White Paper for 5G Cloud VR Service Experience Standards

June 2019

HUAWEI SmartCare | CLOUD OPEN LABS Joint Release

Contents

Acronyms and Abbreviations

Acronyms and Abbreviations	Full Spelling
ACR	absolute category rating
BDR	burst delay rate
BPN	burst pulse number
CCR	comparison category rating
DCR	degradation category rating
DOF	degrees of freedom
DSCQS	Double Stimulus Continuous Quality Scale
EWMA	exponential moving average
FOV	Field of View
FR	full-reference
GPU	Graphics Processing Unit
HMD	head mounted display
IQI	Interaction Quality Index
MBS	maximum burst size
MOS	mean opinion score
MPD	media presentation description
MQI	Media Quality Index
MTP latency	motion-to-photons latency
MTS latency	motion-to-sound latency
NR	no-reference
NR-B	no-reference Bitstream Layer Model
NR-P	no-reference Parametric Packet Layer Model
NR-PL	no-reference Parametric Planning Model
PPD	pixels per degree
PPI	pixels per inch
PQI	Presentation Quality Index

QoE	quality of experience
RR	reduced-reference
SSCQE	Single Stimulus Continuous Quality Evaluation
SDSCE	Simultaneous Double Stimulus for Continuous Evaluation
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution

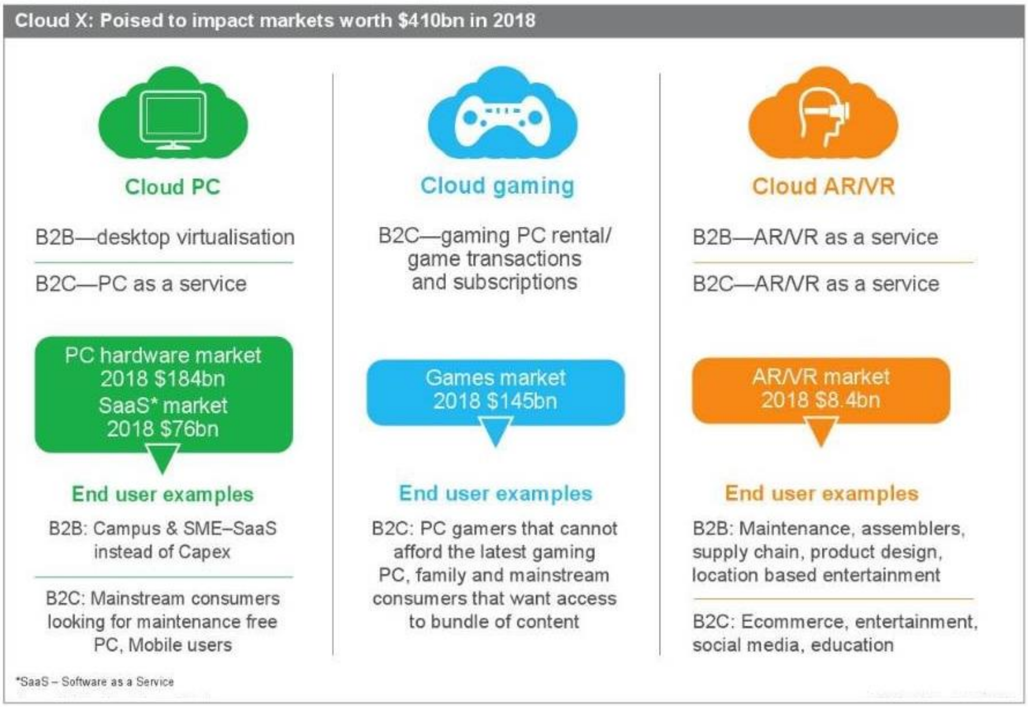
1 Overview

With the faster speed, lower delay, and more connectivity options, 5G will help carriers develop rich services and provide the ultimate network and service experience.

5G will change the business model of carriers and the mode for providing services to enterprises (B2B) and consumers (B2C). The operations mode of carriers is changing from traffic operation with connections as the key factor to service operation with content as the key factor. Network coverage and traffic, will no longer be the key advantage or competition differentiator. Ultimately the experience of consuming 5G services will become the most important standard for customers to choose a carrier.

A recent study from IHS shows that some of the key service value derived from 5G (Figure 1-1) will enable carriers to provide customers with anytime and anywhere connectivity.

Figure 1-1 5G key value services



Cloud Virtual Reality (Cloud VR) may become one of the preferred eMBB services for many 5G commercial carriers. Cloud VR introduces cloud computing and cloud rendering to VR service applications. With fast and stable networks, cloud-based display output and audio output are coded, compressed, and transmitted to user terminals, supporting cloud-based VR service content and content rendering.

This document describes the technology of Cloud VR service experience modeling. There are a few key points we should note, to meet users' requirements:

- Firstly, carriers do not need to provide the same service quality for all AR/VR services.
- Secondly, Cloud VR service experience will be affected by various factors, such as, media quality, network quality, and terminal quality.

Therefore, development of measurable and manageable service experience indicators will be the key basis for carriers to ensure the satisfaction of consumers and enterprise customers alike.

Huawei SmartCare® CEM Solution builds a complete indicator system for Cloud VR service experience and scores user VR experience based on the following indicators:

1. Media Quality Index (MQI),
2. Interaction Quality Index (IQI), and
3. Presentation Quality Index (PQI) to optimize key factors affecting user experience.

These modeling systems can be used in VR games, VR 360 videos, VR live broadcast, VR IMAX, and VR education.

2 Indicator System for Cloud VR Service Experience

2.1 Reference Specifications

During the selection of KPIs for service experience modeling of Huawei 5G Cloud VR, suggestions for the VR experience evaluation model in ITU-T G.QOE-VR and 3GPP TR 26.929 specifications are referenced.

2.1.1 ITU-T G.QOE-VR

The following table describes the evaluation dimensions and main factors that affect the VR service experience in the evaluation model provided by G.QOE-VR.

Category	Influencing Factor
Media immersion	Degrees of freedom (DOF)
	Video quality improved in 3D mode
	Video Field of View (FOV)
	Video pixel per degree (PPD)
	Video frame rate
	Video compression rate
	Number of audio channels
	Audio sampling frequency
	Three-dimensionality audio
	Audio compression rate
Presentation quality	Audio and video fluency
	Audio and video stalling rate
	Audio and video time synchronization
	Audio and video space matching degree

Category	Influencing Factor
Interaction degree	Degree of interaction with the VR environment
	Response delay for interaction with the VR environment
	Interaction space precision

[Reference content from specifications]

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

2.1.2 3GPP TR 26.929

The following table describes the suggestions on quality of experience (QoE) indicators of VR services in 3GPP 26.929 specifications.

Category	Influencing Factor	Description
Transmission impact	Average download throughput	Affects the buffer level and buffer pool status in a measurement period.
	Level of the buffer in milliseconds	Indicates the playout duration for which media data of all active media components is available starting from the current playout time.
	Playback operations	Include users' operations, such as playback, drag, and pause.
	Image delay	Indicates the duration from the time when a user wants to see an image to the time when the user sees the image.
	Motion-to-photons (MTP) delay	Indicates the delay from the time when a motion request is sent to the time when the video image is displayed.
	Motion-to-sound (MTS) delay	Indicates the delay from the time when a motion request is sent to the time when the audio is played.
Device impact	Horizontal FOV	Indicates the horizontal FOV supported by the device.
	Vertical FOV	Indicates the vertical FOV supported by the device.
	Resolution	Indicates the single-eye resolution supported by the device.
	Refresh rate	Indicates the number of screen image updates per second supported by the hardware.
	Decoding capability	Indicates the supported video encoding level and format.

QoE provided by the immersive technologies such 360-degree videos play an important role how many users are going to interact with the technology. Therefore, there is a need to assess the QoE of the new emerging technology, as QoE is one of the contributing factors in making the technology successful.

In this work, the influence of resolutions, camera motion, motion in the content, and simulator sickness on QoE is investigated. Some of the users are prone to simulator sickness, therefore, it is of interest to investigate how the simulator sickness interacts with the QoE and vice-versa.

Preparation of datasets

Six contents were chosen showing significant differences with respect to motion in the scene. Two resolutions, 4K and FHD were chosen which was motivated by the resolution limitation of the HMDs. The resolution of both devices is 2160×1200. The dataset was downloaded from the Internet, because the duration of these video sequences could be chosen much longer compared to the standard dataset [11–13]. We also wanted to use off-the-shelf contents as provided by the services without re-encoding. Table 8.2-1 provides an overview of the content. We downloaded the H.264/AVC encoded video sequences with highest provided bitrates. We proved by visible inspection by experts that the encoding quality for all contents was high. Then the duration was cut to a length of 60-65 seconds [6,7].

Technical setup and equipment

Two HMDs were used from two different companies – named HMD1and HMD2 here. The resolution and field of view (FOV) for both devices are 2160×1200 and 110° respectively. Whirligig player (version 3.89) was used in order to display the 360° videos in both HMDs. The HMDs were connected to a desktop PC equipped with an NVIDIA GTX980 graphics card and an Intel Core i7 processor. The names of the HMDs were hidden to the subjects to decrease contextual effects [6,7].

QoE metrics relevant with network transmission

Average Throughput

Section 10.2.4 in [3] defines the metric for average throughput information. This information could be observed by OP1 of the reference model.

