International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1320

(07/2022)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Telemeeting assessment

Quality of experience assessment of extended reality meetings

Recommendation ITU-T P.1320

ITU-T

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	P.1100–P.1199
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	**P.1300–P.1399**
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	P.1500–P.1599

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.1320

Quality of experience assessment of extended reality meetings

Summary

Recommendation ITU-T P.1320 advises on aspects of importance for quality of experience (QoE) assessment of telemeetings with extended reality elements. The goal is to define the human, context and system factors that affect the choice of the QoE assessment procedure and metrics when extended reality telemeeting systems are under evaluation.

History

Edition	Recommendation	Approval	Study Group	Unique ID[*]
1.0	ITU-T P.1320	2022-07-29	12	11.1002/1000/15011

Keywords

Extended reality, quality of experience, QoE, telemeeting.

---

[*] To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Page

1 Scope 1

2 References..... 1

3 Definitions 2

 3.1 Terms defined elsewhere..... 2

 3.2 Terms defined in this Recommendation..... 2

4 Abbreviations and acronyms 2

5 Conventions..... 3

6 Overview of concepts and technologies used for XR telemeetings 3

 6.1 XR telemeeting..... 3

 6.2 Virtual reality (VR) 4

 6.3 Augmented reality 4

 6.4 Assisted reality 4

 6.5 Mixed reality (MR)..... 4

 6.6 Mediated reality..... 4

 6.7 Metaverse 4

 6.8 QoE constituent 4

 6.9 Quality features 5

7 Use cases..... 6

 7.1 Use cases in 3GPP 6

 7.2 Spatial audio multiparty call..... 8

 7.3 Remote system operation over XR telemeeting..... 8

 7.4 XR conference..... 8

8 Factors influencing QoE of XR telemeetings..... 8

 8.1 Human influence factors (HIFs)..... 8

 8.2 Context influence factors (CIFs) 10

 8.3 System influence factors (SIFs)..... 11

9 Constituents of QoE in XR telemeetings..... 20

 9.1 Simulator sickness..... 20

 9.2 Immersion..... 20

 9.3 Presence, co-presence and social presence..... 20

 9.4 Plausibility..... 21

 9.5 Ethics of XR use..... 22

 9.6 Carving out mental space 22

 9.7 Fatigue and cognitive load..... 22

 9.8 Ability to reach goals..... 22

10 Test methods targeting XR telemeeting QoE..... 22

 10.1 Test paradigms..... 23

Page

Appendix I – Related work in other standardization bodies and industry 25
- I.1 3rd Generation Partnership Project (3GPP) 25
- I.2 Virtual Reality Industry Forum (VRIF) 25
- I.3 Moving Picture Experts Group (MPEG) 25
- I.4 Khronos Group 26
- I.5 Video Quality Experts Group (VQEG) 26
- I.6 European Network on Quality of Experience in Multimedia Systems and Services (Qualinet) 26
- I.7 World Wide Web Consortium (W3C) 26
- I.8 ITU-T SG16/Q8 Immersive Live Experiences 26

Bibliography 27

Recommendation ITU-T P.1320

Quality of experience assessment of extended reality meetings

1 Scope

This Recommendation focuses on aspects of importance for the assessment of quality of experience (QoE) of different types of extended reality (XR) telemeetings, which may be a combination of telemeetings taking place in virtual reality (VR), augmented reality or mixed reality environments. It targets XR services aiming to immerse the user and augment the exchange of information by delivering interactive real-time uni- or multimodal sensory information for two-party and multiparty communication. The services include telemeetings taking place in a virtual location or a combination of virtual and real locations; telemeetings with mixtures of real and virtual participants; telemeetings with augmented elements for collaboration; virtual conferences; and joint teleoperation of equipment.

While multiple QoE evaluation methodologies for telemeetings have been developed, novel XR telemeeting services may result in cognitive effects that are not covered by the existing Recommendations. These effects may include simulator sickness, fatigue, immersion or presence, for example. This Recommendation advises on the key QoE factors affecting user experience of an XR telemeeting service.

The purpose of this Recommendation is to find existing standardized methods that can be applied for assessing the QoE constituents of XR telemeetings, and direct the reader to the appropriate source. When a suitable standardized method is not available, this Recommendation will list the QoE influencing factors related to the QoE constituent in question and guide the reader to relevant scientific literature. Ultimately, this Recommendation specifies the categorization of system, human and context factors influencing XR telemeeting QoE.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.1035]	Recommendation ITU-T G.1035 (2021), *Influencing factors on quality of experience for virtual reality (VR) services*.
[ITU-T P.10]	Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience*.
[ITU-T P.919]	Recommendation ITU-T P.919 (2020), *Subjective test methodologies for 360° video on head-mounted displays*.
[ITU-T P.1301]	Recommendation ITU-T P.1301 (2017), *Subjective quality evaluation of audio and audiovisual multiparty telemeetings*.
[ITU-T P.1305]	Recommendation ITU-T P.1305 (2016), *Effects of delays on telemeeting quality*.
[ITU-T P.1310]	Recommendation ITU-T P.1310 (2017), *Spatial audio meetings quality evaluation*.

[ITU-T P.1311] Recommendation ITU-T P.1311 (2014), *Method for determining the intelligibility of multiple concurrent talkers*.

[ITU-T P.1312] Recommendation ITU-T P.1312 (2016), *Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 QoE [ITU-T P.10]

3.1.2 QoE influencing factors [ITU-T P.10]

3.1.3 telemeeting [ITU-T P.1301]

3.1.4 degree of freedom (DoF) [ITU-T G.1035]

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 virtual reality (VR): An environment that is fully generated by digital means. To qualify as virtual reality, the virtual environment should differ from the local environment.

3.2.2 augmented reality (AR): An environment containing both real and virtual sensory components. The augmented reality continuum runs from virtual content that is clearly overlaid on a real environment (assisted reality) to virtual content that is seamlessly integrated and interacts with a real environment (mixed reality).

NOTE – Based on [b-Rauschnabel].

3.2.3 assisted reality: An environment containing both real and virtual sensory components, where the virtual content is perceived as clearly artificial and overlaid (one end of the augmented reality continuum).

NOTE – Based on [b-Rauschnabel].

3.2.4 mixed reality (MR): An environment containing both real and virtual components that are seamlessly integrated and interact with each other in a natural way (one end of the augmented reality continuum).

NOTE – Based on [b-Rauschnabel].

3.2.5 diminished reality (DR): An environment with deliberately removed contents of the digital representation of the physical environment.

3.2.6 extended reality (XR): An environment containing real or virtual components or a combination thereof, where the variable X serves as a placeholder for any form of new environment (e.g., augmented, assisted, mixed, virtual or diminished reality).

NOTE – Based on [b-Rauschnabel].

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AR	Augmented Reality
CAVE	Cave Automatic Virtual Environment
CfE	Call for Evidence

CfP	Call for Proposal
CIF	Context Influence Factor
CPU	Central Processing Unit
DR	Diminished Reality
DoF	Degree of Freedom
GoP	Group of Pictures
GPU	Graphics Processing Unit
GSR	Galvanic Skin Response
HMD	Head-Mounted Display
HRTF	Head-Related Transfer Functions
MR	Mixed Reality
MTP	Motion-To-Photon
MTS	Motion-To-Sound
NBMP	Network-Based Media Processing
OMAF	Omnidirectional Media Format
PCC	Point Cloud Compression
QIF	Quality Influence Factor
SIF	System Influence Factor
URLLC	Ultra-Reliable Low Latency Communications
VC	VideoConferencing
V-PCC	Video-based Point Cloud Compression
VR	Virtual Reality
VVC	Versatile Video Coding
XR	Extended Reality

5 Conventions

None.

6 Overview of concepts and technologies used for XR telemeetings

This clause introduces concepts and technologies associated with extended reality (XR) telemeetings. A characterization of all types of telemeetings, including XR telemeetings can be found in [b-Skowronek].

6.1 XR telemeeting

The XR telemeeting technology aims to substantially improve the telemeeting experience, for example, through spatial reproduction of audio and video, and to enhance the exchange of information with augmented elements, for example, by using virtual avatars or virtual objects and by better immersing the participants and increasing their sense of (co-)presence. A telemeeting can take place in a virtual location or a combination of virtual and real locations; it can have a mixture of real and virtual participants; or augmented elements for collaboration.

6.2 Virtual reality (VR)

Virtual reality immerses users in a fully artificial (computer-generated) digital environment. The term virtual reality refers to a computer-generated environment in which the user can perceive, feel and interact in a manner that is similar to a physical place. This is achieved by combining stimulation over multiple sensory channels – such as sight, sound and touch – with force-feedback, motion tracking and control devices. In an ideal virtual reality (VR) system, the user would not be able to distinguish an artificial environment from its physical counterpart. While none of the current VR systems would be able to pass this criterion, significant advances in the perceptual fidelity of virtual environments have been achieved over the last few years. The virtual environment itself is a computer-generated three-dimensional model of a physical environment, in which the user can navigate and interact with the objects that it contains.

6.3 Augmented reality

Augmented reality overlays digital objects on the real-world environment with spatial registration that enables geometric persistence concerning placement and orientation within the real world. Augmented reality superimposes a layer of digital content over the real physical world such that users see virtual models at fixed positions and orientations mixed in with the actual world around them.

6.4 Assisted reality

Assisted reality overlays digital information or objects on the real-world environment with the aim to assist the user in obtaining a better understanding of the physical environment rather than to merge virtual objects with the real world [b-Rauschnabel].

6.5 Mixed reality (MR)

Mixed reality is the merging of real and virtual worlds to produce hybrid environments, where physical and digital objects coexist and interact in real-time. Hence, mixed reality (MR) is similar to AR, but the digital objects spatially interact with the real-world objects, rather than being superimposed as 'floating images' on top of the real-world objects.

6.6 Mediated reality

Sometimes the terms mediated reality or computer-mediated reality are used to describe VR, AR, MR and diminished reality (DR) with one encompassing term. In that respect, this Recommendation considers mediated reality and computer-mediated reality as synonyms for XR.

6.7 Metaverse

The metaverse is a digital reality that combines aspects of social media, online gaming, AR and VR to allow users to interact virtually. Users in the metaverse will be able to immerse themselves in a space where the digital and physical worlds converge. An important characteristic of the metaverse is the persistence of its content, meaning that virtual or augmented content is spatially attached to specific physical objects (e.g., a digital vase on a user's physical desk), or attached to a specific geographic location through geocoordinates in order to provide different users a shared experience [b-Rauschnabel].

6.8 QoE constituent

A QoE constituent is formed during an individual's experience of a service and contributes to its QoE. It is considered to be formed by a cognitive process which aggregates a multitude of quality features (see definition of quality features in clause 6.8). This aggregation, however, needs neither to form orthogonal dimensions, as is conventionally the goal of multidimensional quality assessment methods, nor to result in QoE constituents that are independent from each other.

Moreover, a QoE constituent can encompass aspects that are linked to the perception of speech, audio or video signals as well as other aspects such as, for instance, simulator sickness, immersion, feeling of presence, feeling of co-presence or usability. That means the quality features used can encompass perceptual representations of the audiovisual signals, body reactions (e.g., in case of simulator sickness) and other sensory input, contextual information or other characteristics that refer, for instance, to the direct perception, action, interaction or usage instance of the service [b-Möller].

Furthermore, QoE constituents can be differentiated from quality influence factors (QIFs [ITU-T P.10]) by using a process model perspective: QoE constituents are aspects formed inside the QoE formation process in the experiencing person's mind, which is happening during the experience with the service. QIFs can be aspects that influence the QoE formation process from outside the process, and can stem from inside or outside the experiencing person. Alternatively, QIFs can influence the characteristics of the XR meeting and with this all relevant signals and contextual information that form the input to the QoE formation process.

Figure 1 uses this process model perspective to visualize the relationships between the different aspects of QoE, QoE constituents, quality features and QIFs.

6.9 Quality features

A quality feature is a perceivable, recognized and nameable characteristic of an individual's experience of a service which contributes to its quality [b-Brunnström2013], [b-Jekosch].

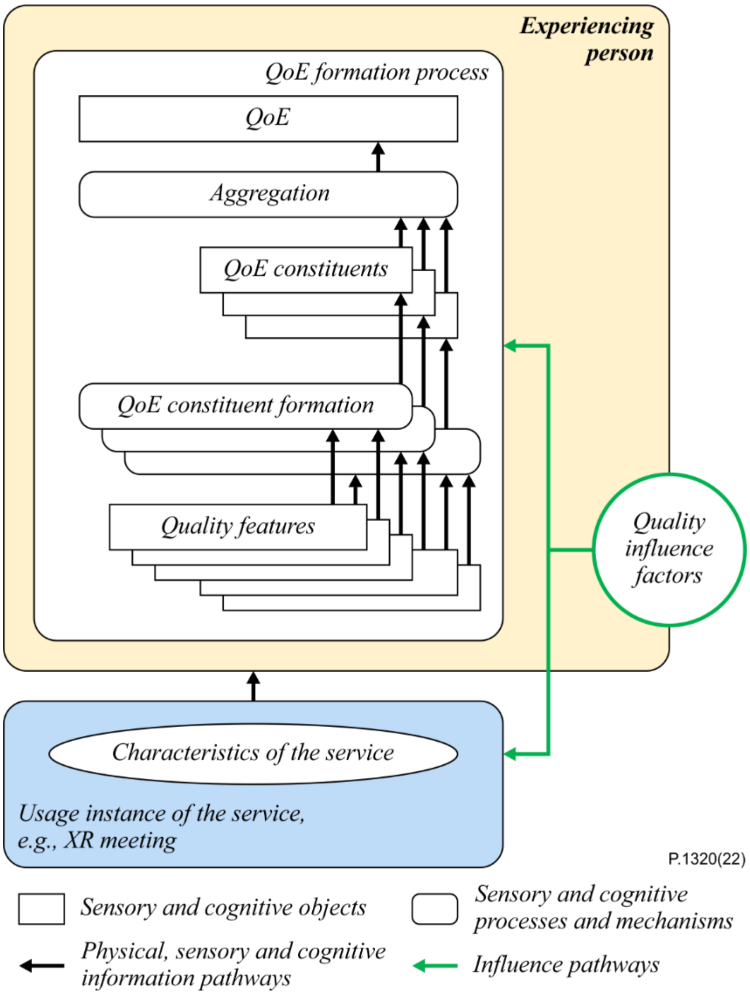

Figure 1 – Visualization of the relationships between QoE, QoE constituents, quality features and QIFs

7 Use cases

7.1 Use cases in 3GPP

The 3rd Generation Partnership Project (3GPP) has published a technical report [b-TR 26.928] on XR technology in 5G. In that report, seven core use cases with a total of 23 individual use cases are specified. A number of those use cases address content delivery for conversational situations and can thus be considered use cases for XR telemeetings. Table 1 summarizes those use cases; for a full description, refer to chapters 4 and 5 of [b-TR 26.928] as well as its appendices.

