|  |  |  |  |
| --- | --- | --- | --- |
| **Question(s):** | 13 | **Meeting:** | Stockholm, 02-04 September, 2019 |
| **Study Group:** | 12 | **Working Party:** | 3 |
| **Source:** | Editor of G.QoE-VR |
| **Title:**  | The baseline of G.QoE-VR |
| **Purpose:** | [Choose a purpose from the dropdown list] |
| **Contact:** | Rachel HuangHuaweiChina | Tel: + 18114717307Fax: + E-mail: rachel.huang@huawei.com |

|  |  |
| --- | --- |
| **Keywords:** | VR, QoE, Factor, Simulator Sickness, Cybersickness, Presence |
| **Abstract:** | This is the new baseline of G.QoE-VR |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Recommendation ITU-T <No.>

Influencing Factors on Quality of Experience (QoE) for Virtual Reality Services

Summary

This draft Recommendation classifies virtual reality services and to identify the QoE key factors of VR.

Keywords

VR, QoE, Factor, Simulator Sickness, Cybersickness, Presence

# 1 Scope

Virtual Reality (VR) is a new type of media different from traditional video and audio. It generates realistic images, sounds and other sensations that replicate a real environment, and simulates 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. The multi-sensory 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 virtual reality equipment is typically able to "look around" the artificial world, move about in 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 or not, benchmarking is critical, which aims to measure user-perceived performance or QoE in that environment. Compared with traditional video and audio, the multi-sensory experience in VR imposes a new set of requirements to QoE assessment. The challenge is to characterize VR’s real-life immersive video, spatial-audio, and interactivity. Before we are able to benchmark the QoE, it’s important to address the requirements and basic factors assessing the VR quality for different VR services.

This draft Recommendation identifies different VR services and their respective requirements for Quality of Experience (QoE). This document also summarizes the key factors affecting user-perceived experience of a VR service, which can help to identify the methodologies for assessing the VR quality.

This Recommendation specifies the categorization of factors influencing VR QoE. The VR QoE assessment methodologies are left for further study.

# 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.

[3GPP TR 26.918] “Virtual Reality (VR) media services over 3GPP”, 3GPP TR 26.918 v15.0.0, September 2018

[ITU-T H.ILE-SS] TD 11-WP3/16, January 2017 *Service scenario of immersive live experience (ILE)*.

 [MPEG-I Part 1] N16918 “Working Draft 0.2 of Technical Report on Immersive Media”, MPEG
 118, April 2017

# 3 Definitions

## 3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

QoE [ITU-T P.910/G.100]

QoE influencing factors [ITU-T P.910/G.100]

TBD

## 3.2 Terms defined in this Recommendation

**Degree of Freedom**

DOF represents the ways an object can move within a space, which is a key element that helps create immersive environment for the user.

**Immersion**

It is 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 [Witmer].

**Presence**

Sense of being there [Witmer].

**Cybersickness / Simulator Sickness**

Cybersickness / Simulator Sickness is a physiological condition arising when exposed to a virtual reality environment [Kennedy, Stanney].

**Motion-to-Photon Latency**

The time it takes between the user moving their head and this motion being reflected on the screen of the HMD [Brandenburg].

# 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

|  |  |
| --- | --- |
| VR | Virtual Reality |
| HMDSSQ | Head mounted displaySimulator Sickness Questionnaire  |

# 5 Conventions

TBD

# 6 Virtual Reality Overview

Virtual Reality (VR) is a technology that uses certain devices to create artificial environments which makes people interact with them while having little or no awareness of the interference. It 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 in a reality completely detached from their physical, real-world surroundings. This is different from Augmented Reality (AR) or Mixed Reality (MR), which enhances 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 3D image by angling the two 2D images to mimic how human eyes see the world.

Achieving an immersive Head tracking or eye tracking is used in HMDs to create the right camera angle and perspective so as to reach a natural viewing experience. In addition to that, tracking may also include button presses, or capturing of the movements of any other body parts.

Depending on the VR service and degrees of freedom permitted for the users, audio hardware should consider headphone or loudspeaker reproduction. The headphones may be standalone or integrated into the HMD, both allowing open and closed acoustic design. When using standalone headphones additional hardware such as a soundcard or wireless technology may be involved.