Buffer Level

Annex D.4.5 in ISO/IEC 23009-1 [4] defines the metrics for buffer level status events. This information could be observed by OP1 of the reference model.

Play List

Section 10.2.7 in [3] defines the metrics for event that may happen due to user action, the end of the content or a permanent failure. This information could be observed by OP1, OP2, OP3 and OP4 of the reference model.

Presentation Delay

VR device impact on QoE

Introduction

Compared with traditional streaming video, the key feature of VR service is to create immersive experience and enable smooth interactivity between user and the environment, in which VR device would play an important role. This contribution proposes device information relevant to user experience of VR service. All the device property information could be collected by OP5 of the reference model described in Section 6.1.

QoE metrics relevant with VR device

Field of View

One of the factors that contribute to the uniqueness of 360 video experience is the level of immersion induced by the wider FOV of HMD, which represents the extent of observable environment at any given time. A wider FOV could help provide a more authentic feeling of immersion. Thus FoV of the HMD is an important parameter that helps evaluate to what extent a VR device could help create immersive experience.

Resolution

Resolution here is defined as for per eye. An appropriate screen resolution would provide the best and comfortable experience.

Refresh Rate

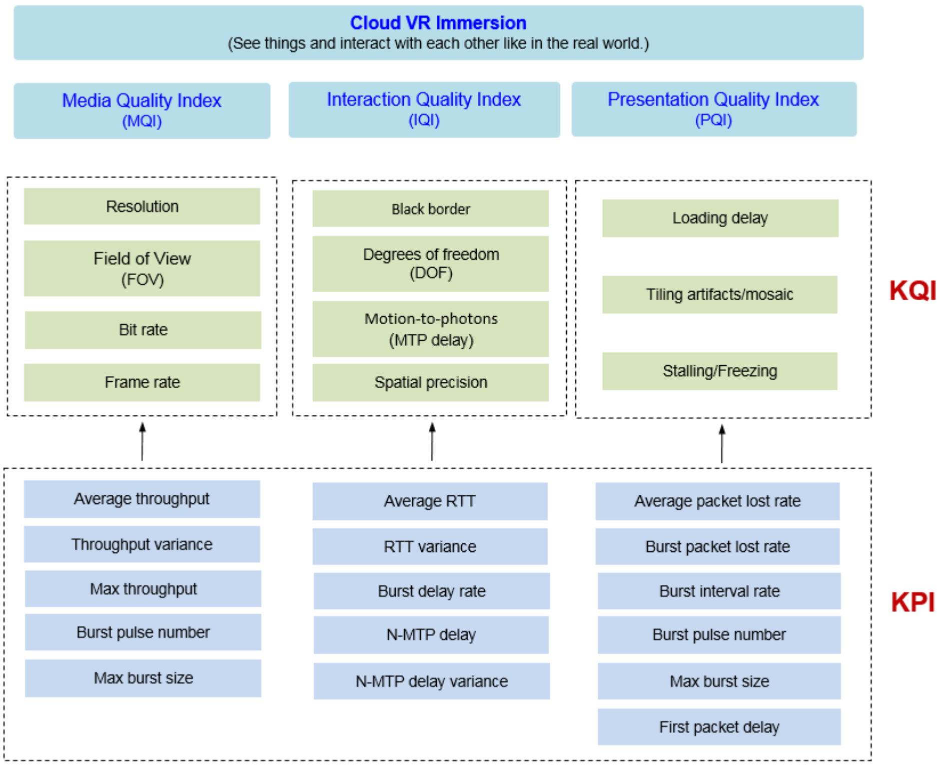
Refresh rate is the number of times per second the display grabs a new image from the graphic processing unit. Lower refresh rate would contribute to processing latency and lead to VR sickness, i.e. viewing glitches on the screen. While higher refresh rate adds to the sense of presence in virtual worlds.

Decoder capability

The support of codec profile and level is an important property that decides the content types it could decode.

2.2 Indicator System for Huawei Cloud VR Service Experience

According to the protocol specifications, Huawei SmartCare® CEM Solution scores user experience of Cloud VR services based on MQI, IQI, and PQI.



2.2.1 MQI

MQI: Media quality score (0–100). This indicator indicates whether the sensory stimulation to users brought by the VR content, including the audio, video, and content DOF, has been close to the sensory effect in the real world.

Subjective Evaluation	Description	MOS	MQI
Excellent	The screen does not look granular, and the video is clear. The basic condition is that the value of PPD is larger than 60.	5	80–100
Good	The video is clear, but the video quality is worse than that at the excellent level.	4	60–80
Fair	The screen looks granular. Users can see colors or lines in some details, and the image is distorted.	3	40–60
Poor	The granular phenomenon of the video image is serious.	2	20–40
Very poor	Users cannot watch or experience the VR videos because the image is rather unclear. Users do not want to watch the VR video completely or recommend the VR video of this resolution to other users.	1	0–20

Key factors

Factor	Impact
Bit rate	Bit rate refers to the number of audio or video bits transmitted or processed per time period. Bit rate is a more common indicator for measuring audio and video quality. A high resolution, high frame rate, or low compression usually increases the bit rate in the same encoding environment. For VR services, bit rate is not a key indicator. However, it is a basic indicator that can ensure that VR services provide high-quality images.
Frame rate	The frame rate indicates the frequency at which frame-based images are continuously displayed on a display. The frame rate of VR content must be compatible with the frame rate attribute of the display device. The frame rate of VR services is higher than that of common 2D video services. This is because the video stalling is one of the reasons why users feel dizzy when using VR services. VR game applications have higher requirements on frame rates because images in VR games are rendered by Graphics Processing Units (GPUs) instead of shot by cameras.

Factor	Impact
Resolution	Resolution refers to the number of pixels in the video content. The video resolution must be compatible with the resolution of the display device. Otherwise, the video resolution may decrease or even the video cannot be displayed. To achieve better VR video quality, a 4K or higher resolution is required, because the VR has a 360 panoramic display function, and the single-eye resolution of the VR panoramic display determines the VR image quality. The low resolution of the VR content is magnified during VR near-eye display.
FOV	The FOV measures the range of visual environments at any given time. A wider FOV makes users feel more immersive. Therefore, FOV is an important parameter that can evaluate the immersive experience creation capability of the VR device. In the model, FOV is considered as an important factor for the spatial influence module and occlusion influence module.
PPD	PPD is a core technical specification that is more suitable for measuring the pixel density of VR near-eye display than (PPI) per inch pixel. A larger value of PPD indicates better VR image quality.

2.2.2 IQI

IQI: Interaction quality score (0–100). This indicator indicates the experience of interaction between a user and VR content operations when the user uses the VR service. Interaction delay can cause dizziness, nausea, and isolation from the feeling in the real world.

Subjective Evaluation	Description	MOS	IQI
Excellent	The VR system responds to users' actions smoothly, which is the same as that in the real world. Users do not feel any drag, pause, or asynchrony. The interaction DOF is the same as that of the real world, and users feel comfortable, free, and flexible in the VR system.	5	80–100
Good	Occasionally, user can feel slight asynchrony and pause. Additionally, when moving the game handles or rotating heads violently, uses can feel the black border or asynchrony between the image and game handles for very short time.	4	60–80

Subjective Evaluation	Description	MOS	IQI
Fair	Users can often feel slight asynchrony between the image and game handles, the black border, and that the game video responds slowly after performing operations. The experience is not good, but no continuous black border or asynchrony occurs. It is uncomfortable to wear the VR device for a long time, or the interaction experience is normal.	3	40–60
Poor	Users feel continuous asynchrony between the image and game handles, the black border, and that the game handles are out of control.	2	20–40
Very poor	In most of time, the game handles are out of control, black border occupies a large area of the screen, and games cannot be played. Users feel bad, do not want to continue games, and will not recommend VR services to other users.	1	0–20

Key factors

Factor	Impact
MTP delay	The MTP delay refers to the response duration of the video and audio after a user performs an action during VR experience. The value of this indicator must be less than 20 ms for cloud VR games and less than 60 ms for cloud VR videos.
DOF	DOF indicates the mode in which an object can move in space. It is a key factor that helps users create an immersive environment.

2.2.3 PQI

PQI: Presentation quality score (0–100). This indicator indicates the continuous and smooth sensory experience of users when users use the VR service. Poor user experience refers to tiling artifacts/mosaic and stalling.

Subjective Evaluation	Description	MOS	PQI
Excellent	The VR system responds to user operations smoothly. Users do not feel tiling artifacts/mosaic, stalling, or intermittent voice during the audio service, video service, and interaction operations.	5	80–100
Good	Users feel slight tiling artifacts/mosaic, stalling, or discontinuous voice occasionally.	4	60–80
Fair	Users feel tiling artifacts/mosaic or stalling for several times. The duration of tiling artifacts/mosaic or stalling lasts for longer than 1 second once, but no continuous tiling artifacts/mosaic or stalling occurs frequently.	3	40–60
Poor	Continuous tiling artifacts/mosaic, stalling, or discontinuous voice occurs frequently.	2	20–40
Very poor	Users can feel continuous tiling artifacts/mosaic, stalling, and discontinuous voice at most of the time. They do not want to continue experiencing the service or recommend the VR service to other users.	1	0–20

Key factors

Factor	Impact
Loading delay	For Cloud VR video services, the loading delay refers to the initial buffering delay. Generally, the delay must be less than 10 seconds to ensure user experience. For Cloud VR games, the loading delay refers to the delay from the time when a user chooses to play a game to the time when the game starts. Generally, the delay must be less than 3 seconds to ensure user experience.

