International Telecommunication Union

ITU-T G.1035

(11/2021)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Multimedia Quality of Service and performance – Generic and user-related aspects

Influencing factors on quality of experience for virtual reality services

Recommendation ITU-T G.1035

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	**G.1000–G.1999**
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

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

Recommendation ITU-T G.1035

Influencing factors on quality of experience for virtual reality services

Summary

Recommendation ITU-T G.1035 classifies virtual reality (VR) services and identifies the key quality of experience (QoE) factors of VR.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T G.1035	2020-05-29	12	11.1002/1000/14274
2.0	ITU-T G.1035	2021-11-29	12	11.1002/1000/14826

Keywords

Cybersickness, presence, quality of experience, QoE, simulator sickness, virtual reality, VR.

* 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

Recommendation ITU-T G.1035

Influencing factors on quality of experience for virtual reality services

1 Scope

This Recommendation categorizes and summarizes the factors affecting the user-perceived experience of a virtual reality (VR) service, with the intention of helping to identify the methodologies for assessing VR quality. VR quality of experience (QoE) assessment methodologies are left for further study. Since VR technologies are still evolving, this Recommendation mainly addresses omnidirectional video services, while leaving others, e.g., point-cloud or volumetric video types, for further study.

VR is a new type of media which is different from traditional video and audio. It generates realistic images, sounds and other sensations that replicate a real environment and simulate a user's physical presence in this environment by enabling the user to interact with this space and any objects depicted therein using specialized display screens or projectors and other devices. These multisensory experiences, which can include sight, hearing and, less commonly, touch and smell, are well coordinated and synchronized through the user's interaction and feedback. A person using VR equipment is typically able to "look around" the artificial world, move about within it and interact with features or items that are depicted on a screen or in goggles as in the real world.

In order to understand whether QoE or user-perceived performance of the VR service is good enough or not, benchmarking is critical. Benchmarking aims to measure user-perceived performance or QoE in the VR environment. Compared with traditional video and audio, the multisensory experience in VR imposes a new set of requirements for QoE assessment. The challenge is to characterize VR's real-life immersive video, spatial audio and interactivity. Prior to being able to benchmark the QoE, it is important to address the requirements and basic factors that are relevant for assessing the VR quality for different VR services.

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.

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 frame rate [b-ITU-T H.262]: The rate at which frames are output from the decoding process.

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

3.1.3 QoE influencing factors [b-ITU-T P.10]: Include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service

and their fulfilment, the user's cultural background, socio-economic issues, psychological profiles, emotional state of the user, and other factors whose number will likely expand with further research.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 cybersickness or simulator sickness: A physiological condition arising when exposed to a virtual reality environment.

NOTE – Definition based on combined information in [b-Kennedy] and [b-Stanney].

3.2.2 degree of freedom (DoF): Represents the ways an object can move within a space, which is a key element in helping create an immersive environment for a user.

3.2.3 immersion: A psychological state characterized by perceiving oneself to be enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences.

NOTE – Definition based on [b-Witmer].

3.2.4 motion-to-photon latency: The time it takes between the user moving their head and this motion being reflected on the screen of the head-mounted display (HMD).

NOTE – Definition based on [b-Brandenburg].

3.2.5 presence: The subjective experience of being in one place or environment, when one is physically situated in another place or environment.

NOTE – Definition based on [b-Witmer].

3.2.6 refresh rate: The frequency at which a display updates its image, expressed in hertz.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

2D Two Dimensional

3D Three Dimensional

AR Augmented Reality

DoF Degree of Freedom

FoV Field of View

GPU Graphic Processing Unit

HD High Definition

HMD Head-Mounted Display

HRTF Head-Related Transfer Function

MPEG Moving Picture Experts Group

MR Mixed Reality

PPD Pixel Per Degree

PPI Pixel Per Inch

SSQ Simulator Sickness Questionnaire

UHD Ultra High Definition

VR Virtual Reality

VoD Video on Demand

VVC Versatile Video Coding

5 Conventions

Within the scope of this Recommendation, a person interacting with a VR head-mounted display (HMD) is referred to as a user, viewer or player, all with equivalent inference.

6 Virtual reality overview

Virtual reality (VR) is a technology that uses game engines (e.g., Unity) to create artificial environments that enable people to interact in six degrees of freedom (DoF). VR generates realistic images, sounds and other sensations that emulate a real environment or create a synthetic one. VR services aim to provide users with high levels of immersion and presence wherein users may feel detached from their physical, real-world surroundings. This is different from augmented reality (AR) or mixed reality (MR), which enhance user experiences by adding virtual components such as digital images, graphics or sensations as a new layer of interaction with the real world.

6.1 Devices

A VR display device is usually a typical head-mounted display (HMD) with two goggle-size miniature screens – one per eye. These displays focus and reshape the picture for each eye and can create a stereoscopic three-dimensional (3D) image by angling the two dimensional (2D) images to mimic how human eyes see the world.

To achieve an immersive experience, head tracking or eye tracking are used in HMDs to create the correct camera angle and perspective so as to attain a natural viewing experience. In addition, tracking may include capturing of the movements of any other body parts.

Depending on the VR service and DoF permitted for the users, audio hardware needs to take into account headphone or loudspeaker audio reproduction. The headphones may be standalone or integrated into the HMD, both of which allow open and closed acoustic design. When using standalone headphones, additional hardware such as a soundcard or wireless technology may also be involved.

6.2 Content

There are two types of content to consider when constructing a virtual environment for the VR experience.

The first consideration is synthetic, which is completely invented from geometric primitives and simulated physics. This is common in VR games and VR social services. Synthetic representations of individuals, called avatars, enable users to interact and can provide a level of anonymity in some contexts.