## 6.2 Content

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

One 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 ourselves called avatars enable us to interact and provide a level of anonymity if desired in some contexts.

The other is captured using imaging techniques. For example, panoramic images and videos are captured to allow users to see them from any viewpoint in a VR system, which could be seen as an extension of traditional video streaming application.

These types of content can be mixed in the same VR experience.

## 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 devices to provide an acceptable user experience.

Cloud VR is a new clouding computing technology, of which VR content is stored and rendered in the cloud. 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 QoE influence introduced by the additional network processing time.

## 6.4 Services

VR services can be classified into two types: weak-interaction VR and strong-interaction VR.

Weak-interaction VR services mainly comprise but not limited to 360-degree panoramic video, VR theatre, and VR live broadcast. User can select the view and location by moving the head but do not interact with, e.g. touch, entities in the virtual world.

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

Appendix A provides information for some typical VR services.

# 7 Virtual Reality QoE Influence Factors

The QoE influence factor categories (see also [Reiter]) are illustrated in the Fig. 1.



Figure 1 Virtual Reality QoE Influence Factor Categories

##  Human influence factors

## Vision and hearing

Visual abnormalities, like farsightedness, nearsightedness, astigmatism, chromatic aberration, may occur in the human eye. Each eye may be affected differently. Such vision problems may negatively affect the user experience. In case that the vision problems can be corrected by lenses, having the user wear their normal glasses, or using an HMD with adjustable lenses may be a solution. The former may be uncomfortable and could reduce the field of view.

Hearing impairments may result in attenuated hearing over the full audible frequency range or at specific frequencies. The 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 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 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.

## Simulator sickness

Cybersickness is also known as Simulator Sickness or VR sickness or Visually Induced Motion Sickness and triggered only by visual stimuli. It is an undesirable phenomenon that is caused by the sensory conflict arising between visual and vestibular system. While watching the 360° videos in HMD, users may experience symptoms of simulator sickness such as fatigue, sweating, vertigo, nausea, etc. [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 [Singla]. There are different factors such as resolution, audio, time, Field of View, the orientation of users, HMD, player, type of video sequences, etc. by which simulator sickness scores can get affected. For example, some sequences which lead to the highest simulator sickness scores and the lowest QoE scores. Inversely, those sequences which have the lowest simulator sickness scores and the highest QoE. These observations indicate that simulator sickness interacts with QoE when 360° videos are watched in HMDs [Singla].

Although, it is unrealistic to eliminate the simulator sickness totally for everyone, VR service providers still need to consider it as one of the most important factors for VR QoE and trying to reduce the uncomfortable symptoms as much as possible. For example, trying to reduce the delay between motion and the rendering of the corresponding sensory information, improve the screen refreshing rate, or keep the sensory information as synchronized as possible.

## Immersion

Tendency to experience immersion and level of expertise in using VR systems varies individually. People with a high tendency to be immersed in a story, for example, may more likely accept small impairments in reproduction. How immersion affects VR QoE is still for further study. Relevant investigations will be done in the work item P.360-VR.

## Expectations and Expertise

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. 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 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.

##  System influence factors

## Content related

VR content is crucial for the user’s experience. It has more requirements compared with traditional multimedia content. Besides good quality of video and audio, VR content requires stitching, special effects, stereoscopic 3D and composition. To ensure an immersive experience, it is important that VR content is generated in good quality and methods in each step, and then delivered as perfectly as possible. This section lists the aspects related to VR content which will influence the quality of a VR service.

**Spatial audio**

Spatial audio involves the use of 3D audio reproduction to produce sound sources at any point within the 3-dimensional space. Sound reproduction for VR is most often done via headphones, but also loudspeaker setups are possible especially in 3 DoF and 6 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.

**Spatial depth (3D)**

It also is possible to playout stereoscopic video content, which simulates the way of the human ability

to view with both eyes in similar, but in slightly different ways. This allows humans to judge distance and have a perception of depth. To minimize the crosstalk effect, the left view of the content has to be only displayed to the left eye and the right view of the content has to be only displayed to the right eye (cf. [Woods]). By the system design of most of the head-mounted based systems, which are commonly used for playing out VR content, both views already are separated from each other. In the area of stereoscopic omnidirectional video content, a pre-study has been done.

