UHD World Association
世界超高清视频产业联盟

UHD World Association
世界超高清视频产业联盟

White Paper on Chinese Video Consumer User Experience

Mobile short video and mobile live broadcast (viewing end)

(V1.0)

UHD World Association

November 30, 2022

About This Document

The Video Experience Working Group of the UHD World Association (UWA) oversaw efforts to formulate this document and is responsible for its interpretation.

This document was published on March 29, 2023.

This document is proposed and managed by the UWA.

This document is mainly drafted by:

China Information and Communication Research Institute, China United Network Communications Co., Ltd., China Telecom Group Co., Ltd., China Mobile Communications Co., Ltd., Huawei Technology Co., Ltd., Youku Information Technology (Beijing) Co., Ltd., Beijing Fast Hand Technology Co., Ltd., Beijing Volcano Engine Technology Co., Ltd., Beijing Aiqiyi Technology Co., Ltd., Beijing Mijing Hefeng Technology Co., Ltd., Shanghai Zhaoyan Network Technology Co., Ltd Jugood-looking Technology Co., Ltd., 263 Network Communication Co., Ltd., Wise People Sihai (Beijing) Technology Co., Ltd., Beijing Liujinnian Media Technology Co., Ltd., Shanghai Aice Communication Technology Co., Ltd., China Mobile (Hangzhou) Information Technology Co., Ltd., ZTE Communication Co., Ltd., Shanghai Jiaotong University, Beijing University of Technology, Beijing University of Posts and Telecommunications, Dekeshi Communication (Shanghai) Co., Ltd Zhonghengda (Beijing) Software Testing Technology Co., Ltd., Beijing Bohui Technology Co., Ltd., Zhongguancun Modern Information Consumption Application Industry Technology Alliance.

Main drafters of this document:

Song Li, Jia Wu, Luo Chuanfei, Yang Kun, Zhang Shijun, Tang Conghua, Song Zuping, Wu Liangping, Wang Yajun, Miao Xuyang, Zhu Jiayue, Zhang Chunlei, Wu Jingfang, Li Jing, Mei Dawei, Du Yifei, Ma Yingwu, Zheng Chuanchuan, Yu Li, Xu Jingwen, Wang Zhihang, Gao Weize, Tang Geng, Qian Lei, Duan Tao, Ren Zijian, Shi Dongping, Liu Shuai, Zhao Yi, Ai Fang, Wu Jing, Guo Jincheng, Zhou Kaixuan, Liu Wenquan, Wang Yong, Yang Chen, Huang Ting Yin Qin, Wang Jindong, Tao Guojun, Weng Dongdong, Yan Shi, Wu Xuebo, Chen Jielian, Zhong Lin, Zhang Jiabin, Jiang Weiping, Chen Hong, Liu Xuan.

Disclaimer:

1. This document is available for use free of charge. It is used for reference only and is not liable for any products developed by referencing this document.
2. Updates to this document will be uploaded to the official UWA website without notice.

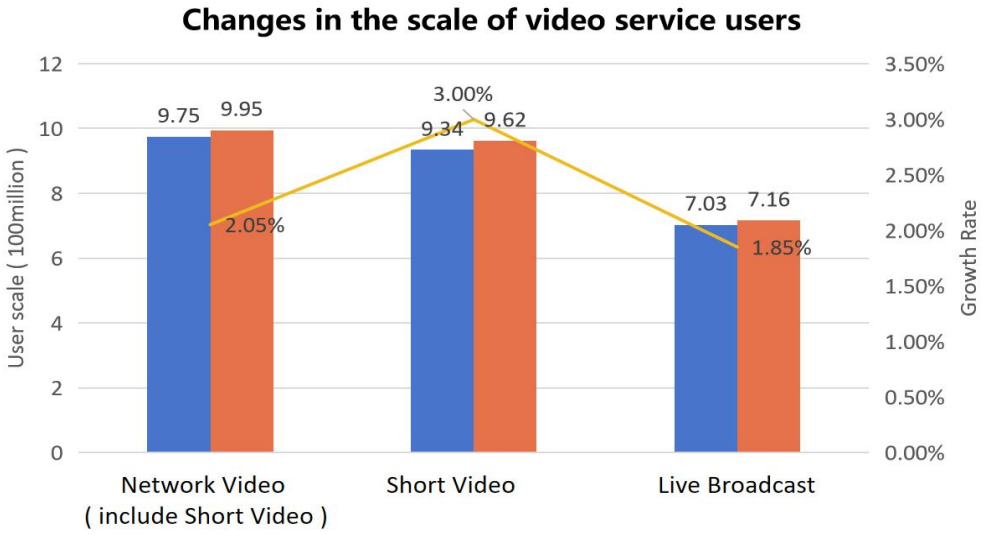
Table of Contents

1. Development status and trend of mobile video

1.1 Mobile video has become the main application of mobile Internet

Mobile Internet access has become the main form of Internet users in China. As of June 2022, the number of Internet users in China is 1.051 billion, with 1.047 billion of those users using mobile devices, accounting for 99.6% [1] of all Internet users. In terms of duration, the per capita online duration of Internet users is 29.5 hours per week, an increase of 1.0 hours compared with December 2021.