The second consideration is captured using 360° cameras. For example, 360°/omnidirectional images and videos are captured to allow users to explore the scene in 360° × 180° from a given viewpoint in a VR system, which could be seen as an extension of traditional video streaming applications.

6.3 Platform

When using online VR applications, VR contents are stored in servers, streamed by requests and rendered locally in user's devices. Local rendering requires high-performance terminal devices to provide an acceptable user experience.

Cloud VR [b-GSMA AR/VR] is a new cloud computing technology, where VR content is stored and rendered in the server. Therefore, video and audio outputs are coded, compressed and transmitted to

user terminals. Servers in cloud VR are required to be as close as possible to end-users to reduce the influence introduced on QoE by the additional network processing time.

6.4 Services

VR services can be classified into two types: weak-interaction VR and strong-interaction VR [b-Huawei VR].

Weak-interaction VR services mainly comprise but are not limited to 360° video, VR theatre and VR live broadcast. In this kind of VR services, users can explore the scene by turning their head; however, they do not interact with the objects present in the scene. For example, touching the entities in the virtual world is not possible.

Strong-interaction VR services include VR games, VR home fitness, VR social networking, etc. Users can interact with these virtual environments through interactive entities (e.g., controllers) in addition to HMD head tracking.

Appendix I provides information for some typical VR services.

7 Virtual reality QoE influencing factors

QoE influencing factor categories are illustrated in Figure 1. (See also [b-Reiter].)

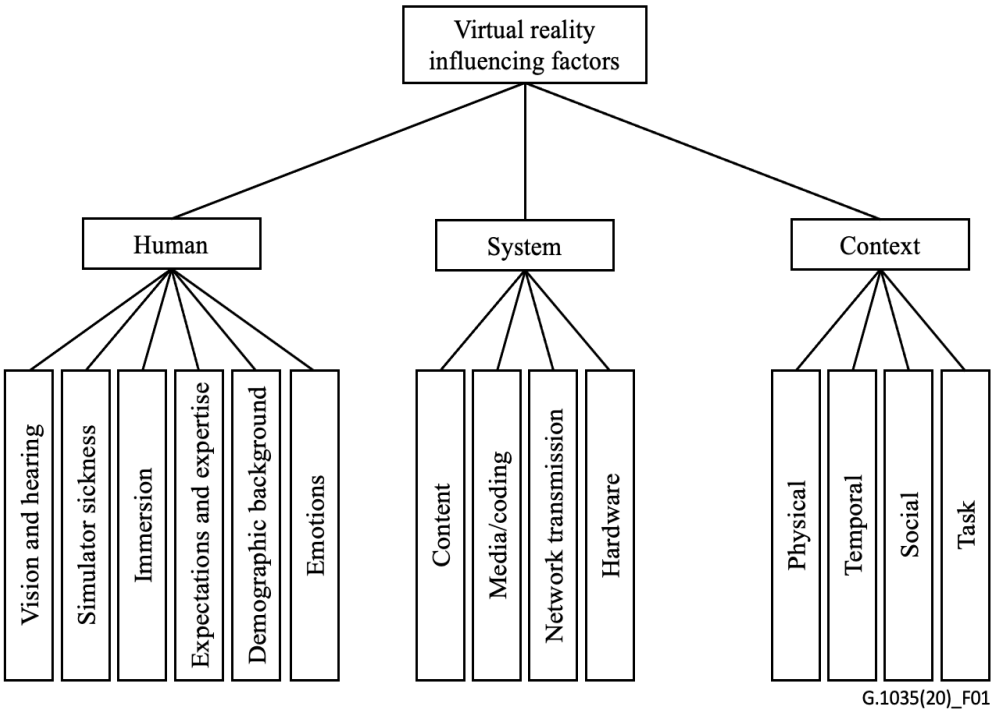

Figure 1 – Virtual reality QoE influencing factor categories

Furthermore, influencing factors can occur on different levels of magnitude (micro vs. macro), behaviour (static vs. dynamic), and patterns of occurrence (rhythmic vs. random), either independently or as usual mixtures of all three levels [b-Reiter]. That is because this is not a binary state; all factors can be static or dynamic, but the time-frequency of changes can determine whether a factor is dynamic or static.

7.1 Human influencing factors

7.1.1 Vision and hearing

Visual abnormalities such as longsightedness, shortsightedness, astigmatism and chromatic aberration may occur in the human eye. Each eye may also be affected differently. Such vision problems may negatively affect the user experience. When vision problems can be corrected by lenses, having the user wear their normal glasses may be a solution, although this may be uncomfortable for the user.

Hearing impairments may result in attenuated hearing over the full audible frequency range or at specific frequencies. An impairment may also occur asymmetrically in only the left or the right ear, which has consequences on spatial hearing. Loss of sensitivity at high frequencies is a normal age-related hearing impairment, but different types of impairments are present in populations of all ages. Often, individuals are not aware of having a hearing impairment as they develop over time and the auditory system adapts to the lowered sensitivity.

It is important to consider that each individual hears differently, and that audio reproduction has to take this into account in order to provide a good experience. Head-related transfer functions (HRTFs) describe the individual filtering process that occurs when sound travels from a point in space to the two ears. HRTFs depend on the shape of the pinna and ear canal, and the size and shape of the head and upper torso. They vary individually and to achieve perfect binaural reproduction the HRTFs should be individually measured and applied in the audio rendering process. In practice, generalized HRTFs are often used, resulting in possible degradation in sound localization.

7.1.2 Simulator sickness

Cybersickness is also known as simulator sickness, VR sickness or visually induced motion sickness and is triggered only by visual stimuli. This undesirable phenomenon is caused by the sensory conflict arising between the visual and vestibular system. While watching 360° videos in an HMD, a user may experience symptoms of simulator sickness such as fatigue, sweating, vertigo or nausea [b-Kennedy]. The most popular questionnaire for assessing simulator sickness is the simulator sickness questionnaire (SSQ), published in 1993.

