In 2018, the world witnessed the first trial commercial deployment of operator-grade Cloud VR services, opening a strategic window for Cloud VR industry development. During the trial, we found that the fast integration of service systems and acceptance, standardization of products and solutions, and satisfaction of user experience requirements were key issues in Cloud VR development.

This document is created and published jointly by Huawei Technologies Co., Ltd. and Video Experience Alliance. It describes the industry's first Cloud VR user experience and evaluation white paper. Based on typical cloud VR application scenarios, this document describes the key processes in Cloud VR service solutions, defines and breaks down the key factors affecting Cloud VR experience, determines the Cloud VR user experience model and quantifies component evaluation items and indicators to promote standardization and help operators improve user experience and develop Cloud VR services.
01 Research Purpose ......................................................... 03

02 Research Idea ......................................................... 05

03 Cloud VR Application Scenarios and Solutions ....................... 07
   3.1 Cloud VR Solution Architecture ..................................... 08
   3.2 Key Service Processes Affecting Cloud VR User Experience .... 09

04 Cloud VR User Experience Model .................................... 13
   4.1 Key Cloud VR User Experience Requirements Based on User Survey 14
   4.2 Cloud VR User Experience Modeling Based on Research Results .... 15

05 Cloud VR Service Evaluation Baseline ................................ 21
   5.1 Mapping Between Cloud VR User Experience Requirements and Evaluation Items 22
   5.2 Strong-Interaction Cloud VR Service .................................. 23
   5.3 Cloud VR Video Services ............................................... 28

06 Appendix ............................................................... 33
01 Research Purpose
Simplify Cloud VR Solution Integration, Acceptance, and Evolution

During the commercial delivery of Cloud VR, product specifications and implementation methods vary with vendors, adversely affecting the project progress and user experience. Therefore, a unified set of product specifications and service acceptance criteria need to be established.

Determine the Specifications of Cloud VR Components and Promote Standardization of the Industry

The key VR technologies have become mature, and the industry is developing fast. However, the further development of Cloud VR services and large-scale commercial use still require unified products and solution standards.

Make Cloud VR Services Evaluable and Measurable to Ensure User Experience

With the popularity of Cloud VR and the continuous development of the industry, users are expecting better experience. Cloud VR used to attract only curiosity of users; now it has their full attention and is expected to offer superb experience in content presentation and operation response. Evaluating and assuring Cloud VR user experience is a key factor for the development and wide commercial use of Cloud VR services.

This document analyzes and studies key technologies and user experience requirements of Cloud VR services, and provides reference and guidance for operators and industry partners to standardize service experience elements and product performance requirements and jointly promote the openness and development of the industry.
02 Research Idea
This document breaks down Cloud VR experience into indicators of functional components and modules based on the key experience-related service processes learned from VR cloud application scenarios and solutions, user surveys and feedback, and research about user experience evaluation in the industry. This document also provides evaluation indicators, tools, and methods in line with industry development, lab test results, and project delivery experience.

In the future, we hope to improve user experience, guide project integration acceptance, and facilitate industry standardization by quantifying and optimizing service and component indicators.

**Application Scenario:** Cloud VR services are classified as strong-interaction services and video services. Strong-interaction services allow users to interact with cloud applications in real time. Typical applications include games and education applications. Video services meet user demand for surround view, with on-demand panoramic video and VR live broadcast being typical applications. Differentiated solutions are developed for strong-interaction and video services.

**Technology Solution:** Asynchronous rendering between the cloud and terminals relieves motion sickness in strong-interaction services. Compared with full-view transmission, field of view (FOV) transmission can reduce needed transmission bandwidth and terminal decoding performance in video services. The solutions depend on collaboration between key components, such as the content, platform, network, and terminal.

**Experience Model:** In Cloud VR user experience evaluation, user experience requirements are classified into immersive experience requirements and interactive experience requirements. These requirements focus on content, presentation quality, operation, and response quality.

**Evaluation Baseline:** The complete Cloud VR service evaluation baseline includes specific quantifiable service and component specifications, interface specifications, and tools and methods required to evaluate these specifications.

VR OpenLab defines three phases of Cloud VR development: fair-experience, comfortable-experience, and ideal-experience. Each phase has its own application scenarios, solutions, and user experience requirements. The Cloud VR service evaluation baseline will evolve with industry development.

**NOTE:**
The research objects in this document are those in black in the previous figure. The contents in blue are planned for later versions.
03

Cloud VR Application Scenarios and Solutions
As shown in the preceding figure, a Cloud VR service solution consists of the following system components:

**Content:** This component provides high-quality content that meets specific scenarios and requirements for the service platform. Currently, Cloud VR provides panoramic videos, live broadcast, games, and educational content.

**Platform:** Based on service scenarios, the Cloud VR video and cloud rendering platforms can be used. The Cloud VR video platform is responsible for uploading, transcoding, storing, and distributing VR videos. Its system structure is similar to that of operators’ IPTV and OTT video platforms, and multiple components can be reused. The Cloud VR cloud rendering platform is responsible for computing, rendering, encoding, and streaming of user operation instructions in games and education scenarios.

**Network:** Networks include the backbone, metro, access, and home networks. They provide stable, high-bandwidth, and low-latency transmission for Cloud VR services. The home Wi-Fi network is a key challenge in Cloud VR services.

**Terminal:** The terminal is responsible for service access, user authentication, content presentation, and spatial positioning of Cloud VR. Currently, most standalone headsets provide only three degrees of freedom (3DOF) spatial positioning and require positioning suites to support 6DOF.
### 3.2 Key Service Processes Affecting Cloud VR User Experience

During research, we analyzed various application scenarios and solutions and extracted key service processes that affect user experience.

#### 3.2.1 Strong-Interaction Cloud VR Service

Strong-interaction Cloud VR services allow users to interact with cloud applications in real time through terminal sensors. Cloud applications calculate, render, compress, and encode interactive instructions, then stream the video to terminals for decoding and display.

#### Asynchronous Rendering Between the Cloud and Terminal

Motion sickness and interactive latency are key experience problems faced by all strong-interaction Cloud VR services. Asynchronous rendering between the cloud and terminal mitigates this issue, but causes black edges and smear. The following figure shows the process.
**Motion Sickness:** The Motion-to-Photon (MTP) latency is the time needed for a user's physical movement to be reflected in the virtual world on the display screen. The consensus in the industry is that an MTP shorter than 20 ms is necessary to minimize motion sickness. The cloud rendering used in traditional solutions to refresh images displayed on screens cannot meet this requirement. Asynchronous rendering uses asynchronous timing and space warping technologies, performing predictive adjustment and frame insertion processing (header motion rendering) on the pre-rendered images according to the posture or position change of the head. In this way, the MTP process can be completed on the terminal, no longer depending on cloud rendering and streaming to minimize user dizziness. Cloud rendering and streaming is a process in which a terminal submits an instruction to the cloud, and the cloud performs rendering, encoding, and streaming. Then, the terminal receives and decodes the streams. This is an extra process of Cloud VR relative to local VR during service processing.

