User Datagram Protocol
VR	Virtual Reality

TABLE OF CONTENTS

1 Introduction

1.1 Scope and objectives of this deliverable

In this deliverable, a short report on requirements for the development of the Quality of Experience (QoE) model for ACCORDION use cases is described. The document targets the development of parametric-based monitoring models that can predict the mean opinion scores (MOS) on a 5-point ACR scale (see [1], [2]) based on the impact of impairments introduced by typical networks and compression on the quality experienced by users using the OVR, PLEX and ORBK applications.

The main objective of this deliverable is to plan the necessary steps before the development of QoE models, including defining the scope of the models, the range of the parameters, planning subjective tests, and developing the structure of the model. The targeted models are monitoring tools that can be used by service/network providers for purposes such as edge and cloud resource allocation and configuration of IP-network transmission settings.

The goal of Task 6.3 is to provide information describing all necessary steps to conduct the subjective tests for the development of the QoE model for each use case of ACCORDION. The deliverable must include information on the scope of the model, list of influencing factors, range of parameters per use case, model structure, test structure, and information about test setup, test participants, selection of stimuli, considered quality features, test procedure, condition plan and the tool that will be used for the subjective test. Up to this point of the project, a full description of mention steps are provided for ORBK use case, whereas due to the challenges discussed in Section 1.3, some steps are provided for other two use cases of the project from OVR and PLEX.

1.2 Relation to other work packages

This deliverable is the baseline for work package 5, where the model must be developed. Task 6.3 must provide details for the subjective experiment to collect data for the development of the model. In addition, the core model structure must be defined in this deliverable.

1.3 Challenges and Issues

The main process of task 6.3 is to collect a large amount of data through subjective tests. Due to the Covid-19 pandemic with the increase of high infection and stricter restrictions in Europe, the subjective tests for OVR and PLEX applications are postponed to 2021. However, it is planned to monitor the situation and replan the activities accordingly. In addition, for planning the subjective test, pilot tests are typically employed for selecting assessment tools (e.g., questionnaires) before conducting the actual experiment. Similar to the main subjective tests, pilot tests are also postponed, which impact the project further, especially the planning of subjective tests for two use cases, OVR and PLEX.

1.4 Structure of the document

In this report, short descriptions of each use case are given in Section 2. The scope of each model is discussed in detail in Section 3. A list of influencing factors that may impact the QoE of each use case of ACCORDION is provided in Section 4. Section 5 gives an overview of the range of parameters that each model works. This includes the parameters that will be used for testing the model as well as the range, value, or mode of each parameter. The experimental setup is described in Section 6, where details about the subjective test, data collection (for use case 2), and assessment tools are given. Section 7 and 8 provide information about the structure of the quality model for each use case. Finally, the plan for future work and conclusion of the deliverable is given in Section 9. Details about the assessment tools developed for the ORBK use case are given in Appendix A.

2 Use Cases

2.1 OVR application

OVR use case aims to support collaborative cloud VR training applications specially formulated for untethered HMDs, and the adaptation of OVR's networking layer to edge computing will optimize the current status of the cooperative mode, ensuring lower latency and higher performance on average network conditions and ultimately a higher number of CCUs. To exploit the functionalities of the ACCORDION platform, the application is redesigned to enable offloading of certain computations across the edge minicloud or in central clouds eliminating dependencies on proprietary API's linked to specific HMD vendors.

2.2 ORBK application

The ORBK use case is a multiplayer mobile online game. Game servers will be deployed on top of the ACCORDION system to meet the requirements of NextGen mobile online gaming, which aims to lower latency between servers and clients and highly improve user experience. It will also take advantage of AI-based network orchestration to dynamically and automatically deploy new servers based on performance metrics and the player's geographical localization.

The ORBK game system will consist of two elements: game server and mobile application. Mobile application will be run by the end-users – players on their own mobile devices. The game will be available for download to the Play App Store. The game system must be able to handle up to 100 players, handle a huge number of in-game events while performing a full simulation of the game world, and generate responses with a minimal possible delay.

2.3 PLEX application

The PLEX use case consists of two sub use cases. For this document the first use case, a mobile augmented reality game, has been analysed. The second sub use case will be taken into account in the next phase of the project.

1. Mobile Augmented Reality game
   - The game is an augmented reality customer loyalty game where points can be won based on clients playing against each other. An 3D object needs to be captured in order to gain points. Depending on the interaction between the user and their coordinates the game scenario changes. The following image shows the 3D object to be captured by the user. The 3D object will be based on the logo of the client (hospital, mall, train station, etc.). In this case the 3D object is based on the logo of Plexus.

Figure 1 Mobile augmented reality game with 3D object

2. Client identification and adapted content displayed on screens (digital signage)

– Based on the characteristics obtained from the mobile device information from the clients in a certain space like a mall, content reproduced on the information screens within that space is adapted. For example, in case there are more young players compared to elderly players, the content can be adapted to their interest. The following image shows digital signage solution ´Anblick´´from Plexus.

Figure 2 Digital signage solution Anblick

3 Scope of QoE models

In this section, the scope of the QoE models for ACCORDION use cases is described. The scope of each model needs to describe the focus of the model, the output of the model, the range of parameters it works, and explain boundaries that the model is not functional. The models under development are parametric based models that can be used by network providers and service providers to build their infrastructure, the allocation of resources (distribution of users based on the distance or load of servers), as well as monitoring the quality based on the network, client and compression parameters.

3.1 OVR Use Case

The focus of the OVR QoE model is to predict the Quality of Experience (QoE) for Virtual Reality (VR) service provided by OVR, considering relevant factors that are identified and discussed in ITU-T Recommendation G.1035 [8].

The OVR application supports collaborative VR training through the online, real-time collaboration of remote participants in a shared virtual environment. There are two modes of operation that will be explored as part of the ACCORDION project. On the first mode, all processing takes place on the client device of each user (i.e., untethered HMD) and a relay server based on Photon networking handles multicasting, host migration, and game state continuity between the clients. State changes that are linked to interactions with the virtual environment are transmitted through the network, and the state of the application is updated locally at each client node. The model aims to address impairment factors based on network parameters, delay, and packet loss considering the round-trip time (RTT) of the network transmission, the incoming /outgoing message, position, and rotation rate (see also section 7.1). On the second mode, the processing is partitioned between the HMD and the edge node following the ''full'' offloading option, and information is encoded, streamed, and decoded on the HMD over user datagram protocol (UDP), while the communication between distant clients resides on the networking solution as in the first mode. In addition to the aspects related to network impairments of the first mode, the QoE model targets the investigation of various video encoding parameters to adapt to the network throughput, packet loss, and end-to-end delay.

In this report, only the requirement for streaming mode is described, and the collaborative online mode will be investigated in the second phase of the project.

The core model is an impairment model inspired by the E-model (ITU-T Recommendation G.107), which will be driven by analyzing the suitability of a variety of quality features for the prediction of the QoE. The model will be trained using data collected from participants that have experience with medical training tasks of the OVR application. This experience can be gain through short training sessions.

3.2 ORBK Use Case

The focus of the ORBK QoE model is to predict gaming QoE for online mobile gaming services by considering relevant factors that are identified and discussed in ITU-T Recommendation G.1032 [3]. The impairment factors are driven based on network parameters, delay, and packet loss. The model can be used by network providers and mobile online gaming service providers to monitor the quality, build their infrastructure, allocate the resources (distribution of users based on the distance or load of servers) under the assumption of a network that is prone to packet loss and delay. While the focus of the model is on mobile online gaming, it may also be applicable to traditional online gaming on desktop computers or consoles. The presented model targets game with a high level of sensitivities towards network degradations.

