|  |  |  |  |
| --- | --- | --- | --- |
| **Question(s):** | Q13 | **Meeting:** | Geneva, February 2018 |
| **Study Group:** | SG12 | **Working Party:** | WP3 |
| **Source:** | Editor of G.QoE-VR |
| **Title:**  | Updated version of G.QoE-VR baseline |
| **Purpose:** | [Choose a purpose from the dropdown list] |
| **Contact:** | Rachel HuangHuawei Technologies Co. Ltd.China | Tel: +86 25 56623633Fax: E-mail: rachel.huang@huawei.com |

|  |  |
| --- | --- |
| **Keywords:** | Insert keywords separated by semicolon (;) |
| **Abstract:** | This contribution proposes to update the baseline of G.QoE-VR to have a adjusted structure. |

**Introduction**

This is the output of Q13 interim meeting in Geneva, February 2018, regarding G.QoE-VR. The main change is to modify the structure of the document and include some new content in it, comparing to the baseline document, i.e., TD362R1.

**Proposal**

**CONTENTS**

|  Page |
| --- |
| [1 Scope 4](#_Toc507611014)[2 References 4](#_Toc507611015)[3 Definitions 5](#_Toc507611016)[3.1 Terms defined elsewhere 5](#_Toc507611017)[3.2 Terms defined in this Recommendation 5](#_Toc507611018)[4 Abbreviations and acronyms 5](#_Toc507611019)[5 Conventions 5](#_Toc507611020)[6 Virtual Reality Taxonomy and Overview 5](#_Toc507611021)[6.1 Hardware 5](#_Toc507611022)[6.2 Software 5](#_Toc507611023)[7 Virtual Reality Services Use Cases 6](#_Toc507611024)[7.1 Use Scenario 6](#_Toc507611025)[7.2.1 Live 6](#_Toc507611026)[7.2.2 VoD 6](#_Toc507611027)[7.2.3 Gaming 7](#_Toc507611028)[7.2.4 Social 7](#_Toc507611029)[7.2.5 Shopping 7](#_Toc507611030)[8 Virtual Reality Influence Factors 7](#_Toc507611031)[8.1 Human influence factors 7](#_Toc507611032)[8.1.1 Vision and hearing 7](#_Toc507611033)[8.1.2 Simulator sickness 7](#_Toc507611034)[8.2 System influence factors 7](#_Toc507611035)[8.2.1 Content related 7](#_Toc507611036)[8.2.2 Media/coding related 8](#_Toc507611037)[8.2.3 Network/transmission related 9](#_Toc507611038)[8.2.4 HMD related 9](#_Toc507611039)[8.3 Context influence factors 10](#_Toc507611040)[8.3.1 Physical context 10](#_Toc507611041)[8.3.2 Temporal context 10](#_Toc507611042)[8.3.3 Social context 10](#_Toc507611043)[8.3.4 Task context 10](#_Toc507611044)[9 Virtual Reality QoE Assessment 10](#_Toc507611045)[9.1 Quality of Immersive media 10](#_Toc507611046)[9.2 Presentation quality 11](#_Toc507611047)[9.3 Interaction quality 11](#_Toc507611048) |

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

Recommendation ITU-T <No.>

Overview of Quality of Experience (QoE) in Virtual Reality Services

Summary

This draft Recommendation identifies service scenarios by analysing virtual reality applications, in order to classify virtual reality services and to clarify the QoE key factors of VR.

Keywords

Insert keywords separated by semicolon (;)

# 1 Scope

Virtual Reality (VR) is a new type of media different from the traditional video and audio media. 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, touch, hearing, and, less commonly, 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.

The scope of this Recommendation includes:

* VR service use cases
* Categorization of influence factors
* VR QoE assessment

# 2 References

The following ITU-T Recommendations and other references contain provisions, which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T 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:

TBD

## 3.2 Terms defined in this Recommendation

TBD

# 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

|  |  |
| --- | --- |
| VR | Virtual Reality |
| HMD | Head mounted device |

# 5 Conventions

TBD

# 6 Virtual Reality Taxonomy and Overview

Virtual Reality (VR) is a technology that uses certain devices to create artificial environments hijacking the sense of people who interact with them while have little or no awareness of the interference. It generates the realistic images, sounds and other sensations that replicate a real environment or create an imaginary one. VR blocks the real world out entirely, and creates its own reality that is completely computer generated and driven. This is different from Augmented Reality (AR), 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 Hardware

A VR display device is usually a typical Helmet-mounted Display (HMD) with 2 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 each of our two eyes views the world.

VR is heavily dependent on motion-tracking to present visuals that behave in relation to the user’s body movements. Head tracking or eye tracking is applied to HMDs to create the right camera angel and perspective so as to reach the natural viewing experiences.

## 6.2 Software

TBD

# 7 Virtual Reality Services Use Cases

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 user can experience is limited by the capacity of local storage device. Offline VR services are not in the study scope of this document.

## 7.1 Use Scenario

According different use scenarios, 5 different types of VR services are introduced.

## 7.2.1 Live

Live VR is a kind of broadcast in real-time, as events happen, in the present. The difference from traditional live program is that live VR is panoramic and interactive. Live VR can provide 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 camera 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.

## 7.2.2 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 relative smaller than live VR as viewers can consume 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.

## 7.2.3 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.

## 7.2.4 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 very bad experience for people to use the services.

## 7.2.5 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.

# Virtual Reality Influence Factors

[Editor’s note]: There should be a figure here to show all the influence factor categories.