Some new use cases including AR remote cooperation, AR Conferencing and shared AR conferencing experience can be found in [b-TR 26.998].

Table 1 – Overview of use cases from 3GPP's technical report [b-TR 26.928] that can be considered XR telemeetings

Core use case	Individual use case (parentheses: use case number according to TR 26.928)	Aspect of use case relevant for present Recommendation
Real-time XR sharing	(7) Real-time 3D communication	Video telephony call with partially captured (e.g., face) and partially reconstructed (e.g., back of head) 3D representation of conversation partner
Real-time XR sharing	(8) AR guided assistant at remote location (industrial services)	Remote expert helps local technician to work on a physical object/machine, both use audio video communication and AR overlays
Real-time XR sharing	(11) Real-time communication with a shop assistant	Client and shop assistant use audio video communication and AR overlays to place 3D representations of items for sale in the client's environment
Real-time XR sharing	(17) AR animated avatar calls	Similar to use case "Real-time 3D Communication": telephony call in which one conversation partner does not transmit video data, hence that partner is represented by a full 3D reconstruction
Real-time XR sharing	(23) 5G shared spatial data	Co-located people wearing an XR head-mounted display (HMD) collaboratively interact with a detailed 3D virtual model from their own perspective into a shared coordinate system (using a shared map)
XR mission critical	(9) police critical mission with AR	Police officers on site and in a police station use audio video communication and additional data using AR overlays and control of drones is also considered
XR conference	(12) 360-degree conference meeting	Multiparty telemeeting in which all conversation partners communicate in a virtual 3D representation of a real meeting room; in a variation of this use case mediated representations of the remote participants are projected into a real meeting room by means of AR for those who are actually in that room

Table 1 – Overview of use cases from 3GPP's technical report [b-TR 26.928] that can be considered XR telemeetings

Core use case	**Individual use case (parentheses: use case number according to TR 26.928)**	**Aspect of use case relevant for present Recommendation**
XR conference	(13) 3D shared experience	Conversation partners collaborate in a virtual environment to perform a task
XR conference	(14) 6DOF VR conferencing	Virtual world in which conversation partners are represented by avatars, avatars can move in the virtual world and can communicate with others when they are near enough
XR conference	(15) XR meeting	Remote participants experience a virtual representation of a real environment and can communicate with local participants who are physically present in that environment, local participants can communicate with remote participants using AR
XR conference	(16) Convention / poster Session	Similar to XR Meeting: local and remote participants join a physical poster session, presenters and listeners may be physically present or participate remotely
Spatial audio multiparty call	(18) AR avatar multiparty calls	Multiparty call in which remote participants are placed in each local environment using AR, whereas spatial audio is used (a) for rendering each conversation partner in the direction of his/her geolocation and (b) for sharing individual acoustic scenes when desired (e.g., "hear what I hear" feature)
Spatial audio multiparty call	(19) Front-facing camera video multiparty calls	Multiparty video telephony in which spatial audio is used to keep the acoustic location of the conversation partners at the same spot relative to the video display (e.g., mobile phone)
XR multimedia Streaming	(20) AR streaming with localization registry	While this is mainly a streaming-only scenario based on the example of an AR museum guide, a variation allows for a communication with a travel guide avatar
XR multimedia streaming	(21) Immersive six degrees of freedom (6DoF) streaming with social interaction	An extension of a streaming-only scenario ("immersive game spectator mode") based on the example of watching in real-time a game (or other event) as a remote spectator: in this extension, however, the spectator can communicate with the avatars of other spectators

Related work on XR meetings has been conducted by ITU. For that reason, additional information concerning some use cases is provided in the following to clarify the relation of XR telemeetings to telemeetings considered in existing Recommendations.

7.2 Spatial audio multiparty call

In [b-TR 26.928], two use cases (Nos 18 and 19) are grouped into a core use case "spatial audio multiparty call". The detailed descriptions in [b-TR 26.928] actually reveal two aspects that qualify this core use case as an XR meeting.

First, according to the definition in clause 3.2.1, using spatial audio technology for a telemeeting already qualifies as an XR meeting. In terms of quality assessment, it should be noted that a number of assessment methods for such "spatial audio meetings" are described in [ITU-T P.1310], [ITU-T P.1311] and [ITU-T P.1312]. Also, the combination of spatial audio with video (over ordinary displays) has been addressed in some of these methods. That means the core use case "spatial audio multiparty call" comprises additional individual use cases that are addressed in the aforementioned Recommendations.

Second, the combination of spatial audio with other XR elements such as a visual representation by means of VR or AR also qualifies as an XR meeting. In terms of quality assessment, however, this is a new use case that has not been covered in the aforementioned Recommendations and thus requires further study.

7.3 Remote system operation over XR telemeeting

This is in general a new field, a combination of telemeeting technology with physical objects and machines which leads to a cyberphysical system remotely controlled by multiple persons.

7.4 XR conference

A virtual conference experience typically happens with multiple participants. As an example, almost 2000 participants participated in IEEE VR 2020 [b-Ahn]). The experience may involve different types of events, which are sub-parts of the XR conference with different requirements, such as plenary talks, parallel sessions, workshops, poster sessions or demonstrations [b-ACM].

Hybrid conferences may be held with virtual participants represented by remotely operated telepresence robots [b-Neustaedter].

8 Factors influencing QoE of XR telemeetings

This clause identifies the XR telemeeting specific influencing factors affecting the QoE constituents listed in clause 9. They are divided into three categories: human (clause 8.1), context (clause 8.2) and system (clause 8.3) influencing factors. These factors can also be mixed [b-Reiter], [b-Skowronek-3]. Other relevant XR-related QoE influencing factors collected in [ITU-T G.1035] are mentioned here, but not explained in detail.

8.1 Human influence factors (HIFs)

Human influence factors (HIFs) refer to any characteristics of a user that have an influence on QoE, including the background and the mental, psychophysiological and physiological state of a user [b-Brunnström, [b-Reiter]. Next to the fidelity of the representation of a mediated environment and the persons therein, the experienced quality of an XR meeting experience may also depend on highly subjective secondary factors such as personal relevance [b-Toet-2] and the user's context (e.g., task, available information: [b-Hägni], [b-Lee-2]), current (mental and physical) state, personality [b-Alsina], [b-Hofer], [b-Sacau], engagement and involvement (e.g., enjoyment, flow, and mental absorption or attention [b-Lee-2]).

8.1.1 Characteristics of the perceptual and cognitive processes

A user's visual, auditory, olfactory and tactile acuity can all affect the perceived quality of an XR telemeeting [b-Skowronek-3], [ITU-T G.1035]. For example, impaired visual or hearing acuity will influence the perception of any degradations in the audio and video signals.

QoE is a complex cognitive construct, resulting from the technical aspects of a telemeeting system. During an XR telemeeting, users can be confronted with different technical possibilities to communicate remotely, resulting in high cognitive load and fatigue. The user's cognitive process (i.e., the way concepts and information are formulated and processed) will determine the level of cognitive load and thus ultimately the quality of the experience.

8.1.2 Internal state of individual participants

One aspect of the internal state of individual participants is their tendency to experience simulator sickness [ITU-T G.1035]. If a participant is experiencing simulator sickness during an XR meeting, then the quality of the overall experience is correspondingly degraded [b-Toet-1].

Another such aspect is their tendency to experience immersion [ITU-T G.1035]. Immersion is the extent to which a system can replace the user's natural sensory input with mediated input, while presence refers to the user's interpretation of immersion. Since it appears likely that presence requires a tendency for immersion, Witmer and Singer [b-Witmer-1] developed the Immersive Tendency Questionnaire (ITQ) to measure the propensity for immersion. However, no correlation has been found between immersion and ITQ ratings [b-Agrawal], and only a partial negative correlation between the ITQ and presence measured with the Presence Questionnaire (PQ) [b-Witmer-1], [b-Nystad, b-Youngblut] and a partial positive correlation between the ITQ and presence measured with the Slater-Usoh-Steed (SUS) [b-Slater] questionnaire [b-Youngblut].

The level of expertise can also influence QoE, for example in the sense of technical expertise about the system (when a user understands the technology, the experience is rated differently than it is by a naive user), the expertise to use the system to achieve a goal [ITU-T G.1035], and the expertise about the subject of the XR meeting, which influences cognitive load etc. [b-Skowronek-3].

Spatial intelligence and introversion both affect the experience of spatial presence, such that higher degrees of spatial intelligence and introversion correspond to higher degrees of spatial presence within the sense of presence [b-Jurnet-1], [b-Jurnet-2].

8.1.3 Conversation behaviour

The conversation behaviour of participants will influence the structure and flow of the conversation during a meeting and thus they will influence the QoE perceived by the participants. Here, aspects such as, for instance, the formality of the conversation or communication discipline (see [b-Mengis]) determine conversation behaviour.

Moreover, the degree of involvement, that is the extent to which participants actively contribute to the XR meeting, can also influence the QoE. Passive participants tend to perceive quality degradations more critically than actively engaged participants, see e.g., [b-Schmitt], [b-Skowronek-1].

8.1.4 Relations between participants

People in close relationships typically experience higher levels of social presence when communicating through XR, particularly when the communication medium affords interpersonal affective social touch [b-Longa], [b-Fermoselle], [b-Gavrilovska], [b-Haritaipan], [b-Huisman].

The arrival of the tactile Internet will stimulate the integration of the haptic modality into digital communication devices and interfaces, with remote personal communication. health care and teleoperation becoming the primary market domains [b-Jewitt]. Mediated social communication including touch can contribute significantly to a feeling of social presence in the context of collaborative tasks performed in shared virtual environments. However, since mediated social touch can cause feelings of discomfort between strangers [b-Smith], a closer (e.g., romantic) relationship may be preferred for this kind of tactile stimulation [b-Rantala], [b-Suvilehto].

8.1.5 Language and body language aspects

The QoE of an XR telemeeting depends on the extent to which the system conveys linguistic social cues and body language. Linguistic cues refer to the choice of words and structure of sentences, i.e., dialect and syntax, that users employ during communication. While linguistic social cues are essential in mediated social communication, they are far more effective when paired with other non-verbal social cues such as body language [b-Sharan]. Body language consists of two key elements: (1) body posture and movements and (2) head movements and hand gestures. While the former is subtle, less definite and can indicate some feelings and attitudes, the latter is deliberate and important to communicate meaning without words. Some examples of body posture movements are arm- or leg-crossing. Head and hand gestures could include movements such as nodding, pointing and waving [b-Sharan].

8.2 Context influence factors (CIFs)

Context influence factors (CIFs) refer to the physical, temporal, social, economic, task and any technical and information contexts that influence QoE [b-Brunnström], [b-Reiter]. With respect to XR meetings, CIFs essentially refer to the overall situation in which the XR meeting takes place. [b-Döring], clause 3.2.2, provides an overview of context (and content) factors that (are likely to) contribute to videoconferencing (VC) fatigue. It remains to be investigated to which extent XR lightens or aggravates those effects.

While there a lot of related work exists, as described in the following paragraphs, it still needs to be validated in the context of XR meetings.

8.2.1 Communication environment

Corresponding CIFs are given by the *acoustical* and *optical/lighting* situation, see e.g., [b-Shimizu], [b-Porcu].

Tests in lab environments demand a detailed specification and validation of the lab setup and procedures, which is a challenging task in face of the increased number of degrees of freedom provided (cf. clause 8.3.3.1). Out-of-the-lab testing complements tests in lab environments, allowing for real use cases while limiting the possibility to derive ground-truth-like insights. Crowdsourcing goes one step further by leaving the execution of the experiment to the user, but there are special methods to validate the results as described in [b-ITU-T PSTR-CROWDS].

8.2.2 Use case

One CIF in this category addresses the *purpose of the meeting*. For meetings with a focus on the accomplishments of tasks, the two-dimensional circumplex model of group tasks [b-McGrath] can be used to characterize the purpose of the meeting. Complementary distinctions of meeting purposes can be made between:

- professional and private/business and leisure meetings;
- accomplishing tasks, fulfilling social needs (including belongingness and co-presence) and exchanging information;
- distributed collaborative work with shared *virtual workspaces* and related perceptual, cognitive and usability aspects.

Dependent on the purpose, the user can be focused on:

- participants (i.e., socializing);
- objects (in space);
- tasks (at hand);

These in turn place demands on the degrees of realism and performance that are required for the various elements. More information can be found in [b-Skowronek-2].

8.2.3 Communication scenario

Related CIFs are the size (number of participants [ITU-T P.1301], [b-Skowronek-4) and nature of a group (closed versus open), the number of connected sites [ITU-T P.1301], as well as potential mixes of face-to-face and mediated conversations [ITU-T P.1301], [b-Skowronek-2].

8.2.4 Time aspects

Telemeetings can bridge time zones, whereby the feasibility of the resulting time zone constitutes a CIF for the affected user. Section 3.2.1 of [b-Döring] provides an overview of temporal-organizational CIFs in VC, such as the number, duration and timing of sessions. Also, social uses and habits represent CIFs.

8.3 System influence factors (SIFs)

System influence factors (SIFs) take into account the various impacts of parts of the system, such as hardware, software, application and network, on QoE. In a VR or XR scenario, the system has to ensure that the flow of video and sensor data is transmitted, processed, combined and viewed in a timely manner.

An SIF may determine the QoE and allow for indirect QoE monitoring once a formal relationship between QoE and SIF is established.

However, a SIF-originated *cause* of QoE degradation, such as data packet delay or losses in an Internet context, might not manifest itself immediately to the user, but transform throughout the communication and application stack [b-Minhas] into various perceivable *effects* on features. An example of such an effect is wireless transmission errors due to fading entail retransmission delays and reductions of throughput, yielding lower quality and/or playback glitches. Such effects can, however, be mitigated by additional measures, e.g., packet loss concealment.

In the following clauses, SIFs will be dealt with in a top-down fashion, from human/world-related factors (clause 8.3.1), to rendering aspects (clause 8.3.2), to network and compression (clause 8.3.3).

8.3.1 Human/world-related factors

These high-level factors address degrees of freedom (clause 8.3.1.1), representation of users (clause 8.3.1.2) as well as aspects of realism and style (clause 8.3.1.3), locomotion (clause 8.3.1.4), positioning (clause 8.3.1.5) and proxemics (clause 8.3.1.6).