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. With respect to simulator sickness scores and how a stereoscopic representation is influencing the VR QoE is for further study.

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

**Spatiotemporal complexity**

Spatiotemporal complexity is not directly affecting people’s experience. Spatial perceptual information indicates the complexity of a video picture. With some high complexity content, the viewer may be distracted, while with some less complexity content, the subject may be more focus on the main objects. Temporal perceptual information indicates the amount of changing of the video picture.

Spatiotemporal complexity is the feature of a content which itself does not affect VR QoE. However, when going through network transmission, the factor may have its influence. When VR content is delivered by unreliable transport protocols, e.g., UDP, different spatiotemporal complexities may present different impact video pictures. Generally, less complexity may have more tolerance on network packet loss, which may result in better final QoE. But when the VR content is delivered by reliable protocols like TCP, this fact may not be one that should be considered when evaluating QoE. Other aspects of the influence on QoE is for further study.

## Media/coding related

**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. There are different codecs that have been developed in the industry and widely used for traditional media coding. These codecs may be used for VR media, however, some codecs are unsuitable for certain scene representations.

**Video**

Traditional video codecs (e.g., H.264, 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, which can cause different information loss when 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.

The common shortcoming for traditional video codecs being used in VR is that compression rate is still too low, which means bandwidth consumption is still a big problem for streaming the VR content when dealing with the full VR streaming. New video coding technologies such as Versatile Video Coding (VVC) are in the progress to well improve the transport quality for virtual reality 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 rest in a minimal quality. This is called viewport-adaptive or tile-based streaming of omnidirectional video,

**Audio**

For non-interactive 3 DoF and 6 DoF VR scenes, requirements on audio meta data 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 listeners movements. For interactive and 6 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 and 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, both requiring various formatting in order to implement and deliver immersive audio. An audio codec would need to be adaptive in both the input data stream of content, as seen in the current MPEG-H codec for 3 DoF scenes, along with a consistent input stream of user actions.

**Storage and Transport**

Besides that, there are also other technologies used in VR to reduce bandwidth and resources consumption. MPEG now has developed a standard called “Omnidirectional media format” [ISO/IEC 23090-2], which intends to standardize the storage and transmission of virtual reality content, mainly for 360-degree videos. There are multiple media profiles supported, one of which is to divide the entire 360-degree video into independently coded tiles and the HMD has to recompose the image from tiles that are required according to the user’s viewing directions.

**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.

**Resolution**

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

To achieve a good VR quality, 4K+ resolution is required. It is because VR features 360-degree panoramic display, while the monocular field of view resolution determines the image quality of VR. The low resolution of the original VR content will be enlarged on VR near-eye display.

**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 with the panel’s refresh rate leads to artefacts like frame fluctuation, frame drops and frame manipulation using black frame insertion (BFI). These artefacts are mostly leading to jerkiness, leading to a lower QoE. [Hofmeyer]

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

In the area of 360° videos, applying motion interpolation to contents having 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. [Hofmeyer, Fremerey]

**Audio sample rate**

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

**Coding delay**

As the codec standards only describe the algorithmic decoding procedure and profile features, there are still many options in selecting coding modes, 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 TV broadcasting (e.g. [ITU-T J.248]), but VR brings new challenges due to the immersive experience and sensorimotor coupling in 6 DoF, where additionally synchronizing the rendered content with self-movement is essential. Perceptual thresholds in 6 DoF scenarios are topics for future study.

## Network/transmission related

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

**Delay**

In VR environments, stringent latency requirements are of most importance for providing a pleasant immersive VR experience. Delay includes the queuing delay, over-the-air delay, and buffering delay. Delay is usually the main reason resulting in high motion-to-photon latency leading to simulator sickness (see Clause 7.1.2). It is also the reason causing presentation quality degradation, e.g., long initial loading delay, stalling. Some VR 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.