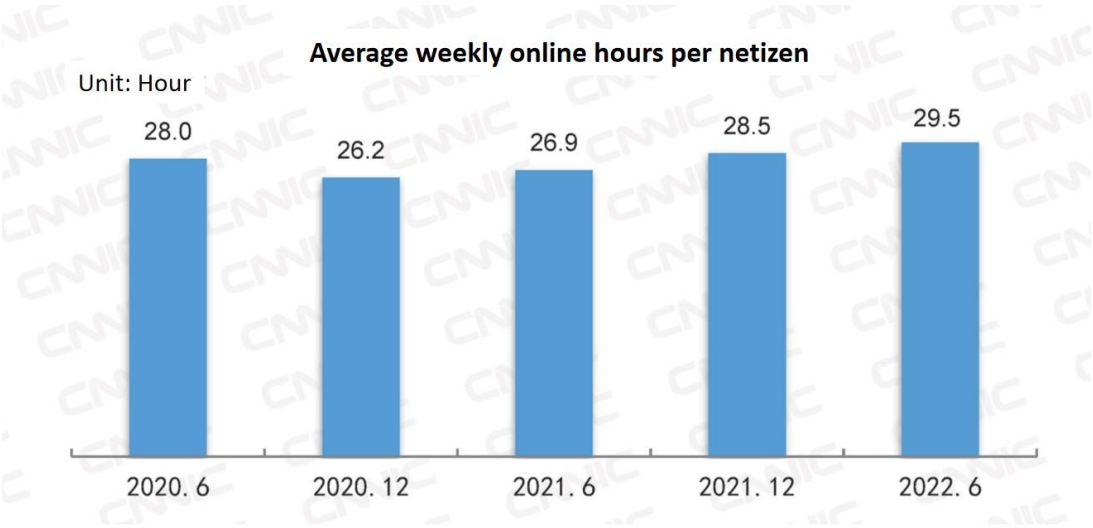
Video service is the main application next to instant social communication. As shown in Figure 1, as of June 2022, the total number of video users (including short videos) in the video business will be 995 million, an increase of 2.05% compared with December 2021; The number of short video users is 962 million, up 3% from December 2021; The number of webcast users is 716 million, an increase of 1.85% over December 2021 [1] [2].



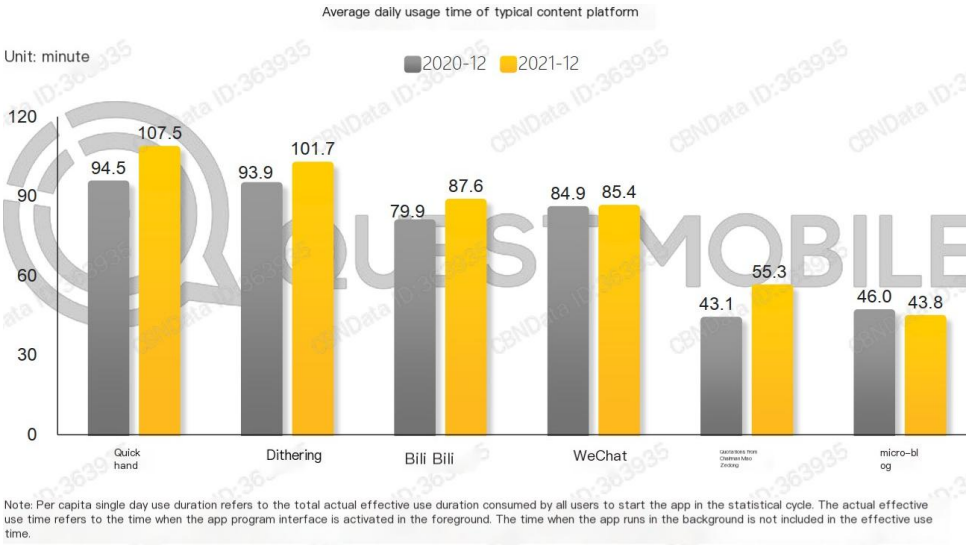
Source: CNNIC's 49th and 50th Statistical Report on China's Internet Development

Figure 1 Statistics of the number of video service users

Video is the application with the longest user immersion time. As shown in Figure 2, as of June 2022, the per capita online time of Internet users in China will be 29.5 hours per week, an increase of 1.0 hours compared with December 2021. Chinese netizens spend more than 4 hours online on average every day), and nearly half of the time is spent on short video platforms, such as TikTok and Kwai.

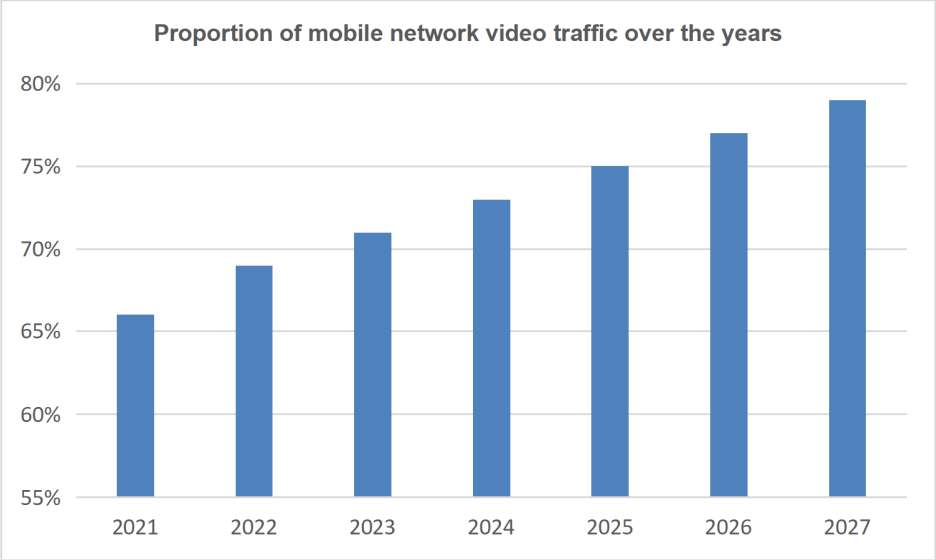


Source: CNNIC's 50th Statistical Report on Internet Development in China
Figure 2 Statistics of monthly active users of video service



Source: QuestMobile2021 China Mobile Internet Annual Report
Figure 3 Statistics of user daily use duration platform

In addition, video services have become the main source of mobile network traffic at this stage. According to Ericsson Mobile Market Report in June 2022, video traffic currently accounts for 66% of all mobile traffic, and is expected to grow to 79% by 2027 [3].