Simulator sickness is an important factor that affects QoE for 360°/VR videos [b-Singla-1]. There are different factors such as resolution, audio, time, field of view (FoV), the orientation of users, HMD, player, type of video sequences, etc., by which simulator sickness scores can be affected. (These and other factors are discussed in clause 7.2.) When considering QoE, the video sequences that are noted to lead to the highest simulator sickness scores also lead to the lowest QoE scores. Inversely, those video sequences that have the lowest simulator sickness scores also have the highest QoE scores. These observations indicate that simulator sickness interacts with QoE when 360° videos are watched in HMDs [b-Singla-1].

Apart from technical factors, there are individual factors such as age, gender, postural stability, etc., contextual factors such as duration of task, orientation of users, etc. and co-variate constructs such as QoE, presence and exploration behaviour that can impact the severity of simulator sickness. The symptoms of simulator sickness can be reduced but cannot be eliminated for every user. If everything is performed correctly and the needs of the user are met, then the symptoms of simulator sickness will become motion sickness [b-Kopyt].

Simulator sickness can also be caused by vergence-accommodation conflict. This symptom occurs when the brain receives mismatching information between a focusing distance and the distance to the virtual object [b-George]. Vergence-accommodation conflict can cause eye strain, fatigue, focusing problems, etc.

7.1.3 Immersion

The tendency to experience immersion and the level of expertise in using VR systems vary individually. People who tend to be more immersed in a virtual environment may not notice small impairments in reproduction. How immersion affects VR QoE is for further study. Relevant investigations are continuing in ITU-T.

7.1.4 Expectations and expertise

The attitude towards VR creates varying contexts for experiencing it. Some people may dislike a VR experience regardless of its technical quality based on their beliefs and fears of using such systems. The level of expertise in using VR systems may affect how capable the users are in using the systems to achieve a certain goal, which in turn affects the QoE. Some users may be awed by the novel experience, while more experienced users can focus on the task at hand.

The influence of a subject's internal reference, provided by their interactions and experiences in the real world, will also influence the QoE of a VR service.

7.1.5 Demographic background

Demographic data is a statistical representation of variables such as age and gender, as well as socio-economic information such as employment, education, income, race, marital status and others. When it comes to VR services in general, depending on the demographic background, users could have different experiences and performances. Solutions for VR services in a number of areas (from gaming over telemeeting, e-learning and sports) have been researched in order to determine which user attributes have a considerable influence on the experience.

Demographic factors such as age is a user factor that influences one's preference for VR technology. When it comes to how likely individuals are to interact with technology in general [b-Franke], younger participants had a higher affinity for technology, including a comparable inclination for VR services [b-Kojic-1]. Also, gender is another influencing factor where significant differences in technological affinity were reported [b-Franke], in particular for VR [b-Kojic-2]. Further on, there is a gender gap reported while exploring presence in VR services. Significant differences were identified on presence subscales (considering the binary gender scale) [b-Felnhofer], where men reported a different feeling presence than women for VR services.

7.1.6 Emotions

The increasing availability of VR equipment and advances in technology have inspired its use in a variety of research sectors, including human emotional states and responses to VR services. At the level of human emotional states, the effect of moods and emotions on QoE has been researched [b-Felnhofer]; although both have a relatively brief duration, moods typically last longer (varying from hours to days) than emotions (ranging from seconds to minutes) [b-Reiter]. Overall, emotions can have an impact on QoE in general including for VR services in the same way as a stimulus that triggers (among other things) an emotional response in the recipient; secondly, a stimulus may generate an emotional response in the recipient as a result of its meaning [b-Schleicher].

Studies have shown that content and interaction possibilities with VR environments can change arousal and valence levels. Further on, for example, the possibility to interact with the virtual environment can impact various quality and emotion related parameter such as involvement and spatial presence [b-Voigt-Antons-2]. Furthermore, when implementing the most immersive presentation method (such as head-mounted display), 360° video stimuli might be used to evoke the intended emotional reactions in participants with a greater sense of presence [b-Voigt-Antons-1].

7.2 System influencing factors

7.2.1 Content related factors

VR content is crucial for the user's experience and has additional requirements when compared with traditional multimedia content. In addition to the requirements for good quality of video and audio, VR content also requires stitching, special effects, stereoscopic 3D and composition. To ensure an immersive experience, it is important that the VR content is generated at a good quality and then delivered as perfectly as possible. This clause lists the aspects related to VR content which will influence the quality of a VR service.

7.2.1.1 Spatial audio

Spatial audio involves the use of spatial audio reproduction techniques that are loudspeaker- or headphone-based such as multiloudspeaker stereophonic approaches, generalizations such as vector-base amplitude panning, sound-field synthesis (higher-order ambisonics, wave-field-synthesis), headphone-based binaural, or combinations thereof. By using spatial audio, virtual sound sources can be created at any point within the 3D space. Sound reproduction for VR is most often done via headphones, but loudspeaker setups are also possible especially in three DoF and six DoF scenarios. In addition to the direct sound, spatial audio can further include the auditory spatial impression of the room acoustics (e.g., early reflections and reverberation), perceptually plausible acoustic effects of sound sources being occluded by structures in the VR world (e.g., attenuation and diffraction), and sound radiation patterns of individual audio objects. Spatial audio is an important aspect for creating the illusion of immersion for VR services. QoE-related aspects are, in particular, the perceived coloration and spaciousness (including features such as source width, envelopment, locatedness) of the (virtual) sources, as well as additional artefacts due to coding or other types of processing.

7.2.1.2 Spatial depth (3D)

It also is possible to playout stereoscopic video content, addressing human binocular vision, which is based on the depth-dependent disparity stemming from the two slightly different views presented to the two eyes. This allows humans to judge distance and have a perception of depth. To avoid crosstalk effects, the left view of the content has to be displayed only to the left eye and the right view of the content has to be displayed only to the right eye (see [b-Woods]). In the system design of most HMD systems, which are commonly used for playing out VR content, both views are already separated from each other.