##  Human influence factors

## Vision and hearing

TBD

## Simulator sickness

TBD

##  System influence factors

## Content related

VR content is crucial for 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 metrics related to VR content which will help with assessing the quality of VR service.

**Spatial audio**

Spatial audio indicates the 3D audio experience that a user has when listening. Spatial audio helps a lot in improving the immersive experience of VR services.

**Spatial depth (3D)**

This factor indicates the video content is sent slightly differently to different eyes of the human being. It simulates the way of the human ability to view with both eyes in similar, but slightly different ways. This allows humans to judge distance and have a depth perception.

**Spatiotemporal complexity**

For VR, Spatial perceptual information may not be so important to affect people’s experience, as there’s no data loss during transmission in the case of VR application uses HTTP streaming technology to deliver the content. However, it is possible that VR HMD will do some artificial modification to make the current FOV content fit in the monocular display, which is different from traditional screen. In such cases, the content with low SI may have better experience than the one with high SI.

Temporal perceptual information is not so important to affect people’s experience, as there’s no data loss during transmission in the case of VR application uses HTTP streaming technology to deliver the content.

**Genre**

TBD

##  Media/coding related

**Video and audio codec**

TBD

**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. This metrics is not critical to VR services, but it’s a basic metric to make sure VR services to provide high quality images.

**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. Low resolution of the original VR content will be enlarged on VR near-eye display. 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.

**Frame rate**

Frame rate indicates the frequency at which consecutive images called frames are displayed. Frame rate of the VR content should be compatible with the frame rate attribute of the display device. The frame rate in VR services has higher requirement than normal 2D video services, it is because influent scenses is one of the sources causing VR dizziness. It is even more demanding for VR gaming applications, in which the scenes are rendered by GPU instead of those created by video cameras.

**Audio sample rate**

Sample rate is the number of samples of audio carried per second, measured in Hz or kHz. In VR services, this factor does not have no difference from the traditional streaming services.

**Coding delay**

TBD

##  Network/transmission related

**Loss**

TBD

**Delay**

TBD

**Bandwidth**

TBD

**Jitter?**

## HMD related

HMD plays an important role in creating immersive experience for users. Compared with traditional terminals, VR device exhibits more rich and complex attributes. And some of these attributes could help assess the quality of VR service.

**Decoder performance**

The decoder capability has an impact on the overall resolution of video to be transmitted and decoded in the device, e.g. HD, or UHD, and thus decides the final resolution of 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 support quite different content type.

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.

**Head-tracking Latency**

To enable interaction between users and the environment it is important to create believable VR experience. While low head-tracking latency is certainly an important attribute to provide smooth change of view for the user, long head-tracking latency induces discomfort and loss of immersive experience.

**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.

**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. 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 video resolution, would provide the best and comfortable experience.

**Refresh Rate**

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

##  Context influence factors

## Physical context

TBD

## Temporal context

TBD

## Social context

TBD

## Task context

TBD

#  Virtual Reality QoE Assessment

The VR QoE refers to the level of presence perceived by VR users. It is assessed by the level of immersion provided by the VR system combined with the quality of the interaction between the user and the VR environment.

Immersion, regarded as a quality of the VR system’s technology, is commonly described through an objective measure of the extent to which a VR system presents a vivid virtual environment while shutting out physical reality. For the VR systems, the level of immersion perceived by a user can be determined by the quality of the immersive media and the presentation quality delivered by the device.

The interaction quality between a user and the virtual environment can have a significant impact to the overall experience quality. Good interactions can lead to a high-quality experience with a strong sense of presence. On the other hand, an unrealistic interaction may lead to discomfort and distress.

[Editor’s note]: here, we may put a figure to show the relationship between VR QoE and the 3 subsection. Could use the figure of G.360-VR,but with some rewording and omitting some details.

##  Quality of Immersive media

The following features of the immersive media considered in [MPEG-I Part 1] have been identified as key factors to impact the level of immersion:

* Degrees of freedom (DoF): proposed formats include 3DoF, 3DoF+, Windowed 6DoF, Omnidirectional 6DoF, or 6DoF;
* Quality of video:
	+ Three-dimensionality: proposed formats include monoscopic 360 video, stereoscopic 360 video, or full 3D 360 video;
	+ Field of view (FOV): proposed range between 90 ~ 220 degrees
	+ Spatial resolution in pixels per degree (PPD): 12 ~ 60
	+ Frame-rate: proposed formats 60fps, 90fps, or higher
	+ Compression: up to visually lossless
	+ Projection:
* Quality of audio:
	+ Three-dimensionality: proposed formats include 3D, or stereoscopic, etc ,
	+ Channels and sample rate:
	+ Compression: high fidelity

##  Presentation quality

The following system performance measures have been identified as key indicators of the level of immersion provided by the immersive systems in [MPEG-I Part 1]:

* Playback quality: smoothness and number of freezes, video/audio quality change, etc.
* Audio and video synchronization.
* Audio and video spatial alignment.

##  Interaction quality

The following interactive features have been identified as key influence factors to the user’s overall experience quality with the immersive systems described in [MPEG-I Part 1]:

* Intractability with objects in the VR environment
* Response time between human action and adaptation in sound and display: less than 20ms
* Spatial precision between human action and adaptation in sound (3D audio) and visual information