Factor	Impact
Stalling	Stalling is a key factor for users to perceive streaming media smoothness. During VR video playback, there is still a buffer zone with a certain amount of data. If no data is available in the buffer zone, stalling occurs, affecting user experience. Generally, stalling occurs because the download throughput cannot meet the video encoding quality requirements.
Freezing	Freezing is a phenomenon that a user perceives a pause of the game image, and is a key factor for evaluating game smoothness. If a key frame (such as the frame I) is discarded during VR gaming, the key factors of the image will be lost and the image cannot be displayed. As a result, freezing occurs, affecting user experience in VR gaming.
Tiling artifacts/mosaic	For cloud VR games, tiling artifacts/mosaic is a key factor for evaluating the game smoothness because users perceive mosaics in some areas of the game image. During VR gaming, if some video frame information (some block information in the video frame) is lost, the image can be displayed, but mosaics occur in some areas, affecting user experience of the VR game fluency.

3 Modeling Method for Cloud VR Service Experience Indicators

3.1 Model Selection

Service experience quality can be evaluated using subjective or objective methods.

3.1.1 Subjective Evaluation Criteria

Subjective evaluation is that observers evaluate service experience quality based on their subjective perception. Subjective evaluation requires heavy workload and takes a long time, and is difficult to realize in commercial application scenarios. The results of subjective evaluation (testers provide the score based on their subjective experience) are used as reference for model training and verification.

The International Telecommunication Union (ITU) specifies many subjective quality evaluation methods. These methods have different characteristics, and the tester can choose the appropriate method according to the evaluation purpose.

The following table lists the common subjective evaluation methods.

Method		Standard
Category rating		
Absolute evaluation	Absolute Category Rating (ACR)	P.910, P.911, P.913, P.920

Method		Standard
Relative evaluation	Degradation Category Rating (DCR)	P.910, BT.500, P.911
	Comparison Category Rating (CCR)	BT.500
Continuous rating		
Absolute evaluation	Single Stimulus Continuous Quality Evaluation (SSCQE)	BT.500, P.911
Relative evaluation	Double Stimulus Continuous Quality Scale (DSCQS)	BT.500
	Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)	BT.500

The content of Cloud VR video streaming services is perceived. To evaluate video quality subjectively, DSCQS and ACR methods are commonly used in the industry.

- DSCQS is a reference evaluation model. That is, the testee watches multiple original reference videos and distorted videos, and scores each video based on the quality difference between the two types of videos.
- The ACR method does not use the original reference video. The testee scores each video directly. This method is typically used to evaluate the QoE model affected by the codec, network, or clients.

This document describes how subjective video quality evaluation results (MOS-ACR scoring) obtained using the ACR rating standard defined in ITU-T P.913 are used as a reference for model training and verification of the Cloud VR experience indicator modeling.

Subjective Evaluation	Excellent	Good	Fair	Poor	Very Poor
MOS-ACR	5	4	3	2	1

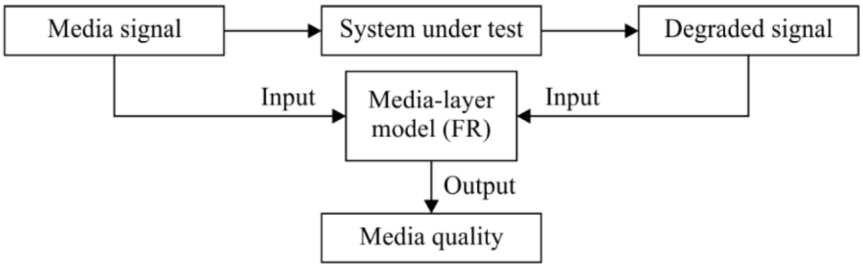
3.1.2 Objective Evaluation Criteria

Objective evaluation is to evaluate the user experience quality based on the quantitative quality indicators provided by the computer algorithm model. This method can select an objective, comprehensive, real, and standard indicator system, which is flexible and can apply to different scenarios. The modeling method of Huawei Cloud VR service experience is an objective evaluation method.

In terms of objective evaluation, the model methods used in the industry are classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) models. According to the ITU-T G.1011 protocol, the description of the models is as follows:

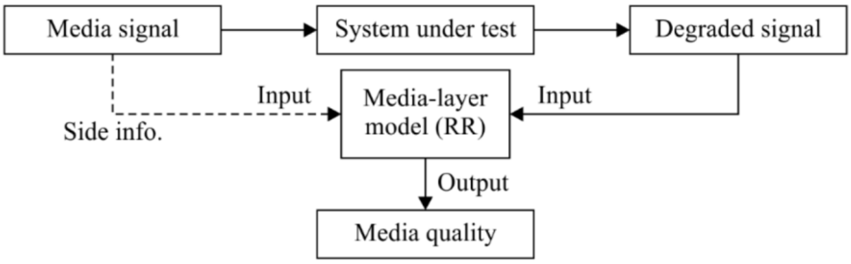
- FR method

  An entire original image or video is used as a reference, and is compared with the degraded image or video.



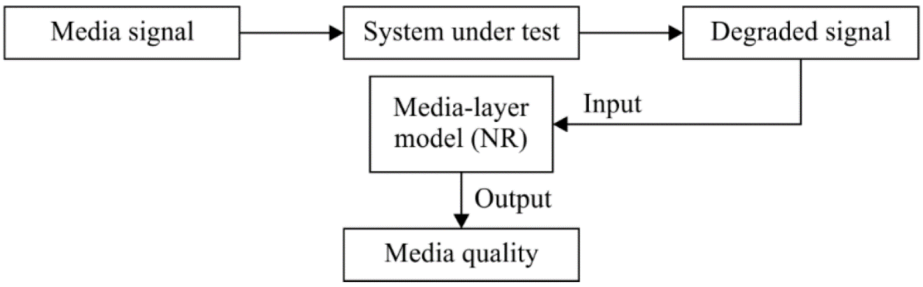
- RR method

  The original image or video does not need to be accessed. Only typical features of the original image or video need to be provided. The input information of this method is the result of comparison between the information of the reduced original image or video and the corresponding information of the degraded image or video.



- NR method

  This objective quality evaluation method does not need to access an original image or video, but evaluates quality by using information included in a related image or video media stream.



With the video streaming media as an example, the following table lists some standards based on which model suggestions on the three methods are provided.

	Type	Standard	Description
FR	Media-based model	[J.144] (SD) [J.247] (QCIF, CIF, VGA) [J.341] (HD)	This model performs intrusive evaluation and applies to DT and CQT scenarios. It is expensive.

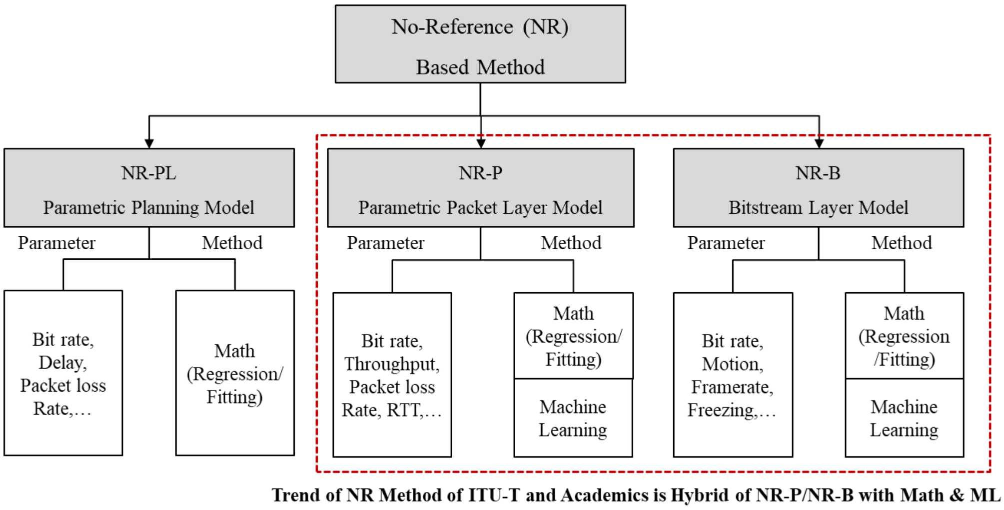
	Type	Standard	Description
RR	Pixel-based model	[J.249] (SD) [J.246] (QCIF, CIF, VGA) [J.342] (HD)	This model can decode video data packets into video signals for image analysis and processing. It is extremely expensive.
NR	Planning model	[G.1070] (NB/WB) [G.1071]	This model performs non-intrusive evaluation and applies to IPTV services. In addition, only network parameters are considered for this model.
	Packet based model/Bitstream-based model	[P.1201] [P.1202] [P.1203]	This model performs non-intrusive evaluation and supports terminal and network models based on the input mode. This model applies for service scenarios.