Encoding the stereoscopic representation with a low bitrate has to be avoided as it decreases the perceived quality even more than using non-stereoscopic content with the same bitrate. The quality advantage of the 3D over the 2D representation is only slightly visible for higher bitrates. Thus, a relatively high bitrate should be used for encoding to assure that the advantage of stereoscopic over non-stereoscopic omnidirectional video content becomes visible. It has to be mentioned that this is strongly dependent on the general stereoscopic quality of the video content. Simulator sickness scores and how a stereoscopic representation influences the VR QoE are for further study.

Note that this factor is not mandatory in VR. Many VR services use 2D content while still allowing people to feel immersed due to the rendered omnidirectional scenes.

7.2.1.3 Spatiotemporal complexity

Spatial perceptual information indicates the complexity of a video picture. With some high complexity content, the viewer may be distracted, while with less complexity content, the subject may be more focused on the main objects. Temporal perceptual information indicates the amount of changing of the video picture. Different spatial and temporal complexity video sequences require a different amount of bandwidth. For example, sequences with higher spatial and temporal information generally require higher bandwidth.

Spatiotemporal complexity is a content feature which may affect VR QoE. For example, if the VR content has high temporal values, it may produce a high amount of simulator sickness which may affect QoE [b-Singla-1].

7.2.2 Media/codec related factors

7.2.2.1 Compression

Video/audio codecs are used to compress original scene data from raw format so that they can be saved offline or streamed via a network, saving bandwidth and resources. Various codecs have been developed in the industry and are widely used for traditional media coding and these codecs may be used for VR media. However, some codecs are unsuitable for certain scene representations.

7.2.2.2 Video

Traditional video codecs (e.g., [b-ITU-T H.264], [b-ITU-T H.265], VP8, VP9) may also be used for VR content, but may not be suitable for certain spatial representations (e.g., point-clouds).

These different codec technologies are based on different compression implementations, and each can cause different information loss when the encoder compresses the raw data, and the decoder renders it back for display. This results in different perceived quality of the experience and decoding speed.

A common shortcoming for traditional video codecs being used in VR is that the compression rate is still too low. This means that the bandwidth consumption is still a big problem for streaming the VR content when dealing with full VR streaming. New video coding technologies in progress such as versatile video coding (VVC) will significantly improve the transport quality for VR content in the industry.

In order to save bandwidth and network resources for 360°/VR videos, many streaming service providers propose to transmit only pixels in the users' FoV in high resolution, and the remaining pixels in minimal quality resolution. This technique is called viewport-adaptive or tile-based streaming of omnidirectional video. The tile-based adaptive streaming architecture is shown in Figure 2. The two main strategies of tile-based streaming are described in Appendix II.

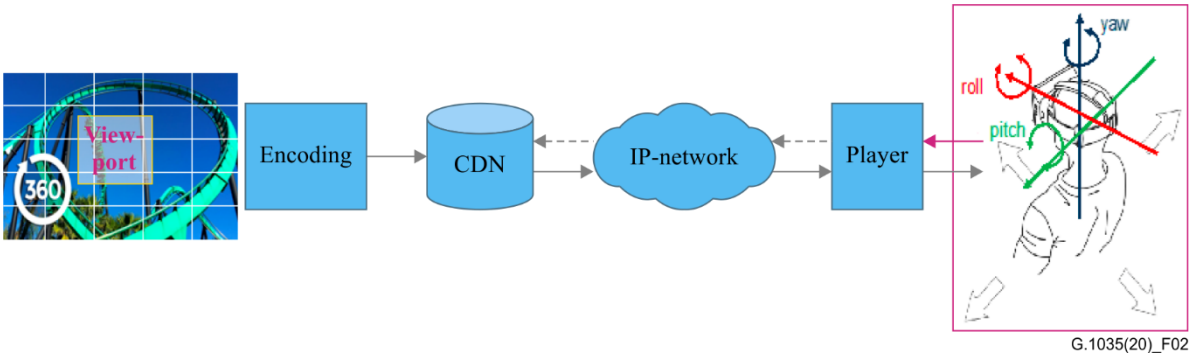

Figure 2 – Tile-based adaptive streaming – behaviour dependent

7.2.2.3 Audio

For non-interactive three DoF and six DoF VR scenes, requirements on audio metadata are typically consistent with current 3D audio content. Where a categorical number of static positions can be authored, additional data for user head rotation should be incorporated to render a spatial auditory scene consistent with listener movements. For interactive and six DoF VR services where an infinite number of source and listener positions are available, translating user movements and geometric data of the auditory environment is essential for setting the requirements for an audio codec. Direct sound, early reflections and late reverberation should be accounted for and be coherent with all sound sources and listener(s) movements as they may influence the QoE.

The production of audio content for VR may also feature various approaches such as channel-based, object-based or ambisonics. These may be recorded via multichannel microphones or single microphones, with both techniques requiring various formatting in order to implement and deliver immersive audio. An audio codec would need to be adaptive in the input data stream of content, as seen in the current moving picture experts group (MPEG) H-series codec for three DoF scenes, along with a consistent input stream of user actions.

7.2.2.4 Storage and transport

In 2016, a new approach was released to encode 360° videos with a pyramid geometry, which saves almost 80% of the bitrate [b-Kuzyakov]. In pyramid projection, the base of the pyramid is always available in the full resolution and the sides of the pyramid decrease gradually in quality until all the sides meet at a point. When a user changes the viewing direction, it is decided which stream should be fetched based on the network condition and the orientation of the user.