The model is not designed to predict the influence of games' design or the motivation of users to play them, even though some information about the game content itself is of relevance to develop a more generalizable model. Furthermore, the influence of social factors, which are arguably important, especially for multiplayer games, will not be covered by the model. In line with ITU-T Recommendation G.1072, the audio quality will not be part of the model, as network degradation can only have a minor impact on this quality aspect due to the local processing/generation on the client phone. The effect of network degradations depends on the implementation of an online gaming service, but much of this information is typically not available to the network operator. These include the server tick rates (i.e., how often the state of the game is updated on the server), methods for FEC, delay compensation methods, buffer strategies, etc. In addition to these service implementation parameters, the user device may influence the display size and touch input responsiveness of the mobile phone. In respect to the screen size, studies in [5, 6] showed that there is no significant impact of the screen size on gaming QoE once a threshold of about 5” is exceeded, as this avoids usability issues. Thus, we aim at using a screen size of at least 5.5”. Additionally, when discussing the impact of delay, we consider the RTT of the network transmission, while the touch display and screen refresh rate also contribute to the overall delay. Thus, we aim to use mobile phones with a low but constant touch delay. Lastly, it will be assumed and controlled that the load on the game server is not degrading the player experience during tests.

It must be noted that on the contrary to cloud gaming, no video streaming is performed. Instead, packets to control the game state on the server(s) are sent from the client, and responses are sent back to the clients after the user input is applied on the server.

3.3 PLEX Use Case

The focus of the PLEX QoE model is to predict gaming QoE for online Augmented Reality (AR) mobile gaming services by considering relevant factors that are identified and discussed in ITU-T Recommendation G.1032 [3]. The scope of the model follows the ORBK due to the similarity of applications with differences in the range of parameters, which is described in Section 5.3. Thus, for brevity, we refer to the scope described for ORBK.

4 Influencing Factors

The COST Action Qualinet has defined a QoE influence factor as follows [2]:

“Any characteristic of a user, system, service, application, or context whose actual state or setting may have an influence on the Quality of Experience for the user.”

In this report, the factor groups proposed in the Qualinet White Paper is followed. These are Human influence factors, System influence factors, and Context influence factors.

The list of relevant influence factors for the ACCORDION use cases is shown in Table 1. Detailed information on each factor is presented in ITU-T Recommendation G.1032 [3] in the context of gaming services, and ITU-T Recommendation G.1035 [8] for virtual reality application.

Table 1: List of relevant influencing factors.

Factors	OVR UC	ORBK UC	PLEX UC
Experience		X	X
Intrinsic and extrinsic motivation		X	X
Static and dynamic human factors		X	X
Simulator Sickness	X		
Expectations	X		
Human vision	X	X	X
Task of User	X		
Temporal and spatial features	X		
Aesthetics and design characteristics	X		
Learning difficulty	X	X	
HMD Specification	X		
Head-tracking	X		
Field of View	X		
Game Genre		X	X
Game mechanics and rules		X	X
Dynamics and maximum successful reaction time	X	X	X
Pace		X	X

Visual perspective of the player		X	X
Aesthetics and design characteristics		X	X
Learning difficulty	X	X	X
Device Portability	X	X	X
Handheld Device Size	X	X	X
Input modalities	X	X	X
Output modalities	X	X	X
Display	X	X	X
Delay	X	X	X
Jitter	X	X	X
Physical environment factors	X	X	X
Social context	X	X	X
Service factors	X	X	X
Novelty	X	X	X

5 Area of Application

5.1 OVR Use Case

The focus of the OVR QoE model is on an interactive virtual reality application in the medical domain. It must be noted that depends on the task of the OVR application, the exchange of information between the server and clients could be frequent or limited. In a case of very limited exchange between the client and server, the task needs to be labeled as a low sensitive task. In addition, the task of the user could be critical (e.g., surgery) or less critical (body temperature check), which must be labeled for building the model.

The model will be developed based on a series of subjective quality studies using interactive paradigms. For the subjective test, head-mounted displays (HMD) are used with the specification that is described in Table 2.

Table 2: HMD screen specification used in development of the model

Resolution	Angular Resolution	Field of View	Refresh rate
1440 x 1600	16 average PPD	95 °	72

In particular, the factors and application range summarized in Table 3 are considered. It has to be noted that the model only works in the range of parameters that are listed in Table 3.

Table 3: Factors and application ranges of the model

Application information	Value range, unit
Stimulus duration	90 seconds
Resolution	1600x1440
Input devices	---
Packetization	RTSP (over RTP/UDP/IP)
Video codec	h.264
Coded video bitrate (kbps)	1000-50000
Frame rate (fps)	30, 60, 72
Group of Pictures *(Note 1)*	---
Pre-set	---
Encoding Mode	CBR
Video Compression	Standard H.264, Main 4.0
Audio codec	AC3

Coded audio bitrate (kbps)	192 (stereo)
Audio sample rate (Hz)	48,000
Packet loss degradation	uniform loss (0-2%)
Delay Range	0 – 200 ms

The following factors were not considered during model design:

- Audio/video sync distortions
- Different levels of audio quality
- Packet loss and delay distribution
- Video codecs for which the model is not validated (MPEG2, HEVC, VP9, AV1, etc.)
- Transcoding solutions
- The effects of noise and color correctness in a video
- Different ranges of parameters than the ones tested, e.g. delay of 800 ms

5.2 ORBK Use Case

The focus of the ORBK QoE model is the interactive mobile online gaming services. It must be noted that a variety of mobile games are only played locally without a frequent exchange of information between the server/node and clients. As these games are not impacted by network degradations, they will not be considered for the development of the model. The model will be developed based on a series of subjective quality studies using the interactive paradigms proposed in ITU-T Rec. P.809 [4]. For the subjective test, mobile devices are used with the display size of larger than 5 inches, as it was shown in [5] that there is no impact of the screen size on mobile gaming QoE once a threshold of about 5” is exceeded, as this avoids usability issues. Characteristics of mobile devices that are used in the experiments to develop the model are presented in Table 4.

Table 4: Mobile screen specification used in development of the model

Display size	Pixel resolution	Pixels per inch	Brightness	Refresh rate
6.1- 5.7	1792x828, 2280x1080	326, 443	625,445 nits	60 Hz

In particular, the factors and application range summarized in Table 5 are considered. In order to have a realistic simulation of network parameters, also considering time variation of delay and non-uniform packet loss, a large number of cellular network traces was collected using a crowdsourcing approach. Based on the measurements, a relation of delay and jitter, as well as packet loss rate and burst rate, was derived, which represents the most typical pattern of these pairs (likelihood).

Table 5: Factors and application ranges for mobile online gaming experiments.

Application information	Value range, unit
Stimulus duration (secs)	90 (interactive)
Screen size	5.7, 6.1”
Input devices	Touch display
Jitter (ms)	0 - 235
Delay (RTT in ms)	0 - 1000
Packet loss (%)	0 - 50

5.3 PLEX Use Case

The focus of the QoE model is interactive online mobile AR gaming services provided by PLEX. It must be noted that the PLEX AR game is only played locally without a frequent exchange of information between the server/edge node and clients. Since the game is not severely impacted by network bandwidth, network bandwidth is not be considered for the development of the model. The model will be developed based on a series of subjective quality studies using the interactive paradigms proposed in ITU-T Rec. P.809 [4]. For the subjective test, similar to ORBK use case, mobile devices are used with display size of larger than 5 inches, as it was shown in [5] that there is no impact of the screen size on mobile gaming QoE once a threshold of about 5” is exceeded, as this avoids usability issues. Characteristics of mobile devices that are used in the experiments to develop the model are presented in Table 6.