Source: Ericsson Mobile Market Report from 2019 to 2022

Figure 4 Statistics of video traffic proportion in mobile network

1.2 Mobile video service has undergone great changes

As shown in Figure 5, since 2020, various video application platforms have launched differentiated functions to improve users' viewing experience. We can see that the definition of video content is constantly improving, and the interactivity is also increasingly strong. At the same time, the entire video content is gradually changing from plane to 3D.

Source: Internet public data sorting

Figure 5 Development Trend of Video Applications

The development of mobile video formats can be summed up in the following aspects:

A. **Ultra high definition:** The performance of ultra high definition is the improvement of video resolution. For example, the mobile terminal of the mobile phone is gradually transitioning from viewing 1080p content source to 2K, 4K or higher. In particular, the mobile video application platform supports the playback of 180 °/360 ° panoramic video and free angle video. Users also have strong demands for viewing content sources with resolution above 1080P on mobile phones.

B. **Spatialization**: Spatialization of video is the most significant change compared with plane video. Mobile terminal viewing plane video is often based on a fixed limited perspective provided during production and shooting of content source. Spatial videos, such as 180 °/360 ° panoramic videos and free angle videos, can be operated by users through mobile terminal sliding screens and gyroscopes to select the desired angle of view in the video content source. Of course, the spatialization also brings about the interactivity of video, and the degree of user interaction is gradually improving.

C. **Interactive**: video interactive segmentation. The first is the interaction between users and video content, that is, the interaction between people and content brought by the spatial development of video mentioned above. On the other hand, it is the interaction between users, such as Youku's "watch together" and Tencent's short video content social networking. This kind of video content is more used as a carrier to break through the social barriers between different users, improve the interest of video watching, and further improve the viewing experience of users.

1.3 Mobile video plays an important role in promoting industrial upgrading

In 2021, the number of short video users will continue to grow, and the industry will still maintain a steady growth trend. On the one hand, the short video platform accelerates the layout of knowledge fields and promotes knowledge dissemination; On the other hand, it constantly integrates with traditional industries to create greater economic value. Short videos promote knowledge dissemination and become an important channel for information dissemination. Since 2020, all major short video platforms have, on the one hand, vigorously supported content creators and encouraged pan knowledge content output; On the other hand, actively develop new forms such as new functions of video collections and live classes to create a multi-level and three-dimensional knowledge map. In terms of breadth, the knowledge content of the platform has covered life, education, humanities, finance, military and many other fields, fully meeting the diversified needs of users; In terms of depth, the platform promotes the systematic dissemination of knowledge and

enhances the depth of knowledge learning by introducing video collection and other functions, creating live public classes for famous teachers in famous schools and other forms. In 2021, Diaoyin launched four "Mengzhi Plan", invested 10 billion traffic to support knowledge creators, and encouraged the creation of more knowledge content suitable for young people to learn; Fast Hands launched two seasons of large-scale live broadcast activity "Fast Hands New Knowledge Broadcast" to provide users with a new cognitive perspective and access to knowledge. Short videos are deeply integrated with agricultural products and cultural and tourism industries to stimulate economic vitality. First, short video applications help agricultural product sales. Source farmers and businesses publicize and promote high-quality agricultural products through short videos and live broadcasts, opening up a market for agricultural products in cities. The data shows that from January to October 2021, there will be more than 420 million orders for agricultural products sent from rural areas to all parts of the country through live broadcast e-commerce. Compared with the same period in 2020, the sales and orders of agricultural products have increased by 88% and 99% respectively. In addition, the short video platform also provides professional training for farmers and rural entrepreneurs to ensure the sustainable development of short video and live broadcast sales mode of agricultural products. Second, short video applications stimulate the vitality of the cultural tourism industry. At the level of cultural industry, the short video platform has fostered the understanding and curiosity of the young generation on intangible cultural heritage and helped to explore the cultural and market value of "intangible cultural heritage" by strengthening traffic support, improving liquidity, building an open platform and carrying out urban cooperation. At the tourism level, the short video platform has continuously strengthened cooperation with Xi'an, Chongqing, Nanjing and other cities to attract tourist attractions to settle in and promote, help the city image dissemination and promotion, and drive the development of tourism.

1.4 The current industry lacks mobile video experience evaluation system and standards

At present, the industry has formed a relatively complete video user experience evaluation system and standards in the field of DVB, IPTV and OTT Internet TV (such as YD/T 3776-2020, YD/T 3777-2020, YD/T 3778-2020, YD/T 3779-2020 and other broadband video service user experience evaluation series industry standards), and the monitoring probe products and systems based on this standard have been widely deployed in the existing networks of radio, television and telecommunications operators, Effectively ensure that large screen TV users can obtain good experience quality. However, in the field of mobile short video and UGC live broadcast services, the industry has not yet formed a relatively complete user experience evaluation system, test scheme and related standards. This white paper aims to propose a set of effective test indicators, test plans

and test data for mobile short video and UGC live broadcast services, so as to provide strong data support and practical basis for promoting the formulation of relevant standards.