3.1.3 Huawei Model Selection

The FR and RR methods have great limitations in commercial application scenarios. Their limitations are as follows:

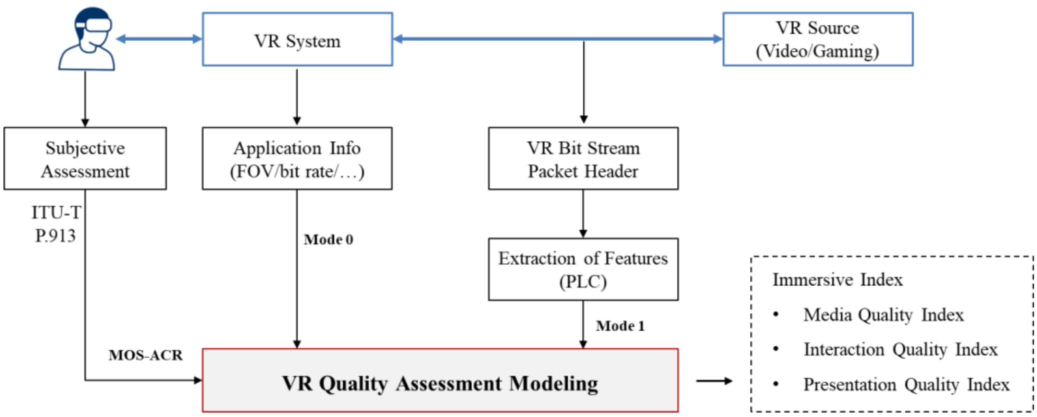
- Feasibility: Original videos and images need to be obtained, and the methods do not apply to the scenario where experience management needs to be performed for all videos.
- Security: The video content needs to be parsed, which causes network security risks.
- Effectiveness: The cost of FR and RR is high, their efficiency is low, and a long time is required to achieve the expected effect, which does not meet the requirements of network carriers and content providers for quick evaluation and improvement of network and video quality.

NR does not require original video or image information, and has better flexibility and universality. The NR method, driven by objective data and supported by the mathematical theory, is a popular application and research field. Modeling of the NR video quality evaluation method is generally classified into Parametric Planning Model (NR-PL), Parametric Packet Layer Model (NR-P), and Bitstream Layer Model (NR-B), as shown in the following figure.



It is noticed that NR-P and NR-B models are the mainstream development directions based on the NR development trend in ITU-T and academia, and machine learning methods are gradually incorporated in terms of modeling methods.

Huawei develops the NR-P and NR-B hybrid model on the basis of the P.1203 framework and VR service characteristics.



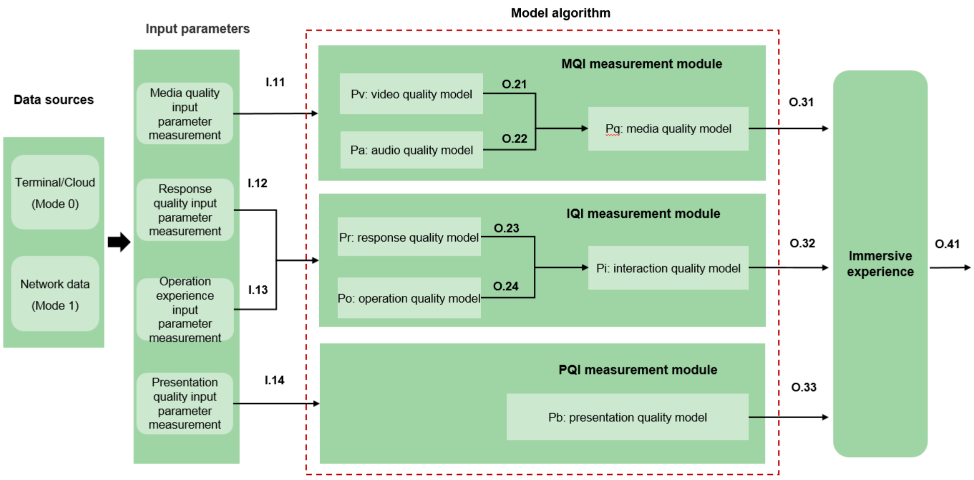
This model has the following advantages:

- This model is based on the existing specifications standards and framework (G.QOE-VR/TR 26.929/P.1203).
- With the data availability of network carriers and content providers considered, this model supports the device-cloud (Mode 0) and carrier network (Mode 1) data collection modes.
- To better meet service modeling requirements when carrier data acquisition is limited, multiple types of input parameters at the media layer (video source description, such as resolution, bit rate, frame rate, and FOV), packet header layer (frame type and frame size), and transport layer (packet loss rate, delay, and RTT) can be used.
- Based on the key factors provided by specifications, key parameters, such as network burst deterioration and continuous deterioration, are added to model input so that the model can better apply to the problem detection and optimization scenarios of carriers.

- In terms of the algorithm, this model not only uses the traditional mathematical fitting method, but also integrates the mature machine learning method to realize better reliability and generalization ability.

3.2 Modeling Framework

3.2.1 Overall Framework



- **[Model input]**
  - I.11: Input information of the media quality measurement module
  - I.12: Input information of the response quality measurement module
  - I.13: Input information of the operation experience quality measurement module
  - I.14: Input information of the presentation quality measurement module
- **[Model output]**
  - O.21: Video quality score in each measurement period. The video quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.22: Audio quality score in each measurement period. The audio quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.23: Response quality score in each measurement period. The response quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.24: Operation experience quality score in each measurement period. The operation experience quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.31: Media quality score in each measurement period. The media quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.32: Interaction quality score in each measurement period. The interaction quality score is displayed in each window for each session. The score ranges from 0 to 100.
  - O.33: Presentation quality score in each measurement period. The presentation quality score is displayed in each window for each session. The score ranges from 0 to 100.

- O.41: Immersion quality score in each measurement period. The immersion quality score is displayed in each window for each session. The score ranges from 0 to 100.

3.2.2 Data Sources and Input Parameters

3.2.2.1 Data Source Mode

Mode	Data Source	Applicable Scenario
0	Terminal/Cloud: Obtain application- or media-layer data from a terminal or a cloud server.	This mode applies to the scenario where the SDK can be embedded in the terminal and cloud server and real user experience needs to be measured.
1	Network: Obtain pipe-layer data from the pipe of the carrier's network.	This mode applies to the scenario where pipe probes can be deployed and only pipe quality needs to be analyzed.

3.2.2.2 Input Parameters

- I.11–Media quality input parameters

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Screen refresh rate	Integer	One session	**Mode 0** Device data	Pv: video quality model
Screen resolution	Length x Width (2880 x 1600)	One session		
FOV	Integer	One session		
Video bit rate	Float, kbps	Per segment	**Mode 0** Media information	
Video frame rate	Integer	Per segment		
Measurement period	Float, ms	Per segment		
Video resolution	Length x Width (2880 x 1600)	Per segment		
Video codec and profile	One of the following formats: H264-baseline, H264-high, and H264-main	One session		
Video frame number	Integer, starting at 1, denoting the frame sequence number in encoding order	Per frame		

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Video frame duration	Float, ms	Per frame		
Frame presentation timestamp	Frame presentation timestamp	Per frame		
Frame decoding timestamp	Frame decoding timestamp	Per frame		
Video frame size	Size of the encoded video frame in bytes	Per frame		
Video frame type	"I" or "Non-I" for mode 1 "I"/"P"/"B" for modes 2, 3	Per frame		
Average throughput	Float, kbps	Per segment	**Mode 1** Network data	
Throughput variance	Float, kbps	Per segment		
Max throughput	Float, kbps	Per segment		
Max burst size	Float, Kbyte	Per segment		
Burst pulse number	Integer	Per segment		
Measurement period	Float, ms	Per segment	**Mode 0** Media information	Pa: audio quality model
Audio bit rate	Float, kbps	Per segment		
Audio frame number	Integer	Per segment		
Audio frame size	Float, Kbyte	Per frame		
Audio frame duration	Float, ms	Per frame		
Audio codec	One of the following formats: AAC-LC, AAC-HEv1, AAC-HEv2, and AC3	Per segment		
Audio sampling frequency	Integer, HZ	Per segment		
Number of audio channels	Integer	Per segment		

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Audio bit-stream	Float, Kbyte	Per frame		

- I.12–Response quality input parameters

Input Parameter	Value	Acquisition Frequency	Data Source	Module
MTS delay	Float, ms	Per segment	**Mode 0** Media information	Pr: response quality model
MTP delay	Float, ms	Per segment		
Black border rate (BLR)	Float, %	Per segment		
N-MTP delay	Float, ms	Per segment	**Mode 1** Network data	
N-MTP delay variance	Float, ms	Per segment		
Burst delay rate	Float, %	Per segment		
Average UL RTT	Float, ms	Per segment		
Average DL RTT	Float, ms	Per segment		
Burst RTT rate	Float, %	Per segment		
RTT variance	Float, ms	Per segment		

- I.13-Operation experience input parameters

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Head DOF	Integer,0,3,6	One session	**Mode 0** Device data	Po: interaction quality model
Hand DOF	Integer,0,3,6	One session		
Body DOF	Integer,0,3,6	One session		
Weights of head mounted display (HMD)	Integer, gram	One session		
Spatial precision	Integer, centimeter	One session		

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Interaction device	ENUM, Gamepad, key board and mouse, No controller	One session		