Table 6: Mobile device specification used in development of the model

Display size	Pixel resolution	Pixels per inch	Brightness	Refresh rate
5.80-inch	720x1560 pixels	aspect ratio of 19.5:9(296 ppi)	473 cd/m^2	60

In particular, the factors and application range summarized in Table 7 are considered. It has to be noted that the model only works in the range of parameters that are listed in Table 7.

Table 7: Factors and application ranges of the model

Application information	Value range, unit
Stimulus duration	90 seconds
Resolution	720x1560
Input devices	Touch display
Frame rate (fps)	60, 30, 20
Video Compression	Standard H.264, Main 4.0
Audio codec	AC3
Coded audio bitrate (kbps)	192 (stereo)
Audio sample rate (Hz)	48,000
Packet loss degradation	(0-30 %)
Jitter	0 - 235
Delay Range	0 – 1000 ms

6 Experimental Setup

6.1 OVR Use Case

Contrary to the gaming quality assessment, there is no validated/standardized subjective evaluation method to be followed for VR experiments. Thus, the necessary information that must be considered in the VR experiment is discussed in this section. In general, the ITU-T Recommendations of P.809 and P.910 are considered for the setup of the experiment, while the new aspects of VR will be taken into account not only for test setup but also for assessment tools and instruction of participants. In the following, the important points that must be considered for the VR test setup are discussed.

6.1.1 Test environment

A special test room must be set up for VR quality assessment. This room should let participants use the HMD freely without any distraction while ensuring that the participant will not hit the wall or fall due to obstacles on their path.

Before starting any experiment, participants should be instructed on how to set-up the HDM properly. This includes the fixation of the device, the settings of the lens's distance, and the calibration of the display brightness. In addition, a user interface for the questionnaire, such as an overall quality rating interface in the VR environment, is recommended to avoid distractions due to taking off the HMD. This might also be relevant when judging immersion and presence.

As a choice of HMD, for the subjective test, it is decided to use Oculus Quest due to compatibility with the use case.

6.1.2 Participant Interruption

One of the challenging parts of the subjective assessment in the virtual reality environment is deciding on how to interrupt a participant whilst he/she is using a VR headset. Therefore, we plan to define some communicating codes, to interrupt the participants in case of an emergency before starting the test, such as agreement on calling the participant loudly. In addition, it is strongly recommended not tabbing on the participant's shoulder in order to stop the test. Decreasing the volume of the application remotely could be a good approach.

The way to end the stimulus is an important point that must be taken into consideration since a surprising event may affect the quality of judgment significantly.

6.1.3 Instructions for participants

The experience with HMD might be new to many participants. Therefore, since not everyone has experience with HMD prior to the subjective assessment, a training session including instructions on how to use the OVR application is necessary.

Pre-questionnaires need to include information linked to prior user experience with VR technology, such as experiences with 360° videos. This offers a possibility to control for the impact of subjects' expertise discrepancies on the conclusions drawn. Additionally, it is recommended to analyze the data from experienced and inexperienced participants separately to exclude the impact of the "Novelty" effect on the final scores.

6.1.4 Ethics issue

There are a few ethical issues concerning VR applications which need to be considered carefully before conducting any experiment. The participants should not experience dread, pain, distress, or other emotions associated with a criminal act. In addition, VR applications with possible shocking events such as sudden falls should be avoided as well.

Furthermore, VR experience may cause symptoms similar to that of motion sickness. In the case of any motion sickness symptom such as discomfort, nausea, and headache, the test must be ended immediately.

6.1.5 Assessment Method

In order to assess the OVR Quality of Experience in the subjective experiment, a list of questionnaires is compiled for assessing different dimensions of VR quality. Table 8 gives a list of the well-known questionnaire that will be considered for the assessment of the OVR application. To select a questionnaire, some pilot tests are planned to find the most relevant quality aspect for the OVR application.

Table 8: Comparison of candidate questionnaires attributes for VR experiment.

Questionnaire	Dimensions	Scale Details	Domain
Temple Presence Inventory [9]	Spatial Presence, Social Presence-Actor Within Medium, Passive Social Presence, Active Social Presence, Presence as Engagement, Presence as Social Richness, Presence as Social Realism, Presence as Perceptual Realism	42 items using 7-point Likert scales	Passive Video Consumption
Slater Usoh Steed Presence Questionnaire (SUS) [10] [11]	-	6 long statements rated on a 7-point Likert scale	Simulation of room environments (living room, office, kitchen, bar)
Presence Questionnaire [12]	Main: Control, Sensory, Distraction, Realism; Sub: Involvement, Natural, Auditory, Haptic, Resolution, Interface Quality	32 items using 7-point Likert scales	Virtual Environments for learning and many more
Simulator Sickness Questionnaire (SSQ) [13]	Nausea, Oculomotor and Disorientation	16 items (symptoms) on 3 dimensions	Virtual Reality

Igroup Presence Questionnaire (IPQ)[14]	Spatial Presence Involvement Realness	14 Items	Virtual Environments
MEC Spatial Presence Questionnaire (MEC) [15]	Spatial Presence: Self Location Spatial Presence: Possible Actions Cognitive Involvement	3 Versions (Items Per Subscale) Long- 8, Medium- 6, Short- 4	Cross-Media
ITC Sense of Presence Inventory (ITC-SOPI) [16]	Sense of Physical Space Engagement; Naturalness (Ecological Validity)	17 Items	Cross-Media
Game Engagement Questionnaire (GEngQ) [18]	immersion, presence, flow and psychological absorption	19 items using 7-point Likert scales	Gaming
Immersive Experience Questionnaire (IEQ) [17]	3 person factors (cognitive involvement, real world dissociation and emotional involvement), 2 game factors (challenge and control).	33 items using 7-point Likert scales	Gaming
Player Experience of Need Satisfaction (PENS) [19]	3 STD dimensions: *competence, autonomy, relatedness, and intuitive controls* 3 Presence dimensions: narrative, emotional, physical	21 items (3 items per dimension) using 7-point Likert scales	Gaming, Motivation
Spatial Presence Experience Scale (SPES) [20]	self-location (SL) and possible actions (PA)	8 items (4 per dimension)	Cross-Media
Game Experience Questionnaire (GEQ) [21]	Sensory and Imaginative Immersion, Flow, 5 player experience dimensions	2 or 6 item version available, 5-point Likert scale	Gaming

6.2 ORBK Use Case

An example structure of an interactive study design is shown in Figure 3. The structure and assessment methods are in line with those used for the interactive tests carried out for the data collection for ITU-T Rec G.1072 [6], which are described in detail in ITU-T C.293 [7]. In respect to the assessed quality features, only the spatial video quality items (fragmentation and unclearness) will be removed. In line with the measurements performed for the development of ITU-T Rec. G.1072 [6], the following quality features were assessed: Overall Quality of Experience, Input Quality (Controllability, Responsiveness, Immediate Feedback), Audio Quality, Video Quality, Video Discontinuity, Fairness, Player Experience (Immersion, Competency, Negative Affect, Flow, Tension, Positive Affect, Challenge), Player Performance, as well as an Acceptance rating. Details about the questionnaires, items, and scales are provided in Appendix A.