8.3.1.1 Degrees of freedom

In XR, *degrees of freedom* are defined as the basic ways that an object or avatar can move. There are six degrees of freedom in 3D space, rotational movement around the x, y and z axes (pitch, yaw and roll), and translational movement along those axes (forward/backward (surge), left/right (strafe) and up/down (elevation)).

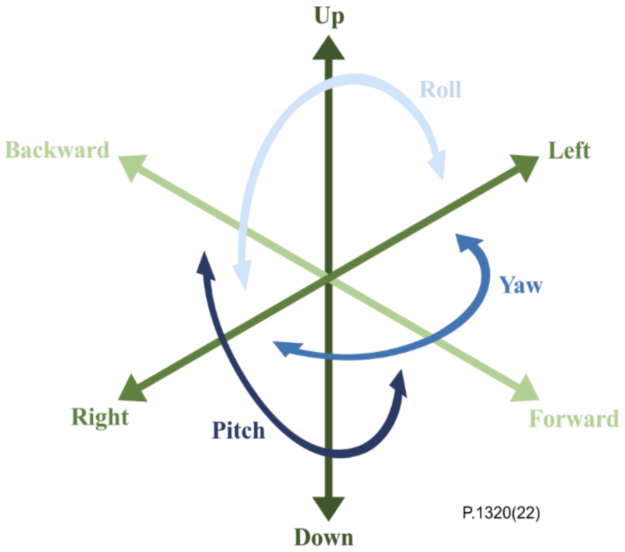

Figure 2 – Illustration of six degrees of freedom

A common distinction in XR is between three degrees of freedom (3DoF) or fully six degrees of freedom (6DoF). This both relates to the tracking of the physical movement of a device, but also to how the avatar and objects controlled by the devices can move in the virtual environment. Usually, 3DoF means that rotational movement can be tracked and acted upon, but the object or avatar controlled by the user does not change its position.

Comparison between subjective experiments conducted in 3DoF and 6DoF environments showed a statistical effect of degree of freedom (DoF) on the collected scores for at least one out of two test contents under study [b-Subramanyam]. Moreover, on average more time was spent on observing contents in 6DoF, compared with the time spent in the 3DoF counterpart.

DoF influences and is influenced by the content chosen to represent the XR environment: having 6DoF in a 360° video will not be useful, as the content does not change even if the user changes its position (actually, it might shatter the illusion of presence, since there is no parallax effect). Visualizing a 3D content in 3DoF will mean that some parts will never be visible/accessible.

8.3.1.2 Representation of users

Degree of user modelling is a characteristic feature of collaborative virtual environments with effects on co-presence [b-Steed]. A first distinction in user representation is whether it is achieved through the insertion of a real-time video stream in the scene or through an animated avatar.

Real-time video user representation (sometimes described as photorealistic avatars) display a real-time capture of the remote user within the scene. This can be done using conventional 2D video or a 3D representation, typically based on point clouds.

Animated avatars may show different levels of realism with respect to its animation capabilities and its appearance. **Puppeteered avatars** present the first stage of modelling, where users can customize the avatar but controlling them happens only through an interface. **Tracked avatars** involve tracking a few or multiple points such as the limbs and joints of the user's body in addition to head tracking in order to create a virtual representation that is able to convey more information about the behaviour and mood of the user. **Reconstructed avatars**, or captured avatars, require a full 3D capture of the user's appearance and movement in real time. Ideally the facial expression and eye gaze of the animated avatar will also match the users in real time.

Comparisons between different types of user representation in XR meetings have been reported in [b-Gunkel], [b-Li].

Additionally, the representation of the user's own avatar within the virtual environment is relevant, particularly with respect to the user's sense of embodiment [b-Gonzalez-Franco]. The representation of the user may be provided as a **computer-generated avatar** or as a **segmented video stream** from an egocentric capture. In either case, the representation may include only the hands, or the full user body [b-Gonzalez-Sosa]. A comparison between computer-generated and segmented self-representations can be found in [b-Villegas].

8.3.1.3 Realism/style

Realistic video background enhances the feeling of co-presence and realistic avatars increase the sensation of trust in a teleconference [b-Jo-1]. Mismatching levels of realism between background and avatar may reduce the feeling of body ownership and presence [b-Jo-2].

Various properties of an avatar contribute to realism, such as eye contact, body movement, facial expression, gestures, proxemic cues [b-Sharan], and temporal aspects such as a repetition pattern of simulated movements. With respect to the degree of realism, an 'uncanny valley' effect can arise when there is a mismatch between the visual and behavioural degrees of realism (e.g., a highly photorealistic avatar showing an unnatural behaviour) [b-Diel].

8.3.1.4 Locomotion

Locomotion is defined as the act or power of moving from place to place. This concept also transfers into XR locomotion being an essential interaction mode [b-Bozgeyikli]. XR locomotion needs to support either active or passive navigation in the virtual world through walking, driving, flying or other modes. Because of the increased degrees of freedom in viewing virtual worlds, different postures also need to be considered such as seated, standing and walking [b-Zielasko]. In [b-Nguyen-Vo], four locomotion interfaces allowing for limited translation combined with full rotation were used in a mixed-method experiment to compare the associated four levels of translational cues in terms of measures such as task performance, task load and simulator sickness. Further, seating arrangements include fixed chair, swivel chair, couch, and cave automatic virtual environment (CAVE). The influence of standing and seated viewing of 360° degree videos on an head-mounted display (HMD) regarding head movements and viewing behaviour changes over time has been studied in [b-Elwardy]. The analysis of the results from a subjective experiment indicate that each participant has their own distinct exploration behaviour for standing viewing which becomes less different among the participants for seated viewing. Further, amplitude of head movements were higher in standing viewing compared with when seated with high rotations in yaw followed by pitch and then roll.

8.3.1.5 Positioning

For two-way immersive telepresence communication, the local user is represented by an avatar to the remote user. The placement and movements of the avatar of the local user need to adhere to the semantics of the remote user's environment. If the configurations of the two remote spaces are identical, this can be achieved through a rigid transformation between both spaces, and the local user's placement and motion can be captured and applied to the remote avatar [b-Orts-Escolano]. However, users are likely to be in rooms with different layout, size and furniture arrangement, so a simple mapping between two such spaces cannot exist. In this case, a non-trivial method that determines the placement and movement of the avatar according to the configuration of the local environment is needed so that the local user can properly recognize what the remote user is doing and interact through an avatar [b-Yoon].

8.3.1.6 Proxemics

Interpersonal distances should adhere to real-life conventions. People utilize space to communicate comfort, anger, friendliness, and standoffishness through four distance zones [b-Sharma].

1) intimate (< 0.45 m);
2) personal (0.45-1.2 m);
3) social (1.2-3.6 m); and
4) public (> 3.6 m).

Each distance zone has a specific range of proximity affording certain types of communication. For example, the intimate zone is common for communicating through physical contact activities such as expressing affection, comfort, physical stress and protection. People show similar proxemic spacing behaviour in virtual worlds as they do in the physical world [b-Williamson], [b-Yee]. Perceived interpersonal distance (proximity) is a significant determinant of social presence and quality of communication in immersive VR [b-Ennis]. As in real life, invasions of one's personal space by strangers (represented by their avatars) may feel highly uncomfortable in VR/XR. Social communication systems can impose personal boundaries that prevent avatars from coming within a set distance of each other, creating more personal space for people and making it easier to avoid unwanted interactions.

8.3.2 Rendering

This clause starts from the VR-perspective, with focus on single-user rendering including time warp and audio (clause 8.3.2.1), which is extended to a multiparty rendering scenario (clause 8.3.2.2) and short reviews of rendering errors (clause 8.3.2.3), resolutions (clause 8.3.2.4), foveated rendering (clause 8.3.2.5) and overlaying of rendered images (clause 8.3.2.6).

8.3.2.1 Rendering per client

Figure 3 illustrates the main components of the reference system for single-user rendering per client.

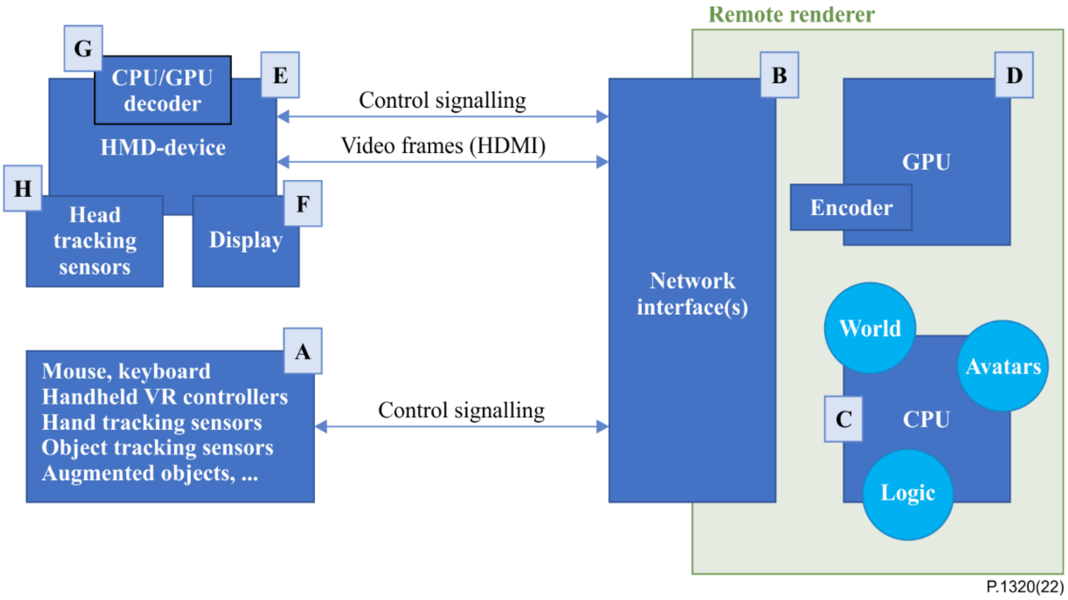

Figure 3 – Reference system for single-user rendering per client

Referring to Figure 3, rendering on both a distributed and a local system can be split into two parts: (1) Create an image on the graphics processing unit (GPU) based on control signals from the client, descriptions of models, physics, textures, viewport and lighting that are already available in the renderer's central processing unit (CPU) and GPU, and eventually (2) adjust the image to be displayed remotely, which is usually not needed for local rendering.

In VR or XR, this image is to be shown on the HMDas quickly as possible. However, given that miniscule movements of the head must coincide with what is displayed in order to avoid motion sickness, the motion-to-photon (MTP) delay needs to be minimal, ideally not exceeding one interframe time, which for frame rates above 50 fps is an even stricter requirement than the 20 ms reported in [b-TR26.918].

There are two possibilities; either content and viewport updates can be synchronized, or they can be separated.

In the synchronized case, MTP applies to all movements together. This MTP is the time it takes from position sensing via rendering to displaying the frame in Figure 3 described by path P1: {H & A}→B→C→D→B→E→G→F. There are many contributions by various entities along this path to the total delay by the sensing logic, application logic (virtual world and avatar calculations), CPU/GPU processing, viewport rendering and lighting calculation, image encoding, network transmissions, sensor buffers, etc. Any delay affects both content and viewport latencies in the same way. In a case in which the actual and perceived instants of movements grow beyond one inter-frame time, the risk for nausea increases [b-Norman].

Breaking synchronicity by separating content and viewport updates makes it possible to adjust the viewport based on head movement alone, before all content-related input has been processed. This delay is then described by path P2: H→G→F within E. Accordingly, the head-MTP delay (path P2) is much shorter than the content delay (path P1).

Time warp

One approach to speeding up the viewport update process is called *time warp* [b-van Waveren]. This implies increasing the content delay by one frame in order to be able to slightly adjust the image laterally to cater for small head movements, as seen from the most recent sensor data (H). To enable this, the first rendering step (D) is actually performed with a resolution that is higher than that of the display so that the adjusted image is a subset of the originally rendered image.

These slight adjustments improve QoE and allow for higher MTPs by smoothing out first stage rendering hiccups and frame losses. Note that time warp, and similar techniques such as space warp, reprojection, etc., are also used in locally rendered devices to handle unplanned jitter from the renderer (see Figure 4 for an example).

In both synchronized and separated cases, barrel distortion to minimize lens distortion effects [b-Watson] and other display specific adjustments are performed in step G to optimize the perception of the flat image on an HMD.

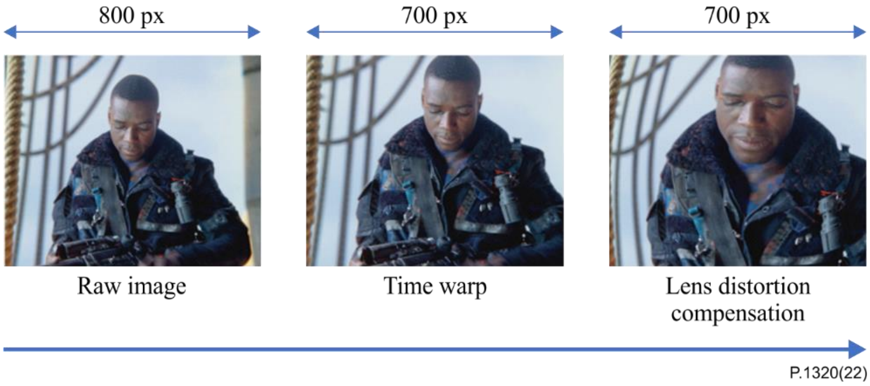

Figure 4 – Transition from flat to lens compensated and time warped image (performed on a local GPU)[1]

Audio rendering

Various degrees of spatial audio rendering may have implications on the QoE. Providing spatial audio within a single shared acoustic space, as opposed to each participant's local acoustic background, may increase QoE constitutes such as immersion and presence, and potentially also task performance. More complex representations of sound sources using directivity can also be modelled which describe the 360° propagation sound level as a function of the source's size, shape and material properties. Combined with an avatars' head orientation driven by the tracked user, source directivity may provide more clarity towards emphasizing the intended recipient of a dialogue exchange in a multiuser scenario. Conversely, the use of *head-related transfer functions* (HRTF) models the directivity of the receiver (listener) used to deliver binaural spatial audio. Sound mediation depends on the placement of the speakers, e.g., over the ears (HTC Vive) or close to them (Oculus Quest 2).

This will probably be handled in the same way as the "split" rendering described in clause 8.3.2.2, with the main audio "space" rendered on the server and the HRTF filtering done in the client. Ambisonics in the AR/VR environment can be used to create a "recording" from the virtual space that then can be delivered to the HRTF algorithm to create the sound that is presented to the ears via headphones or speakers using crosstalk-cancellation. A key performance measure in this context is the motion-to-sound (MTS) delay.