2. Defining a unified mobile video experience evaluation scheme is critical to the healthy development of the industry

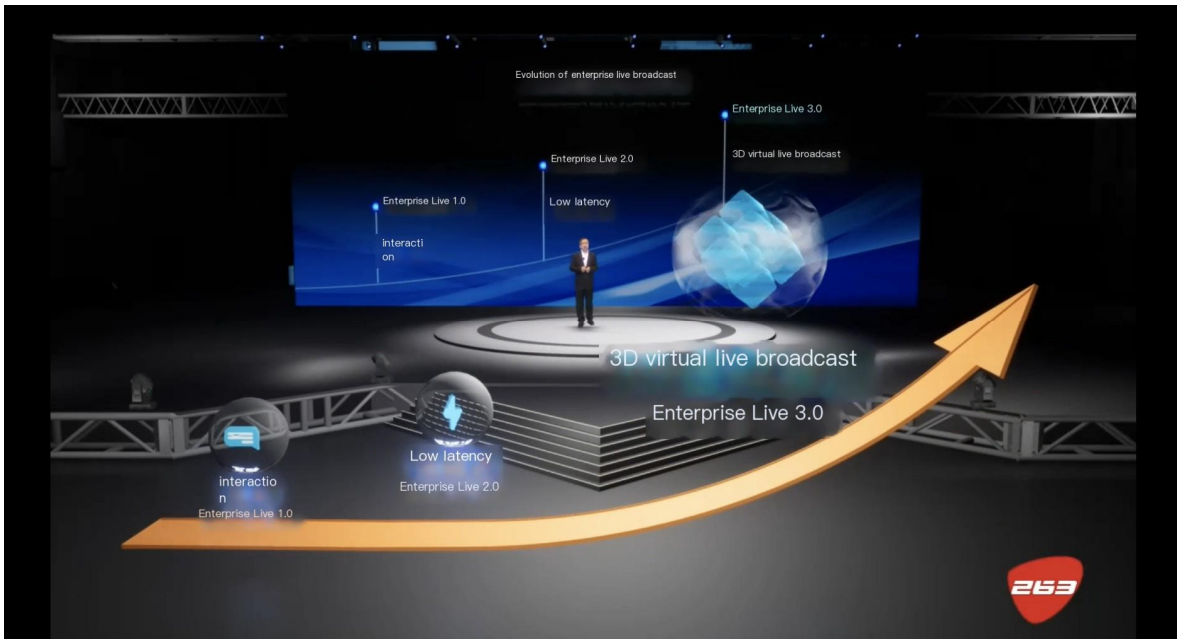
2.1 Diversity of factors affecting user retention under mobile video service

With the development of the 5G network and the modernization of Internet infrastructure, various video applications have sparked a wave of video Internet over the past ten years, and video production and sharing are migrating from conventional radio and television (radio and television) to the Internet and OTT on a large scale. Due to the variety of mobile video services and video formats, there is a multi-faceted trend of elements influencing user retention and experience.
On the ToC (that is, for mass consumers) side, users have a higher quality pursuit for the viewing experience of ultra-high definition video. Taking "Frame Sharing" as an example, it is an ultra-high definition solution jointly launched by Youku, video platform, film and television producers and terminal equipment manufacturers. Through the technology of "Frame Sharing Ultra High Definition Cloud+Terminal+Intelligence", it links video production and terminal playback to create the ultimate audio-visual playback effect. Statistics show that compared with users watching 1080P resolution, the conversion rate of members of users watching "Frame Sharing" UHD has increased by 67%.

The ability of the interested short film to be opened fast is a direct determinant of users' retention and playback of data in the field of short videos. Youku enhanced and improved the initial screen latency in the brief video sequence. After optimization, the single user VV (video view video playback times) rose on both the iOS and Android sides by 20%+ and both ends by 29%+, reaching a total of 60%+.

In the field of ToB (that is, for enterprise users) live broadcast, as shown in Figure 6, enterprise live broadcast has evolved from the original one-way live broadcast to interactive live broadcast. Whether it is the anchor answering the audience's questions about the products during e-commerce live broadcast, or the one-on-one interaction between exhibitors and potential customers in digital exhibitions, it breaks through the limitation of watching by multiple people. The continuous enhancement of interactivity has led to a variety of play methods for enterprise live broadcast activities. Enterprise live broadcast builds a bridge for two-way communication between enterprises and users, and becomes a new interactive window. With the help of live broadcast,

enterprise customers can show their own image, communication ability and positioning, and deliver vivid "business cards" to the audience. Many elements in ToB live broadcast will affect the retention and renewal of corporate customers, such as the smoothness of the picture, whether the functions are simple and clear, and the image and quality of the anchor.



Source: Internet public pictures

Figure 6 Evolution of enterprise live broadcast

In the era of full network distribution of content, it's crucial to consider whether the platform fosters the development of video content if you want to break into the video market. As far as Zhihu is concerned, as a well-known knowledge community in China, the videos in the explanation category with question and answer as the main form are highly consistent with the community atmosphere. The strengthening of Zhihu's video content trend has promoted the creation and efficient circulation of excellent content. The daily active user penetration rate in the area of high-quality video with typical "sense of gain" features has reached 40%.

In the era of mobile Internet, the live broadcast industry has become more systematic. At present, the field of live broadcasting is in full swing, involving entertainment, games, e-commerce, social networking and other live broadcasts that people see and hear, involving signal collection, transmission, and ultimately presentation on mobile phones. Each link affects the user's audio-visual experience. For example, in the e-commerce industry, a considerable number of anchors and institutions have realized the importance of live broadcast experience for transactions, and have made key investment in improving and optimizing the resolution, code rate, first screen delay, end-to-end delay, and stuck rate.