**Interactive Latency:** In interactive applications such as Cloud VR games and education, when a user moves, rotates their headset, pulls the trigger, or moves the handle, the user hopes to experience an instant visual and audio response. Based on the experience of most online shooting games, people feel as though response is instantaneous if the interactive latency is less than 100 ms. In terms of the processing flow, operation response requires motion capture and rendering display processes in addition to cloud rendering and streaming. In the future, the latency in these processes will be optimized.

**Black Edge and Smear:** Asynchronous time warping is used in asynchronous rendering. The technology can predict and offset pre-rendered images based on the user head rotation information, but cannot generate images out of nothing. Therefore, when the user rotates the head quickly, the areas beyond the original field of view are displayed as black edges or smears. The black edge and smear effect is more severe when the rotation is fast or the streaming and cloud rendering delay is long. One way to minimize the black edge and smear effect is to perform extra-perspective rendering, but this consumes more end-to-end system resources.

Another way is to reduce cloud rendering and streaming latency. According to test results, users find the smear or black edge acceptable if the streaming and cloud rendering delay is limited to 70 ms. In the future, the delay will be further reduced to 30–50 ms.
3.2.2 Cloud VR Video Services

On-demand panoramic video and VR live broadcast are typical application scenarios of Cloud VR. Based on the transmission mode and content, panoramic video transmission solutions can be classified as full-view transmission and FOV transmission.

Full-View Video Transmission Solution

Full-view transmission is widely used in the fair-experience phase of Cloud VR video services. As shown in the following figure, the cloud pushes 180°/360° panoramic content to the terminal. The terminal tracks the user’s posture change and parses, renders, and outputs the audio and video data cached locally. This process is the same as that of OTT video streaming, except that the process must also meet MTP latency requirements. A major problem with the process is insufficient download throughput, which causes frame freezing.

In addition to easy implementation, another advantage of this solution is that the displayed images are of the same quality regardless of the viewing angle. The downside is that content beyond the viewing angle is also transmitted (a waste of bandwidth) and the terminal must support the same resolution and bitrate as the full-view content. However, most VR terminals can only hard-decode 4K panoramic videos, and not 8K panoramic videos.
**FOV Video Transmission Solution**

With continuous improvement of the resolution, frame rate, and bitrate of video, full-view transmission will overwhelm the network transmission bandwidth and terminal decoding capabilities. To solve this problem, the FOV transmission solution was developed to transmit and decode only partial tiles based on the viewing angles. This solution supports 8K and 16K video. However, it also complicates service processing flow and introduces experience problems caused by delayed image quality switching. The following figure shows the process.

---

### Delayed Image Quality Switching

In Huawei TWS FOV transmission, the video source file is divided into multiple tiles for storage after special encoding. Each tile corresponds to a different HD viewing area. When the user’s head posture changes, the cloud selects corresponding high-quality tiles based on the information reported by the terminal to match the user’s view area. In the meantime, the cloud also transmits low-quality full-view video content. The terminal fills in the view with low-quality panoramic content until the high-quality tiles arrive and replace the low-quality parts. In this process, if the update duration of a high-quality tile exceeds 200 ms, the user may see obviously low-quality images, resulting in poor experience.
Cloud VR User Experience Model
4.1 Key Cloud VR User Experience Requirements Based on User Survey

We have received a total of over 30,000 users and distributed some questionnaires at exhibitions and the lab. By analyzing the most representative 300 questionnaires using the Kano model, we distilled the core requirements of Cloud VR user experience, as shown in the following figure:

The Kano model was first developed in 1984 by Noriaki Kano, a professor at the Tokyo University of Science. The Kano model is common for classifying and sorting user requirements.

- **Basic requirements**: Basic requirements are also called mandatory requirements. When a product or service does not meet these requirements, users will be very dissatisfied. However, the converse does not hold: merely meeting these requirements does not indicate that users will be satisfied. In Cloud VR, such requirements may include sharp definition, smooth streaming, fast response, no dizzy feeling, no black edges, and no distortion.

- **Expected requirements**: Expected requirements are not mandatory, but are expected. Such requirements may draw the attention of competitors or manufacturers and determine product competitiveness. They may include screen mirroring, sharp color, and spatial audio.

- **Delight requirements**: Delight requirements may not be expected by users, and users will not be dissatisfied if such requirements are not met. Such requirements may include smell and temperature. Users would understand if they could not smell any scent, because this is not typically implemented.

After analysis using the Kano model, we find that most key user experience factors are consistent with our understanding. Basic requirements and expected requirements are the focus of our research and evaluation.
4.2 Cloud VR User Experience Modeling Based on Research Result

In addition to regularly collecting user feedback, we also actively participated in industry organizations, researched VR user experience, and made some achievements. For example, at the ITU-T SG12 plenary meeting held in Geneva on January 19, 2017, the Quality of Experience for Virtual Reality (QoE for VR) project proposed by Huawei Network Research Dept and Huawei iLab was approved. We also participated in research on VR experience standards in the video industry alliances in China.

Based on this research, we recommend that Cloud VR user experience requirements be classified into immersive and interactive experience requirements. Key requirements for user experience can then be extracted from four dimensions: Content quality, presentation quality, operation quality, and response quality.

- Immersive experience requirements focus on the content quality perceived by users through their own sight and hearing systems when using VR devices. In the future, VR experience will be further extended to sense of smell, force, and temperature. However, the experience may be affected by both terminals and networks.

- The interactive experience requirements focus on space and time matching of the information obtained by the user’s sensory systems. Positioning accuracy and response speed are core factors.

Cloud VR user experience requirements model
4.2.1 Glossary and Breakdown of Immersive Experience Requirements

Sharpness
Unclear images are the most frequently reported problem of Cloud VR users. Sharpness is determined by the quantity and quality of video pixels in the user visual area. The main factors affecting sharpness include content resolution, encoding method, terminal resolution, decoding capability, and FOV.

▲ Insufficient content resolution is the main factor affecting sharpness of panoramic video.

▲ Insufficient terminal display resolution is a major factor limiting the sharpness of rendered images.

▲ When the resolution is insufficient, it is unnecessarily sharper when FOV is larger, because there are fewer pixels in unit angle, resulting in screen-door effect.

▲ The bitrate determines the average number of bits per pixel. Therefore, a higher bitrate usually means sharper images.