From a physiological point of view, the location of audio sources is strongly influenced by visual information [b-Hershey].

8.3.2.2 Multiparty effects on rendering

Figure 5 illustrates the coordination of several parties and entities. Multiparty communication extends the rendering scenario across several dimensions. On the one hand, different parties may render the same scene using their own instance of the remote renderer system. This means that the information used for rendering, particularly regarding the global status of the system, might differ between parties. This might cause relative latency differences in the update of several graphical objects of the scene, such as avatars.

---

[1] This figure has been captured from the film Tears of Steel by the (CC) Blender Foundation | mango.blender.org, issued under the Creative Commons licence CC BY 3.0.

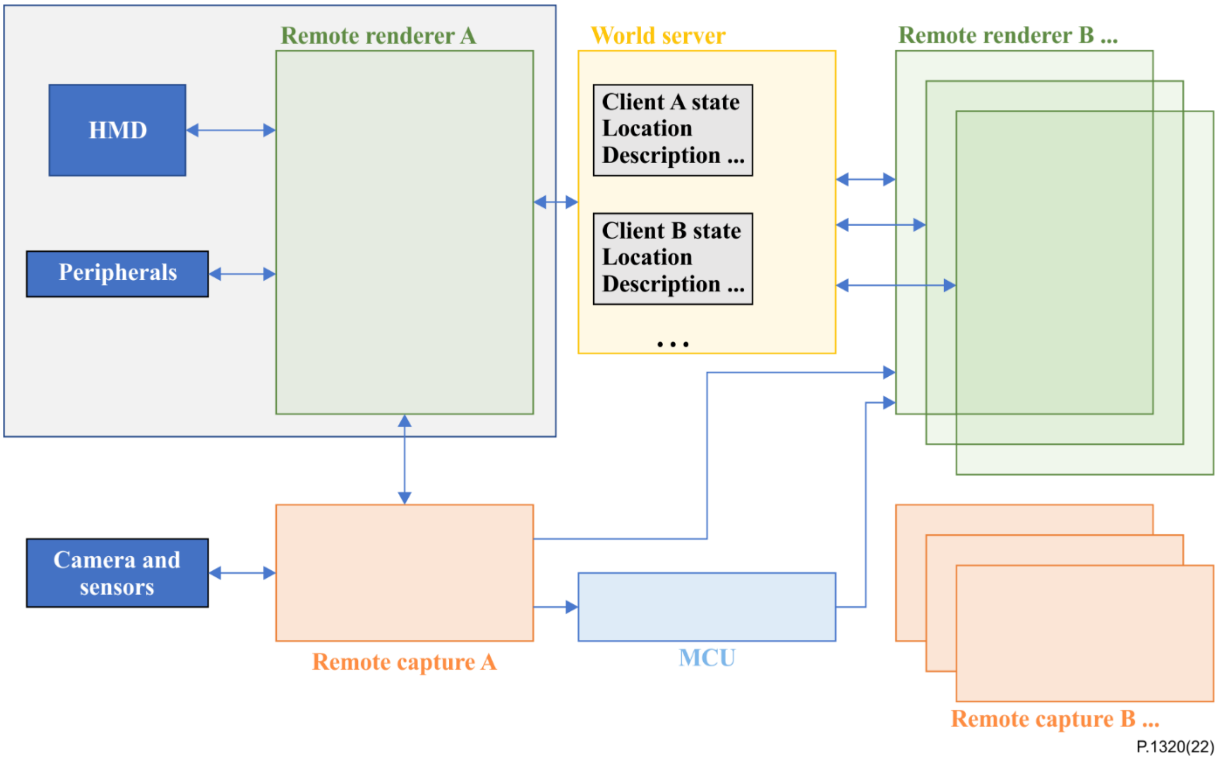

Figure 5 – Multiparty rendering scenario

Obviously, the coordination affects only the processes between entities B-C-D (see Figure 3), but not in the traditional delay sense. Only the client and avatar descriptions are shared together with the complete world description. Each individual renderer then takes care of rendering the scene for each client, based on their individual points of view. Discrepancies between the "true" world state, according to the server, and information in the clients will cause visible effects. Clients will try to estimate the true state, and differences between the true and estimated state are repaired once the information has been shared. This repair is sometimes visible and will give cause to artefacts in the form of temporal rendering errors.

8.3.2.3 Rendering errors

A temporal rendering error is *rubber banding*, which means temporary slow-downs and back-and-forth jumps due to network delays and errors combined with untimely game engine predictions. This is a symptom of the clients "catching up" to the true state of the game or environment server.

Spatial rendering errors include *texture transition errors*, emanating from scalability issues of textures due to varying distances to the corresponding object.

Multiuser AR-specific rendering issues spans both the spatial and temporal VR rendering errors but also a group of errors that originates from the situation of trying to match a virtual object to an already lighted scene, with non-virtual entities that also can affect the rendered items. These errors can be both analogue, due to lens issues or capture settings, but also purely digital.

The digital environment and augmentation errors can originate from mismatch of lighting conditions on objects and the real lighting, occlusion errors, resolution mismatch, errors in reproducing reflections of both real and virtual items, capturing and rendering rate mismatches and viewing angle offset. Not matching colour depth between what is rendered and what is recorded is also something that can make items look out of place in the real world.

User specific errors are errors where there are discrepancies in location, orientation etc. of objects between the different users. This is related to rubber banding and object location prediction, but can also be unsolved object positioning changes making items appear in different locations for different users [b-Kruijff].

8.3.2.4 Resolutions

The spatial resolution, determined by the size of the pixel matrix per eye and the covered viewing angle, impacts the QoE. If video encoding and rendering does not match the display resolution, additional artefacts can occur due to scaling. As the time needed to render, encode and transmit video frames grows with their size, there might be a need for compromising between size and framerate.

The temporal resolution reflects the supported number of video frames per second, which is either given by the minimal latency deadline that can be met when adding up latencies from rendering, encoding, transmitting, decoding and displaying the frame, or by the refresh rate of the head-mounted device.

The delay along the rendering loop reveals to which extent a specific combination of spatial resolution, specific hardware and eventual networks is able to maintain a given temporal resolution, i.e., enable a desired framerate and latency [b-Kelkkanen-3]. At sufficiently high resolutions (2K and above), the delay is dominated by encoding and decoding times.

8.3.2.5 Foveated rendering

Foveated rendering is a technique used together with an eye-tracking device for dynamically adjusting the visual quality depending on the peripheral vision area. The technique will reduce the quality where the human vision has less resolution while preserving a higher quality inside the foveal area.

Another technique that does not require eye-tracking is fixed foveated rendering. In this case, the foveal area is previously fixed assuming a specific point in the rendering. For telemeetings, the foveated rendering may reduce the use of the bandwidth. However, the delay associated with the foveated algorithms may lead to a deterioration of the QoE.

8.3.2.6 Overlaying rendered images

Combining rendered synthetic images with video as well as projecting them on glass or real objects might imply resolution, scaling and location and depth mismatches, subpixel aliasing and (temporal) synchronization issues.

8.3.3 Network and compression

This clause addresses low-level factors related to encoding and decoding (clause 8.3.3.1), latency (clause 8.3.3.2), bandwidth requirements (clause 8.3.3.3) and synchronization (clause 8.3.3.4) requirements.

8.3.3.1 Media encoding/decoding

Textures, images or videos are commonly compressed before being sent over a network to minimize the necessary bandwidth. This is also true for audio, but audio will usually have smaller bandwidth requirements anyway.

There are very fast encoders implemented directly on most modern GPUs today. Having access to the same GPU memory as the rendering process itself makes this a very efficient process, reading from the same frame/texture that the GPU has produced from the game directly. Off-the-shelf solutions can produce encoding times below 2-3 ms per video frame for common full viewport VR headsets, and proprietary solutions can even reach sub-ms algorithmic delays.

Video frames are encoded and transmitted with various techniques, such as varying lengths of group of pictures (GoP), infinite GoP but with video encoding referencing only received P-frames (making it very good at coming back after an error) and trying to balance delay and encoding efficiency for network error resilience in its own way.

Realistic human avatars may be acquired in real-time using volumetric video capture and then represented as point clouds. Point clouds must be encoded to be transmitted through the network. The most widespread compression solution to compress dynamic point clouds is video-based point cloud compression (V-PCC), which was standardized by MPEG [b-Mekuria]. V-PCC encoding is computationally costly, and therefore other solutions such as octree representations [b-Kammerl] are used as an alternative.

VR images contain unseen areas commonly referred to as the hidden area mesh that make images unnecessarily large. In experiments with remote VR, an overall speed increase of up to 10% was achieved by remapping the rendered images in a way that cuts out the unseen areas, thus reducing image size and therefore codec time consumption [b-Kelkkanen-2].

8.3.3.2 Latency

In the VR context, constant latency stemming from physical distance and various processing steps affects MTP and thus reaction times, but may not need to have a negative impact on the frame rate [b-Kelkkanen-3]. On the other hand, delay variations due to communication outages and/or resource competition introduce temporary reductions of video frame rates [b-Kelkkanen-3] and freezes [b-Fiedler9]. Recent measurements revealed significant and varying network delays in case of 4G, 5G without ultra-reliable low latency communications (URLLC), Wi-Fi (802.11ac/n) and wired connections across the public Internet [b-Kelkkanen-3].

Relationships between injected delays and QoE ratings in VR remote control and game scenarios have been reported in [b-Brunnström] and [b-Vlahovic].

MTP delays in immersive teleconferencing environments are mainly classified into those affecting the consistency of the world with regard to the user's movements and those related to the user's embodiment or self-delay. Some studies have addressed the influence on the QoE of the consistency delay [b-Singla-2], [b-Cortés]. Other studies have evaluated the influence of the self-delay [b-Caserman], [b-vanDam]. These studies suggest that the consistency delay is more restrictive than the self-view delay in terms of QoE and should remain below 150 ms.

8.3.3.3 Bandwidth requirements

The bandwidth requirements in order to maintain a given spatial and temporal resolution depend on the scene complexity and HMD rotation speed. [b-Kelkkanen-1] reported that average head moving speeds required a steadily available bandwidth of more than 20 Mbit/s for supporting both channels at 2K; this requirement basically doubled when head moving speeds were tripled.

For point clouds, the common range of bitrates for typical representations of 800K-1M points is between 5 and 100 Mbit/s [b-Schwarz], [b-Essaili]. User-adaptive strategies can be adopted to reduce bandwidth requirements while prioritizing the user's QoE. As [b-Subramanyam-1] shows, user-adaptive allocation of bandwidth in tiled point cloud contents leads to significant gains in perceived quality compared with non-adaptive streaming at the same bitrates. In [b-Subramanyam-2], the quality of live-captured point clouds of approximately 130K points at 15 fps was evaluated in an XR meeting system, demonstrating that tiled user-adaptive point cloud streaming at 14 Mbit/s leads to similar visual quality as uncompressed streaming at ca 300 Mbit/s. Also, the end-to-end delay can differ in the order of 70 ms depending on whether the point cloud compression/decompression is carried out in the edge or on the device itself, according to [b-Essaili].

Similarly to delay variations, bandwidth variations cause quality variations over time. [b-Orduna] evaluated the QoE of 360° video with time-varying qualities, simultaneously with high-level features (presence, empathy, attention), in a simulated telemeeting environment. For 360° video transmission of XR meeting scenes, where the camera is normally in a fixed position, 5-10 Mbit/s are sufficient for a 4K resolution equirectangular panorama [b-Tran], [b-Krogfoss].

8.3.3.4 Synchronization requirements

Regarding audio and video synchronization, [b-ITU-R BT.1359-1] reports on *detectability* thresholds between 45 ms (audio advanced) and 125 ms (video advanced) and *acceptability* thresholds between 90 ms (audio advanced) and 185 ms (video advanced). These figures relate to broadcasting in the first place, and need to be validated in the context of XR meetings. There appears to be a significant positive correlation between sensory synchronization and social presence [b-Becker].

9 Constituents of QoE in XR telemeetings

This clause surveys the constituents contributing towards QoE that are especially relevant for XR telemeetings.

9.1 Simulator sickness

Simulator sickness (also known as cybersickness) refers to the unpleasant sensations which can occur among users during and after exposure to XR. The symptoms are likened to those of motion sickness and usually increase with time of exposure. After a threshold time the symptoms may stop increasing or even decrease due to an adaptation effect following multiple exposures to XR. For a review, see [b-Duzmanska]. Simulator sickness is negatively related to presence (feelings of presence decrease with increasing simulator sickness) [b-Weech], and this relationship appears to be mediated by factors including illusions of self-motion (vection), navigation control and display factors.

Simulator sickness is currently most often assessed by using self-report methods such as the Simulator Sickness Questionnaire (SSQ) [b-Kennedy]. Physiological, objective indicators are under active investigation. Measures such as fatigue, for example, may have already manifested to a degree within participants. This is then unproportionally attributed to the system via subjective evaluation and can be better determined objectively [b-Iskander]. Understanding and measuring simulator sickness is considered relevant to the evaluation of XR telemeeting QoE; see related findings in [b-Singla-1], [b-Singla-3], [ITU-T P.919].

9.2 Immersion

The term immersion is closely related to the sensation of presence (clause 9.3). In this Recommendation, immersion is understood along three axes: (1) as a property of a system, (2) a subjective response to narrative contents, or (3) a subjective response to challenges within the virtual environment following the taxonomy presented in [b-Nilsson]. The first is objectively measurable from the properties of the system, whereas the last two are related to subjective attention and focus on the events in the XR service manifested by mental absorption with the mediated experience. These three dimensions span a space which allows immersion to arise equally well from services requiring no technology (e.g., reading a book) to complex XR systems. Understanding the narrative properties, sensorimotor skill requirements and intellectual challenges of an XR telemeeting service in addition to technical fidelity may become defining constituents of the QoE.

9.3 Presence, co-presence and social presence

Presence has been defined as: "a psychological state in which the individual perceives himself or herself as existing within an environment" [b-Blascovich]. XR has the potential to embody users in another reality, where they feel present and have agency and first-person perspective heightening the sense that the experience is really happening. Today, assessment of presence is usually done by using post-test questionnaires, which can be considered basic tools in the measurement of presence [b-Laarni]. These include the Immersive Experience Questionnaire (IEQ) [b-Jennet], Igroup Presence Questionnaire (IPQ) [b-Schubert] and Presence Questionnaire (PQ) [b-Witmer-1] among others. All of the above questionnaires measure different aspects of presence such as spatial presence, realism, involvement and distraction. Presence has also been measured with psychophysical methods, behavioural measures, performance measures and physiological measures, to name a few, but comparison of the results remains challenging due to their different types and the diversity of the XR applications used [b-Laarni].