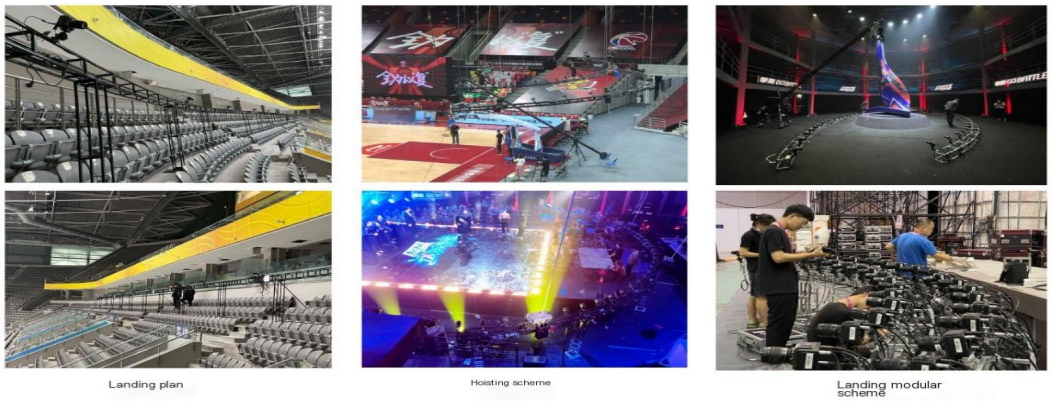
2.2 OTT video service innovation to improve user experience

According to the 2021 China Online Audiovisual Development Research Report [4] released in June 2021, 28.2% of China's 944 million online audiovisual users will choose to watch videos at double speed, especially nearly 40% of post 2000 users will choose to watch videos at double speed. At the same time, as shown in Figure 7, the research results released by Volcano Engine show that when the user's first screen delay exceeds 200ms, the probability of the user quitting the current playing video will be increasing. Ideally, the first screen delay should be controlled within 100ms. The above data shows that video viewing forms are changing in more and more diverse forms, so it is of positive practical significance to study the video experience indicators under interactive circumstances.

Source: Internet public data

Figure 7 Volcano Engine Video First Frame Delay and User Behavior Investigation

In order to improve the user experience and innovate the content form, different platforms will take different measures to promote the retention and innovation of users. For example, Youku took the lead in landing the ability of "free perspective" on the end side, and users can freely choose the viewing angle during the viewing process. Taking this is the fourth season of hip-hop dance as an example, the total free perspective VV (video view, video playback) is nearly 1.5 million, and the total PV (page view, page views) is 1.4 million+, which significantly improves the interest and immersion of users. As shown in Figure 8, the key special project of "Science and Technology Winter Olympics" participated by Youku - "Ice and snow project - interactive multi-dimensional experience technology and system" was successfully put into use in the Beijing Winter Olympics.



Source: Internet public pictures

Figure 8 Youku's free perspective live shooting scheme

Users are becoming more and more accustomed to the interactive form of "bullet screen" when watching videos. It has gradually become a habit of watching movies to express their feelings and communicate with other users through bullet screens. However, when the number of bullet screens is too large, the screen will be blocked, resulting in users who mainly watch movies are not as experienced as before. Based on this, many platforms have also introduced the function of "bullet screen piercing people". Taking a video as an example, according to user feedback, more than half of users believe that bullet screen penetration meets their own needs; From the perspective of objective data performance, users generally increased the number of bullet screen opening lines after the bullet screen was penetrated online, and were more willing to share interactively through the bullet screen.

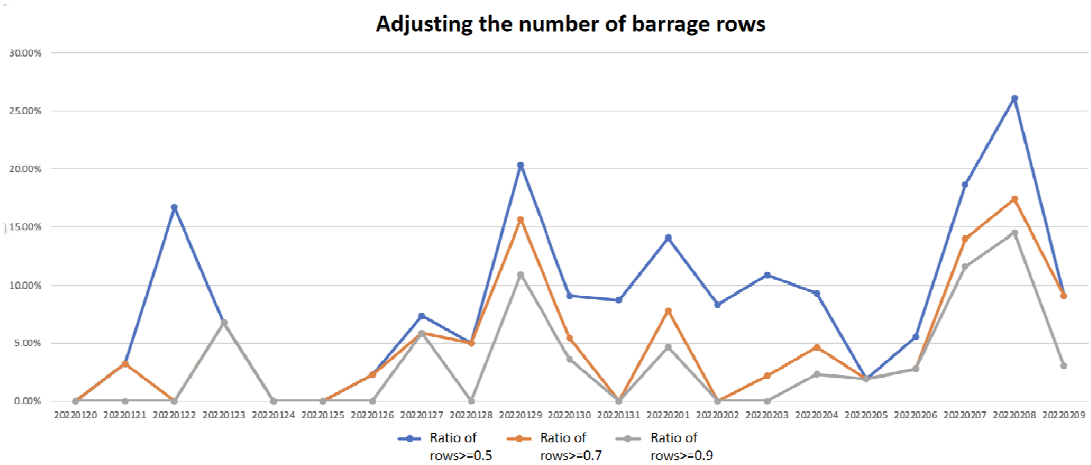


Image source: Internet public data sorting

Figure 9 After the bullet screen penetrates the release time point (January 27, 2022), users are more likely to select more bullet screen lines after opening the panel to adjust the number of bullet screen

Users' pursuit of ultra-high definition is becoming more and more urgent and popular. In terms of UHD video experience, "Frame Share" has not only evaluated the quality of video sources, but also built a complete set of device authentication system to authenticate the chip decoding and

computing capabilities of mass devices, screen refresh rate, resolution, brightness, color gamut and display effect. In view of the current situation that different manufacturers have different post-processing capabilities for terminal devices, especially when the image quality processing of low-end devices may cause the final display effect to become worse, Framing Share uses the massive computing power and complex algorithms on the cloud to analyze and send the analyzed algorithm information to the terminal, and the terminal device performs image quality processing in combination with the content information on the source. Frame sharing ensures that users can see the best effect on the terminal through the cloud+end+intelligent cloud linkage mode. Data shows that the conversion rate of frame sharing UHD members is higher than 1080p.