Panoramic video is the main VR videos form. When a VR terminal plays a video, all pixels are projected onto a sphere, then the content in the user’s view area is extracted and displayed on the screen. The aspect ratio of the panoramic video is usually 2:1, and the resolution of current mainstream 4K panoramic video is approximately 3840 x 1920. If the headset FOV is 90 degrees, the resolution of the displayed image is only 960 x 960, and there are only 10 to 11 pixels per degree (PPD), equivalent to regular 240P video. 8K content improves image sharpness, but raises requirements on screen resolution. For details, see the following table.

<table>
<thead>
<tr>
<th>VR Panoramic Video Definition</th>
<th>VR Single-Eye Definition</th>
<th>Equivalent TV Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>4K</td>
<td>1K</td>
<td>240P</td>
</tr>
<tr>
<td>8K</td>
<td>2K</td>
<td>480P</td>
</tr>
<tr>
<td>16K</td>
<td>4K</td>
<td>960P</td>
</tr>
</tbody>
</table>
Smoothness
The video frame rate determines the smoothness of video and game images. Frame rate is determined by the content source. Beyond a particular frame rate, generally about 24 FPS, the human eye cannot perceive the discrete pictures comprising a video. Almost all VR videos meet this requirement. However, the image generation principle of games is different from that of video. Video images are usually shot by cameras, and capture movement of objects. In contrast, game images are completely rendered by graphics cards, and there is no motion tracking information of the object in the picture. Therefore, to obtain the smoothness of video content, rendered game images must have a higher frame rate. The smoothness is also directly related to the refresh rate of the display, which refers to the number of times the monitor obtains images from the GPU per second.

![Impact of frame rate on the smoothness of images]

**Color Fidelity**
The higher the color depth (number of bits used to store color information of a pixel), the more vivid and realistic the colors, pictures, and contrast. However, color fidelity is less helpful if images are not sharp. Sharpness takes precedent. In the short term, 8 bits per pixel can meet requirements.

**Three Dimensional Sense**
VR applications are often used to construct and restore three-dimensional space. The visual three-dimensional effect is achieved by providing image content generating binocular disparity. The stereoscopic auditory effect is achieved by controlling the number of audio sources during recording and the number of speakers during playback. Dual audio tracks are the minimum requirement.

**Audio-to-Video Synchronization**
When processing multimedia content, the player needs to separate video and audio data for independent decoding and rendering. If the audio and video separated from the multimedia file are not synchronized, the unsynchronized playback is unbearable to users.
Image Distortion

The combined visual field of both eyes is 200°–220° horizontally, with 120° overlapped. Imaging in the overlapped area is clearer. When the FOV of VR devices is smaller than that of human eyes, users feel that they are viewing through a window instead of being in an immersive environment.

![FOV of human eyes](image)

To ensure that the images presented by the VR device are immersive and consistent with human physiology, most VR terminals increase the FOV by adding a lens, but at a cost: pincushion distortion occurs if the original images are not processed. Currently, a popular solution is to apply barrel distortion to the images to make them look neutral.

![Pincushion distortion and Barrel distortion](image)

Artifact/Frame Freezing

Video is especially sensitive to packet loss and jitter during network transmission (particularly during transmission over Wi-Fi air interfaces), which causes artifacts and frame freezing. On-demand video services are usually carried over TCP. During playback, if the bandwidth is insufficient for the video bitrate, the content in the buffer is used up, causing frame freezing. Multicast video and game services are mainly carried over UDP. In this case, the consequence might be frame freezing or artifacts, depending on the mechanism for processing lost packets. For example, when the decoder uses the freezing mechanism, the image freezes when data is lost until the next complete I-frame arrives. When the decoder uses the ignorance mechanism, the decoder ignores the lost frame and continues decoding, causing artifacts.
4.2.2 Glossary and Breakdown of Interactive Experience Requirements

Degrees of Freedom
Degrees of freedom (DOF) refers to the movement of a body in space. There are 6 DOF in total, which can be classified as translations and rotations. A body can translate in 3 DOF: forward/back, up/down, and left/right. A body can also rotate with 3 DOF: pitch, yaw, and roll. 3DOF headsets only support rotation movements, whereas 6DOF headsets support both rotation and translation movements. 3DOF headsets allow users to watch 360° movies, and 6DOF headsets support strong interaction applications where users can move in a virtual environment.

Precision
A virtual world is a simulation of the physical world. Providing accurate mapping and response in the virtual world requires accurately obtaining the location and action information of the user in the physical world. Currently, an inertial measurement unit (IMU, which usually includes a gyroscope, accelerometer, and magnetometer) is widely used in VR headsets to track rotations. Other technologies used to measure movements include lasers, infrared light, ultrasonic sound waves, and image recognition-based (IRB) positioning technologies. Some of these spatial positioning technologies require external sensors. Such technologies are referred to as outside-in tracking, which is mature and highly precise, but can only be used in a fixed area. There are also inside-out tracking technologies, which do not require any external sensors. The headsets use a built-in time-of-flight (TOF) camera to detect environment change, then uses algorithms to calculate its own motion track. This poses higher requirements on terminal processing performance and has a lower positioning accuracy. However, it supports a wider range of activities and is less affected by obstacles. It is therefore suitable for mobile scenarios.

<table>
<thead>
<tr>
<th></th>
<th>Outside-in Tracking</th>
<th>Inside-out Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion capture precision</td>
<td>High</td>
<td>Slightly lower</td>
</tr>
<tr>
<td>Processing delay</td>
<td>Relatively short</td>
<td>Relatively long</td>
</tr>
<tr>
<td>External sensor</td>
<td>Required</td>
<td>Not required</td>
</tr>
<tr>
<td>Movement range</td>
<td>Restricted by sensor coverage</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Terminal cost</td>
<td>Reasonable</td>
<td>High</td>
</tr>
<tr>
<td>Occlusion</td>
<td>Susceptible</td>
<td>Not susceptible</td>
</tr>
</tbody>
</table>

To balance performance and cost, outside-in tracking may become the mainstream solution for home scenarios. However, most VR terminal manufacturers will use multiple positioning methods to support 6-DOF movement.

Comfort
VR headsets must be comfortable. Therefore, the headset design must consider weight, dimensions, heat dissipation, seal, and breathability.

Battery Life
The analysis of user habits shows that IMAX and gaming are the most popular service scenarios of Cloud VR. An IMAX movie is about 2-3 hours long. The average duration per use in gaming is about 1 hour. The batteries of VR terminals are usually not removable. Therefore, the battery life of headsets, handles, and position trackers is a major usage bottleneck.
**Friendliness**
To meet the requirements of nearsighted users, most VR devices either allow users to wear glasses within a headset or have a built-in object distance adjustment function, which adjusts the distance between the display and eyes to accommodate naked eyes with myopia under 600 degrees.