Co-presence refers to the "experience of being with others" as a physical mode and social sense [b-Zhao], [b-Steed]. The topic is less studied than presence and it may be best understood via a situation where co-presence is absent or reduced such as when using instant messaging or making a phone call. In these situations, keeping attention is not as easy as it is during real interaction and misunderstandings easily occur. As [b-Steed2015] found, novel lines of investigation are needed to understand the sensation of being there with others compared to the more studied spatial experience of being there (communication versus spatial use of collaborative XR).

Social presence[2] builds on the experience of co-presence and adds the dimension of affective and intellectual connection with others. Like presence above, it has been defined as: "a psychological state in which the individual perceives himself or herself as existing within an interpersonal environment" [b-Blascovich]. Social presence in virtual environments is typically measured and compared with face-to-face meetings through questionnaires [b-Greenwald], [b-Li], [b-Toet-4].

9.4 Plausibility

Plausibility can be considered a component of presence, and it may be thought to refer to the illusion that perceived events in the virtual environment are really happening [b-Bergström]. [b-Slater-1] defined plausibility as the extent of an immersive system producing realistic responses with regard to the user, as well as the overall credibility of the virtual scenario taking place. The simulation has to match expectations where these are relevant, e.g., if the system simulates a real-world event, that system has to conform to what is expected to happen in reality. Measuring plausibility is a less researched area than that of presence, and no consensus about the methodology exists. Attempts are made to estimate the relative plausibility of a system by comparing it with another system and iteratively adjusting technical properties to achieve the same level of realism in both presentations [b-Slater-2], [b-Bergström].

Perceived realism, plausibility and coherence are the central outcomes of the sensory, perceptual and cognitive processing layers in the human brain that determine the quality of a mediated experience [b-Latoschik], [b-Parola], [b-Weber].

At the sensory level, the perceptual or *sensory fidelity* of the experience is the relevant quality factor for telepresence. Sensory fidelity determines the extent to which users fail to perceive or acknowledge the fact that (part of) their sensory input is mediated. Users should preferably experience the feeling that their sensory input originates directly from the represented environment (the illusion of non-mediation [b-Lombard-1], [b-Lombard-2]).

At the affective or emotional level, the *internal plausibility* (sensory congruity or internal consistency; [b-Cahill]) of a mediated experience is an essential factor contributing to the sense of telepresence [b-Wirth]. Internal plausibility determines the extent to which users have the feeling that their multisensory input is coherent [b-Skarbez] and agrees (is congruent and consistent) with their mental model (expectations or memories) of the represented environment [b-Cahill], [b-Hofer], [b-Wirth].

At the cognitive level, the relevant quality factor for telepresence is the *external plausibility* or environmental and thematic congruity [b-Cahill] of the experience. External plausibility determines the perceived fidelity [b-Alexander], realness [b-Khenak], [b-Latoschik], [b-Lindau] or illusion that the represented environment is authentic [b-Gilbert] and a place that can actually be visited [b-Gonçalves], [b-Slater-1].

---

[2] Presence relates to the feeling of being in an environment, while social presence refers to the feeling of being with others.

Suspension of disbelief refers to the avoidance of critical thinking in explaining an experience for the sake of enjoyment.

9.5 Ethics of XR use

XR telemeetings enable individuals to interact with virtual characters representing other real humans in a myriad of possible experiences. In this setting, the reciprocity of human interaction, also known as the golden rule "treat others as you would have them treat you", may become less clear and open a channel for immoral behaviour. Considering the potential harm of immoderate use, minimizing content-induced risk, and controlling the level of realism may become ways to ensure the quality and safety of an XR telemeeting service for all populations. For a review, see [b-Slater-3].

9.6 Carving out mental space

XR conferences do away with physical dislocation, which makes it harder for participants to "be" at a virtual conference [b-ACM2020]. Part of the QoE might arise from the service being able to encourage real-time participation and casual interactions among the participants.

Regarding surrounding people, the social acceptability of being "physically present, but socially absent" is relevant [b-ACM2020].

9.7 Fatigue and cognitive load

In our digital age, human social interaction is often mediated. Technologies such as videoconferencing software (e.g., Zoom, Microsoft Teams, Skype) are becoming increasingly popular as they afford a new form of virtual togetherness by facilitating shared and synchronous social activities, thereby substituting real-life interactions [b-Hacker], [b-Shah]. To provide the affective experience of in-personal social interactions these tools should reliably and intuitively transmit the social and ambient affective cues that people typically need to build trust, empathy, and confidence in real-life (face-to-face) social settings [b-Nadler]. Communication through systems that distort or fail to transmit these cues will result in a lack of social presence (i.e., the feeling of being in the presence of, and having an affective and intellectual connection with, other persons [b-Biocca), and may lead to physical and mental exhaustion [b-Hacker], a phenomenon that is also known as Zoom fatigue [b-Nadler].

9.8 Ability to reach goals

The system should provide means to achieve the targets set for the telemeeting in an effective, efficient, and satisfying manner [b-ISO 9241-11]. It is still for further study to understand which goals +are easier to reach in an XR meeting than in an ordinary video conference.

10 Test methods targeting XR telemeeting QoE

Any test method aiming at evaluating QoE factors in XR telemeetings should control confounding human, context and system influencing factors, possibly by isolating each factor or designing the test accordingly so all factors can be considered as non-aliased. In particular, system influencing factors, such as rendering technology and bandwidth allocation, are more likely to be considered independent variables, whereas human and context influencing factors should be controlled and balanced to ensure a wide applicability of the obtained results.

The test should aim at revealing the effects of the selected independent variables on the dependent variable(s) associated with QoE constituents, as described in clause 8. In particular, two main objectives can be envisioned for a test evaluating the QoE of XR telemeetings:

(a) Systematically control an independent variable and observe the effects on QoE.

(b) Test complete black box systems without exploring individual variables.

The test methodology should define a suitable task for the participants that will allow for the examiner to observe the effects of the independent variables on the dependent variable(s) associated with QoE constituents. In this regard, tasks that mainly focus on audio synchronization and that have a limited visual and interaction component might not be appropriate to study independent and dependent variables connected to the visual and interactive properties of the XR telemeeting system. As such, the tasks should be carefully selected, making sure that they are suitable for assessing the independent and dependent variables under test. [ITU-T P.1301] provides a summary of sample tasks for QoE assessment in telemeetings. In the Recommendation, tasks are classified according to whether they use audio and video (audiovisual) or audio only. However, these tasks have not yet been validated for XR.

Each test should be subject to an ethics review, following the guidelines of the institution where such tests will be performed.

10.1 Test paradigms

10.1.1 Conversation and behaviour analysis

The following items need to be taken into account:

- Conversation analysis, to examine for example turn taking, and effects of delay [ITU-T P.1305], [b-Berndtsson].
- How well collaborators understand the intention of the others [b-Steed]. Being able to tell what the other person intends to do based on the subtle gestures and eye gaze alongside their speech [b-Bailenson].
- Remaining misunderstandings and personality bias arising from collaborative situations. [b-Steed]. This should display an improvement over traditional telemeetings.
- "Waiting room" paradigm and gaze behaviour [b-Bailenson].
- Taxonomy (cues, tasks, metrics) [b-Skowronek].

10.1.2 Post-experience questionnaires

QoE constituents are typically measured through dedicated post-experience questionnaires. More information about constituents of QoE in XR telemeetings can be found in clause 9.

Some examples of measurements used for assessing QoE constituents are:

- Simulator sickness: SSQ [b-Kennedy], VRSQ [b-Kim], VSR [b-Perez].
- Immersion [b-Georgiou].
- Presence [b-Benyon], [b-Schubert], [b-Slater-4], [b-Vorderer], b-Witmer-2], co-presence and social presence [b-Bailenson], [b-Basdogan], [b-Harms], [b-Lessiter], [b-Li], [b-Makransky].
- Plausibility [b-Lindau].
- Fatigue and cognitive load: NASA-TLX [b-Cao], [b-Hart].
- Engagement [b-O'Brien].
- Emotional involvement (valence, arousal) [b-Toet-3].

Additional assessment tools are free form questionnaires and (semi-structured) interviews.

10.1.3 Physiological measurements

Psychophysical measurements reveal participants' quality perception while physiological measurements relate to covert processes that may induce emotional arousal. Prominent measurement options that have been used for physiological-based QoE assessment include the following:

- Pupillometry: Deals with measuring the pupil of the eye such as pupil size and reactivity. Specifically, in psychology, the recordings of the diameter of the eye in reaction to visual stimuli are explored to determine a participant's interest in the stimuli. The work reported in [b-Beatty] indicates that the pupil diameter can be correlated to processing load, memory, emotion and arousal. The pilot study reported in [b-Hu] on the effect of standing and seated viewing of 360^{o} videos on subjective quality assessment recorded slightly higher pupil dilation for standing viewing. This suggests a slightly more pleasant immersive experience during standing viewing compared with seated viewing.

- Galvanic skin response (GSR): The GSR measures the resistance on the surface of human skin, which can be translated to the emotional arousal of a person elicited by visual stimuli. GSR signals consist of a tonic component and a phasic component. The tonic component relates to slow changes in the GSR signal which can be attributed to changes, e.g., the level of stress, and other general changes in autonomic arousal. The phasic component relates to rapid changes or peaks in the GSR signal on top of the tonic level. GSR peak analysis is often used to relate changes in emotional arousal to specific stimuli or unexpected events. In [b-Engelke], it was conjectured that, e.g., immersive environments and interactive systems that have an impact on arousal would benefit more from GSR measurements than conventional viewing environments.
- Electroencephalography (EEG): EEG measures the changes of electrical potentials on the scalp induced by the electrical activity on the surface layer of the brain. Because EEG has excellent time resolution, it can be used to pinpoint event-related cognitive and emotional processing. The work in [b-Kroupi] studied the impact of QoE on brain oscillations and the EEG signal was decomposed into six frequency bands. The results indicate that perceived high/low quality induces positive/negative emotional processes. In [b-Engelke], it was reported that brainwaves in the 4-8 Hz band can be used to reveal decreased attention and increased drowsiness while brainwaves in the 8-13 Hz band can be related to decreased alertness and increased relaxation.
- Electrocardiography (ECG): ECG measures the heart's electrical activity as voltage over time using electrodes placed on the skin. The work reported in [b-Keighrey] observed a slightly higher average heart rate for participants in VR environments compared with non-VR environments. In [b-Egan], a physiology-based QoE comparison of interactive AR, VR and tablet-based applications was provided. The analysis of the heart rate conducted in this work revealed no statistically significant results among the considered environments.

Appendix I

Related work in other standardization bodies and industry

(This appendix does not form an integral part of this Recommendation.)

I.1 3rd Generation Partnership Project (3GPP)

Frameworks for XR have been analysed by 3GPP [b-TR 26.928]. After defining key terms and outlining the QoE/QoS issues of XR-based services, the delivery of XR in the 5G system is discussed. In addition to the conventional service categories, conversational, interactive, streaming and download, split compute/rendering is identified as a new delivery category. A survey of 3D, XR visual and audio formats is provided.

Use cases and device types have been classified, and processing and media centric architectures are introduced. This includes viewport independent and dependent streaming, as well as different distributed computing architecture for XR. Core use cases of XR include those unique to AR and MR in addition to those of VR discussed in [b-TR 26.918], ranging from offline sharing of 3D objects, real-time sharing, multimedia streaming, online gaming, mission critical applications, and multiparty calls and conferences.

In [b-TR 26.998] the findings of [b-TR 26.928] were further analysed with specific focus on augmented reality (AR) experiences and AR glasses.

3GPP SA4 has a number of work items on immersive services, for example, "Immersive Voice and Audio Services (IVAS)" and "Immersive Real-time Communication for WebRTC (iRTCW)".

I.2 Virtual Reality Industry Forum (VRIF)

VRIF will rely on, and liaise with, standards development organizations for the development of standards in support of VR services and devices. The guidelines work is divided in four parallel tracks: VR360, Volumetric, Social VR and 5G Cloud. The latest version of the guidelines can be found at [b-VR Industry Forum].

I.3 Moving Picture Experts Group (MPEG)

MPEG is a working group of authorities that was formed by ISO and IEC to set standards for audio and video compression and transmission.

I.3.1 MPEG-I: Coded Representation of Immersive Media, ISO/IEC 23090

MPEG-I currently comprises a number of parts including the recently finalized versatile video coding (VVC), omnidirectional media format (OMAF) for coding of 360° video, point cloud compression (PCC), and network-based media processing (NBMP). In the immersive audio coding project, a VR/AR evaluation platform is used to render the visual scene to a head-mounted display or AR glasses. The evaluation process is controlled through a hand-held controller and an AB evaluation panel is rendered to the user through the head-mounted display.

I.3.2 ISO/IEC JTC 1/SC 29/AG 5 MPEG Visual Quality Assessment

AG5 is an advisory group devoted to support needs for quality assessment testing in close coordination with the relevant MPEG Working Groups, dealing with visual quality. The activities range from assisting in evaluating visual quality of new technologies, to be considered for new standardization activities, to support the definition of calls for proposals (CfPs) for new standardization work items, to design subjective quality methodologies and objective quality metrics for assessment in calls for evidence (CfE) and CfPs, to the evaluation of new standards compared with existing solutions. An ad hoc group on the quality of immersive visual media has recently been created, and an overview of quality metrics and methodologies for immersive visual media (AG 05/N00013) is currently under study in the group.

I.4 Khronos Group

Khronos Group maintains an API called OpenXR which aims to make it easier for software developers and device manufacturers to connect in a platform-agnostic way [b-OpenXR1].

I.5 Video Quality Experts Group (VQEG)

The mission of the VQEG Immersive Media Group (IMG) is quality assessment of immersive media, including virtual reality, augmented reality, stereoscopic 3DTV and multiview.

The VQEG Psycho-Physiological Quality Assessment (PsyPhyQA) WG is directed towards the development of new psycho-physiological methodologies for subjective assessment and objective measurement of interactive communications services [b-Bosse].

VQEG collaborates with ITU-T SG12 and has contributed to corresponding standards, for example [ITU-T G.1035] "Influencing factors on quality of experience (QoE) for virtual reality (VR) services" and [ITU-T P.919] "Subjective test methodologies for 360° video on head-mounted displays". The 5G KPI has recently submitted GSTR-5GQoE, "QoE requirements for real-time multimedia services over 5G networks", which includes an XR use case. There is also a collaboration regarding assessment of XR meetings with the IMG group.