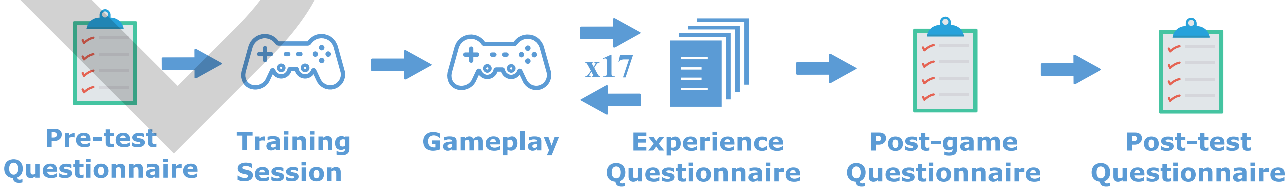


Figure 3: Structure of interactive test assessing gaming QoE (cf. [6]).

6.2.1 Network Parameter emulation

In real networks, especially on cellular networks, which are typically used for mobile online gaming, neither network delays are constant nor are packet losses uniform. Thus, jitter and burst rates should also be considered for the model. However, a full factorial design of multiples levels of all four parameters would lead to such a great number of conditions that carrying out subjective tests for all would not be feasible. Therefore, we used for each delay and packet loss rate, the jitter and burst rate values, respectively, which is present with the highest probability in real networks. By using a crowdsourcing approach of network performances of cellular networks (i.e., LTE), these values were derived and a combination of four delay combined with jitter and four packet loss combined with burst rate are used for the tests. For the distribution of delay, a Pareto distribution is used due to its resemblance with real networks. As an emulator tool, we used NetEm.

The condition plan for the test is designed, as shown in Table 9. However, for one game, we had a few more conditions, as shown in Table 10.

Table 9: Condition plan for all games (except one)

Condition/ Parameters	Delay	Jitter	Packet loss	Burst rate
1	14	3	0	0
2	50	12	0	0
3	100	24	0	0
4	200	47	0	0
5	400	94	0	0
6	1000	235	0	0
7	100	24	20	47
8	14	3	30	50
9	100	24	30	50
10	200	47	30	50
11	14	3	40	53.3
12	100	24	40	53.3
13	200	47	40	53.3
14	14	3	50	57
15	100	24	50	57
16	200	47	50	57

Table 10: Additional condition plan for one tested game

Condition/ Parameters	Delay	Jitter	Packet loss	Burst rate
1	100	0	0	0

2	100	0	20	47
3	100	0	30	50
4	100	0	40	53
5	100	0	50	56
6	100	66	0	0
7	100	66	20	47
8	100	66	30	50
9	100	66	40	53
10	100	66	50	56
11	200	0	0	0
12	200	0	30	50
13	200	0	40	53
14	200	0	50	57
15	200	142	0	0
16	200	142	30	50
17	200	142	40	53
18	200	142	50	57
19	400	0	0	0
20	400	234	0	0

6.2.2 Selected Games

The importance of game characteristics on gaming QoE has been studied and established in work presented in [6]. For the first draft of the model, it is planned to use six games for the training dataset, while some are low delay-sensitive and some are high delay-sensitive, and two games for the test dataset. In addition, the selected will be very well-known state-of-the-art games from shooting, sport, and MOBA genres.

6.3 PLEX Use Case

The PLEX use case follows the ORBK experimental setup with differences in condition plan which requires to be decided in the future after the application is tested under pilot experiments.

7 Data Collection

7.1 OVR Use Case

Due to COVID-19 pandemic, all subjective experiment of OVR application is postponed to the time with lower risk of infection. However, a latency test was conducted to assess the reliability of the service before conducting experiment.

The latency test was comprised of six multiplayer VR sessions involving four users. These sessions were conducted on the same VR application, and on each session, the server's message sends rate varied as well as the host's network speed. In these tests, multiple statistics were recorded form each user:

- Incoming/Outgoing Message rate
- Logic-Level Byte Rate
- Incoming/Outgoing Position rate
- Incoming/Outgoing Rotation rate
- Round Trip Time
- CPU Frame Time

In the recorded data, significant deviations were observed from session to session. We believe this was due to the nature of the VR technology offering the users with multiple options for altering these statistics, e.g., in session A, users moved around, talked with each other, and interacted with the environment significantly more than in session B, which resulted to an increased incoming/outgoing message rate for that session, even when in session A the server's send rate was configured to be lower than in session B.

This observation led us to plan more detailed scenarios, each focusing on specific interactions the users have with the virtual environment. Again, in these tests, the server's message send rate will vary from session to session, as well as the users' internet connection speed. In more detail:

- On the 1st scenario, the users will be tasked to continuously move in the virtual environment for a limited amount of time.
- On the 2nd scenario, users will be tasked to interact with specific objects in the virtual environment.
- On the 3rd scenario, users will be tasked to product VOIP data, by talking with each other and limiting their movement as much as possible.
- On the 4th scenario, users will repeatedly exchange 3D objects with each other in order to observe the overhead of ownership change.

By performing these tests, we hope to conclude on the optimal message send rate as well as specifications on the minimum required network speed. In addition, we believe that these tests can also be used to determine further improvements on the VR networking service.

7.2 ORBK Use Case

7.2.1 Assessed Data

To find games suitable for building the required dataset for the model, we downloaded close to 100 popular games from the PlayStore/AppStore. Games with very low impact of the network (offline games, single-player games, no perceivable impact) as well as games with overly complex game scenarios were removed. In total, we collected data from 7 different games. For six of those games, we tested 16 different conditions of delay and packet loss, whereas we tested one game in more detail using a greater number of jitter values. The selected games cover a large portion of the overall market share and cover different genres, such as MOBA, Action, and Sport.

7.2.2 Impact of game

During the selection process of the games, it became clear that the impact of network delay and packet loss is very different among the different games. While this is partly due to different gameplay rules and mechanics, for online gaming, the implementation of the game in terms of error concealment (i.e., latency compensation and retransmission of lost packets) is also very important. To consider the impact of rules and mechanics for ITU-T Rec. G.1072 [6] classification of game content was proposed. However, it turned out that this classification created for cloud gaming is not suitable for online gaming, as the error concealment methods are very dominant. Games with very similar rules and mechanics can be impacted very differently by network issues depending on the error handling. This is a major issue for the development of the opinion model, as information about the concrete game implementations are not available for network operators. In Figure 4, we visualize the difference between the tested games for the reference condition, a delay of 200 ms, and a packet loss rate of 40 % for overall gaming QoE, Input Quality, and Video Discontinuity. The data shows that the MOS for all three dependent variables varies by more than 1 for the different games. Additionally, it is visible that the error handling of the games outweighs the mechanics, as game 1 and game 4 are very similar in terms of rules, mechanics, and challenges, but there is a significant difference in the ratings between them.