- I.14–Presentation quality input parameters

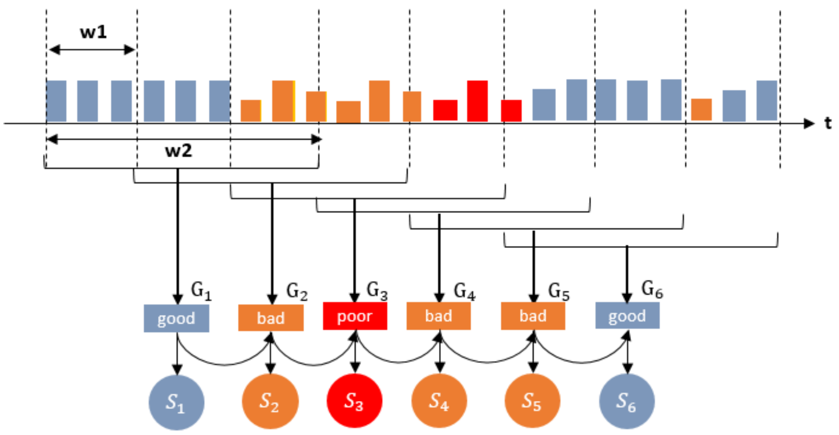
Input Parameter	Value	Acquisition Frequency	Data Source	Module
Total duration	Float, ms	One session	**Mode 0** Application cache data	Pb: presentation quality model
Loading delay	Float, ms	One session		
Stalling/freezing start time	Float, ms	Per event		
Stalling/freezing end time	Float, ms	Per event		
Tiling artifacts/mosaic start time	Float, ms	Per event		
Tiling artifacts/mosaic end time	Float, ms	Per event		
Delay from the time when the SYN message is sent to the time when the ACK message is received	Float, ms	One session	**Mode 1** Network data	
Delay from the time when the SYNACK message is sent to the time when the ACK message is received	Float, ms	One session		
Delay from the time when the ACK message is received to the time when the first data packet is sent	Float, ms	One session		

Input Parameter	Value	Acquisition Frequency	Data Source	Module
Average packet loss rate	Float, %	Per segment		
Burst packet loss rate	Float, %	Per segment		
Burst interval rate	Float, %	Per segment		
Average DL RTT	Float, ms	Per segment		
Burst RTT rate	Float, %	Per segment		
RTT variance	Float, ms	Per segment		
Burst pulse number	Integer	Per segment		
Average throughput	Float, kbps	Per segment		
Throughput variance	Float, kbps	Per segment		
Max throughput	Float, kbps	Per segment		

3.2.3 Model Algorithm

3.2.3.1 Universal Measurement Mechanism

This section describes the universal measurement mechanisms used by three quality indicators: sliding window mechanism, recency effect mechanism, and experience residual effect.

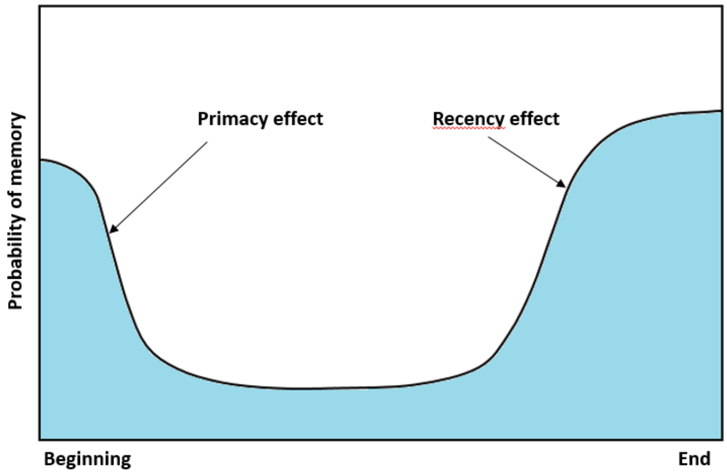


- Sliding window mechanism description

  "Window" in the "sliding window mechanism" refers to the time range (w2 in the preceding figure) selected during the measurement. The entire experience process of a user is divided into multiple windows (time ranges) for measurement, which can more comprehensively reflect user experience quality. "Slide" in the "sliding window mechanism" refers to the time step (w1 in the preceding figure) of each movement of the measurement window. The sliding distance of the window can be adjusted as required.

- Recency effect description

  Recency effect means that people have deeper impression of more recent information. Thus, the more recent information has a higher weight than the information presented earlier when judgment is made. The sliding window mechanism measures the service quality in each time range and provides the score (G_i). However, as time passes, the score (G_1) in the first time range has the minimum impact on the subjective experience when a service ends.



As a classical method for evaluating the recency effect, EWMA indicates that the weighting coefficient of each value decreases exponentially with time. The closer current time results in a larger weighting coefficient. Therefore, experience of the current moment t may be calculated by using the following formula:

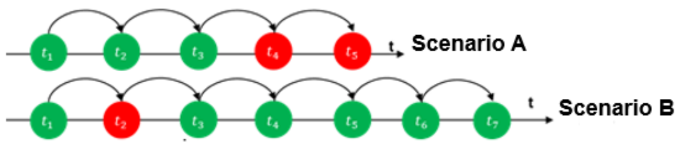
$$S_t = \beta S_{t-1} + (1-\beta)G_t$$

In the preceding formula, G_t is an experience quality index measured by window t (time t), β indicates a weighting coefficient, and S_t indicates the overall experience quality index by time t.

- Experience residual effect description

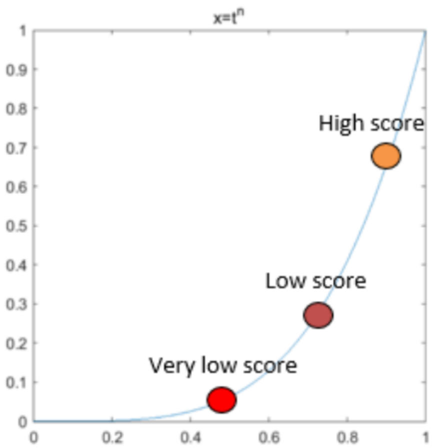
  The following scenarios are involved:

  - Scenario A: Although user experience is good at the beginning, the user tends to give up after experiencing continuous poor quality services.
  - Scenario B: Although the overall evaluation is good, occasional poor experience in the past degrades the final user experience evaluation.



Users get more stimulation if user experience is poor. This phenomenon is called residual experience effect. Therefore, when calculating the final S_t experience result, you need to amplify the impact of previous poor experience on the final experience.

The n function $f(x) = x^n, x \in [0,1]$ is used to process the periodic calculation result. Poor experience is amplified, and the impact on good user experience is not great.



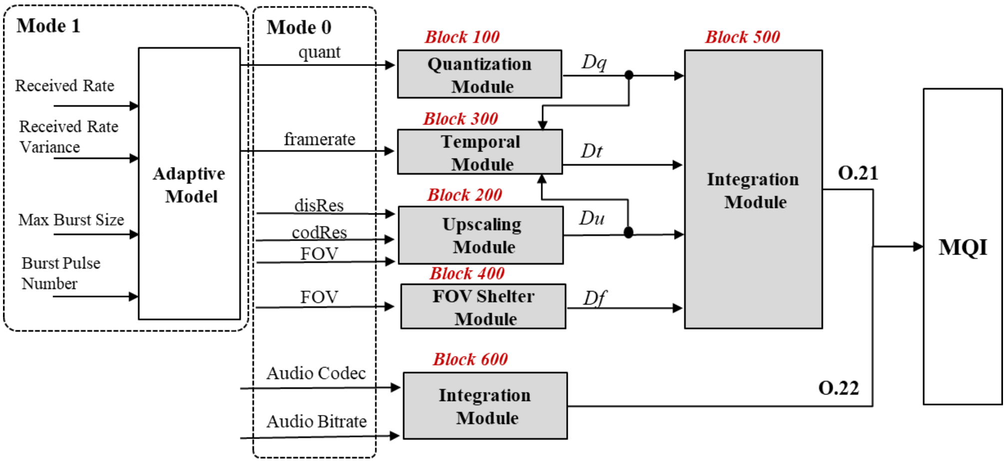
The EWMA function optimized by the n function is as follows:

$$S_t = \beta S_{i-1}{}^n + (1 - \beta) G_t{}^n$$

The preceding formula incorporates the advantages of the EWMA algorithm and combines the experience residual effect and recency effect.

3.2.3.2 MQI Measurement Module

The video quality and audio quality are modeled based on the idea of ITU-T P.1203 specifications. PPD and FOV are added to the core parameters, and the formulas are adjusted.



- Mode 0. If carriers can obtain input parameters from terminals and cloud servers, this mode can be used to calculate the media quality experienced by users.

  In video and game scenarios, the head-end needs to provide information, such as the resolution, FOV, bit rate, and frame rate.

- Mode 1. In most cases, carriers have only objectively measured input parameters and cannot obtain input information from the player. They can evaluate the VR media quality based on network measurement indicators, including the throughput, throughput fluctuation, and pulse information. The measurement data is used in mode 1. The measurement data passes through the adaptation layer, some information equivalent to mode 0 is generated. Data except for the equivalent information uses the default values.