I.6 European Network on Quality of Experience in Multimedia Systems and Services (Qualinet)

The mission of Qualinet is to create a network for multidisciplinary QoE research in Europe. Task Forces are used for collaboration around targeted research questions. One of the Task Forces is "Immersive Media Experiences (IMEx)". Qualinet also authored a white paper on the definition of immersive media experience [b-Perkis].

I.7 World Wide Web Consortium (W3C)

W3C develops inclusive XR solutions to help create new standards for inclusive XR on the web. W3C identifies accessibility gaps and architecture issues in existing web XR technologies.

I.8 ITU-T SG16/Q8 Immersive Live Experiences

Current work items:

- Draft Recommendation Requirements of Interactive Immersive Services: H.IIS-Reqts
- Update service scenarios of immersive live experience in [b-ITU-T H.430.3]

Bibliography

[b-ITU-T H.430.3]	Recommendation ITU-T H.430.3 (2018), *Service scenario of immersive live experience (ILE)*.
[b-ITU-T PSTR-CROWDS]	Technical report PSTR-CROWDS (2018), *Subjective evaluation of media quality using a crowdsourcing approach*.
[b-ITU-R BT.1359-1]	Recommendation ITU-R BT.1359-1 (1998), *Relative timing of sound and vision for broadcasting*.
[b-ISO 9241-11]	ISO 9241-11: 2018, *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*.
[b-TR 26.918]	3GPP (2022), TR 26.918, *Virtual reality (VR) media services over 3GPP (Release 17).*
[b-TR 26.928]	3GPP TR 26.928 (2022), *Extended reality (XR) in 5G (Release 17).*
[b-TR 26.998]	3GPP TR 26.998 (2022), *Support of 5G glass-type augmented reality/ mixed reality (AR/MR) devices* (Release 17).
[b-ACM2020]	ACM presidential task force (2020), *Virtual conferences: A guide to best practices*, Version 1.1, 1-30. https://people.clarkson.edu/~jmatthew/acm/VirtualConferences_GuideToBestPractices_CURRENT.pdf
[b-Agrawal]	Agrawal, S., Bech, S., Bærentsen, K., De Moor, K., and Forchhammer, S. (2021), *Method for subjective assessment of immersion in audiovisual experiences*, Journal of the Audio Engineering Society, Vol. 69, No. 9, 656–671. doi: 10.17743/jaes.2021.0013.
[b-Ahn]	Ahn S.J., Levy L., Eden A., Stevenson Won A., MacIntyre B., and Johnsen K. (2021). *IEEEVR2020: Exploring the First Steps Toward Standalone Virtual Conferences*. Frontiers in Virtual Reality, 2:648575. doi: 10.3389/frvir.2021.648575.
[b-Alexander]	Alexander, A.L., Brunyé, T., Sidman, J., and Weil, S.A. (2005), *From gaming to training: a review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games*, The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC'05), DARWARS Training Impact Group, pp. 1–14.
[b-Alsina]	Alsina-Jurnet, I., and Gutiérrez-Maldonado, J. (2010), *Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments*, International Journal of Human-Computer Studies, Vol. 68, No. 10, 788–801. doi: 10.1016/j.ijhcs.2010.07.001.
[b-Bailenson]	Bailenson, J.N., Blascovich, J., Beall, A.C., and Loomis, J.M. (2003), *Interpersonal distance in immersive virtual environments*. Personality and Social Psychology Bulletin, Vol. 29, No. 7, 819–833. doi: 10.1177/0146167203029007002.
[b-Bailenson]	Bailenson, J.N. (2021), Nonverbal Overload: *A Theoretical Argument for the Causes of Zoom Fatigue*. Technology, Mind, and Behaviour, Vol. 2, No. 1. doi: 10.1037/tmb0000030.

[b-Basdogan] Basdogan, C., Ho, C.-H., Srinivasan, M.A., and Slater, M. (2000), *An experimental study on the role of touch in shared virtual environments*. ACM Transactions on Computer-Human Interaction, Vol. 7, No. 4, 443–460. doi: 10.1145/365058.365082.

[b-Becker] Becker, S., Schrills, T., and Franke, T. (2021). *Social presence despite isolation - Insights into the relation between psychological distance and sensory synchronization in computer-mediated communication*, Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021), eds. N.L. Black, W.P. Neumann and I. Noy. Cham, Switzerland, Springer International Publishing, pp. 145-153.

[b-Benyon] Benyon, D., Smyth, M., O'Neill, S., McCall, R., and Carroll, F. (2006), *The place probe: Exploring a sense of place in real and virtual environments*, Presence: Tele-operators and Virtual environments, Vol. 15, No. 6, 668–687.

[b-Bergström] Bergström, I., Azevedo, S., Papiotis, P., Saldanha, N., and Slater, M. (2017), *The plausibility of a string quartet performance in virtual reality*, IEEE Transactions on Visualization and Computer Graphics, Vol. 23, No. 4, 1332–1339. doi: 10.1109/TVCG.2017.2657138.

[b-Berndtsson] Berndtsson, G., Schmitt, M., Hughes, P., Skowronek, J., Schoenenberg, K., and Raake, A. (2018), *Methods for human-centered evaluation of MediaSync in real-time communication*, M. Montagud, P. Cesar, F. Boronat, and J. Jansen (Eds.), MediaSync: Handbook on Multimedia Synchronization (pp. 229–270), Cham, Switzerland, Springer International Publishing. doi: 10.1007/978-3-319-65840-7_9

[b-Biocca] Biocca, F., Harms, C., and Burgoon, J.K. (2003), *Toward a more robust theory and measure of social presence: Review and suggested criteria*, Presence: Teleoperators and Virtual Environments, Vol. 12 No. 5, pp. 456–480. doi: 10.1162/105474603322761270.

[b-Blascovich] Blascovich, J. (2002). *Social influence within immersive virtual environments*, The social life of avatars (pp. 127–145). London, Springer.

[b-Bosse] Bosse, S., Brunnström K., Arndt S, Martini MG, Ramzan N, Engelke U (2019), *A common framework for the evaluation of psychophysiological visual quality assessment*. Quality and User Experience, Vol. 4, No. 1, pp. 1–10. doi:10.1007/s41233-019-0025-5.

[b-Bozgeyikli] Bozgeyikli, E., Raij, A., Katkoori, S., and Dubey, R. (2016), *Point and teleport locomotion technique for virtual reality*, Proceedings of the ACM Annual Symposium on Computer-Human Interaction in Play, Austin, TX, USA, pp. 205–216.

[b-Brunnström] Brunnström, K., Beker, S. A., de Moor, K., Dooms, A., Egger, S., et al. (2013), *Qualinet White Paper on Definitions of Quality of Experience*. https://hal.archives-ouvertes.fr/hal-00977812 [b-Brunnström] Brunnström, K. et al. (2020), *Latency impact on Quality of Experience in a virtual reality simulator for remote control of machines*. Signal Processing: Image Communication, Vol. 89, Elsevier, https://doi.org/10.1016/j.image.2020.116005.

[b-Beatty] Beatty, J. (1982), *Task-evoked pupillary responses, processing load, and the structure of processing resources*. Psychological Bulletin, Vol. 91, pp. 276–229.

[b-Cahill] Cahill, T.J. (2018), *Dimensions of congruity in immersive virtual environments: A framework for the schematic processing of multimodal sensory cues*, 18th Conference of the International Society for Presence Research: International Society for Presence Research. http://matthewlombard.com/ISPR/Proceedings/2018/P2018-Cahill.pdf

[b-Cao] Cao, A., Chintamani, K.K., Pandya, A.K., and Ellis, R.D. (2009), *NASA TLX: Software for assessing subjective mental workload*. Behavior Research Methods, Vol. 41, No. 1, 113–117. doi: 10.3758/BRM.41.1.113.

[b-Caserman] Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S., and Steinmetz, R. (2019), *Real-time body tracking in virtual reality using a Vive tracker*, Virtual Reality, Vol 23, pp. 155–168. https://doi.org/10.1007/s10055-018-0374-z.

[b-Cortés] Cortés, C., Pérez, P., Gutiérrez, J., and García, N. (2020, May), *Influence of video delay on quality, presence, and sickness in viewport adaptive immersive streaming*, 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1-4, IEEE. https://doi.org/10.1109/QoMEX48832.2020.9123114

[b-Diel] Diel, A., Weigelt, S., and Macdorman, K.F. (2021), *A meta-analysis of the uncanny valley's independent and dependent Variables*, Journal of Human-Robot Interaction, Vol. 11, No. 1, Article 1. doi: 10.1145/3470742

[b-Döring] Döring, N., De Moor, K., Fiedler, M., Schoenenberg, K., and Raake, A. (2022), *Videoconference fatigue: A conceptual analysis*, International Journal of Environmental Research and Public Health, Vol. 19, No. 4. https://www.mdpi.com/1660-4601/19/4/2061

[b-Duzmanska] Duzmanska, N., Strojny, P., and Strojny, A. (2018), *Can simulator sickness be avoided? A review on temporal aspects of simulator sickness*. Frontiers in Psychology, Vol. 9, pp. 1–14. https://doi.org/10.3389/fpsyg.2018.02132

[b-Egan] Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., and Murray, N. (2016), *An evaluation of heart rate and electrodermal activity as an objective QoE evaluation method for immersive virtual reality environments*, Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, June 2016, pp. 1–6. doi: 10.1109/QoMEX.2016.7498964

[b-Elwardy] Elwardy, M., Zepernick, H.-J., and Hu, Y. (2021), *On head movements in repeated 360° video quality assessment for standing and seated viewing on head-mounted displays*, Proceedings of the IEEE Virtual Reality – Workshop on Seated VR and Embodiment, Lisbon, Portugal.

[b-Engelke] Engelke, U., Darcy, D. P., Mulliken, G. H., Bosse, S., Martini, M. G., Arndt, S., Antons, J.-N., Chan, K. Y., Ramzan, and Brunnström, K. (2017), *Psychophysiology-based QoE assessment: A survey*. IEEE Journal of Selected Topics in Signal Processing, Vol. 11, No. 1, 6–20.

[b-Ennis] Ennis, C., and O'Sullivan, C. (2012), *Perceptually plausible formations for virtual conversers*, Computer Animation and Virtual Worlds, Vol. 23, No. 3–4), pp. 321–329. doi: 10.1002/cav.1453

[b-Essaili] Essaili, A., Thorson, S., Jude, A., Ewert, J., Tyudina, N., and Caltenco, H. (2022), *Holographic communication in 5G networks*. https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/holographic-communication-in-5g-networks

[b-Fermoselle] Fermoselle, L., Gunkel, S., Haar, F.t.t., Dijkstra-Soudarissanane, S., Toet, A., Niamut, O., et al. (2020), *Let's get in touch! Adding haptics to social VR*, ACM International Conference on Interactive Media Experiences. Association for Computing Machinery, pp. 174–179. doi: 10.1145/3391614.3399396

[b-Fiedler] Fiedler, M., Zepernick, H.-J., and Kelkkanen, V. (2019), *Network-induced temporal disturbances in virtual reality applications*, Proceedings of the 11th International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, June 2019.

[b-Gavrilovska] Gavrilovska, L., Rakovic, V., and Dixit, S. (2017), *General Concepts Behind Human Bond Communication*, Human Bond Communication: The Holy Grail of Holistic Communication and Immersive Experience. Hoboken, NJ, Wiley.

[b-Georgiou] Georgiou, Y., and Kyza, E.A. (2017), *The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings*, International Journal of Human-Computer Studies, Vol. 98, 24–37. doi: 10.1016/j.ijhcs.2016.09.014

[b-Gilbert] Gilbert, S.B. (2016), *Perceived realism of virtual environments depends on authenticity*, Presence: Teleoperators and Virtual Environments, Vol. 24, No. 4, pp. 322–324. doi: 10.1162/PRES_a_00276

[b-Gonçalves] Gonçalves, G., Melo, M., Vasconcelos-Raposo, J., and Bessa, M. (2020), *Impact of different sensory stimuli on presence in credible virtual environments*, IEEE Transactions on Visualization and Computer Graphics, Vol. 26, No. 11, pp. 3231–3240. doi: 10.1109/TVCG.2019.2926978

[b-Gonzalez-Franco] Gonzalez-Franco, M., and Peck, T. C. (2018), *Avatar embodiment. towards a standardized questionnaire*, Frontiers in Robotics and AI, Vol. 5, No. 74. doi: 10.3389/frobt.2018.00074

[b-Gonzalez-Sosa] Gonzalez-Sosa, E., Perez, P., Tolosana, R., Kachach, R., and Villegas, A. (2020), *Enhanced self-perception in mixed reality: Egocentric arm segmentation and database with automatic labeling*. IEEE Access, Vol. 8, pp. 146887–146900. doi: 10.1109/ACCESS.2020.3013016 [b-Greenwald] Greenwald, S. W., Wang, Z., Funk, M., and Maes, P. (2017), *Investigating social presence and communication with embodied avatars in room-scale virtual reality*, International Conference on Immersive Learning, Cham, Switzerland, Springer, pp. 75–90.

[b-Gunkel] Gunkel, S., Stokking, H., Prins, M., Niamut, O., Siahaan, E., and Cesar, P. (2018), *Experiencing virtual reality together: Social VR use case study*, Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video (pp. 233–238). doi: 10.1145/3210825.3213566

[b-Hacker] Hacker, J., vom Brocke, J., Handali, J., Otto, M., and Schneider, J. (2020), *Virtually in this together – how web-conferencing systems enabled a new virtual togetherness during the COVID-19 crisis*. European Journal of Information Systems, Vol. 29, No. 5, pp. 563-584. doi: 10.1080/0960085X.2020.1814680

[b-Hägni] Hägni, K., Eng, K., Hepp-Reymond, M.-C., Holper, L., Keisker, B., Siekierka, E., and Kiper, D.C. (2007), *The effect of task and ownership on time estimation in virtual environments*, Presence 2007: 10th Annual International Workshop on Presence, The International Society for Presence Research (ISPR), pp. 145–150.

[b-Haritaipan] Haritaipan, L., Hayashi, M., and Mougenot, C. (2018), *Design of a massage-inspired haptic device for interpersonal connection in long-distance communication*, Advances in Human-Computer Interaction. doi: 10.1155/2018/5853474

[b-Harms] Harms, C., and Biocca, F. (2004), *Internal consistency and reliability of the Networked Minds measure of social presence*, M. Alcaniz, and B. Rey (Eds.), Seventh Annual International Workshop: Presence 2004, Valencia, Spain, Universidad Politecnica de Valencia, pp. 246–251.