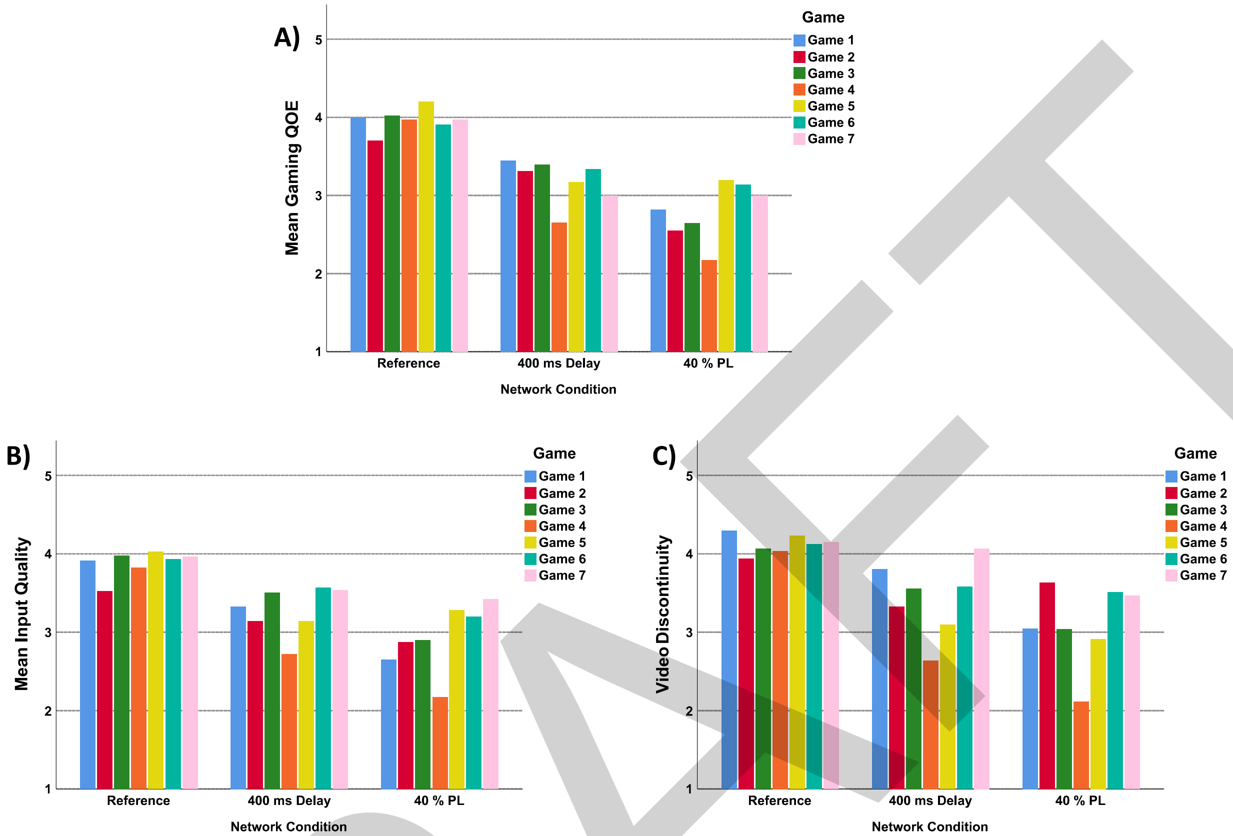


Figure 4: Gaming QoE (A), Input Quality (B), and Video Discontinuity (C) for three example conditions.

7.2.3 Impact of player

Based on feedbacks during the conducted tests it became clear that some participants perceived the impairments due to delay and packet loss while others did not.

Having in mind that the model should aim to predict gaming QoE for at least intermediate players, we defined a variable, participant sensitivity, based on the following rule: if the mean of the four worst conditions (according to a priori known network settings) is not at least 0.5 lower than the mean of the best four conditions, or the reference condition is rated lower than 3, the participant is labelled as low sensitive, else as normal. The labelling was applied to all 215 participants. The rule can also be expressed as shown in Eq. 1.

$$sensitivity = \begin{cases} Mean_{Best\_4} - Mean_{Worst\_4} > 0.5, & normal \\ else & , low \end{cases} \qquad (1)$$

Within the normal sensitivity group, so far, 23 female and 97 male participants took part in the tests. Their age was between 19 and 35 years with a mean of 26 years. 43 participants declared themselves to be highly experienced gamers, 51 judged their gaming expertise to be intermediate, and the remaining 26 were unexperienced players. 59 stated to play more than 10 hours per week and 32 participants reported to frequently used mobile phones for gaming. Therefore, it can be summarized that the recruited participants had enough gaming experience to control the games and to learn them in a reasonable amount of time, but they are no expert players and play more frequently on PC or consoles than on the phone.

Within the low sensitivity group, so far 28 female and 64 male participants took part in the tests. Their age was between 19 and 36 years with a mean of 26. Of the participants, 25 declared themselves to be highly experienced gamers, 49 judged their gaming expertise to be intermediate, and the remaining 18 were rather unexperienced players. 36 stated that they play more than 10 hours per week and 17 participants reported to frequently used mobile phones for gaming. Thus, it can be argued that only based on the self-judgment of expertise and hours played per week, the different sensitivities cannot be explained.

For game one, the ratings of the resulting two user groups are shown in Figure 5. It becomes clear that the low sensitive group could not perceive the network impairment for most of the conditions (apart from condition 16, which was an extremely bad condition), whereas the normal group delivered reasonable and more expected ratings.

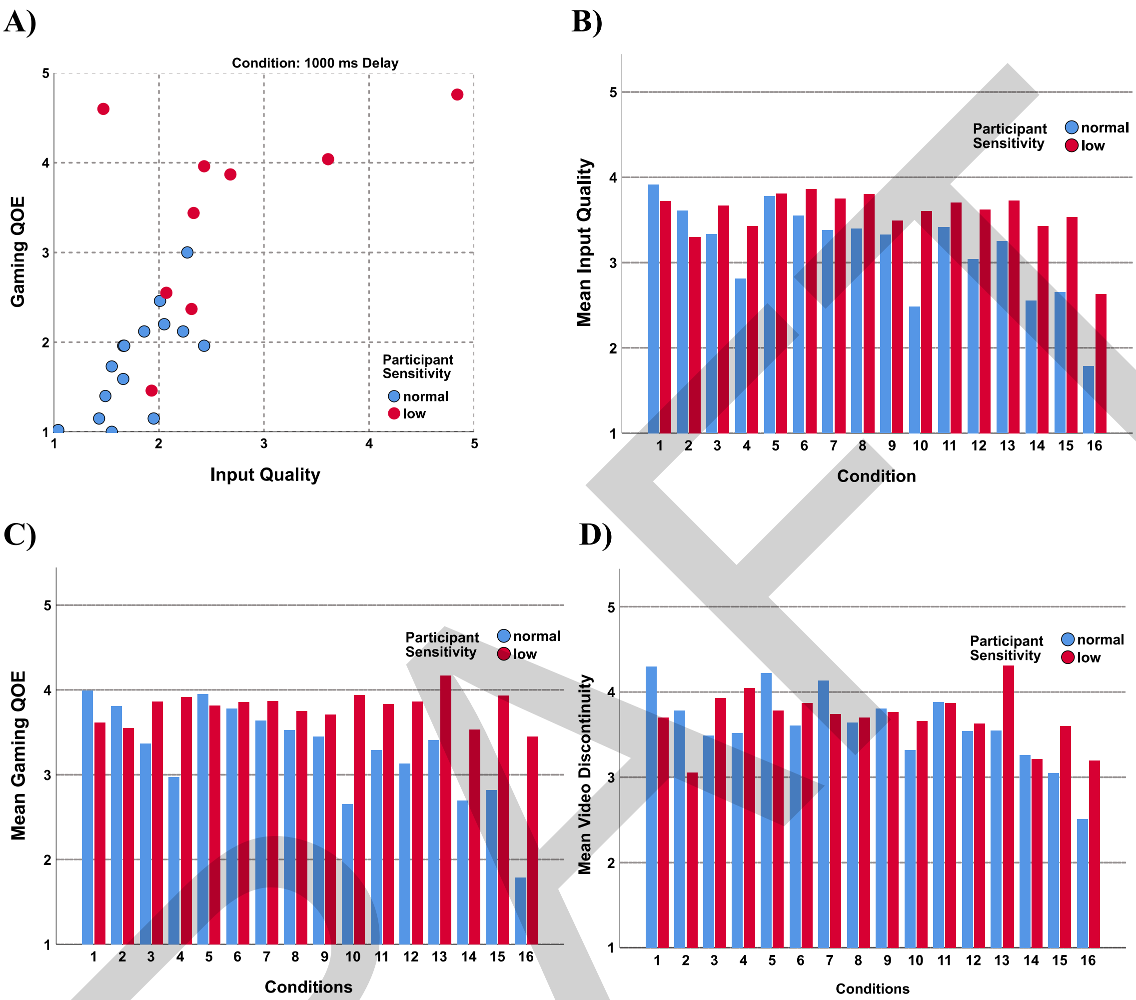


Figure 5: Difference between subjective ratings of sensitive and low sensitive participants for Game 1.