**Motion Sickness**
Motion sickness is caused by inconsistency between a user’s visual observations and physical motion sensed by the vestibule system. There are two causes of motion sickness:

- The user’s head is still, but the displayed image is not. This situation is similar to carsickness and seasickness, and can only be overcome through familiarity.

- The user’s head is moving, but the displayed image lags the movement. Such latency is referred to as MTP latency. Long MTP latency may cause motion sickness. The industry consensus is that an MTP latency must be shorter than 20 ms to minimize motion sickness. Causes of MTP latency include excessively long motion capture duration, low display refresh frequency, and long display response latency after asynchronous time warping is used on terminals.

**Interactive Latency**
The virtual reality system needs to quickly reflect the location and actions of users in the physical world. If processing is too slow, the interaction experience will be poor.

**Loading Waiting**
In the Cloud VR video application, the loading wait duration refers to the duration from when a user clicks the video play button to when the screen starts playing the video. The experience of online video users demonstrates that users have high requirements on the initial buffer time of VR videos. It is recommended that the initial loading be no longer than 2s.

In the Cloud VR game application, the loading wait duration refers to the duration from when a user clicks the game start button when the game images are displayed. The loading duration of common Cloud VR games is between 30–60s. In PC VR, games are loaded locally. However, in Cloud VR, the loading duration is increased due to scheduling of rendering server resources, video encoding and decoding, and transmission. However, the increased delay is more milliseconds and has little impact on the loading duration, measured in seconds. What requires attention is the increased loading duration in strong-interaction Cloud VR services caused by insufficient platform resources in high concurrency.

**Black Edge and Smear**
For details, see the strong-interaction Cloud VR service solution.

**Operation Response**
For details, see the strong-interaction Cloud VR service solution.

**Image Quality Switching**
For details, see the FOV video transmission solution.
05
Cloud VR Service Evaluation Baseline
5.1 Mapping Between Cloud VR User Experience Requirements and Evaluation Items

Based on the research and analysis in chapters 3 and 4, we break down or map key Cloud VR user experience requirements into components and modules, as shown in the following figure.

Based on current industry development, lab test results, and project delivery experience, we provide some evaluation items, tools, and methods.
### 5.2 Strong-Interaction Cloud VR Service