Besides that, other technologies (e.g., tile-based streaming) are also used in VR to reduce bandwidth and resource consumption. MPEG has now developed a standard called omnidirectional media format [b-ISO/IEC 23090-2] which intends to standardize the storage and transmission of VR content, mainly for 360° videos. There are multiple media profiles supported, one of which divides the entire 360° video into independently coded tiles that the HMD has to recompose into the image from the tiles required according to the user's viewing directions.

7.2.2.5 Bitrate

Bitrate is the number of audio or video bits that are conveyed or processed per unit of time. Bitrate serves as a more general indicator of quality. Higher resolution, higher frame rates and lower compression usually lead to an increased bitrate under the same encoding environment.

7.2.2.6 Resolution

Video resolution represents the number of distinct pixels contained in the video content that can be displayed in each dimension. Resolution of a video should be compatible with the resolution of the display device, otherwise the video resolution might have to be reduced or cannot even be displayed.

Higher resolution for VR is required as compared with 2D video in order to have similar visual quality because pixels are spread in a 360° viewing sphere around the viewer. Depending upon the field of view of the HMD, a viewer sees around 25% of the total pixels. For example, if a 4K video is shown to a user in an HMD, it would appear as if the user were watching a 1K video. In order to provide a 4K experience to the viewers, 16K video should be displayed.

7.2.2.7 Frame rate

Frame rate indicates the frequency at which consecutive images, called frames, are displayed. For improving the QoE, the frame rate of the VR content should be exactly the same as the refresh rate of the HMD's display. Playing back the content in a frame rate not matching the panel's refresh rate leads to artefacts such as frame fluctuation, frame drops and frame manipulation using black frame insertion. These artefacts mostly lead to jerkiness, which leads to a lower QoE [b-Hofmeyer].

The frame rate in VR services has higher requirements than normal 2D video services because jerkiness in the motion may lead to simulator sickness in the VR environment. Frame rate is even more demanding for VR gaming applications where scenes are rendered by a graphic processing unit (GPU) instead of those created by video cameras.

In the area of 360° videos, applying motion interpolation to contents with a lower frame rate than the HMD's display refresh rate is a suitable method for increasing the QoE. This especially applies to videos with a higher amount of motion. (See [b-Hofmeyer] and [b-Fremerey].)

7.2.2.8 Audio sample rate

The sample rate is the number of samples of audio carried per second, measured in hertz (Hz). In VR services, this factor has no difference from the traditional streaming services.

7.2.2.9 Coding delay

As the codec standards only describe the algorithmic decoding procedure and profile features, there are still many options in selecting coding modes and parameters when designing a system. VR related applications typically require low and even extremely low delay. Therefore, how to effectively reduce the coding delay which contributes to the final end to end delay should be considered. Extremely low coding delays will also satisfy the need to synchronize audio and video presentation.

Perceptual thresholds exist for television (TV) broadcasting (e.g., [b-ITU-T J.248]), but VR brings new challenges due to the immersive experience and sensorimotor coupling in six DoF where additionally synchronizing the rendered content with self-movement is essential. Perceptual thresholds in six DoF scenarios are topics for future study.

7.2.3 Network/transmission related factors

Network/transmission related factors only exist in online VR services.

7.2.3.1 Delay

In VR environments, stringent latency requirements are of utmost importance for providing a pleasant immersive VR experience. Delay includes queuing delay, over-the-air delay and buffering delay. Delay is usually the main reason for high motion-to-photon latency leading to simulator sickness (see clause 7.1.2). It is also the cause of presentation quality degradation, e.g., long initial loading delay and stalling. Some VR services may offload computing tasks, such as rendering capability, to remote cloud servers to significantly relieve the computing burden from the user's HMDs, which is at the expense of incurring additional communication delay.

The influence of resolution, bandwidth and network round-trip delay on QoE aspects of tile-based streaming of 360° videos was studied in [b-Singla-2]. These experimental results showed that lowering delay by up to 50 ms has a minimal effect on QoE ratings. The video quality degrades significantly for higher values (>100 ms) of delay. The effect of network delay on the simulator sickness scores cannot be seen. This may be explained by the fact that the background is always visible in the low resolution and always moves consistently with head motion [b-Singla-2].

7.2.3.2 Bandwidth

Immersive experience with VR streaming application requires a lot of data. If the required bandwidth is not guaranteed for specific VR applications, the quality of the content will be degraded during network transmission. That is, congestion can cause long delays and packet loss which can then degrade the perceived immersive QoE of the VR system.

7.2.3.3 Loss

The impact of packet loss on the VR experience depends on the method of transmission. In reliable transmission protocols, packet loss incurs packet retransmissions which increase the overall delay. With unreliable transmission, packet loss may result in the loss of parts of frames or entire frames and thus degrade audiovisual quality, which may be presented as phenomena such as video freezing and tiling artefacts.

7.2.4 Hardware related

Hardware plays an important role in creating an immersive experience for users. VR hardware comprises HMDs, headphones, haptic feedback devices, input controllers and tracking systems with various possibilities to bring real-world objects into the VR domain.

7.2.4.1 Head-mounted display

Unlike traditional terminal devices, HMD wearing comfort may also greatly impact the final VR QoE. To improve this, it is important to consider device weight, size, heat dissipation, resolution, refresh rate, etc.

7.2.4.2 Headphones

The frequency response of headphones, when used, is a factor that affects QoE. Neutral headphones or headphones whose frequency response have been compensated may be able to better convey the spatial audio experience of the listener. Additionally, the ability to block outside noise may be of importance for VR, with closed headphones or in-ear headphones being best suited for this. However, the use of such headphones that entirely block outside noise may cause the user to feel disoriented in their real environment and may be a safety hazard.