Image source: Internet public pictures

Figure 10 Chang'an Twelve Hours 1080p vs Frame Sharing (need to zoom in)

Compared with other video applications such as pan entertainment, e-commerce live broadcasting has more complex imaging environments, including lighting, and more uneven imaging equipment. Users will obviously prefer live broadcasting with low bit rate but high image quality. Now Taobao, as the head e-commerce platform, is committed to providing consumers on the platform with broadcast and television level video quality experience through long-term technology investment and innovation. Through the full link optimization in coordination with the imaging environment, specific scenes, hardware and transmission in the live broadcast room, Taobao live broadcast technology supports high-definition 1080P live broadcast with a very low bit rate, and can even control the average bit rate at about 1Mbps.

Image source: Internet public pictures

Figure 11 The case of a tea set industry in TaoLive: 1080p@30fps , bit rate~0.8Mbps

The first screen performance of the live stream in the live scene is also a direct factor that affects the user's retention and playback of data. Zanthoxylum bungeanum Live has specially optimized the video's second opening performance. The time required to open the video live broadcast has been reduced to less than 25% before the optimization, and the user experience has been greatly improved. One month after the launch, data analysis shows that the per capita live viewing number of iOS/Android users has increased by 7.2% month on month.

In the field of ToB, from the first year of enterprise live broadcasting in 2018, live broadcasting has become an essential capability of enterprises; By 2021, the first year of live broadcast marketing, a large number of enterprises have recognized the value of live broadcast and started live broadcast marketing. In just a few years, live broadcast has accelerated its rapid development. At the same time, corporate customers have increasingly high requirements for interactive experience of live broadcast, especially for strong interactive live broadcast scenes such as e-commerce live broadcast and online education. Interactive experience can almost be called the most critical link to measure the success of live broadcast. 263 Cloud Communication publishes low latency live broadcast through self-developed NRTC real-time audio and video bottom technology, and deeply optimized live broadcast can control the end-to-end delay at least within 200ms. After optimization, the millisecond level delay free live broadcast of 263 Cloud Communications has seen a significant increase in the number of customers in education and training, live broadcast with goods, live broadcast of events and other fields.

In 2020, under the influence of the new crown prevention and control policy, online life has become normal from the original short-term exception. Human real life has gradually migrated to the virtual

world, and the boundary between virtual and real life has become increasingly blurred. Virtual live broadcast also came into the public's view at this time. Virtual live broadcast will not only present the sound and picture of the real world, but also live broadcast based on real characters superimposed with virtual environment, and even XR live broadcast composed of digital characters and virtual environment. In addition to having strong visual impact, the highly immersive live broadcast experience can turn the audience from a bystander to a participant. There will be more possibilities for interaction between enterprises and target groups. In the 3D virtual live broadcast scene created by 263 Cloud Communication, it uses virtual reality technology to build and monitor virtual scenes, combines the behavior of characters with virtual scenes, and realizes the simultaneous output of virtual imaging and live broadcast scene through green screen cutout, real-time rendering, ultra-low delay interactive tracking and other technologies, creating a virtual reality environment that integrates virtual reality, and realizing the important transformation of audience from "online" to "on-site", At the same time, it also sets up an interactive entrance for the virtual world and the real world, so that the two worlds can have more integration and connection. As shown in Figure 12, the virtual live broadcast has been widely promoted and applied in medicine, finance, education and other fields.

Image source: Internet public pictures

Figure 12 Application of 3D virtual live broadcast in the field of education

With the further integration of virtual live broadcasting with VR/AR, the Internet, cloud computing, the Internet of Things, artificial intelligence and other emerging technologies, it is bound to usher in the great integration of interactive hardware+immersive content+full scene applications, achieve deeper virtual reality interaction, and open a new era of commercialization of the virtual reality industry.

2.3 Unified experience evaluation standards are common and urgent needs of the industry

To sum up, in the current situation where mobile video forms are so diverse, users' demands for viewing experience in different dimensions show diversity. At the same time, the platform policy's research and development and provision of various innovative capabilities with different needs has turned the current user experience into a broader and multidimensional indicator. The subjective level includes interactivity, ease of use, immersion, clarity, etc., while the objective level includes first screen delay, stuck rate, etc. Based on this, a unified experience evaluation system and a unified experience evaluation standard are common and urgent needs of the industry.

3. Mobile video service experience evaluation scheme and status quo

3.1 Short video service

3.1.1 Short video service process

As shown in Figure 13 below, the short video business experience includes three parts: acquisition, transcoding and consumption.

In the collection part of short video, the resolution code rate uploaded by the collection end determines the highest quality of the video. Currently, 1080P is the main source of the uploaded video on the terminal side, and the highest 4K resolution of high dynamic range (HDR) is supported.

Acquisition part: different resolution versions are formed on the server side for different player conditions.

Transcoding part: the video content collected in large quantities is stored on the cloud server. At the same time, the same video content is converted to different resolution and bit rate versions based on cloud computing power. At the same time, the video content is audited with the help of the artificial intelligence (AI) computing power in the cloud.