#### 5.2.1 Evaluation Baseline for Cloud-Terminal Asynchronous Rendering Solution

#### Evaluation Items

**NOTE:** In the following table, indicators are ranked by service impact.

<table>
<thead>
<tr>
<th>Component</th>
<th>Module</th>
<th>Item</th>
<th>Whether</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen</td>
<td>Screen</td>
<td>Screen resolution</td>
<td>Mandatory</td>
<td>$\geq 1440 \times 2 \times 1600-1020 \times 2 \times 2160$</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>Screen refresh rate</td>
<td>Mandatory</td>
<td>$\geq 90$</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>Screen type</td>
<td>Optional</td>
<td>TFT-LCD/OLED</td>
</tr>
<tr>
<td></td>
<td>Screen</td>
<td>Screen size</td>
<td>Optional</td>
<td>$\geq 3.5$ inch x 2</td>
</tr>
<tr>
<td>Lens</td>
<td>Lens</td>
<td>Lens material</td>
<td>Optional</td>
<td>Fresnel lens, PMMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FOV</td>
<td>Mandatory</td>
<td>$100-120^\circ$</td>
</tr>
<tr>
<td>Terminal</td>
<td>Video</td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>H.264</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>Decoding performance</td>
<td>Mandatory</td>
<td>$\geq 4096 \times 2160$ at 60 FPS, 100 Mbps@H.264</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>AAC/AAC+/eAAC+, MP3, AMR/AMR-WB, MIDI, PCM, OGG, FLAC</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
<td>Number of rendering channels</td>
<td>Mandatory</td>
<td>Dual audio channels, built-in dual stereo speakers</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
<td>Audio-to-video synchronization</td>
<td>Mandatory</td>
<td>Audio in sync with video</td>
</tr>
<tr>
<td>Correction</td>
<td>Correct</td>
<td>Distortion correction</td>
<td>Mandatory</td>
<td>Supported. The image has no pincushion or barrel distortion.</td>
</tr>
<tr>
<td>Comfort</td>
<td>Comfort</td>
<td>Headset type</td>
<td>Optional</td>
<td>Standalone VR headsets</td>
</tr>
<tr>
<td></td>
<td>Comfort</td>
<td>Headset heat dissipation</td>
<td>Mandatory</td>
<td>Good heat dissipation performance, and the contact temperature of the skin does not exceed 40(^\circ).</td>
</tr>
<tr>
<td></td>
<td>Comfort</td>
<td>Headset seal</td>
<td>Mandatory</td>
<td>Good seal with no obvious light leakage.</td>
</tr>
<tr>
<td></td>
<td>Comfort</td>
<td>Headset breathability</td>
<td>Mandatory</td>
<td>Good breathability, and normal breathing is not affected.</td>
</tr>
<tr>
<td></td>
<td>Comfort</td>
<td>Headset weight (including battery)</td>
<td>Mandatory</td>
<td>$\leq 500$ g</td>
</tr>
<tr>
<td>Battery</td>
<td>Battery</td>
<td>Battery capacity</td>
<td>Optional</td>
<td>3500–6000 mAh</td>
</tr>
<tr>
<td></td>
<td>Battery</td>
<td>Battery life (while watching video)</td>
<td>Mandatory</td>
<td>$\geq 2.5$ hours</td>
</tr>
<tr>
<td>Friendliness</td>
<td>Friendliness</td>
<td>Eye comfort mode</td>
<td>Optional</td>
<td>Low blue light certification</td>
</tr>
<tr>
<td></td>
<td>Friendliness</td>
<td>Myopia friendliness</td>
<td>Mandatory</td>
<td>Allows users to wear glasses, or supports focusing</td>
</tr>
<tr>
<td>Terminal</td>
<td>Performance</td>
<td>Mandatory/Optional</td>
<td>Requirement</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>--------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MTP delay</td>
<td>Mandatory</td>
<td>≤ 20 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Terminal decoding delay</td>
<td>Mandatory</td>
<td>≤ 15 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Motion rendering delay</td>
<td>Mandatory</td>
<td>≤ 5.5 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Screen refresh delay</td>
<td>Mandatory</td>
<td>≤ 11 ms (90 FPS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Screen response delay</td>
<td>Mandatory</td>
<td>≤ 5 ms</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Wi-Fi</td>
<td>Mandatory</td>
<td>Supports 2x2 MIMO and 802.11ac@5 GHz.</td>
<td></td>
</tr>
<tr>
<td>System</td>
<td></td>
<td>Optional</td>
<td>Android</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Version</td>
<td>Optional</td>
<td>7.1–8.1</td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td></td>
<td>Optional</td>
<td>Qualcomm 835/Samsung Exynos 8895</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPU</td>
<td>Optional</td>
<td>Adreno 540 GPU/ARM Mali G71 MP18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPU</td>
<td>Optional</td>
<td>4G RAM, LPDDR4X, 1866M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>Optional</td>
<td>64 GB UFS2.1, supporting 256 GB Micro-SD card</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Storage</td>
<td>Optional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positioning suite</td>
<td>DOF (headset)</td>
<td>Mandatory</td>
<td>6DOF</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>DOF (handle)</td>
<td>Mandatory</td>
<td>6DOF</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Positioning (headset)</td>
<td>Optional</td>
<td>Ultrasonic/laser/infrared combined positioning</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Positioning (handle)</td>
<td>Optional</td>
<td>Ultrasonic/laser/infrared combined positioning</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Positioning precision (headset)</td>
<td>Mandatory</td>
<td>≤ 2 mm</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Positioning precision (handle)</td>
<td>Mandatory</td>
<td>≤ 2 mm</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Position sampling frequency (headset)</td>
<td>Mandatory</td>
<td>≥ 120 Hz</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Position sampling frequency (handle)</td>
<td>Mandatory</td>
<td>≥ 120 Hz</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling mode (headset)</td>
<td>Optional</td>
<td>6-axis/9-axis pose sensor</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling mode (handle)</td>
<td>Optional</td>
<td>6-axis/9-axis pose sensor</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling precision (headset)</td>
<td>Mandatory</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling precision (handle)</td>
<td>Mandatory</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling frequency (headset)</td>
<td>Mandatory</td>
<td>≥ 400 Hz</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Pose sampling frequency (handle)</td>
<td>Mandatory</td>
<td>≥ 100 Hz</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Range</td>
<td>Mandatory</td>
<td>≥ FOV 100°, 5 m radius</td>
<td></td>
</tr>
<tr>
<td>Signal processing</td>
<td>Handle connection mode</td>
<td>Optional</td>
<td>2.4G Wi-Fi/Bluetooth</td>
<td></td>
</tr>
<tr>
<td>Signal processing</td>
<td>Signal transmission delay (handle)</td>
<td>Mandatory</td>
<td>≤ 3 ms</td>
<td></td>
</tr>
<tr>
<td>Signal processing</td>
<td>Signal processing delay</td>
<td>Mandatory</td>
<td>≤ 2 ms</td>
<td></td>
</tr>
<tr>
<td>Battery life</td>
<td>Capacity (base station)</td>
<td>Optional</td>
<td>1800 mAh</td>
<td></td>
</tr>
<tr>
<td>Battery life</td>
<td>Capacity (handle)</td>
<td>Optional</td>
<td>1000 mAh</td>
<td></td>
</tr>
<tr>
<td>Battery life</td>
<td>Battery life (base station)</td>
<td>Mandatory</td>
<td>≥ 7 hours</td>
<td></td>
</tr>
<tr>
<td>Battery life</td>
<td>Battery life (handle)</td>
<td>Mandatory</td>
<td>≥ 2.5 hours</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>Performance</td>
<td>Mandatory</td>
<td>≤ 10e-5</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>Forwarding delay</td>
<td>Mandatory</td>
<td>≤ 15–20 ms</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>Transmission bandwidth</td>
<td>Mandatory</td>
<td>≥ 50 Mbps</td>
<td></td>
</tr>
<tr>
<td><strong>Video encoding</strong></td>
<td><strong>Resolution</strong></td>
<td>Mandatory</td>
<td>≥ 1440 x 2 x 1600–1920 x 2 x 2160</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>-----------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Bitrate</strong></td>
<td>Mandatory</td>
<td>≥ 40 Mbps</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Frame rate</strong></td>
<td>Mandatory</td>
<td>≥ 50 FPS</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Encoding format</strong></td>
<td>Mandatory</td>
<td>H.264</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Color depth</strong></td>
<td>Mandatory</td>
<td>≥ 8 bits</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>FOV</strong></td>
<td>Mandatory</td>
<td>≥ 100–120°</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Stereoscopic vision</strong></td>
<td>Mandatory</td>
<td>3D</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Audio encoding</strong></th>
<th><strong>Bitrate</strong></th>
<th>Optional</th>
<th>128 Kbps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Code</strong></td>
<td>Optional</td>
<td>AAC</td>
</tr>
<tr>
<td></td>
<td><strong>Number of source audio channels</strong></td>
<td>Optional</td>
<td>Dual audio channels</td>
</tr>
<tr>
<td></td>
<td><strong>Sampling rate</strong></td>
<td>Optional</td>
<td>48 kHz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Performance</strong></th>
<th><strong>Platform processing delay</strong></th>
<th>Mandatory</th>
<th>≤ 30 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Computing &amp; rendering delay</strong></td>
<td>Optional</td>
<td>= 11 ms</td>
</tr>
<tr>
<td></td>
<td><strong>Preprocessing delay</strong></td>
<td>Optional</td>
<td>≈ 2 ms</td>
</tr>
<tr>
<td></td>
<td><strong>Compression coding delay</strong></td>
<td>Optional</td>
<td>≈ 15 ms</td>
</tr>
<tr>
<td></td>
<td><strong>Streaming delay</strong></td>
<td>Optional</td>
<td>≈ 2 ms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Platform</strong></th>
<th><strong>User and system resource scheduling</strong></th>
<th>Mandatory</th>
<th>Users and resources can be scheduled, and system resources can be locked and released.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Service program release and management</strong></td>
<td>Mandatory</td>
<td>The go-live/go-offline and typesetting of application content can be controlled.</td>
</tr>
<tr>
<td></td>
<td><strong>Rendering encoding parameter adjustment</strong></td>
<td>Mandatory</td>
<td>Parameters such as encoding compression can be dynamically set and adjusted.</td>
</tr>
<tr>
<td></td>
<td><strong>System reliability protection</strong></td>
<td>Mandatory</td>
<td>The application server supports link protection, and the database supports two-node cluster backup.</td>
</tr>
<tr>
<td></td>
<td><strong>User status monitoring</strong></td>
<td>Mandatory</td>
<td>User online duration, resource consumption, and user experience can be viewed.</td>
</tr>
<tr>
<td></td>
<td><strong>Platform performance monitoring</strong></td>
<td>Mandatory</td>
<td>There are alarms for resource overload, system programs, or user operations.</td>
</tr>
<tr>
<td></td>
<td><strong>Screen display</strong></td>
<td>Mandatory</td>
<td>Screen mirroring through an STB is supported.</td>
</tr>
</tbody>
</table>

| **Encapsulation** | **Encapsulation protocol** | Optional | TCP/UDP. Firewall traversal is required for UDP. |