NOTE

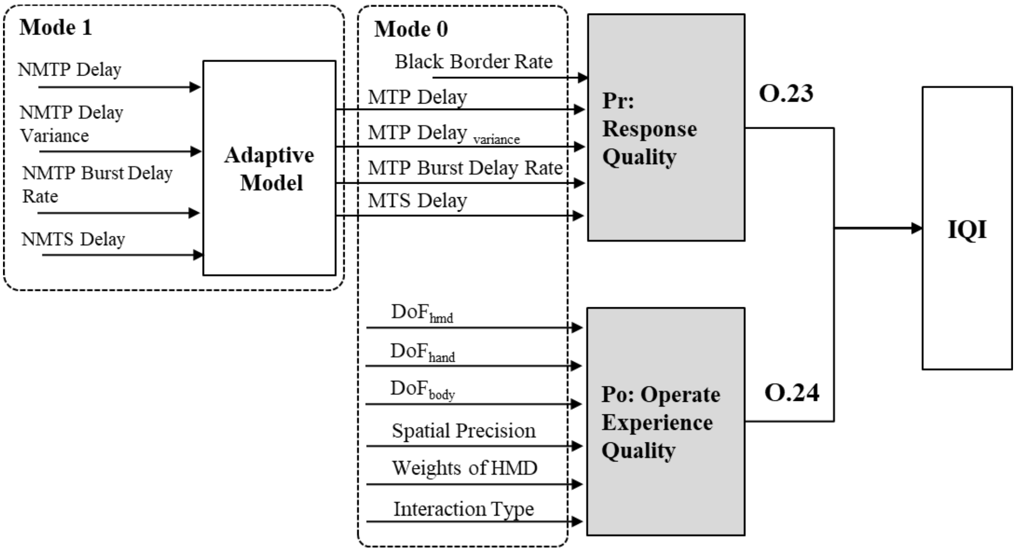
- Block 100: is used to describe the quality impairment caused by the amount of content source information. The input information mainly includes the source resolution, bit rate, and frame rate. D_q indicates the impairment.
- Block 200: is used to describe the media quality impairment caused by the space information. The input information includes the source resolution, device resolution, and FOV. The evaluation method is to calculate the VR media quality impairment caused by the PPD. Perception of human eyes on VR services is in a manner similar to a spherical display effect rather than a plane of conventional videos. A larger value of PPD indicates better VR display quality. D_u indicates the impairment.
- Block 300: is used to describe the media quality impairment caused by the time information. The main input information is the frame rate. In addition, the function of D_q and D_u in time is considered. D_t indicates the impairment.
- Block 400: is used to describe the VR media quality impairment caused by the blocking effect. The main input information is FOV. When FOV is less than the full-coverage field angle sensed by human eyes, human eyes may sense that an image is blocked. D_f indicates the impairment.

  For FOV defined in the specifications, the optimal FOV is 210 degrees. It is divided into three FOVs. The central FOV must be 60 degrees, the auxiliary FOVs on both sides must be 30 degrees, and the single-eye FOV on the edge must be 35 degrees. Currently, the VR HMD is 90 degrees at the primary level and 110 degrees at a higher level. Such a VR HMD only covers the central FOV, but does not cover the entire auxiliary FOVs. This VR helmet cannot reach the entry VR standard of 120 degrees. The maturity standard is 210 degrees.

- Block 500: is used to describe the impact of the preceding impairment on the VR media quality. The input information is the sum of the preceding impairment values, and the output information is the video quality evaluation score O.21.
- Block 600: According to specifications P.1203, it is recommended that the frame rate be used as the input and the audio quality evaluation score O.22 be used as the output.
- In mode 1, if there is no application-layer input information, such as the resolution and frame rate, the adaptation module can be used to convert the input information into the input parameters of blocks 100 to 400 based on the parameters measured on the pipe for quality evaluation.

3.2.3.3 IQI Measurement Module

This indicator applies to scenarios where users interact with the GUI, such as, VR games. Interaction experience involves the following factors: response quality (O.23) and operation experience quality (O.24).

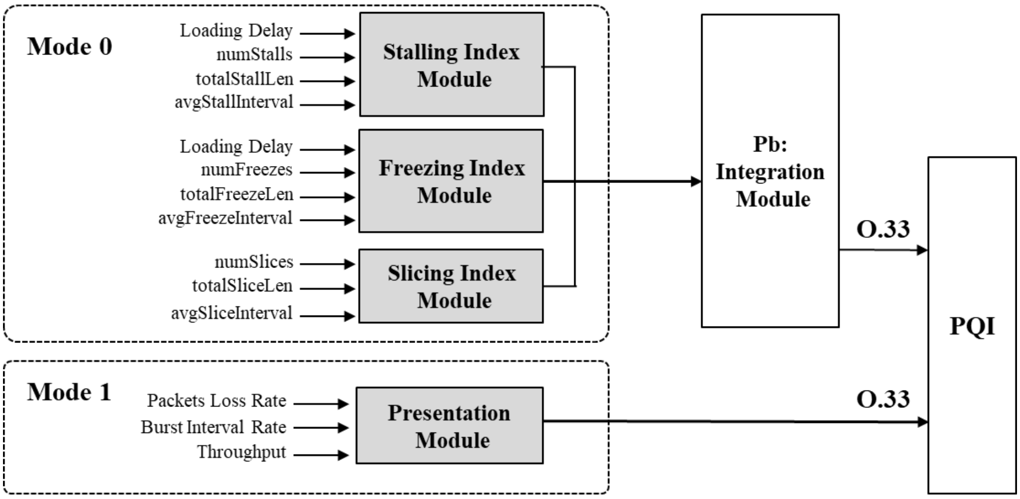


- Mode 0. If carriers can obtain input parameters from terminals and cloud servers, this mode can be used to calculate the quality of interaction between users and the VR system.

  In the game scenario, the head-end needs to provide information, such as the interaction delay, BLR, and DOF.
- Mode 1. In most cases, carriers have only objectively measured input parameters and cannot obtain input information from the player. They can evaluate the user interaction quality based on network measurement indicators, including the network delay information.
- Pr: Response quality. It reflects the impact of network delay on user experience.

  In mode 1, if there is no input information from terminals and cloud servers, the adaptation module can be used to convert the input information into the input parameters corresponding to the response quality model based on the parameters measured on the pipe for quality evaluation.
- Po: Operation experience quality. It is used to measure the usability and comfort of terminals. The operation experience quality cannot be measured from the network side and needs to be measured based on terminal information used as the input information.

3.2.3.4 PQI Measurement Module



The calculation of this indicator supports the following data source modes:

- Mode 0. If carriers can obtain input parameters from terminals and cloud servers, this mode can be used to calculate the visual experience of users when users watch videos or game images.
  - In the video scenario, the player needs to provide the initial buffering delay and stalling information.
  - In the game scenario, the head-end needs to provide freezing and tiling artifacts/mosaic information (freezing events are evaluated with stalling events).
- Mode 1. In most cases, carriers have only objectively measured input parameters and cannot obtain input information from the player. Therefore, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) algorithm can be used to model the presentation quality.

3.2.3.5 Immersive Experience Score

The immersive experience score is a combination of the values of MQI, IQI, and POI.

$$0.41 = \omega_1 * 0.31 + \omega_2 * 0.32 + \omega_3 * 0.33$$

The weight ω is obtained using the random forest algorithm and can be manually modified.

4 Model Verification

4.1 Test Procedure

The test has been conducted by more than 30 Huawei Open Lab engineers for more than 600 times and lasts longer than 6500 minutes.

4.2 Verification Objective

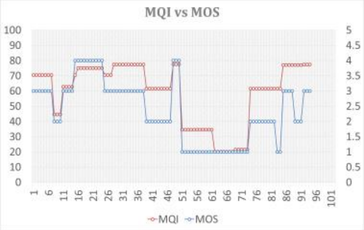
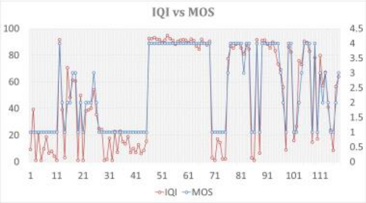
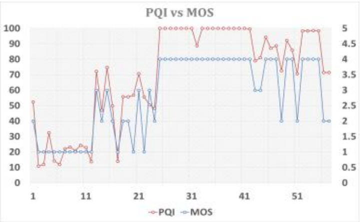
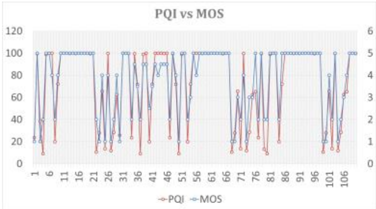
Model verification aims to:

- Verify the model accuracy. The accuracy is measured based on the correlation between the experience indicators (MQI, IQI, and PQI) calculated by the model and the subjective experience MOS.
- Provide reference values of key factors. The relationship between the experience indicators (MQI, IQI, and PQI) and each key factor is established, and the reference value of each key factor is provided.

4.3 Model Accuracy Test Result

Select two types of Cloud VR services (VR games and VR videos), and ask the evaluator to evaluate the Cloud VR services under different conditions to obtain the MOS-ACR score. In addition, calculate the MQI, IQI, and PQI of Cloud VR services based on the training model, and analyze the correlation between the subjective MOS and the MQI, IQI, and PQI.

After scoring and verification for many times, the experience indicators calculated using the current Huawei model can accurately reflect subjective experience changes. The following table describes the correlation between the three indicators and MOS.