**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 the network transmission, e.g., congestion can cause long delay and packet loss, which then degrades the perceived immersive QoE of the VR system.

**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 loss of parts of frames or entire frames and thus degrade audiovisual quality, in which case, the quality degradation may be presented as the phenomenons like video freezing and tiling artifacts

## 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.

**Head-mounted Display**

Different from traditional terminal devices, the HMD wearing comfort may also greatly impact the final VR QoE. To improve it, it is important to consider the following factors: the device weight, size, heat dissipation, fit, shortsightedness/farsightedness-friendly, and so on.

**Headphones**

Frequency response of the headphones is a quality factor. Neutral headphones or headphones whose frequency response has been compensated may be able to better convey the spatial audio experience to the listener. Additionally, the ability to block outside noise may be of importance for VR; closed headphones or in-ear headphones are best suited for that. However, the use of such headphones that entirely block outside noise may cause the user to feel disoriented in their real environment and maybe a safety hazard.

**Decoder performance**

The decoder capability has an impact on the overall resolution of the video to be transmitted and decoded in the device, e.g. HD, or UHD, and thus decides the final resolution of the video that could be displayed to the user. Also, the codec, e.g. H.264 or H.265, decoder support is also important since different codecs have difference decoding performance.

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

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

Comparing to the software, the decoders implemented in hardware have much faster speed 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 which can fix the error during the transport or encoding. This can also increase the quality of the final VR experience.

**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 (IMU) implemented inside HMD, which uses the combination of accelerometers, gyroscopes, and sometimes also magnetometers to track objects’ motions.

There are two tracking technologies so far: outside-in and inside-out. Outside-in tracking indicates the headset and accessories rely on some external devices, e.g., lighthouse sensor or computer display. It has more accuracy and better latency but limited by the environment. Inside-out tracking doesn’t rely on external devices. It uses the HMD sensor to determine how the position is changing in relation to the external environment. Right now, accuracy and latency are not so good comparing to outside-in tracking.

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

**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. When it comes to VR, it is a solid angle that is visible by a human through HMD lenses. The limiting factor is the lenses instead of the pupils. To get a better FoV, the VR HMD lenses are always fairly close to the user’s eyes [VR Lens Lab].

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

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

**Display resolution**

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

Pixel per degree (PPD) is a core technology specification that is better suited for measuring the pixel density of VR near-eye display than pixel per inch (PPI). The higher the PPD is, the better of the image quality will be. The lower PPD may make the screen-door effect occurs.

**Refresh Rate**

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

##  Context influence factors

Context influence 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.

## 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 to that, 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 would restrict their movement and possibilities to explore a VR scene. Room lights may not affect users’ experience so much when they’re using VR HMDs comparing to traditional video environments since the devices are so close to the eyes. The environment temperature should be set to typical room temperature. To achieve the best immersive experience, the environment temperature should be consistent with the virtual world. For example, people will feel “being there” when they are watching or playing a skiing scene while the environment temperature is turning cold enough for them to be in the real world.

It should be payed attention that only a low amount or ideally no direct sunlight can 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 (like e.g. virtual walls or a pass through mode using cameras) should be activated to avoid any collisions of the VR user with his environment. Ideally, another person not consuming VR services should pay attention on the physical actions of the user consuming the VR service.

## Temporal context

Temporal context factors include the frequency and duration of use. A VR device may not be able to support long usage periods: simulator sickness symptoms like dizziness, loss of spatial awareness, nausea, eye soreness typically get worse as the duration grows, which will greatly reduce the quality of experience.

## Social context

Social context factors include VR content popularity and how the 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 the program. And it may still be 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. How and what degree does this factor affect the overall Quality of Experience of VR services needs to be further investigated.

## Task context

VR experience depends on the goals that a user is using this VR service for. These kinds of 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. Another example is that the quality of experience for streaming type VR, e.g. 360-degree VR, would be quite different from the 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.

# Bibliography

[Brandenburg] R. van Brandenburg, R. Koenen, and D. Sztykman. 2017. CDN optimization for VR streaming. <https://www.ibc.org/tech-advances/> cdn-optimisation-for-vr-streaming-/2457.article 2018/06/12). (Oct. 2017).