Consumption part: The consumer side users get the source from the content delivery network (CDN) side to play. Since it is an active push, the CDN will be put in priority for those to be pushed, and the hit rate of CDN in the province will reach more than 90%. CDN miss scenes will go back to the source station to get the film source. CDN hot videos in different regions are different, and the provincial/city level will not sink and subdivide.

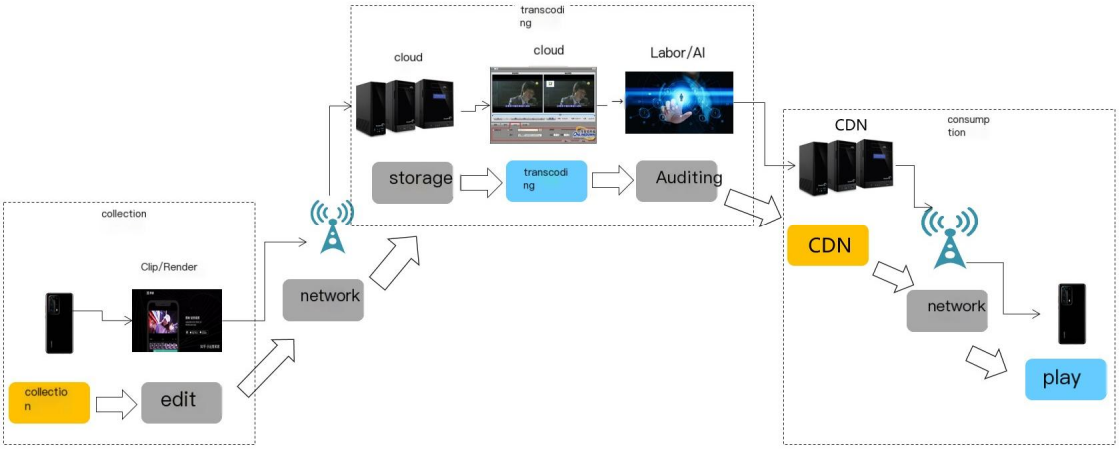


Figure 13 Short video production and consumption process

When a user sends a viewing request, the streaming media service system will actively push the program information stored in the source library according to the user's preferences, and transmit it to the user terminal through high-speed transmission network in the form of video and audio streaming files. Short video is generally realized by unicast network, which uses a protocol to transmit IP packets from an information source to a destination. At this time, information is received and transmitted only between two nodes.

In this paper, the evaluation of short video service mainly focuses on the user experience evaluation of the consumption part. OTT platform acquires user experience and behavior by collecting experience indicators in terminal applications. When a user watches each short video, the application side will record the current user's experience and report it to the OTT data server. The experience data indicators in this article are the user monitoring data from such OTT applications.

3.1.2 Key indicators of short video experience evaluation

Short video is realized through the mobile video service system of Internet service providers. The short video involved in this document covers the following mainstream network transmission protocols: HTTP/TCP, QUIC/UDP, etc. Streaming media transmission protocols include DASH, HLS, etc; The following video coding formats are adopted: H.264, H.265, AV1, etc; Display device: mobile terminals with screen size less than 10 inches, such as smart phones and tablets.

A. Experience evaluation based on single KQI indicator

The process of using short video includes starting the terminal/player, selecting viewing content through sliding screen, viewing and other operations. Therefore, the key indicators affecting short video user experience are extracted from the following three aspects: audio-visual experience

quality, presentation experience quality and interactive experience quality.

a) Audiovisual experience quality is mainly related to content clarity, coding parameters, etc. Here, the code rate is extracted as the key evaluation index of short video audio-visual quality.

b) The quality of presentation experience, that is, the user's viewing/listening experience, depends on the quality of the program signal that appears during the viewing process. The influencing factors include the jam, the non synchronization of audio and video, etc., which can be measured by the objective indicators such as the transmission performance and quality damage of video and audio information. Here, the jam is proposed as the key evaluation indicator of the quality of short video presentation experience.

c) The quality of interactive experience, that is, the user's experience of the interactive process, depends on the response speed of the system to the user's interactive operation, and covers the platform, network, and terminal performance indicators. Here, the first screen delay in the short video viewing process is taken as the key evaluation indicator of the quality of interactive experience.

From the micro model of short video service and interaction mode, we can normalize the jam in the presentation experience quality and the first screen delay in the interaction experience quality.

B. Experience evaluation based on comprehensive experience score

Of course, in addition to the experience evaluation based on a single indicator, the short video user experience evaluation can also comprehensively consider the audio-visual experience quality, presentation experience quality, interactive experience quality and other factors. The user's video experience quality score, which combines the audio-visual experience quality, presentation experience quality and interactive experience quality, is the average subjective score (MOS) that truly reflects the user's experience perception. Follow the standard T/INFOCA6-2022- mobile short video user experience quality (QoE) evaluation method in detail.

3.1.3 Status quo of short video experience evaluation results

A. Overall situation of 5G short video service experience evaluation

The 5G short video service looks at the overall performance of the current user from the distribution of three single KQI dimensions: first screen delay, bit rate and resolution. The user experience data of short video service comes from several short video service providers of the Alliance, covering 22 provinces, 5 autonomous regions and 4 municipalities directly under the Central Government (two special administrative regions and Taiwan Province of China are not involved for the time being), and the overall scale of data volume reaches 10000. On the basis of the median bit rate of 1.1M, the first screen delay of short video still exceeds 50%, which cannot meet the basic experience requirements of 400ms. In terms of resolution, 70% of users can experience 720P and above UHD video, but there is still room for improvement.