<table>
<thead>
<tr>
<th><strong>Hardware</strong></th>
<th><strong>CPU</strong></th>
<th>Optional</th>
<th>22 cores, 2.6 GHz or higher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Memory</strong></td>
<td>Optional</td>
<td>60 GB</td>
</tr>
<tr>
<td></td>
<td><strong>Network adapter</strong></td>
<td>Optional</td>
<td>10G</td>
</tr>
<tr>
<td></td>
<td><strong>Graphics card</strong></td>
<td>Optional</td>
<td>NVIDIA Tesla M60/Tesla V100</td>
</tr>
<tr>
<td></td>
<td><strong>System</strong></td>
<td>Optional</td>
<td>Windows 7/Windows 10</td>
</tr>
<tr>
<td></td>
<td><strong>Virtual platform</strong></td>
<td>Optional</td>
<td>Fusion sphere 6.2/VMware sphere 6.7</td>
</tr>
</tbody>
</table>

| **Service Experience** | **Program loading wait duration** | Mandatory | The application loading duration should not exceed 60 seconds. |
### Evaluation Tool

**Cloud Rendering and Streaming**: During cloud rendering and streaming, a terminal submits instructions to the cloud, and the cloud performs rendering, encoding, and streaming. The terminal then receives and decodes the streams. This is an extra process of Cloud VR relative to local VR during service processing. Latency is measured through timestamping. When the terminal sends an instruction, timestamp T1 is added. When the platform receives the instruction, timestamp T1' is added. After the platform processes the data, timestamp T2' is added to the frames sent out. When receiving frames, the terminal records timestamp T2, and parses the timestamps T2', T1', and T1. After decoding the frame data, the terminal records timestamp T3, calculates the delay values, and records the result in a log.

Cloud rendering and streaming duration = T3 – T1
Cloud rendering processing duration = T2' – T1'
Terminal decoding duration = T3 – T2
Network transmission duration = T2 – T1 – (T2' – T1')
NOTE: The instruction sent by the terminal to the platform includes high-precision ($10^{-4}$) view information, which is reserved throughout the cloud rendering and streaming process. The information can be used as a unified identifier of the instruction and corresponding video frames in a complete strong interaction process.

Frame Freezing: If a frame freeze occurs during a game, the image is suddenly suspended, then is restored to normal after a period of time. In the Cloud VR solution, the main cause of this problem is packet loss and network latency. In the terminal-cloud asynchronous rendering solution, in addition to the basic frames generated by the cloud rendering, the terminal performs local frame compensation by using asynchronous time warping technology to maintain or improve the display frame rate. However, only basic frames generated by cloud rendering are considered effective frames. During actual experience, some obvious frame freezing occurs due to insufficient effective frames received by the terminal over a certain period of time. The terminal can predict and display only the latest effective frames. Inserted intermediate frames are highly similar to effective frames. This means that inserted frames cannot solve the problem of frame freezing. If the display interval between adjacent effective frames exceeds a specified period (for example, 50 ms), we can reasonably conclude that a frame freeze occurs during the game experience. The interval between displayed adjacent effective frames is the duration of single frame freezing event.

Artifacts: In Cloud VR services transmitted using UDP, packet loss can cause artifacts. The artifact type depends on the type of lost frames. Loss of I-frame data affects the decoding of the entire group of pictures (GOP), whereas loss of P-frame data affects all the subsequent P-frames until the GOP ends. The loss of an I-frame has a greater impact than the loss of a P-frame. In the actual tests, lost frames are measured, and their types are considered for calculation of artifact duration.

Black Edges: Local asynchronous warping technologies can only perform prediction based on current images and cannot generate images out of nothing. Therefore, when a user rotates their head quickly, the area beyond the original field of view is displayed as a black edge or smear. Theoretically, this problem can be resolved by performing extra-perspective rendering on the cloud. However, the actual effect is determined by head rotation speed and delay of cloud rendering and streaming. The effectiveness is uncertain and needs to be monitored. Based on the implementation principle, the black edge is actually the difference between the input user angle information and the angle of the warped frames during asynchronous warping. The black edge angle is calculated as follows:
Black edge angle = Head angle after asynchronous time warping – angle in effective frames – (extra-perspective rendering FOV – terminal FoV)/2

Evaluation Method

If a component passes all mandatory items, the component passes the evaluation.

5.3 Cloud VR Video Services

5.3.1 Full-View Video Transmission Solution Evaluation Baseline

Evaluation Items

NOTE: In the following table, indicators are ranked by service impact.