	VR Game	**VR Video**
MQI	Correlation with MOS: 0.882 	Correlation with MOS: 0.944 
IQI	Correlation with MOS: 0.96 	Correlation with MOS: N/A
PQI	Correlation with MOS: 0.925 	Correlation with MOS: 0.948 

4.4 Model and Key Factor Test Results

To ensure good user experience of the Cloud VR services, the recommended values of the related factors are as follows:

- Cloud VR videos

Code Rate	Screen Resolution	Video Resolution	Loading Delay	Stalling Proportion
> 30 Mbit/s	4K	IMAX-4K 360 video – 8K	< 2s	< 5%

- Cloud VR games

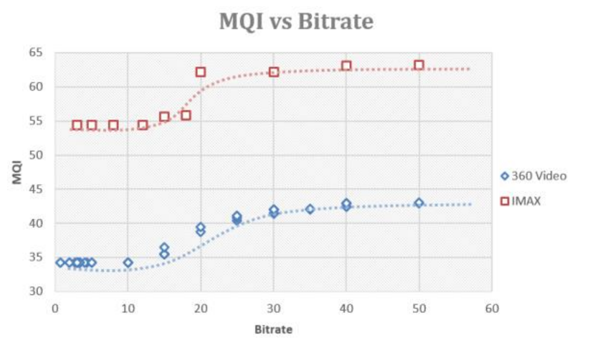
Frame Rate	> 60 FPS
Code Rate	> 50 Mbit/s
FOV	> 100°
Screen Resolution	4K
Video Resolution	4K
MTP Delay	< 100 ms
BLR	< 1%
DOF	> 11
Throughput	> 52 Mbit/s
Throughput Fluctuation	< 2 Mbit/s
Packet Loss Rate	< 0.3%

4.4.1 Cloud VR Videos

4.4.1.1 MQI

Relationship between the MQI and Bit Rate

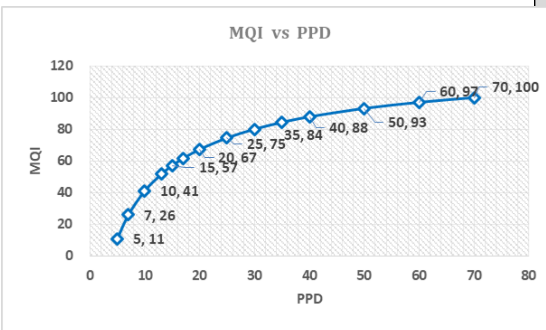
The MQIs of 360 videos and IMAX videos do not bring significant gains when the bit rate exceeds 30 Mbit/s. The bit rate 30 Mbit/s is recommended to meet user experience requirements.



MQI		Bit Rate (Mbit/s)
360 Video	IMAX Video	
43	63	50
41	62	30
39	60	20
35	56	15
34	54	10

Relationship between the MQI and PPD

The value of PPD corresponds to some video pixel scenarios. Therefore, it is recommended that the screen resolution and video resolution be set to 4K and 8K for 360 videos to obtain better user experience, respectively. For IMAX videos, both the screen resolution and video resolution need to be set to 4K to ensure better user experience.

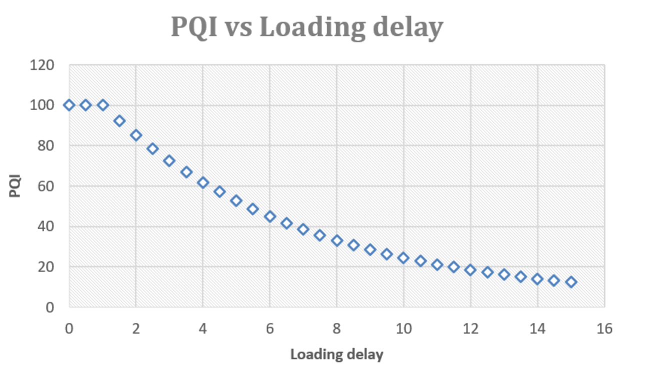


360 Video		IMAX Video		Screen Resolution–Video Resolution
MQI	PPD	MQI	PPD	
18	3	40	12	2K-2K
36	4	56	16	3K-3K
48	5	67	21	4K-4K
67	11	67	21	4K-8K
74	11	88	43	8K-8K

4.4.1.2 PQI

Relationship between the PQI and Loading Delay

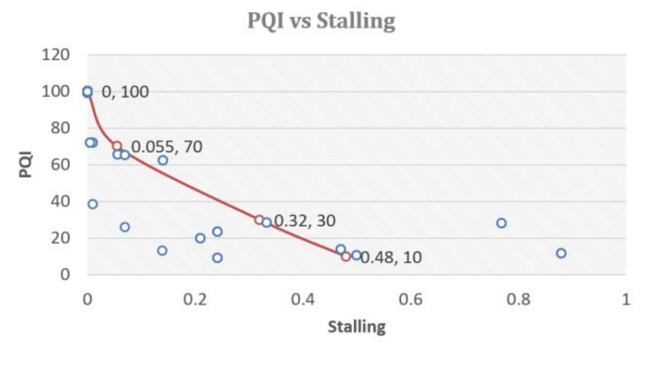
The prerequisite for good PQI is that the loading delay is less than 2s, as shown in the figure of the following table.



PQI	Loading Delay (s)
100	1
80	2
60	4
40	6
20	10

Relationship between the PQI and Stalling Proportion

The following table lists the typical values of the PQI and stalling proportion of Cloud VR videos. **The stalling proportion less than or equal to 5%** ensures good user experience.



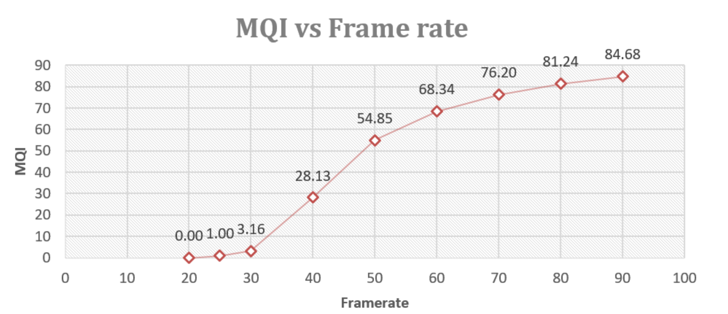
PQI	Stalling Proportion
100	0%
80	5%
60	14%
30	33%
10	48%

4.4.2 Cloud VR Games

4.4.2.1 MQI

Relationship between the MQI and Frame Rate

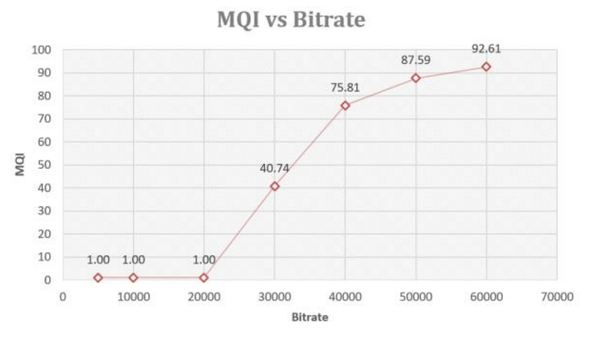
The bit rate is fixed at 50 Mbit/s, the encoding resolution and display resolution are both 2880 × 1600, and the FOV is 110°. It is recommended that the frame rate is 60 FPS,for basic user experience to be ensured.



MQI	Frame Rate
85	90
81	80
76	70
68	60
55	50
28	40

Relationship between the MQI and Bit Rate

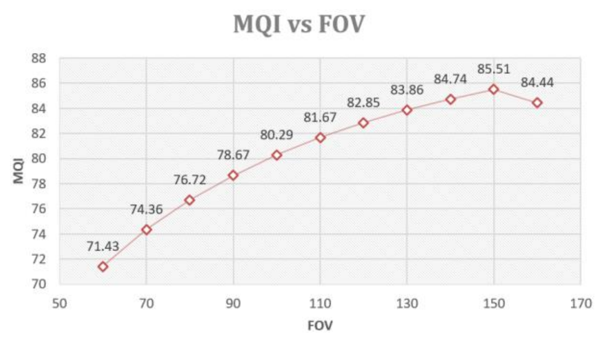
The frame rate is fixed at 25 FPS, the encoding resolution and display resolution are both 2880 × 1600, and the FOV is 110°. It is recommended that the bit rate be set to 50,000 kbit/s to ensure good user experience.



MQI	Bit Rate (kbit/s)
93	60000
88	50000
76	40000
60	35000
41	30000

Relationship between the MQI and FOV

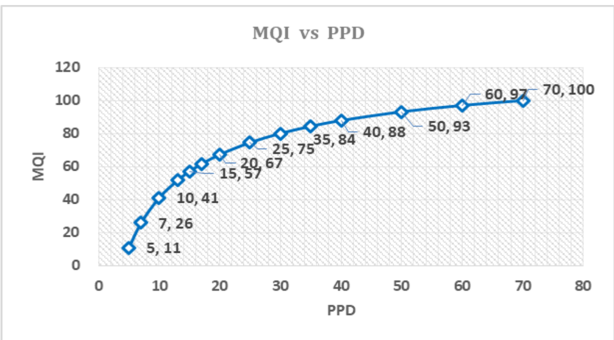
The frame rate is fixed at 25 FPS, the bit rate is 50 Mbit/s, and the encoding resolution and display resolution are both 2880 × 1600. The FOV ranges from 100° to 120°, and good user experience can be obtained within this FOV range.



MQI	FOV
86	150
85	140
83	120
80	100
77	80
71	60

Relationship between the MQI and PPD

It is recommended that both the screen resolution and video resolution be set to 4K for Cloud VR games to ensure good user experience.