[Brunnström] Brunnström, K., K. Wang, S. Tavakoli and B. Andrén (2017). "Symptoms analysis of 3D TV viewing based on Simulator Sickness Questionnaires." Quality and User Experience 2(1): 1-15.

[Fremerey] Fremerey, Stephan, Hofmeyer, Frank, Göring, Steve and Raake, Alexander "Impact of Various Motion Interpolation Algorithms on 360° Video QoE." 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2019.

[Hofmeyer] Hofmeyer, Frank, Fremerey, Stephan, Cohrs, Thaden and Raake, Alexander "Impacts of internal HMD playback processing on subjective quality perception." In Electronic Imaging 2019.12 (2019): 219-1.

[Reiter] Reiter, Ulrich, et al. "Factors influencing quality of experience." Quality of Experience. Springer, 2014. 55-72.

[Kennedy] Kennedy, R. S., N. E. Lane, K. S. Berbaum and M. G. Lilienthal (1993). "Simulator Sickness Questionnaire: An Enhanced Method of Quantifying Simulator Sickness." The International Journal of Aviation Psychology 3(3): 203-220.

[Singla] A. Singla, S. Fremerey, W. Robitza, A. Raake, “Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays”, in 9th International Conference on Quality of Multimedia Experience (QoMEX), May 2017.

[Stanney] Stanney, K. M., Kennedy, R. S., & Drexler, J. M. Cybersickness is Not Simulator Sickness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1997 41(2), 1138–1142.

[VR Lens Lab] <https://vr-lens-lab.com/field-of-view-for-virtual-reality-headsets/>

[Witmer] B. G. Witmer and M. J. Singer, Measuring Presence in Virtual
Environments: A Presence Questionnaire, Presence, June 1998.

[Woods] Woods, Andrew. "Understanding crosstalk in stereoscopic displays." Keynote Presentation at the Three-Dimensional Systems and Applications Conference, Tokyo, Japan. 2010.

# Appendix A

**Virtual Reality Services Use Cases**

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

Generally, VR applications can be divided into 2 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 occurred in this type of VR applications may affect the experience of users. However, it can save the local storage of VR terminal devices and expand the range of content that user can experience.

**Offline VR:**

VR applications of this type work offline. To do this, the users need to download the VR content completely to their devices in advance. While running, these applications usually don’t have any network delay issues and don’t need any 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 document.

## A.1 Use Scenario

There are a lot of different types of VR services. The 5 types are listed below:

## A.1.1 Live

Live VR is a kind of broadcast in real-time, as events happen, in the present. The difference from the traditional live program 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 [ITU-T H.ILE-SS] is a service of this kind, which constructs 360° panoramic view in real-time via multiple cameras feeds from the site. The users can observe the live event around the spot with proper VR HMDs which constantly process and stitch multiple images to project the real world.

Live VR services require extremely stringent delay so that the users can smoothly change the viewpoints when watching it. They also require ultra-high definition resolutions to make people feel that they are in the real venue. The bandwidth consumption issue is very challenging when a massive number of user watch 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 around your head.

## A.1.2 VoD

Video on Demand VR services allow users to select and experience the content at any preferable time of 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 OTT providers like Youtube which allow you 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 relativelysmaller than live VR as viewers can consume the same content at different time.

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

## A.1.3 Gaming

Virtual reality gaming is where a person can 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 HMD. For example, some data glove with small sensors to capture the movements made by that person 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. Besides that, other aspects of gaming discussed in [ITU-T G.QoE-game] should also be considered.

## A.1.4 Social

Virtual reality social is a service which allows people using a virtual reality 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 your profile and spend time with other people in a virtual space.

Like a VR game, social VR also requires extremely sensitive interaction so that people can 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 people to use the services.

## A.1.5 Shopping

Besides above the listed virtual reality services, there are other applications which may be promising in the future when using VR devices. For example, VR shopping allows people to purchase through a virtual reality headset by virtually transporting them to international retail outlets and enabling them to experience the entire shopping experience from finding products to payment. VR shopping is similar to VoD, which records the content in advance, but requires more interaction and less data consumption than VoD streaming VR.