<table>
<thead>
<tr>
<th>Component</th>
<th>Module</th>
<th>Item</th>
<th>Whether Mandatory</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen</td>
<td></td>
<td>Screen resolution</td>
<td>Mandatory</td>
<td>≥ 1440 x 2 x 1600–1920 x 2 x 2160</td>
</tr>
<tr>
<td>Screen</td>
<td></td>
<td>Screen refresh rate</td>
<td>Mandatory</td>
<td>≥ 90</td>
</tr>
<tr>
<td>Screen</td>
<td></td>
<td>Screen type</td>
<td>Optional</td>
<td>TFT-LCD/OLED</td>
</tr>
<tr>
<td>Screen</td>
<td></td>
<td>Screen size</td>
<td>Optional</td>
<td>3.5 inch x 2</td>
</tr>
<tr>
<td>Lens</td>
<td></td>
<td>Lens material</td>
<td>Optional</td>
<td>Fresnel lens, PMMA</td>
</tr>
<tr>
<td>Lens</td>
<td></td>
<td>FOV</td>
<td>Mandatory</td>
<td>≥ 100–120°*</td>
</tr>
<tr>
<td>Terminal</td>
<td>Video</td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>H.264 and H.265</td>
</tr>
<tr>
<td>Terminal</td>
<td>Video</td>
<td>Decoding performance</td>
<td>Mandatory</td>
<td>≥ 4096 x 2160 at 60 FPS, 100 Mbps H.264 (4K H.264) ≥ 7680 x 3840 at 30 FPS, 200 Mbps H.265 (8K H.265)</td>
</tr>
<tr>
<td>Audio</td>
<td></td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>AAC/AAC+/eAAC+, MP3, AMR/AMR-WB, MIDI, PCM, OGG, FLAC</td>
</tr>
<tr>
<td>Audio</td>
<td></td>
<td>Number of rendering channels</td>
<td>Mandatory</td>
<td>Dual audio channels, built-in dual stereo speakers</td>
</tr>
<tr>
<td>Audio</td>
<td></td>
<td>Audio-to-video synchronization</td>
<td>Mandatory</td>
<td>Subjective synchronization without obvious time difference</td>
</tr>
<tr>
<td>Positioning</td>
<td></td>
<td>DOF (headset)</td>
<td>Mandatory</td>
<td>3DOF</td>
</tr>
<tr>
<td>Positioning</td>
<td></td>
<td>DOF (handle)</td>
<td>Optional</td>
<td>3DOF</td>
</tr>
<tr>
<td>Positioning</td>
<td></td>
<td>Pose sampling mode (headset)</td>
<td>Mandatory</td>
<td>6-axis/9-axis pose sensor</td>
</tr>
<tr>
<td>Feature</td>
<td>Requirement</td>
<td>Specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------</td>
<td>----------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>Mandatory</td>
<td>• Angle information collection mode (handle)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Pose sampling frequency (handle) ≥ 100 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Pose sampling precision (handle) ≤ 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Pose sampling precision (headset) ≤ 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Pose sampling frequency (headset) ≥ 400 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction</td>
<td>Mandatory</td>
<td>• Distortion correction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Headset type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Headset heat dissipation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Headset seal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Headset breathability</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Headset weight (including battery) ≤ 500 g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminal</td>
<td>Optional</td>
<td>• Battery capacity 3500–4000 mAh</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Battery life (while watching video) ≥ 2.5 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Eye comfort mode</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Myopia friendliness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Mandatory</td>
<td>• MTP delay ≤ 20 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Terminal decoding delay ≤ 15 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Motion rendering delay ≤ 5.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Screen refresh delay ≤ 11 ms (90 FPS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Screen response delay ≤ 5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Mandatory</td>
<td>• Wi-Fi Supports 2x2 MIMO and 802.11ac@5 GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Optional</td>
<td>• System Android</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Version 7.1–8.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td>Optional</td>
<td>• CPU Qualcomm 835 (4K)/Samsung Exynos 8895 (4K–8K)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• GPU Adreno 540 GPU/ARM Mali G71 MP18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Memory 4G RAM, LPDDR4x, 1866M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optional</td>
<td>• Storage 64GB UFS2.1, supporting 256 GB Micro-SD card</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Performance</td>
<td>Mandatory</td>
<td>• Packet loss rate ≤ 9E−5 (4K@H.264) ≤ 1.7E−5 (8K@H.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Forwarding delay ≤ 20–30 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Transmission bandwidth ≥ 60Mbps (4K@H.264) ≥ 120–180Mbps (8K@H.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Mandatory</td>
<td>• Resolution 3840x1920 (4K@H.264) 7680x3840 (8K@H.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video encoding</td>
<td>Mandatory</td>
<td>• Bitrate ≥ 40Mbps (4K@H.264) ≥ 80–120Mbps (8K@H.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Frame rate ≥ 30 FPS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Encoding format 4K@H.264, 8K@H.265</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mandatory</td>
<td>• Color depth ≥ 8 bits</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Evaluation Tool

**Service Probe:** Frame freezing: During playback when transmitted over TCP, if the bandwidth is lower than the video bitrate, the content in the buffer is used up, causing frame freezing. During evaluation, the player can be used to monitor the volume of buffered data to determine whether frame freezing occurs.

### Evaluation Method

If a component passes all mandatory items, the component passes the evaluation.