[b-Hart] Hart, S.G., and Staveland, L.E. (1988), *Development of NASA-TLX (task load index): Results of empirical and theoretical research*, P.A. Hancock, and N. Meshkati (Eds.), Human mental workload, Amsterdam, Elsevier, pp. 139–183.

[b-Hershey] Hershey, J., Movellan, J. (1999), *Audio vision: Using audio-visual synchrony to locate sounds*. Advances in Neural Information Processing Systems.
https://proceedings.neurips.cc/paper/1999/file/b618c3210e934362ac261db280128c22-Paper.pdf

[b-Hofer] Hofer, M., Wirth, W., Kuehne, R., Schramm, H., and Sacau, A. (2012), *Structural equation modeling of spatial presence: the influence of cognitive processes and traits*, Media Psychology, Vol. 15, No. 4, 373–395. doi: 10.1080/15213269.2012.723118

[b-Hofer] Hofer, M., Hartmann, T., Eden, A., Ratan, R., and Hahn, L. (2020), *The role of plausibility in the experience of spatial presence in virtual environments*, Frontiers in Virtual Reality, Vol. 1, Art. 2. doi: 10.1080/15213269.2012.723118

[b-Hu] Hu, Y., Elwardy, M., Zepernick, H.-J. (2021), *On the effect of standing and seated viewing of* 360^{o} *videos on subjective quality assessment: A pilot study*. Computers, Vol. 10, No. 80, pp. 1–28.

[b-Huisman] Huisman, G. (2017), *Social touch technology: A survey of haptic technology for social touch*, IEEE Transactions on Haptics, Vol. 10, No. 3, pp. 391–408. doi: 10.1109/TOH.2017.2650221

[b-Iskander] Iskander, J., Hossny, M., and Nahavandi, S. (2018), *A review on ocular biomechanic models for assessing visual fatigue in virtual reality*, IEEE Access, Vol. 6, pp. 19345–19361. https://doi.org/10.1109/ACCESS.2018.2815663

[b-Jekosch] Jekosch, U (2005), Voice and Speech Quality Perception – *Assessment and Evaluation*. Berlin, Springer.

[b-Jennet] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008), *Measuring and defining the experience of immersion in games*, International Journal of Human Computer Studies, Vol. 66, No. 9, 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

[b-Jewitt] Jewitt, C., Price, S., Mackley, K.L., Yiannoutsou, N., and Atkinson, D. (2020), *Interdisciplinary insights for digital touch communication*, Cham, Switzerland, Springer Open Publishing.

[b-Jo-1] Jo, D., Kim, K., and Kim, G. J. (2017), *Effects of avatar and background ypes on Users' Co-presence and Trust for Mixed Reality-Based Teleconference Systems*, 30th Conference on Computer Animation and Social Agents, Seoul, pp. 27–36.

[b-Jo-2] Jo, D., Kim, K., Welch, G. F., Jeon, W., Kim, Y., Kim, K. H., and Kim, G. J. (2017), *The impact of avatar-owner visual similarity on body ownership in immersive virtual reality*, Proceedings of the ACM Symposium on Virtual Reality Software and Technology, New York, NY, USA, pp. 1–2. https://doi.org/10.1145/3139131.3141214

[b-Jurnet-1] Jurnet, I.A., Beciu, C.C., and Maldonado, J.G. (2005), *Individual differences in the sense of presence*, Proceedings of Presence 2005: The 8th International Workshop on Presence, London, University College London, pp. 133–142).

[b-Jurnet-2] Jurnet, I.A., Beciu, C.C., and Maldonado, J.G. (2005), *Individual differences in the sense of presence*. Paper presented at the Eighth Annual International Workshop on Presence, London.

[b-Kammerl] Kammerl, J., Blodow, N., Rusu, R. B., Gedikli, S., Beetz, M., and Steinbach, E. (2012), *Real-time compression of point cloud streams*, 2012 IEEE International Conference on Robotics and Automation, pp. 778-785. doi: 10.1109/ICRA.2012.6224647

[b-Keighrey] Keighrey, C., Flynn, R., Murray, S. and Murray, N. *A physiology-based QoE comparison of interactive augmented reality, virtual reality and tablet-based applications*, IEEE Transactions on Multimedia, Vol. 23, pp. 333-341. doi: 10.1109/TMM.2020.2982046

[b-Kelkkanen-1] Kelkkanen, V., Fiedler, M., and Lindero, D. (2020), *Bitrate requirements of non-panoramic VR remote rendering*, Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, pp. 3624–3631. https://doi.org/10.1145/3394171.3413681 [b-Kelkkanen-2] Kelkkanen, V., Fiedler, M., and Lindero, D. (2021), *Remapping of Hidden Area Mesh Pixels for Codec Speed-up in Remote VR*, Proceedings of the 13th International Conference on Quality of Multimedia Experience (QoMEX), Montreal, Canada, June 2021.

[b-Kelkkanen-3] Kelkkanen, V., Fiedler, M., and Lindero, D. (2021), *Synchronous remote rendering for VR*, The International Journal of Computer Games Technology, Vol. 2021, Art. ID 6676644. https://doi.org/10.1155/2021/6676644

[b-Kennedy] Kennedy, R.S., Lane, N.E., Berbaum, K.S., and Lilienthal, M.G. (1993), *Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness*, The International Journal of Aviation Psychology, Vol. 3, No. 3, pp. 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[b-Khenak] Khenak, N., Vézien, J., and Bourdot, P. (2020), *Spatial presence, performance, and behavior between real, remote, and virtual immersive environments*, IEEE Transactions on Visualization and Computer Graphics, Vol. 26, No. 12, pp. 3467–3478. doi: 10.1109/TVCG.2020.3023574

[b-Kim] H. K. Kim, J. Park, Y. Choi, and M. Choe (2018), *Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment*, Applied Ergonomics, Vol. 69, pp. 66–73.

[b-Krogfoss] Krogfoss, B., Duran, J., Perez, P., and Bouwen, J. (2020), *Quantifying the value of 5G and edge cloud on QoE for AR/VR*, 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–4. https://doi.org/10.1109/QoMEX48832.2020.9123090

[b-Kroupi] Kroupi, E., Hanhart, P., Lee, J-S., Rerabek, M. and Ebrahimi, T. (2014), *EEG correlates during video quality perception*, Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, September 2014, pp. 2135–2139.

[b-Kruijff] Kruijff, E., Swan II, J. E., and Feiner, S. (2010), *Perceptual issues in augmented reality revisited*, IEEE International Symposium on Mixed and Augmented Reality, Seoul (pp. 3–12). https://doi.org/10.1159/000150000

[b-Laarni] Laarni, J., Ravaja, N., Saari, T., Böcking, S., Hartmann, T., and Schramm, H. (2015), *Ways to measure spatial presence: Review and Future Directions*, M. Lombard, F. Biocca, J. Freeman, W. Ijsselsteijn, and R. J. Schaevitz (Eds.), Immersed in Media: Telepresence Theory, Measurement and Technology, Cham, Switzerland, Springer International Publishing, pp. 139–185 . https://doi.org/10.1007/978-3-319-10190-3

[b-Latoschik] Latoschik, M. E., and Wienrich, C. (2021), *Coherence and plausibility, not presence?! Pivotal conditions for XR experiences and effects, a novel model*. https://arxiv.org/abs/2104.04846v2

[b-Lee-1] Lee, H. (2020), *A conceptual model of immersive experience in extended reality*. PsyArXiv. doi: 10.31234/osf.io/sefkh [b-Lee-2] Lee, J., Kim, M., and Kim, J. (2020), *RoleVR: Multi-experience in immersive virtual reality between co-located HMD and non-HMD users*. Multimedia Tools and Applications, Vol. 79, No. 1, 979-1005. doi: 10.1007/s11042-019-08220-w

[b-Lessiter] Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J.D. (2001), *A cross-media presence questionnaire: the ITC Sense of Presence Inventory*. Presence: Teleoperators and Virtual Environments, Vol. 10, No. 3, 282–297.

[b-Li] Li, J., Kong, Y., Röggla, T., De Simone, F., Ananthanarayan, S., De Ridder, H., El Ali, A., and Cesar, P. (2019), *Measuring and understanding photo sharing experiences in social virtual reality*, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14.

[b-Lindau] Lindau, A., and Weinzierl, S. (2012), *Assessing the plausibility of virtual acoustic environments*, Acta Acustica united with Acustica, Vol. 98, No. 5, pp.804–810. doi: 10.3813/AAA.918562

[b-Lombard-1] Lombard, M., and Ditton, T. (1997), *At the heart of it all: the concept of presence*, Journal of Computer-Mediated Communications, Vol. 3, No. 2, Art. JCMC321. doi: 10.1111/j.1083-6101.1997.tb00072.x

[b-Lombard-2] Lombard, M., Reich, R.D., Grabe, M.E., Campanella, C.M., and Ditton, T.B. (2000), *Presence and television: the role of screen size*, Human Communication Research: Proceedings of the 1995 Annual Conference of the International Communication Association, Vol 26, pp. 75–98. doi: 10.1111/j.1468-2958.2000.tb00750.x

[b-Longa] Della Longa, L., Valori, I., and Farroni, T. (2022), *Interpersonal affective touch in a virtual world: Feeling the social presence of others to overcome loneliness*. Frontiers in Psychology, Vol 12. doi: 10.3389/fpsyg.2021.795283

[b-Makransky] Makransky, G., Lilleholt, L., and Aaby, A. (2017), *Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach*, Computers in Human Behavior, Vol. 72, pp. 276-285. doi: 10.1016/j.chb.2017.02.066

[b-McGrath] McGrath, J.E. (1984), *Groups: Interaction and performance*. Englewood Cliffs, NJ, Prentice-Hall.

[b-Mengis] Mengis, J. and Eppler, M.J., (2008), *Understanding and managing conversations from a knowledge perspective: An analysis of the roles and rules of face-to-face conversations in organizations*, Organization Studies Vol. 29, No. 10, pp. 1287–1313. doi: 10.1177/0170840607086553

[b-Minhas] Minhas, T.N. and Fiedler, M. (2013), *Quality of Experience Hourglass Model*, in Proc. IEEE ComManTel, Ho Chi Minh City, Vietnam, Jan. 2013, DOI: 10.1109/ComManTel.2013.6482371 [b-Mekuria] Mekuria, R., Blom, K., and Cesar, P. (2016), *Design, implementation, and evaluation of a point cloud codec for tele-immersive video*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 27, No. 4, pp. 828–842. doi: 10.1109/TCSVT.2016.2543039

[b-Möller] Sebastian Möller, S., Garcia, M-N., Wältermann, M. (2014), *Features of quality of experience*, Quality of Experience, Advanced Concepts, Applications, Methods, Eds. Möller, S., Raake, A., Cham, Switzerland, Springer, pp. 73–84.

[b-Nadler] Nadler, R. (2020), *Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication*, Computers and Composition, Vol. 58, pp. 102613.

[b-Neustaedter] Neustaedter, C., Singhal, S., Pan, R., Heshmat, Y., Forghani, A., and Tang, J. (2018), *From being there to watching: Shared and dedicated telepresence robot usage at academic conferences*, ACM Transactions on Computer-Human Interaction, Vol. 25, No. 6. https://doi.org/10.1145/3243213

[b-Nguyen-Vo] Nguyen-Vo, T., Riecke, B. E., Stuerzlinger, W., Pham, D. M., and Kruijff, E. (2021), *NaviBoard and NaviChair: Limited translation combined with full rotation for efficient virtual locomotion*, IEEE Transactions on Visualization and Computer Graphics, Vol. 27, No. 1, pp. 165–177.

[b-Nilsson] Nilsson, N. C., Nordahl, R., and Serafin, S. (2016), *Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence*, Human Technology, Vol. 12, No. 2, 108–134. https://doi.org/10.17011/ht/urn.201611174652

[b-Norman] Norman, K.L. (2018), *Evaluation of virtual reality games: simulator sickness and human factors*, 2nd Workshop on Games-Human Interaction (GHItaly 2018), Castiglione della Pescaia, Grosseto, Italy.

[b-Nystad] Nystad, E., and Sebok, A. (2004), *A comparison of two presence measures based on experimental results*, Presence, International Society for Presence Research, pp. 266–273.

[b-O'Brien] O'Brien, H.L., and Toms, E.G. (2010), *The development and evaluation of a survey to measure user engagement*, Journal of the American Society for Information Science and Technology, Vol. 61, No. 1, 50–69. doi: 10.1002/asi.21229

[b-Okura] Okura, F., Kanbara, M., and Yokoya, N. (2010), *Augmented telepresence using autopilot airship and omni-directional camera*, 9th IEEE International Symposium on Mixed and Augmented Reality, Seoul, pp. 259–260. https://doi.org/10.1109/ISMAR.2010.5643596

[b-OpenXR1] Khronos Group (2021), *The OpenXR specification*. https://www.khronos.org/registry/OpenXR/specs/1.0/html/xrspec.html

[b-Orduna] Orduna, M., Perez, P., Gutierrez, J., and Garcia, N. (2022), *Methodology to assess quality, presence, empathy, attitude, and attention in 360-degree videos for immersive communications*, IEEE Transactions on Affective Computing. doi: 10.1109/TAFFC.2022.3149162

[b-Orts-Escolano]	Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P.A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S.B., Kohli, P., Lutchyn, Y., Keskin, C., and Izadi, S. (2016), *Holoportation: Virtual 3D teleportation in real-time*, The 29th Annual Symposium on User Interface Software and Technology, Association for Computing Machinery, pp. 741–754. doi: 10.1145/2984511.2984517
[b-Parola]	Parola, M., Johnson, S., and West, R. (2016), *Turning presence inside-out: Metanarratives*, Electronic Imaging, Vol. 4, pp. 1–9. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-418
[b-Perez]	P. Pérez, N. Oyaga, J. J. Ruiz, and A. Villegas (2018), *Towards systematic analysis of cybersickness in high motion omnidirectional video*, Proceedings of the International Conference on Quality of Multimedia Experience (QoMEX).
[b-Perkis]	Perkis, A., Timmerer, C., et al. (2020), QUALINET White Paper on Definitions of Immersive Media Experience (IMEx), European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online). https://arxiv.org/abs/2007.07032
[b-Porcu]	Porcu, S., Floris, A., and Atzori, L. (2019), *Towards the evaluation of the effects of ambient illumination and noise on quality of experience*, 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6.
[b-Rantala]	Rantala, J., Salminen, K., Raisamo, R., and Surakka, V. (2013), *Touch gestures in communicating emotional intention via vibrotactile stimulation*, International Journal of Human-Computer Studies, Vol. 7, No. 6, pp. 679–690. doi: 10.1016/j.ijhcs.2013.02.004
[b-Rauschnabel]	Rauschnabel, P.A., Felix, R., Hinsch, C., Shahab, H., and Alt, F. (2022), *What is XR? Towards a framework for augmented and virtual reality*, Computers in Human Behavior, Vol. 133, Art. 107289. doi: https://doi.org/10.1016/j.chb.2022.107289
[b-Reiter]	Reiter, U., Brunnström, K., de Moor, K., et al. (2014), *Factors Influencing Quality of Experience*, Quality of Experience - Advanced Concepts, Applications, Methods. Eds S. Möller and A. Raake, New York, Springer, pp. 55–72.
[b-Sacau]	Sacau, A., Laarni, J., and Hartmann, T. (2008), *Influence of individual factors on presence*, Computers in Human Behavior, Vol. 24, No. 5, pp. 2255–2273. doi: 10.1016/j.chb.2007.11.001
[b-Schubert]	Schubert, T., Friedmann, F., and Regenbrecht, H. (2001), *The experience of presence: Factor analytic insights*, Presence: Teleoperators and Virtual Environments, Vol. 10, No. 3, pp. 266-281. https://doi.org/10.1162/105474601300343603