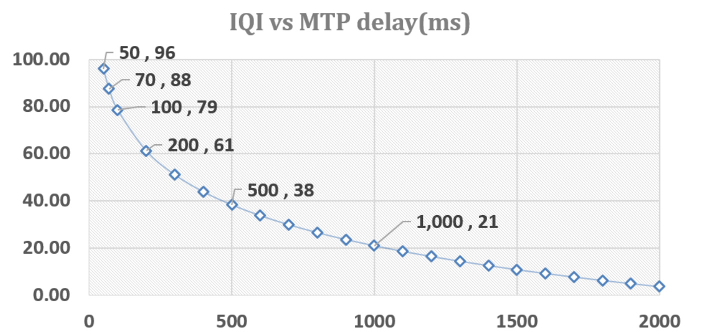


MQI	PPD	Screen Resolution-Video Resolution
40	10	2K-2K
56	14	3K-3K
67	19	4K-4K
88	35	8K-8K

4.4.2.2 IQI

Relationship between the IQI and MTP Delay

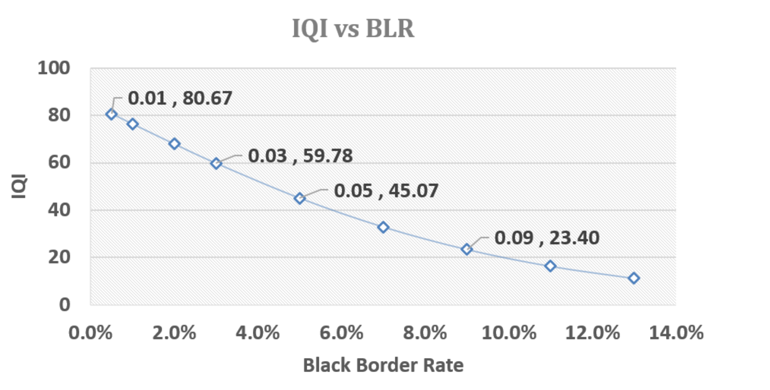
If the MTP delay is less than or equal to 50 ms, the score of the IQI is close to the full score (96 points). If the MTP delay is less than or equal to 70 ms, the score of the IQI is 88 points, indicating good user experience. If the MTP delay is greater than 100 ms, user experience is poor.



IQI	MTP Delay (ms)
96	50
88	70
79	100
61	200
38	500
21	1000

Relationship between the IQI and BLR

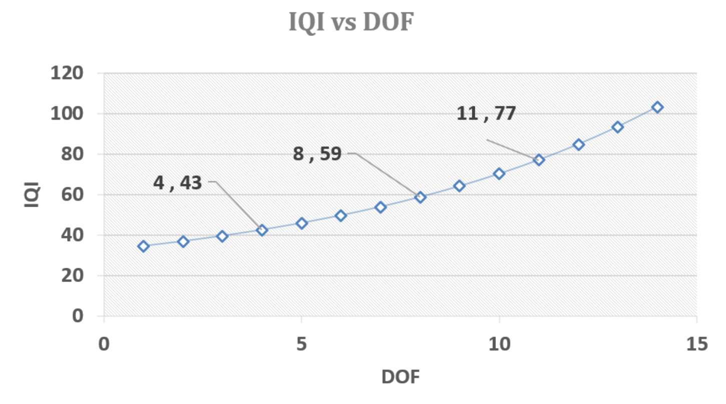
If the BLR is less than or equal to 1%, the score of the IQI is 80 points. In this case, users do not feel the black border.



IQI	BLR
81	1%
60	3%
45	5%
23	9%

Relationship between the IQI and DOF

The following table lists the typical values of the IQI and DOF of Cloud VR games. It is recommended that the DOF be at least 12 degrees to ensure good gaming experience.

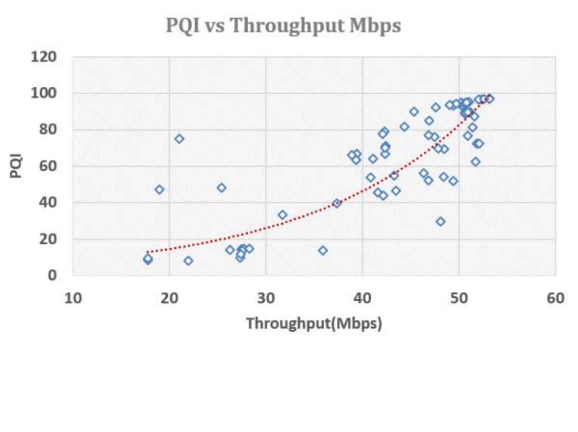


IQI	DOF
92	13
76	11
58	8
42	4
34	1

4.4.2.3 PQI

Relationship between the PQI and Throughput

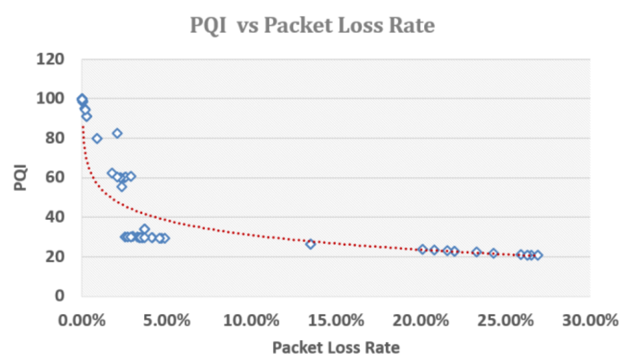
If the throughput is greater than 52 Mbit/s and the throughput fluctuation is less than 2 Mbit/s, excellent user experience can be ensured.



PQI	Throughput (Mbit/s)	Throughput Fluctuation (Mbit/s)
87	52	2
77	51	7
55	43	-
33	32	-
14	26	-

Relationship between the PQI and Packet Loss Rate

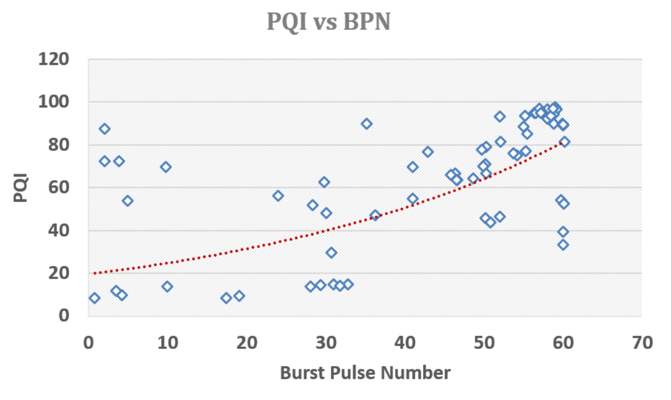
If the packet loss rate is less than 0.02%, user experience is excellent. If the packet loss rate is less than 0.3%, user experience is good.



PQI	Packet Loss Rate
100	0.02%
80	0.3%
60	2.3%
30	3.3%
21	23.9%

Relationship between the PQI and BPN

If the value of BPN is greater than 50, good user experience is ensured. If the value of BPN is 59, excellent user experience is ensured.



PQI	BPN
90	59
71	50
47	36
30	31
14	10

5 References

1. ITU-T P.1203 Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Video quality estimation module.
2. ITU-T WD14-Restructured version of G.QoE-VR baseline.
3. ITU-T P.809 Subjective evaluation methods for gaming quality.
4. ITU-T G.1032 Influence factors on gaming quality of experience.
5. ITU-T P.913 Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.
6. Q. Huynh-Thu, M. Garcia, F. Speranza, P. Corriveau and A. Raake, "Study of rating scales for subjective quality assessment of high-definition video," IEEE Transactions on Broadcasting, vol. 57. No. 1, p. 1-14, Mar. 2011.
7. T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," Quality of Multimedia Experience (QoMEX), Jun. 2010.
8. Huyen T. T. Tran_, Nam Pham Ngocy, Cuong T. Phamy, Yong Ju Jungz, Truong Cong Thang_"A Subjective Study on QoE of 360 Video for VR Communication" in conference: 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP).
9. Selim Ickin, Karel De Vogeleer, Markus Fiedler, David Erman, "The Effects of Packet Delay Variation on the Perceptual Quality of Video", 2010 4th IEEE Workshop On User MObility and VEhicular Networks.
10. Taewan Kim, Jiwoo Kang, Sanghoon Lee, "Multimodal Interactive Continuous Scoring of Subjective 3D Video Quality of Experience" IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 16, NO. 2, FEBRUARY 2014.
11. Fr´ed´eric Guyard and Sergio Beker, "Towards real-time anomalies monitoring for QoE indicators", in annals of telecommunications - annales des télécommunications 65(1-2): 59-71 February 2010.
12. M. Sajid Mushtaq*, Scott Fowler†, Brice Augustin*, Abdelhamid Mellouk* , "QoE in 5G Cloud Networks using Multimedia Services", Conference: IEEE Wireless Communications and Networking Conference (WCNC'16).
13. Victor Clincy and Brandon Wilgor, "Subjective Evaluation of Latency and Packet Lossin a Cloud-based Game", 2013 10th International Conference on Information Technology: New Generations.
14. Kazuhisa Yamagishi NTT Network Technology Laboratories, NTT Corporation, Tokyo, Japan, "QoE-estimation models for video streaming services", Proceedings of APSIPA Annual Summit and Conference 2017.