7.2.4.3 Decoder performance

The decoder capability has an impact on the overall resolution of the video to be transmitted and decoded in the display device, whether high definition (HD) or ultra high definition (UHD), and thus decides the final resolution of the video that could be displayed to the user. In addition, codec support, e.g., [b-ITU-T H.264] or [b-ITU-T H.265], is also important since different codecs have different decoding performances.

The codecs supported by the decoder should be compatible with the encoder; otherwise, the content may not be displayed correctly.

The number of decoders determines the number of streams the device is capable of decoding, e.g., if streams are separately encoded when tiled streaming is applied.

Compared with those implemented in software, decoders implemented in hardware have much faster performance when decoding the same content, which contribute less delay to VR media processing and lead to a better QoE.

Some decoders may also have some error correction mechanisms that are capable of fixing errors that may occur during transport or encoding. This can also increase the QoE of the final VR experience.

7.2.4.4 Head tracking

To enable interaction between users and the environment, it is important to obtain the positions and motion information of users. This is usually done, for example, by the inertial measurement unit implemented inside the HMD. Typical devices use combinations of accelerometers, gyroscopes and sometimes magnetometers to track objects' motions.

There are two tracking technologies that are used thus far: outside-in and inside-out. Outside-in tracking indicates that the headset and accessories rely on some external devices, e.g., a lighthouse sensor or computer display. It has more accuracy and better latency than an inside-out unit but is limited by the environment. Inside-out tracking does not rely on external devices. It uses the HMD sensor to determine how the position is changing in relation to the external environment.

Low head-tracking latency and back tracking accuracy are certainly important attributes to provide a smooth change of view for the user, while long head-tracking latency induces discomfort and loss of immersive experience.

7.2.4.5 Field of view

FoV is the extent of the observable environment at any given time. With a wider FoV, a user is more likely to feel at-the-scene in the experience. FoV is the solid angle that is visible by a human through the HMD lenses. When it comes to VR FoV, the limiting factor is the lenses, not the pupils. To get a better FoV, the user can either move closer to the lenses, as is the case with VR HMD lenses, or increase the size of the lenses [b-VR Lens Lab].

While a wide FoV can increase immersion, it can also more easily cause simulator sickness to certain individuals. This is mainly because some people are sensitive to the flickers or movements of images, and also because the large visual input brought from large FoV may cause conflicts with the vestibular and proprioceptive systems.

Therefore, FoV is an important parameter that helps evaluate to what extent a VR device could create an immersive experience.

7.2.4.6 Display resolution

Display resolution is a basic attribute of the screen that indicates the pixels per inch that a screen supports. An appropriate screen resolution, relative to the resolution of viewpoint shown in the HMD, would provide the best and most comfortable experience.

Pixel per degree (PPD) is a core technology specification that is better suited for measuring the pixel density of a VR near-eye display rather than the more traditional pixel per inch (PPI) value. Typically, the higher the PPD, the better of the image quality will be. A lower PPD may result in the screen-door effect.

7.2.4.7 Refresh rate

The refresh rate is the number of times per second that a display grabs a new image from the GPU. A lower refresh rate can contribute to increased processing latency and lead to VR sickness, i.e., viewing glitches on the screen.

7.3 Context influencing factors

Context influencing factors are related to the setting or situational property which influences a user's environment in terms of direct influences on the signals presented to the user (audio, video, etc.), the goals connected to a certain system usage and the impact on the user’s expectations.

7.3.1 Physical context

Physical context factors are related to the environment where a user is experiencing the VR services. Background noises may affect the user's experience. In addition, the experience of the user may differ depending on whether the HMD device is wireless or connected to a stationary processing device, e.g., a PC, which could restrict movement and the possibilities to explore a VR scene. Room lighting may not affect users' experience as much when using HMDs as compared with traditional video environments since the devices are close to their eyes. The ambient temperature of the room in which the user is engaged with the VR service affects the QoE. Normally, the room temperature is set to a typical room temperature to which the user is accustomed, so that they will feel comfortable enough not to get distracted. However, in some cases, the room temperature can be modulated to approximate the virtual world in order to achieve the best immersive experience. For example, if a user is immersed in a VR skiing scene, they may have a better sense of 'being there' when the room temperature is cold enough to make them feel as if they are in the real world, as simulated in the VR scene.

The amount of sunlight entering the environment can also affect VR QoE. Only a low amount or ideally no direct sunlight should enter the environment. The infrared light could influence the performance of the HMD's tracking system, which could lead to picture outages and other errors.

Furthermore, the respective safety features of the provided HMD system (such as virtual walls or a pass-through mode using cameras) should be activated to avoid any collisions of the VR user with their physical environment. Ideally, a second person who is not participating in the VR services should pay attention to the physical actions of the user consuming the VR service.

7.3.2 Temporal context

Temporal context factors include the frequency and duration of use. A VR device may not be able to support long usage periods. Over time, simulator sickness symptoms such as dizziness, loss of spatial

awareness, nausea and eye soreness typically become worse as the duration of use increases. These effects will greatly reduce the QoE.

7.3.3 Social context

Social context factors include considerations such as VR content popularity and how VR services are consumed (i.e., alone or in a group). A user may be affected by the interaction with a group of other people, e.g., their family, friends or even strangers. For example, co-located co-viewing or co-playing may increase a user's overall satisfaction with a program. This may also hold true for VR services, especially for social VR, in which people use a virtual reality platform to form synthetic societies which contain avatars connected to real people to simulate the physical world. To what extent this factor affects the overall QoE of VR services requires further investigation.

7.3.4 Task context

VR experience depends on the goals of the user of the VR service. These factors are called task context factors. For example, if the task is formal, the participants may pay specific attention to some aspects of perceived influence, while they may ignore such experience when doing a relatively casual task. Additionally, the QoE for streaming-type VR, e.g., 360° VR, would be quite different from gaming VR or social VR. For the former, users may have less tolerance towards video impairments. For the latter, users may have less tolerance towards bad interaction experience.