[b-Schwarz] Schwarz, S., Preda, M., Baroncini, V., Budagavi, M., Cesar, P., Chou, P. A., Cohen, R. A., Krivokuća, M., Lasserre, S., Li, Z., Llach, J., Mammou, K., Mekuria, R., Nakagami, O., Siahaan, E., Tabatabai, A., Tourapis, A. M., and Zakharchenko, V. (2018), *Emerging MPEG standards for point cloud compression*, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol. 9, No. 1, pp. 133–148. https://doi.org/10.1109/JETCAS.2018.2885981

[b-Shah] Shah, S.G.S., Nogueras, D., van Woerden, H.C., and Kiparoglou, V. (2020), *The COVID-19 pandemic: A pandemic of lockdown loneliness and the role of digital technology*, Journal of Medical Internet Research, Vol. 22, No. 11, pp. e22287. https://doi.org/10.2196/22287

[b-Sharan] Sharan, N.N., Toet, A., Mioch, T., Niamut, O., and van Erp, J.B.F. (2022), *The relative importance of social cues in immersive mediated communication*, IHIET 2021, Human Interaction, Emerging Technologies and Future Systems V, Cham, Switzerland, Springer International Publishing, pp. 491–498. https://doi.org/10.1007/978-3-030-85540-6_62

[b-Sharma] Sharma, V. (2022), *Introducing a personal boundary for Horizon worlds and venues*. https://www.oculus.com/blog/introducing-a-personal-boundary-for-horizon-worlds-and-venues/

[b-Shimizu] T. Shimizu, T. and Onaga, H. (2018), *Study on acoustic improvements by sound-absorbing panels and acoustical quality assessment of teleconference systems*, Applied Acoustics, Vol. 139, pp. 101–112. doi: https://doi.org/10.1016/j.apacoust.2018.04.021

[b-Singla-1] Singla, A., Fremerey, S., Robitza, W., Raake, A. (2017), *Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays*, 9th International Conference on Quality of Multimedia Experience (QoMEX) Erfurt, Germany, pp. 1–6.https://doi.org/10.1109/QoMEX.2017.7965658

[b-Singla-2] Singla, A., Ramachandra Rao, R. R., Goering, S., Raake, A. (2019), *Assessing media QoE, simulator sickness and presence for omnidirectional videos with different test protocols*, IEEE Conference on Virtual Reality and 3D User Interfaces, Osaka, Japan, pp. 1163–1164. https://doi.org/10.1109/VR.2019.8798291

[b-Singla-3] Singla, A., Robitza, W., and Raake, A. (2018), *Comparison of subjective quality evaluation methods for omnidirectional videos with DSIS and modified ACR*, Electron. Imag., Hum. Vis. Electron. Imag., Vol. 2018, No. 14, pp. 1–6.

[b-Skarbez] Skarbez, R., Brooks, F., Whitton, M. (2021), *Immersion and coherence: Research agenda and early results*, IEEE Transactions on Visualization and Computer Graphics, Vol. 27, No. 10, pp. 3839–3850. doi: 10.1109/TVCG.2020.2983701

[b-Skowronek-1] Skowronek, J. Schiffner, F. and Raake, A. (2013), *On the influence of involvement on the quality of multiparty conferencing*, 4th International Workshop on Perceptual Quality of Systems (PQS 2013), International Speech Communication Association, Vienna, Austria, Sept. 2013, pp. 141–146.

[b-Skowronek-2] Skowronek, J. (2017), *Quality of experience of multiparty conferencing and telemeeting systems – methods and models for assessment and prediction*, PhD Thesis, Berlin Institute of Technology, Technische Universität Berlin. http://dx.doi.org/10.14279/depositonce-5811

[b-Skowronek-3] Skowronek, J., Raake, A., Berndtsson, G., Rummukainen, O., Usai, P., Gunkel, S., Johansson, M., Habets, E., Malfait, L., Lindero, D., Toet, A. (2022), *Quality of experience in telemeetings and videoconferencing: A comprehensive survey*. https://ieeexplore.ieee.org/document/9777989

[b-Skowronek-4] Skowronek, J., and Raake, A. (2015, *Assessment of cognitive load, speech communication quality and quality of experience for spatial and non-spatial audio conferencing calls*, Speech Communication, Vol. 66, pp. 154–175, doi: 10.1016/j.specom.2014.10.003

[b-Slater-1] Slater, M. (2009), *Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments*, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 364, No. 1535, pp. 3549–3557. https://doi.org/10.1098/rstb.2009.0138

[b-Slater-2] Slater, M., Spanlang, B., Corominas, D. (2010), *Simulating virtual environments within virtual environments as the basis for a psychophysics of presence*, ACM SIGGRAPH, Vol. 29, No. 4, pp. 1–9. https://doi.org/10.1145/1778765.1778829

[b-Slater-3] Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., Watson, Z., Breen, G., Schwarz, R., Steptoe, W., Szostak, D., Halan, S., Fox, D., and Silver, J. (2020), *The ethics of realism in virtual and augmented reality*, Frontiers in Virtual Reality, Vol. 1, pp. 1–13. https://doi.org/10.3389/frvir.2020.00001

[b-Slater-4] Slater, M., Usoh, M., and Steed, A. (1994), *Depth of presence in virtual environments*. Presence: Teleoperators and Virtual Environments, Vol. 3, No. 2, pp. 130–144.

[b-Smith] Smith, J., and MacLean, K. (2007), *Communicating emotion through a haptic link: Design space and methodology*, International Journal of Human-Computer Studies, Vol. 65, No. 4, pp. 376–387. doi: 10.1016/j.ijhcs.2006.11.006

[b-Schmitt] Schmitt, M., Gunkel, S., Cesar, P. et al. (2014), *Asymmetric delay in video-mediated group discussions*, 2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX), 2014, pp. 19–24.

[b-Steed] Steed. A. and Schroeder, R. (2015), *Collaboration in immersive and non-immersive virtual environments*, M. Lombard, F. Biocca, J. Freeman, W. Ijsselsteijn, and R. J. Schaevitz (Eds.), Immersed in Media: Telepresence Theory, Measurement and Technology, Cham, Switzerland, Springer International Publishing Switzerland (pp. 263-282). https://doi.org/10.1007/978-3-319-10190-3

[b-Subramanyam-1] Subramanyam, S., Li, J., Viola, I., and Cesar, P. (2020, March), *Comparing the quality of highly realistic digital humans in 3DoF and 6DoF: A volumetric video case study*, 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 127–136.

[b-Subramanyam-2] Subramanyam, S., Viola, I., Jansen, J., Alexiou, E., Hanjalic, A., and Cesar, P. (2022), *Evaluating the impact of tiled user-adaptive real-time point cloud streaming on VR remote communication*. https://arxiv.org/abs/2205.04906

[b-Suvilehto] Suvilehto, J.T., Glerean, E., Dunbar, R.I.M., Hari, R., and Nummenmaa, L. (2015), *Topography of social touching depends on emotional bonds between humans*, Proceedings of the National Academy of Sciences, Vol. 112, No. 45, pp. 13811–13816. doi: 10.1073/pnas.1519231112

[b-Toet-1] Toet, A., van der Spek, E.D., and Houtkamp, J.M. (2010), *Cybersickness influences the affective appraisal of a virtual environment*. The Open Virtual Reality Journal, 2, 26-31. doi: 10.2174/1875323X01002010026.

[b-Toet-2] Toet, A., Houtkamp, J.M., and Vreugdenhil, P.E. (2016), *Effects of personal relevance and simulated darkness on the affective appraisal of a virtual environment*, PeerJ, Vol. 4, e1743, pp. 1–24. doi: 10.7717/peerj.1743

[b-Toet-3] Toet, A., Heijn, F., Brouwer, A.-M., Mioch, T., and van Erp, J.B.F. (2019), *The EmojiGrid as an immersive self-report tool for the affective assessment of 360 VR videos*, P. Bourdot, V. Interrante, L. Nedel, N. Magnenat-Thalmann, and G. Zachmann (Eds.), EuroVR 2019: Virtual Reality and Augmented Reality, LNCS 11883, Cham, Switzerland:Springer International Publishing pp. 330–335. doi: 10.1007/978-3-030-31908-3

[b-Toet-4] Toet, A., Mioch, T., Gunkel, S. N., Sallaberry, C., van Erp, J. B., and Niamut, O. (2020), *Holistic quality assessment of Mediated Immersive Multisensory Social Communication*, International Conference on Virtual Reality and Augmented Reality, Cham, Switzerland, Springer, pp. 209–215.

[b-Tran] Tran, H.T., Ngoc, N.P., Pham, C.T., Jung, Y.J., and Thang, T. C. (2017), *A subjective study on QoE of 360 video for VR communication*, 2017 IEEE 19th international workshop on multimedia signal processing (MMSP), pp. 1-6. https://doi.org/10.1109/MMSP.2017.8122249

[b-VanDam] Van Dam, L.C., and Stephens, J.R. (2018), *Effects of prolonged exposure to feedback delay on the qualitative subjective experience of virtual reality*. PloS One, Vol. 13, No. 10, e0205145. https://doi.org/10.1371/journal.pone.0205145

[b-Van Waveren] Van Waveren, J.M.P. (2016), *The asynchronous time warp for virtual reality on consumer hardware*, Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, pp. 37–46.

[b-Villegas] Villegas, A., Perez, P., Kachach, R., Pereira, F., and Gonzalez-Sosa, E. (2020), *Realistic training in VR using physical manipulation*, 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 109–118). https://doi.org/10.1109/VRW50115.2020.00025 [b-Vlahovic] Vlahovic, S., Suznjevic, M., and Skorin-Kapov, L. (2019), *The impact of network latency on gaming QoE for an FPS VR game*, 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), IEEE, pp. 1–3.

[b-Vorderer] Vorderer, P., Wirth, W., Gouveia, F.R., Biocca, F., Saari, T., Lutz Jäncke, Böcking, S., Schramm, H., Gysbers, A., Hartmann, T., Christoph Klimmt, Laarni, J., Ravaja, N., Sacau, A., Baumgartner, T., and Jäncke, P. (2004), *MEC Spatial Presence Questionnaire (MEC-SPQ): Short documentation and instructions for application*.

[b-VR Industry Forum] VR Industry Forum (2021), *Volumetric Video Guidelines v0.95*. https://www.vr-if.org/wp-content/uploads/vrif2021.059.02.-VolumetricVideoGuidelines-vo.95_pdraft.pdf

[b-Watson] Watson, B.A. and Hodges, L.F. (1995), *Using texture maps to correct for optical distortion in head-mounted displays*, IEEE Virtual Reality Annual International Symposium (VRAIS) 1995, pp. 172-178.

[b-Weber] Weber, S., Weibel, D., and Mast, F.W. (2021), *How to get there when you are there already? Defining presence in virtual reality and the importance of perceived realism*, Frontiers in Psychology, Vol. 12, p. 1538. https://doi.org/10.3389/fpsyg.2021.628298

[b-Weech] Weech, S., Kenny, S., and Barnett-Cowan, M. (2019), *Presence and cybersickness in virtual reality are negatively related: A review*, Frontiers in Psychology, Vol. 10, No. 158. doi: 10.3389/fpsyg.2019.00158

[b-Williamson] Williamson, J., Li, J., Vinayagamoorthy, V., Shamma, D.A., and Cesar, P. (2021), *Proxemics and social interactions in an instrumented virtual reality workshop*, The 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, Art. 253.

[b-Wirth] Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F.R., Biocca, F., Sacau, A., Jancke, L., Baumgartner, T., and Jäncke, T. (2007), *A process model of the formation of spatial presence experiences*, Media Psychology, Vol. 9, No. 3, pp. 493–525. doi: 10.1080/15213260701283079

[b-Witmer-1] Witmer, B.G., and Singer, M.J. (1998), *Measuring presence in virtual environments: A presence questionnaire*, Presence: Teleoperators and Virtual Environments, Vol. 7, pp. 225–240. https://doi.org/10.1162/105474698565686

[b-Witmer-2] Witmer, B.G., Jerome, C.J., and Singer, M.J. (2005), *The factor structure of the presence questionnaire*, Presence: Teleoperators and Virtual Environments, Vol. 14, No. 3, pp. 298–312. doi: 10.1162/105474605323384654

[b-Yee] Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., and Merget, D. (2007), *The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments*, CyberPsychology and Behavior, Vol. 10, No. 1, pp. 115–121.

[b-Yoon] Yoon, L., Yang, D., Kim, J., Chung, C., and Lee, S.H. (2022), *Placement retargeting of virtual avatars to dissimilar indoor environments*, IEEE Transactions on Visualization and Computer Graphics, Vol. 28, No. 3, pp. 1619–1633. doi: 10.1109/TVCG.2020.3018458

[b-Youngblut] Youngblut, C., and Huie, O. (2003), *The relationship between presence and performance in virtual environments: results of a VERTS study*, IEEE Virtual Reality 2003, pp. 277–278. doi: 10.1109/VR.2003.1191158

[b-Zhao] Zhao, S. (2003), *Toward a taxonomy of copresence*, Presence: Teleoperators and Virtual Environments, Vol. 12, No. 5, pp. 445-455. https://doi.org/10.1162/105474603322761261

[b-Zielasko] Zielasko, D., and Riecke, B. (2020), *Sitting vs. standing in VR: Towards a systematic classification of challenges and (dis)advantages*, Proceedings of the IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops, Atlanta, GA, USA, pp. 297–298.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2022