### 5.3.2 FOV Video Transmission Solution Evaluation Baseline

#### Evaluation Items

**NOTE:** In the following table, indicators are ranked by service impact.

<table>
<thead>
<tr>
<th>Component</th>
<th>Module</th>
<th>Item</th>
<th>Whether Mandatory</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal</td>
<td>Screen</td>
<td>Screen resolution</td>
<td>Mandatory</td>
<td>≥ 1440 x 2 x 1000–1920 x 2 x 2160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen refresh rate</td>
<td>Mandatory</td>
<td>≥ 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen type</td>
<td>Optional</td>
<td>TFT-LCD/OLED</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Screen size</td>
<td>Optional</td>
<td>3.5 inch x 2</td>
</tr>
<tr>
<td>Terminal</td>
<td>Lens</td>
<td>Optional</td>
<td>Fresnel lens, PMMA</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------</td>
<td>----------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lens material</td>
<td>Optional</td>
<td>Fresnel lens, PMMA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOV</td>
<td>Mandatory</td>
<td>≥ 100–120°</td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>H.265</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decoding performance</td>
<td>Mandatory</td>
<td>≥ 4096 x 2160 at 60 FPS, 100 Mbps@H.265</td>
<td></td>
</tr>
<tr>
<td>Audio</td>
<td>Decoding type</td>
<td>Mandatory</td>
<td>AAC/AAC+/eAAC+, MP3, AMR/AMR-WB, MIDI, PCM, OGG, FLAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of rendering channels</td>
<td>Mandatory</td>
<td>Dual audio channels, built-in dual stereo speakers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Audio-to-video synchronization</td>
<td>Mandatory</td>
<td>Audio in sync with video</td>
<td></td>
</tr>
<tr>
<td>Positioning</td>
<td>DOF (headset)</td>
<td>Mandatory</td>
<td>3DOF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DOF (handle)</td>
<td>Optional</td>
<td>3DOF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling mode (headset)</td>
<td>Mandatory</td>
<td>6-axis/9-axis pose sensor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling mode (handle)</td>
<td>Optional</td>
<td>6-axis/9-axis pose sensor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling precision (headset)</td>
<td>Mandatory</td>
<td>≤ 0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling precision (handle)</td>
<td>Optional</td>
<td>≤ 0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling frequency (headset)</td>
<td>Mandatory</td>
<td>≥ 400 Hz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pose sampling frequency (handle)</td>
<td>Optional</td>
<td>≥ 100 Hz</td>
<td></td>
</tr>
<tr>
<td>Correction</td>
<td>Distortion correction</td>
<td>Mandatory</td>
<td>Supported, without pincushion distortion</td>
<td></td>
</tr>
<tr>
<td>Comfort</td>
<td>Headset type</td>
<td>Optional</td>
<td>Standalone VR headsets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Headset heat dissipation</td>
<td>Mandatory</td>
<td>Good heat dissipation performance, and the contact temperature of the skin does not exceed 40°.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Headset seal</td>
<td>Mandatory</td>
<td>Good seal with no obvious light leakage.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Headset breathability</td>
<td>Mandatory</td>
<td>Good breathability, and normal breathing is not affected.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Headset weight (including battery)</td>
<td>Mandatory</td>
<td>≤ 500 g</td>
<td></td>
</tr>
<tr>
<td>Battery life</td>
<td>Battery capacity</td>
<td>Optional</td>
<td>3500–4000 mAh</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Battery life (while watching video)</td>
<td>Mandatory</td>
<td>≥ 2.5 hours</td>
<td></td>
</tr>
<tr>
<td>Friendliness</td>
<td>Eye comfort mode</td>
<td>Optional</td>
<td>Low blue light certification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Myopia friendliness</td>
<td>Mandatory</td>
<td>Allows users to wear glasses, or supports focusing</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>MTP delay</td>
<td>Mandatory</td>
<td>≤ 20 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Terminal decoding delay</td>
<td>Mandatory</td>
<td>≤ 15 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Motion rendering delay</td>
<td>Mandatory</td>
<td>≤ 5.5 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Screen refresh delay</td>
<td>Mandatory</td>
<td>≤ 11 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Screen response delay</td>
<td>Mandatory</td>
<td>≤ 5 ms</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>Wi-Fi</td>
<td>Mandatory</td>
<td>Supports 2x2 MIMO and 802.11ac@5 GHz.</td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>System</td>
<td>Optional</td>
<td>Android</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Version</td>
<td>Optional</td>
<td>7.1–8.1</td>
<td></td>
</tr>
<tr>
<td>Hardware</td>
<td>CPU</td>
<td>Optional</td>
<td>Qualcomm 835/Samsung Exynos 8895</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPU</td>
<td>Optional</td>
<td>Adreno 540 GPU/ARM Mali G71 MP18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memory</td>
<td>Optional</td>
<td>4G RAM, LPDDR4X, 1866M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Storage</td>
<td>Optional</td>
<td>64GB UFS2.1, supporting 256 GB Micro-SD card</td>
<td></td>
</tr>
<tr>
<td>Network Performance</td>
<td>Packet loss rate</td>
<td>Mandatory</td>
<td>≤1.7E–5 (8K)</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forwarding delay</td>
<td>Mandatory</td>
<td>≤20 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transmission bandwidth</td>
<td>Mandatory</td>
<td>≥90 Mbps</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FOV video</th>
<th>Resolution</th>
<th>Mandatory</th>
<th>7680 x 3840</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bitrate</td>
<td>Mandatory</td>
<td>≥60 Mbps</td>
</tr>
<tr>
<td></td>
<td>Frame rate</td>
<td>Mandatory</td>
<td>≥30 FPS</td>
</tr>
<tr>
<td></td>
<td>Encoding format</td>
<td>Mandatory</td>
<td>H.265</td>
</tr>
<tr>
<td></td>
<td>Color depth</td>
<td>Mandatory</td>
<td>≥8 bits</td>
</tr>
<tr>
<td>Stereoactive vision</td>
<td>Optional</td>
<td>2D/3D</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Background video</th>
<th>Resolution</th>
<th>Mandatory</th>
<th>3840 x 1920</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bitrate</td>
<td>Mandatory</td>
<td>≥6 Mbps–15 Mbps</td>
</tr>
<tr>
<td></td>
<td>Frame rate</td>
<td>Mandatory</td>
<td>≥30 FPS</td>
</tr>
<tr>
<td></td>
<td>Encoding format</td>
<td>Mandatory</td>
<td>H.265</td>
</tr>
<tr>
<td></td>
<td>Color depth</td>
<td>Mandatory</td>
<td>≥8 bits</td>
</tr>
<tr>
<td></td>
<td>Stitching angle</td>
<td>Optional</td>
<td>360°/180°</td>
</tr>
<tr>
<td>Stereoactive vision</td>
<td>Optional</td>
<td>2D/3D</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audio encoding</th>
<th>Bitrate</th>
<th>Optional</th>
<th>128 Kbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Optional</td>
<td>AAC</td>
<td></td>
</tr>
<tr>
<td>Number of source audio channels</td>
<td>Optional</td>
<td>Dual audio channels</td>
<td></td>
</tr>
<tr>
<td>Sampling rate</td>
<td>Optional</td>
<td>48 kHz</td>
<td></td>
</tr>
<tr>
<td>Encapsulation</td>
<td>Encapsulation protocol</td>
<td>Mandatory</td>
<td>DASH over TCP</td>
</tr>
<tr>
<td>Number of connections</td>
<td>Optional</td>
<td>5–6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Service Experience</th>
<th>Video loading wait duration</th>
<th>Mandatory</th>
<th>≤2s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MTP latency threshold-crossing count</td>
<td>Mandatory</td>
<td>Every time MTP latency exceeds 20 ms, an MTP latency threshold-crossing event is logged.</td>
</tr>
<tr>
<td></td>
<td>Image quality switching latency threshold-crossing count</td>
<td>Mandatory</td>
<td>If the image quality switching delay exceeds 200 ms, an image quality switching delay threshold-crossing event is logged.</td>
</tr>
<tr>
<td></td>
<td>Frame freezing count/duration</td>
<td>Mandatory</td>
<td>If the interval between displayed adjacent effective frames exceeds 50 ms, a frame freezing event is logged.</td>
</tr>
</tbody>
</table>

## Evaluation Tool

**Image Quality Switching:** Timestamping is used to evaluate image quality switching delay. The terminal records T1 when a tile request is sent and T2 after all tiles are stitched, parsed, and displayed. The difference T2 – T1 is the image quality switching duration. Motion capture and rendering and display delay values are not considered because they are fixed depending on the product specifications.

## Evaluation Method

If a component passes all mandatory items, the component passes the evaluation.
Appendix
<table>
<thead>
<tr>
<th>Acronyms and Abbreviations</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Access point</td>
</tr>
<tr>
<td>ATW</td>
<td>Asynchronous Time Warping</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>BR</td>
<td>Border Router</td>
</tr>
<tr>
<td>BRAS</td>
<td>Broadband Remote Access Server</td>
</tr>
<tr>
<td>BSS</td>
<td>Business support system</td>
</tr>
<tr>
<td>CDN</td>
<td>Content Distribution Network</td>
</tr>
<tr>
<td>CR</td>
<td>Core Router</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer Relationship Management</td>
</tr>
<tr>
<td>DOF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>FPS</td>
<td>Frame per Second</td>
</tr>
<tr>
<td>GOP</td>
<td>Group Of Pictures</td>
</tr>
<tr>
<td>HSI</td>
<td>High-Speed Internet</td>
</tr>
<tr>
<td>IMU</td>
<td>Inertial Measurement Unit</td>
</tr>
<tr>
<td>MTP</td>
<td>Motion to Photons Latency</td>
</tr>
<tr>
<td>ONT</td>
<td>Optical Network Terminal</td>
</tr>
<tr>
<td>OLT</td>
<td>Optical Line Terminal</td>
</tr>
<tr>
<td>OTN</td>
<td>Optical Transport Network</td>
</tr>
<tr>
<td>OTT</td>
<td>Over The Top</td>
</tr>
<tr>
<td>PON</td>
<td>passive optical network</td>
</tr>
<tr>
<td>PPD</td>
<td>Pixels Per Degree</td>
</tr>
<tr>
<td>PSP</td>
<td>Platonic Solid Projection</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>RET</td>
<td>Retransmission</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>STB</td>
<td>Set Top Box</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TWS</td>
<td>Tile Wise</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
</tbody>
</table>