Appendix I

Virtual reality services use cases

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

Generally, VR applications can be divided into two types, online and offline:

– Online VR: VR applications of this type work either partially or primarily through the Internet or another computer network. In this case, VR content is streamed from a server at the time when the user is using it. Obviously, any network delay occurring in this type of VR application may affect the experience of users. However, it can save the local storage of VR terminal devices and expand the range of content that the user can experience.

– Offline VR: VR applications of this type work offline. To do this, users need to download the VR content completely to their devices in advance. While running these applications there are usually no network delay issues and no requirement for network bandwidth. However, the content that the user can experience is limited by the capacity of a local storage device. Offline VR services are not in the study scope of this Recommendation.

I.1 Use scenario

There are many different types of VR services. Five types are listed in this clause.

I.1.1 Live

Live VR is broadcast in real time, as events happen, in the present. The difference between the traditional live program and live VR is that live VR is panoramic and interactive. Live VR can provide an immersive experience of attending the live event at the event venue. The live 360° VR described in [b-ITU-T H.430.3] is a service of this type, which constructs 360° panoramic view in real time via multiple cameras feeds from the site. Users can observe the live event with proper VR HMDs that constantly process and stitch multiple images to project the real world.

Live VR services require extremely stringent delay so that users can smoothly change the viewpoints when watching it. Live VR services also require UHD resolutions to make users feel as if they are in the real venue. The bandwidth consumption issue is very challenging when a massive number of users consume the live VR service at the same time. Interaction is also an important issue, although in 360° VR, only a few actions can be taken, for example, turning your head around.

I.1.2 Video on demand

Video on demand (VoD) VR services allow users to select and experience the content at any preferable time of their choice rather than a specific broadcast time. Live and VoD VR share the same experience. The only difference is that the content of VoD VR is prepared in advance rather than in real time. The typical usage could be that some applications offered by some major over the top (OTT) providers allow users to watch the entire environment in every scene.

VoD VR services have the same requirements as live VR for delay issues and video resolutions. The bandwidth consumption is relatively smaller than live VR as viewers can consume the same content at different times.

Many VR applications in different industries can be seen as VoD VR services. For example, some applications present a user with a cinematic experience with HMD at home and some applications use VR for education, but they are all basically VoD VR services.

I.1.3 Gaming

VR gaming services allow a user to experience being in a 3D virtual entertainment environment through an avatar and interact with the environment during the game. VR gaming services may

require more devices other than an HMD. One example is a data glove with small sensors that can capture the movements made by the user, which are then interpreted by computers and trigger a variety of responses within that space.

VR game services require extremely sensitive interaction to reach the best experience. Also, the immersive experience of "being there" where "there" is not equivalent to the position of one's own body but the place the VR content suggests is what a VR game seeks. Other aspects of gaming discussed in [b-ITU-T G.1032] should also be considered.

I.1.4 Social

VR social is a service that allows users of the VR platform to form synthetic societies which contain avatars connected to real people to simulate the physical world. A typical example would be the Facebook of VR which provides new social VR features for Oculus Rift. Users can create a custom avatar based on photos from their profile and spend time with other people in a virtual space.

Like a VR game, VR social also requires extremely sensitive interaction so that users feel as if they are in a real world. Non-synchronized movement of these synthetic avatars with actual human motion will result in a very bad experience for users of the services.

I.1.5 Shopping

In addition to the VR services listed in this clause, there are other applications which may be promising in the future when using VR devices. For example, VR shopping could allow users to purchase items through a VR headset by virtually transporting themselves to international retail outlets, enabling them to experience the entire shopping experience from finding products to payment. VR shopping is similar to VR VoD, which records the content in advance, but requires more interaction and less data consumption than VoD streaming VR.

Appendix II

Tile-based streaming

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

In tile-based streaming, only the field of view is transmitted in the highest quality and rest of the video is transmitted in lower quality. Based on the client, this can be divided mainly into two categories:

1) Full delivery basic: In this method, the base layer is always available in the low resolution and high-resolution tiles are only available for the current field of view/viewport. When a user changes the viewing direction, the player software finds which of the tiles are in current viewport and fetch those tiles from the network [b-Brandenburg].

2) Full delivery advanced: In this method, the tiles that belongs to the user's viewport are sent in higher quality. In addition, the user head motion is continuously predicted to find where the user's viewport will be in future. These corresponding tiles are requested in a higher quality at the time when it is expected that the user will move into a specific direction [b-Mario].

Bibliography

[b-ITU-T G.1032] Recommendation ITU-T G.1032 (2017), *Influence factors on gaming quality of experience*.

[b-ITU-T H.262] Recommendation ITU-T H.262 (2012), *Information technology – Generic coding of moving pictures and associated audio information: Video*.

[b-ITU-T H.264] Recommendation ITU-T H.264 (2019), *Advanced video coding for generic audiovisual services*.

[b-ITU-T H.265] Recommendation ITU-T H.265 (2021), *High efficiency video coding*.

[b-ITU-T H.430.3] Recommendation ITU-T H.430.3 (2018), *Service scenario of immersive live experience (ILE)*.

[b-ITU-T J.248] Recommendation ITU-T J.248 (2008), *Requirements for operational monitoring of video-to-audio delay in the distribution of television programs*.

[b-ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience*.

[b-ISO/IEC 23090-2] ISO/IEC 23090-2 (2019), *Information technology – Coded representation of immersive media – Part 2: Omnidirectional media format*.

[b-Brandenburg] R. van Brandenburg, R. Koenen, and D. Sztykman (2017). *CDN optimization for VR streaming*. <https://www.ibc.org/cdn-optimisation-for-vr-streaming-/2457.article>

[b-Felnhofer] Felnhofer, A., *et al*. (2012), *Is virtual reality made for men only? Exploring gender differences in the sense of presence*. Proceedings of the International Society on Presence Research, pp. 103–112.