7.2.4 Discussion

For the ORBK use case, a series of test was conducted, including over 200 participants who played a total of 7 games under 16 different network conditions. It must be noted that these games represent the most played games on the market very well, but adding more games is not a trivial task due to the effort and availability of highly interactive multiplayer mobile online games. Furthermore, as shown in the report, we propose to use a player sensitivity screening and cannot use the content classification used in ITU-T Rec. G.1072 [6], as the implementation of the game in terms of error handling, is more dominant than their mechanics regarding the influence of network impairment.

7.3 PLEX Use Case

Due to COVID-19 pandemic, all subjective experiments of PLEX application are postponed to a time with lower risk of infection.

8 Model Structure

8.1 OVR QoE Prediction

The model structure, as illustrated in Figure 6, is composed of two main modules, namely input quality (*IPQ*) and video quality (*VQ*).

The model considers two types of input parameters: network and encoding parameters. For the network parameters, delay and packet loss (*PL*) are used. For the encoding parameters the video resolution (*Res*), encoding framerate (*FR_enc*), and the bitrate (*Br*) used for the video stream are used.

Furthermore, five different estimations of quality impairments expressed on the R-scale, namely *I_VQ_cod*, *I_VQ_trans*, *I_TVQ*, *I_IPQ_frames*, and *I_IPQ_delay*, are calculated based on the previously mentioned input parameters. Their calculation is described in the following subsections.

To predict the overall OVR QoE (*MOS_QoE*), the estimated quality impairments are weighted with the coefficients *a, b, c, d*, and *e*. Next, their sum is subtracted from a reference value, *R_max*, resulting in *R_QoE*. Finally, the predicted *MOS_QoE* is calculated using a conversion to the MOS-scale.

The core model predicting OVR QoE is defined as:

$$\mathrm{R_{QoE}} = \mathrm{R_{max}} - \mathrm{a} \cdot \mathrm{I_{VQ_{trans}}} - \mathrm{b} \cdot \mathrm{I_{VQ_{cod}}} - \mathrm{c} \cdot \mathrm{I_{TVQ}} - \mathrm{d} \cdot \mathrm{I_{IPQ_{frames}}} - \mathrm{e} \cdot \mathrm{I_{IPQ_{delay}}} \quad (2)$$

$$\mathrm{MOS_{QoE}} = \mathrm{MOS\_from\_R}\ (\ \mathrm{R_{QoE}}\ ) \quad (3)$$

where,

R_{QoE} is the overall estimated QoE expressed on the R-scale, where 0 is the worst quality and 100 the best quality,

MOS_{QoE} is the overall estimated QoE expressed on the MOS-scale, where 1 is the worst quality and 5 is the best quality,

R_{max} is the reference value indicating the best possible QoE (= 100) on the R-scale,

$I_{VQ_{trans}}$ is the impairment of spatial video quality caused by video transmission errors on the R-scale,

$I_{VQ_{cod}}$ is the impairment of spatial video quality caused by video compression artefacts on the R-scale,

I_{TVQ} is the impairments on the Temporal Video Quality (discontinuity) caused by compression artefacts on the R-scale,

$I_{IPQ_{frames}}$ is the impairment of Input Quality caused by frame rate reductions,

$I_{IPQ_{delay}}$ is the impairment of the Input Quality caused by network delay and packet loss in combination with delay.

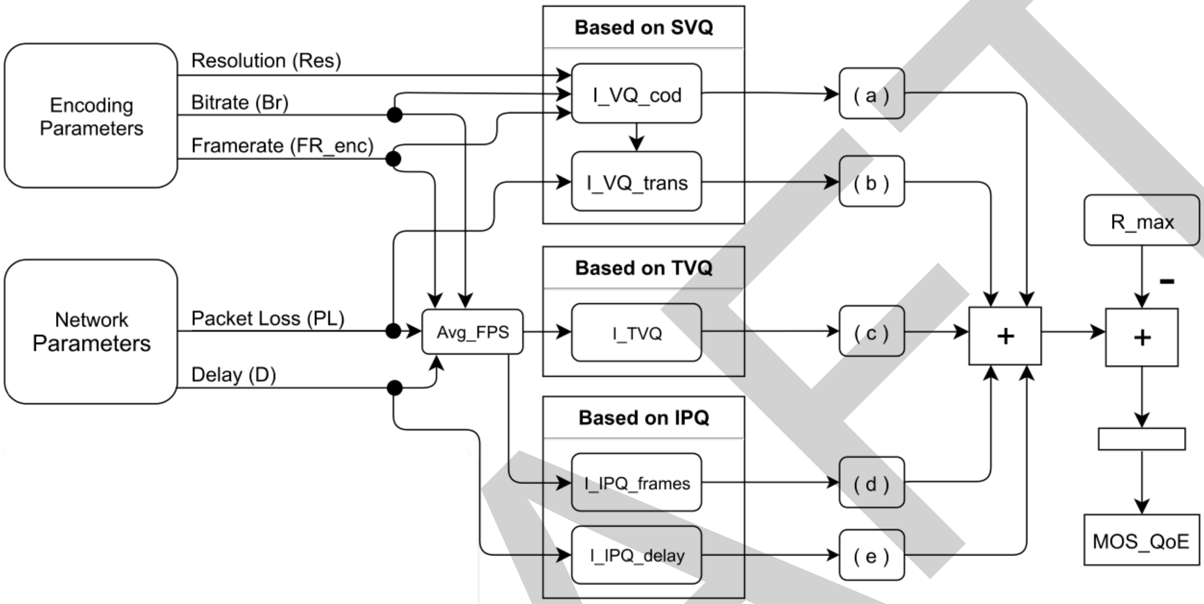


Figure 6: Structure proposed for G.1072. [6].

8.2 ORBK QoE Prediction

Quantitative studies and data gathered from previous studies adhering to ITU-T Rec. P.809 [4] showed a dominant impact of Video Quality (*VQ*) and Input Quality (*IPQ*), often referred to as Playability, on gaming QoE. The structure of the model consists of two modules of Temporal Video Quality (*TVQ*) and Input Quality (*IPQ*). As the model structure illustrated in Figure 7, both modules are influenced by the same network parameters but with different perceptual effects.

The derived impairment factors will be subtracted from a reference point $R_{max,QoE}$ which is based on the highest average ratings for reference condition. The model hereby will have the following structure:

$$R_{QoE} = R_{max,QoE} - a \cdot I_{TVQ} - b \cdot I_{IPQ} \qquad (4)$$

where

$R_{max,QoE}$ = corresponds to the R value that when transformed gives the highest average conditions for the reference condition,

I_{TVQ} is the impairments on the Temporal Video Quality (discontinuity) caused by packet loss in combination with delay

$I_{IPQ_{delay}}$ is the impairment of the Input Quality caused by network delay and packet loss in combination with delay.