[b-Franke] Franke, T., Attig, C., and Wessel, D. (2019), *A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale*. International Journal of Human–Computer Interaction Vol. 35, No. 6, 456–467. <https://doi.org/10.1080/10447318.2018.1456150>

[b-Fremerey] Fremerey, S., Hofmeyer, F., et al. (2019). *Impact of Various Motion Interpolation Algorithms on 360° Video QoE*. 11th International Conference on Quality of Multimedia Experience (QoMEX 2019), IEEE Signal Processing Society.

[b-George] Koulieris, G., Bui, B., et al. (2017). *Accommodation and comfort in head-mounted displays*. ACM Transactions on Graphics, Vol. 36, No. 4, Article 87, July. <https://doi.org/10.1145/3072959.3073622>

[b-GSMA AR/VR] GSMA (2019). *Cloud AR/VR Streaming: Accelerate mass adoption and improve quality of experience of AR/VR using 5G and edge cloud*. GSMA Mobile World Congress, booklet. <https://www.gsma.com/futurenetworks/wp-content/uploads/2019/03/Cloud-ARVR-booklet-for-MWC19.pdf>

[b-Hofmeyer] Hofmeyer, F., Fremerey, S., et al. (2019). *Impacts of internal HMD playback processing on subjective quality perception.* Society for Imaging Science and Technology, International Symposium on Electronic Imaging 2019, pp. 219-1-291-7. <https://doi.org/10.2352/ISSN.2470-1173.2019.12.HVEI-219>

[b-Huawei VR] iLab (2019). *Cloud VR – Service Quality Monitoring White Paper*. Huawei iLab. <https://www.huawei.com/minisite/static/Cloud_VR_Service_QM_WhitePaper.pdf>

[b-Kennedy] Kennedy, R. S., Lane, N. E., et al. (1993). *Simulator sickness questionnaire: An enhanced method of quantifying simulator sickness.* The International Journal of Aviation Psychology, Vol. 3, No. 3, pp. 203–220. <https://www.tandfonline.com/doi/abs/10.1207/s15327108ijap0303_3>

[b-Kojic-1] Kojic, Tanja, et al (2010), *Influence of UI complexity and positioning on user experience during VR exergames*. Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE. <https://doi.org/10.1109/QoMEX.2019.8743273>

[b-Kojic-2] Kojić, Tanja, et al (2020), *Exploring visualisations for financial statements in virtual reality*. IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). IEEE. <https://doi.org/10.1109/AIVR50618.2020.00018>

[b-Kopyt] Koyt, A., Narkiewicz, J. (2013). *Technical factors influencing simulator sickness*. Zeszyty Naukowe Politechniki Rzeszowskiej. Mechanika Vol. 85 [288], No. 4 (2013), pp. 455–467.

[b-Kuzyakov] Kuzyakov, E., Pio, D. (2016). *Next-generation video encoding techniques for 360 video and VR*. Blogpost, January. <https://engineering.fb.com/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr/>

[b-Mario] Graf, M., Timmerer, C., Mueller, C. (2017). *Towards bandwidth efficient adaptive streaming of omnidirectional video over http: Design, implementation, and evaluation*. Proceedings of the 8th ACM on Multimedia Systems Conference. pp. 261–271. <https://dl.acm.org/doi/abs/10.1145/3083187.3084016>

[b-Reiter] Reiter, U., et al. (2014). *Factors Influencing Quality of Experience*. Quality of Experience. T-Labs Series in Telecommunication Services. Springer, Cham, Switzerland. pp. 55–72. <https://doi.org/10.1007/978-3-319-02681-7_4>

[b-Schleicher] Schleicher, R., and Antons, J.-N. (2014) *Evoking emotions and evaluating emotional impact*. Quality of Experience. Springer, Cham, Switzerland. 121–132. <https://doi.org/10.1007/978-3-319-02681-7_9>

[b-Singla-1] Singla, A., Fremerey, S., et al. (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 2017). pp. 1–6. <https://doi.org/10.1109/QoMEX.2017.7965658>

[b-Singla-2] Singla, A., Göring, S., et al. (2019). *Subjective quality evaluation of tile-based streaming for omnidirectional videos*. Proceedings of the 10th ACM Multimedia Systems Conference (MMSys '19). Association for Computing Machinery, New York, NY, USA. pp. 232–242. <https://doi.org/10.1145/3304109.3306218>

[b-Stanney] Stanney, K. M., Kennedy, R. S., Drexler, J. M. (1997). *Cybersickness is Not Simulator Sickness*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 41, No. 2. pp. 1138–1142.

[b-Voigt-Antons-1] Voigt-Antons, Jan-Niklas, et al. (2020), *Comparing emotional states induced by 360° videos via head-mounted display and computer screen*. Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE. <https://doi.org/10.1109/QoMEX48832.2020.9123125>

[b-Voigt-Antons-2] Voigt-Antons, Jan-Niklas, et al (2021), *Don't Worry be Happy - Using virtual environments to induce emotional states measured by subjective scales and heart rate parameters*. IEEE Virtual Reality and 3D User Interfaces (VR). IEEE. <https://doi.org/10.1109/VR50410.2021.00094>

[b-VR Lens Lab] VR Lens Lab (2016). *Field of View for Virtual Reality Headsets Explained*. Website. <https://vr-lens-lab.com/field-of-view-for-virtual-reality-headsets/>

[b-Witmer] Witmer, B. G., Singer, M. J. (1998). *Measuring Presence in Virtual Environments: A Presence Questionnaire*.

[b-Woods] Woods, A. (2010). *Understanding crosstalk in stereoscopic displays*. Keynote Presentation at the Three-Dimensional Systems and Applications Conference, Tokyo, Japan.

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