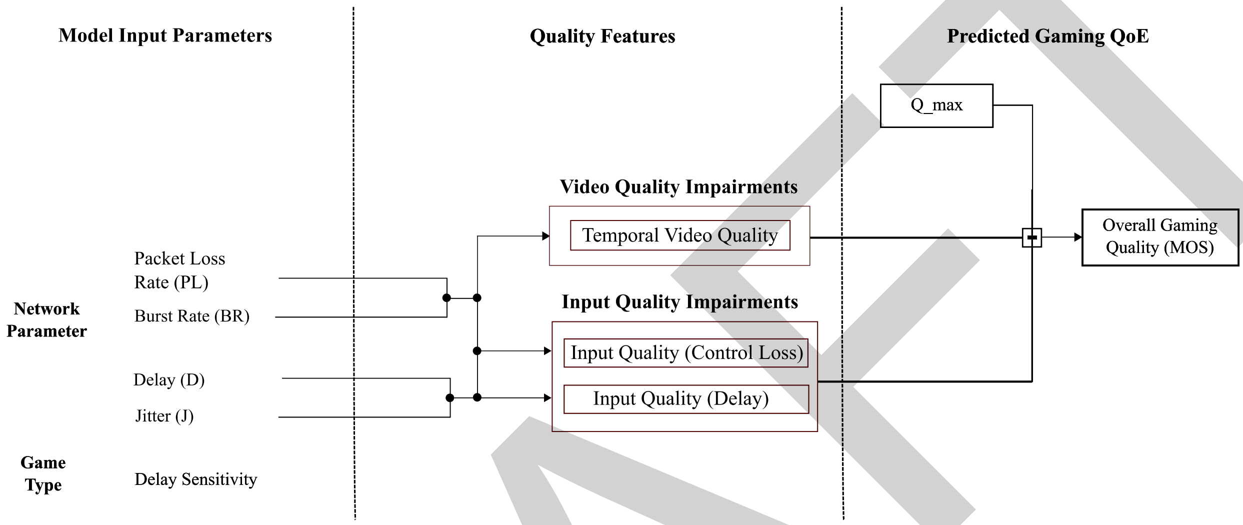


Figure 7: Proposed model structure of the mobile online (AR) gaming model.

8.3 PLEX QoE Prediction

For the PLEX QoE prediction model, it is decided to follow the ORBK model framework. It is expected to see differences in the range of parameters to be selected for the subjective test. The two use cases will still follow a very similar model structure due to the similarity of the use case.

9 Conclusion and Future Works

This report gives a summary of the work that has been done within the ACCORDION project to draw the requirements that must be considered for conducting subjective tests. This requirement includes the necessary steps before conducting any experiment and information about the data collection process as well as QoE model structure.

Due to the COVID-19 pandemic, the subjective experiments are strongly affected. This affects the subjective tests of two use cases of PLEX and OVR. For ORBK, the tests are conducted for seven games, which might not be enough for the development of a reliable quality model. Thus, it is required to conduct more tests for the ORBK use case to develop a reliable model in WP5.

In the next phase, it is planned to continue the planning and conduct subjective tests for all three use cases. The developed models in WP5 will be done in multiple iterations; each iteration, the performance of the model will be evaluated under the ACCORDION platform, and the requirement for QoE models will be updated accordingly.

Appendix A. Questionnaire Information

A.1 Pre-test Questionnaire for ORBK use case

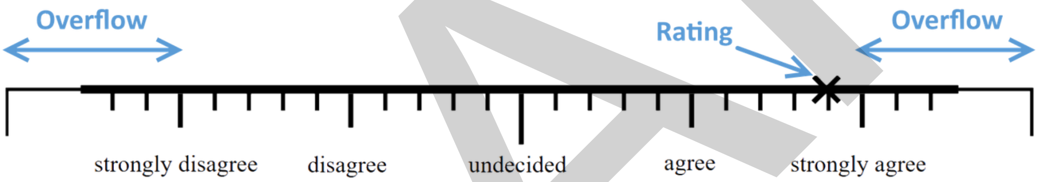
Pre-test Questionnaire:

1. What is your Year of Birth?
2. What is your gender?
    a. Female
    b. Male
    c. Transgender
    d. Prefer not to say
3. Roughly how many hours per week do you spend on playing video games?
    a. Between 0 to 1 hours
    b. Between 1 to 5 hours
    c. Between 5 to 10 hours
    d. Between 10 to 20 hours
    e. More than 20 hours
4. Roughly how often do you play video games in a week?
    a. Never
    b. Between 1 to 3 times a week
    c. Between 3 to 7 times a week
    d. Between 7 to 14 times a week
    e. More than 14 times week
5. How would you describe your gaming experience (expertise)?
    a. 1 – Beginner
    b. 2
    c. 3 – Intermediate
    d. 4
    e. 5 – Expert
6. I like playing video games.
    a. 1 - Strongly Disagree
    b. 2 – Disagree
    c. 3 – Undecided
    d. 4 – Agree
    e. 5 - Strongly Agree
7. On which kind of device do you usually play games?
    a. PC (Desktop)
    b. Smartphone / Tablet
    c. Console (PlayStation, Xbox, ...)
    d. Others
8. What kind of monitor are you typically using when playing?
    a. Television (> 30”)
    b. Desktop Monitor (> 20”)
    c. Laptop (> 12”)
    d. Tablet (> 8”)
    e. Large Smartphone (> 5”)

   f. Small Smartphone (< 5”)
   g. Other
9. How experienced are you in playing the game "*game name*"?
   a. 1 – Unexperienced
   b. 2
   c. 3 – Intermediate
   d. 4
   e. 5 - Expert

A.2 Post-game Questionnaire:

The post-game questionnaire covers the following aspects: Performance Indication (PI), Learnability (LE), Appeal (AP), and Intuitive Controls (IC). Component scores are computed as the average value of its items. The following 7-point continuous scale was used:

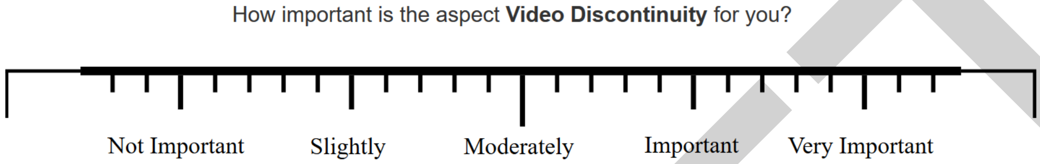


1. I could easily assess how I was performing in the game. - PI1
2. Learning to operate the game is easy for me. - LE1
3. I liked the graphics and images used in the game. - AP1
4. Learning the game controls was easy. - IC1
5. It was clear to me how my performance was going. - PI2
6. It is easy for me to become skillful at using the game. - LE2
7. The game appealed to my visual senses. - AP2
8. The game controls are intuitive. - IC2
9. I was informed about my progress in the game. - PI3
10. I find the game easy to use. - LE3
11. The game was aesthetically appealing. - AP3
12. It was easy to remember the corresponding control. - IC3

Post-test Questionnaire:

For the post-test questionnaire, the following instructions are given to participants:

"In the following, we would like you to tell us about your judgement criteria. Please indicate on the scales below, how important in general the listed aspects are for your rating of the overall quality of your gaming experience."



The bold written aspect (in the example Video Discontinuity) was exchanged with: Video Quality, Audio Quality, Controllability, Responsiveness, Immediate Feedback, Video Fragmentation, Video Unclearness, Video Discontinuity, Suboptimal Video Luminosity, and Playing Performance. The order of items was randomized and an open question about potential other aspects was added.

A.3 Post-condition Questionnaire

Participants are asked to indicate how they felt while using the service for each of the following items by clicking on the 7-point scale below as explained in the introduction. The questionnaire covers the following aspects (component scores are computed as the average value of its items):

Overall QoE: QOE

Input Quality: CN: Controllability, RE: Responsiveness, IF: Immediate Feedback

Output Quality: AQ: Audio Quality, VQ: Video Quality, VF: Video Fragmentation,

VU: Video Unclearness, VD: Video Discontinuity, VL: Suboptimal Video Luminosity

Player Experience (iGEQ): IM: Immersion, CO: Competency, NE: Negative Affect, FL: Flow,

TE: Tension, PO: Positive Affect, CH: Challenge,

Playing Performance: PR, **Service Acceptance**: AC

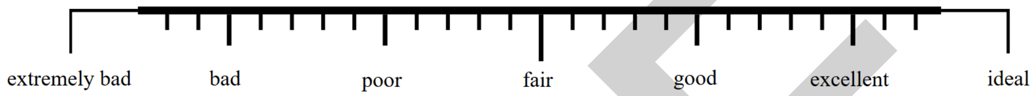
Item	Scale	Question
QOE	1	How do you rate the overall quality of your gaming experience?
CN1	2	I felt that I had control over my interaction with the system.
RE1	2	I noticed a delay between my actions and the outcomes.
CN2	2	I felt a sense of control over the game* interface and input devices.
RE2	2	The responsiveness of my inputs was as I expected.
CN3	2	I felt in control of my game actions*.
IF1	2	I received immediate feedback on my actions.
RE3	2	My inputs were applied smoothly.
IF2	2	I was notified about my actions immediately.
AQ	2	How do you rate the overall audio quality?
VQ	2	How do you rate the overall video quality?
VF	3a	Fragmentation
VU	3b	Unclearness
VD	3c	Discontinuity
VL	3d	Suboptimal Luminosity
IM1	2	I found it impressive.
CO1	2	I felt successful.
NE1	2	I felt bored.
IM2	2	It felt like a rich experience.
FL1	2	I forgot everything around me.
TE1	2	I felt frustrated.
NE2	2	I found it tiresome.
TE2	2	I felt irritable.
CO2	2	I felt skillful.
FL2	2	I felt completely absorbed.
PO1	2	I felt content.

CH1	2	I felt challenged.
CH2	2	I had to put a lot of effort into it.
PO2	2	I felt good.
PR	1	How do you rate your own playing performance?
AC	4	Would you accept using a service under these conditions?

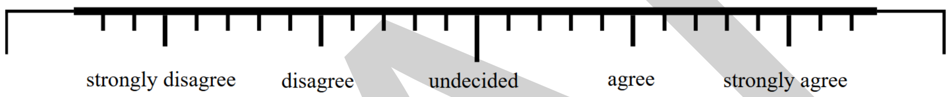
** for non-gaming domain the word "game" was replaced by "application".*

The following scales are used for the post-condition questionnaire:

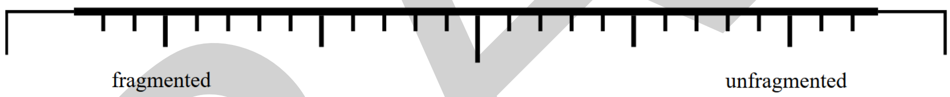
Scale 1:



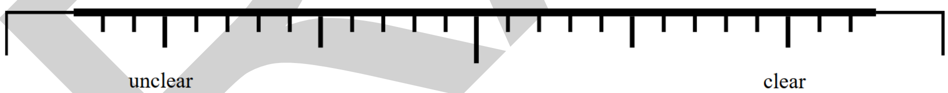
Scale 2:



Scale 3a:



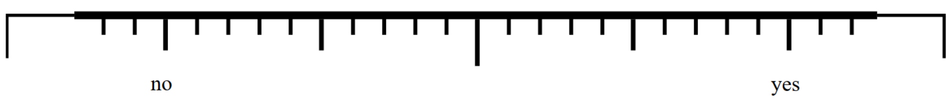
Scale 3b:



Scale 3c:

Scale 3d:

Scale 4:



Bibliography

[1] Recommendation ITU-T P.800.1 (2016), Mean Opinion Score (MOS) terminology, Geneva, Switzerland: International Telecommunication Union.

[2] Recommendation ITU-T P.910 (2008), Subjective video quality assessment methods for multimedia applications, Geneva, Switzerland: International Telecommunication Union.

[3] Recommendation ITU-T G.1032 (2017), Influence factors on gaming quality of experience, Geneva, Switzerland: International Telecommunication Union.

[4] Recommendation ITU-T P.809 (2018), Subjective evaluation methods for gaming quality, Geneva, Switzerland: International Telecommunication Union.

[5] Beyer, J., & Möller, S. (2014). Assessing the Impact of Game Type, Display Size and Network Delay on Mobile Gaming QoE. *PIK-Praxis der Informationsverarbeitung und Kommunikation*, *37*(4), 287-295.

[6] Recommendation ITU-T G.1072 (2020), Opinion Model Predicting Gaming QoE for Cloud Gaming Services, Geneva, Switzerland: International Telecommunication Union.

[7] ITU-T Contribution C.293, "Data Assessment for an Opinion Model Predicting Gaming QoE" Geneva, Switzerland: International Telecommunication Union, 2019.

[8] Recommendation ITU-T G.1035 (2020), Influencing factors on quality of experience for virtual reality services, Geneva, Switzerland: International Telecommunication Union.

[9] Lombard, M., Weinstein, L., & Ditton, T. B. (2011). MEASURING TELEPRESENCE: THE VALIDITY OF THE TEMPLE PRESENCE INVENTORY (TPI) IN A GAMING CONTEXT. Presented at the 2011 annual conference of the International Society for Presence Research (ISPR), Edinburgh, Scotland. [Available here (.pdf)]

[10] Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. Presence: Teleoperators and Virtual Environments, 3, 130-144.

[11] Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. Presence: Teleoperators & Virtual Environments, 9(5), 497-503.

[12] Witmer, B., & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225-240.

[13] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lillienthal, M. G. (1993). A simulator Sickness questionnaire (SSQ): A new method for quantifying simulator sickness. International Journal of Aviation Psychology. 3(3), 203–220.

[14] Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators and Virtual Environments, 10, 266-281.

[15] Vorderer, P et al., Development of the MEC Spatial Presence Questionnaire (MEC SPQ). Unpublished report to the European Community, Project Presence: MEC (IST-2001-37661).

[16] Lessiter, J., Freeman, J., Keogh, E. & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. Presence: Teleoperators and Virtual Environments, 10(3), 282-297.

[17] C. Jennett et al., “Measuring and defining the experience of immersion in games,” Int. J. Hum.-Comput. Stud., vol. 66, no. 9, pp. 641–661, 2008.

[18] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, “The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing,” J. Exp. Soc. Psychol., vol. 45, no. 4, pp. 624–634, 2009.

[19] Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. Motivation and emotion, 30(4), 344-360.

[20] Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., ... & Gouveia, F. (2015). The spatial presence experience scale (SPES). Journal of Media Psychology.

[21] Poels, K., de Kort, Y. A. W., & IJsselsteijn, W. A. (2007). D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games. Eindhoven: Technische Universiteit Eindhoven.