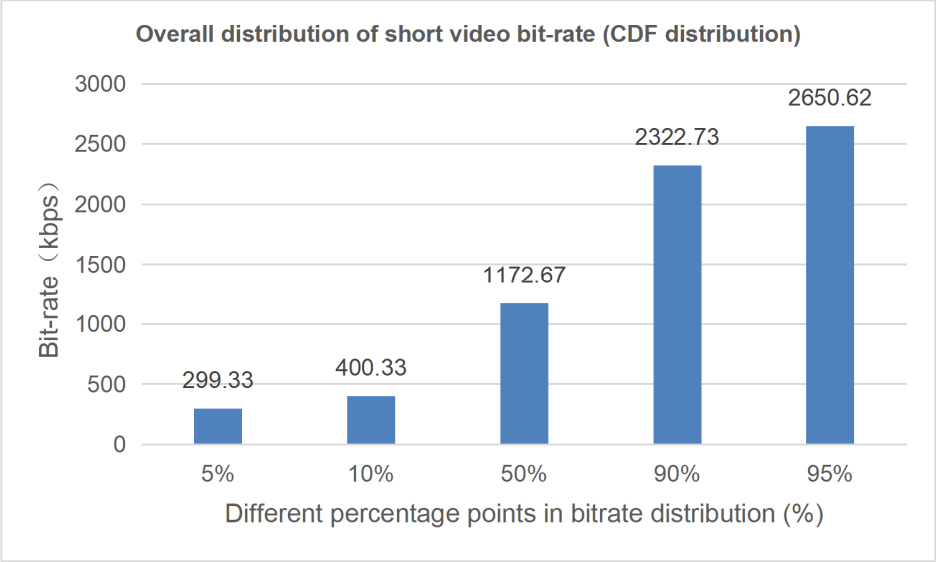


Figure 14 Overall Distribution of Short Videos Bit-rate

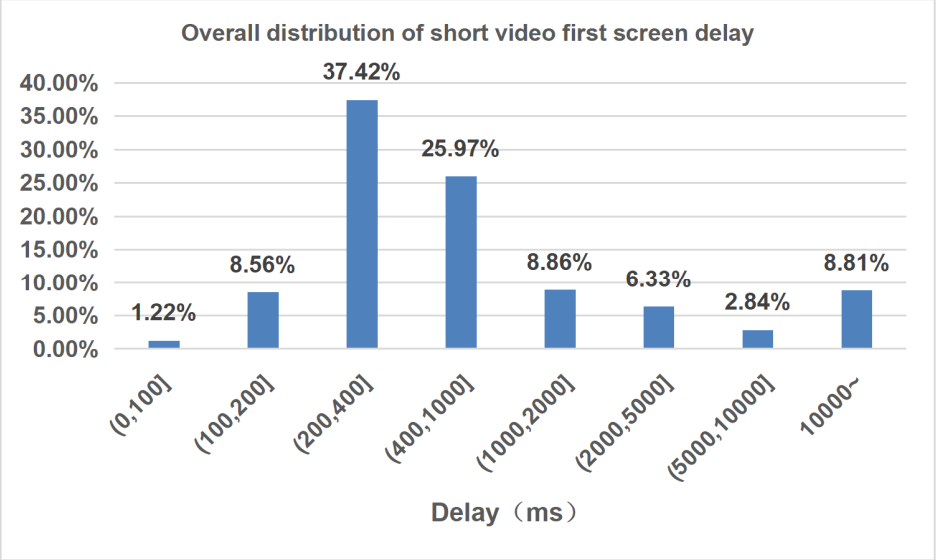


Figure 15 Overall Distribution of Short Video First Screen Delay

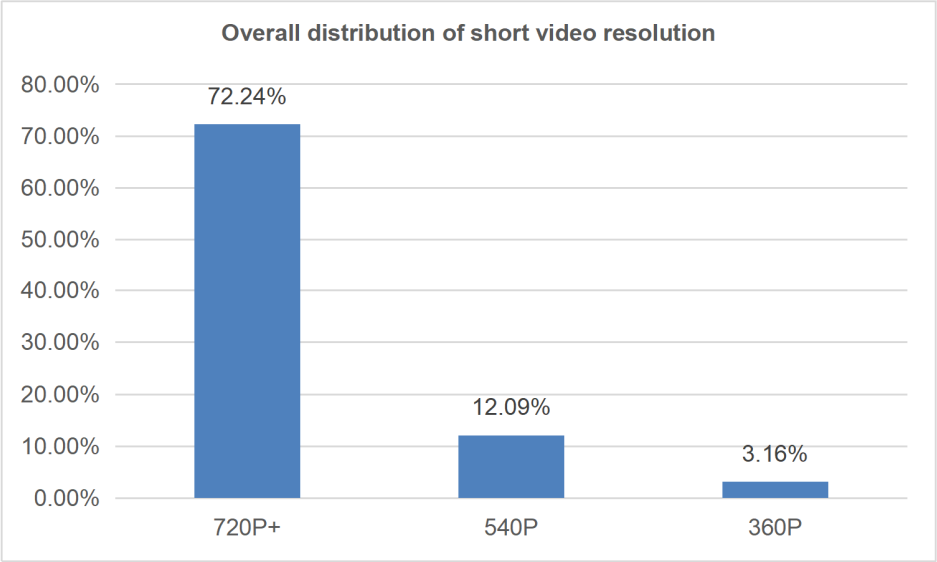


Figure 16 Overall Distribution of Short Video Resolution

B. Comparison of experience differentiation between 4G and 5G networks

In terms of bit rate, there is no significant difference between 4G and 5G networks. The main reason is that the network based adaptive bit rate selection strategies of each short video platform are different.

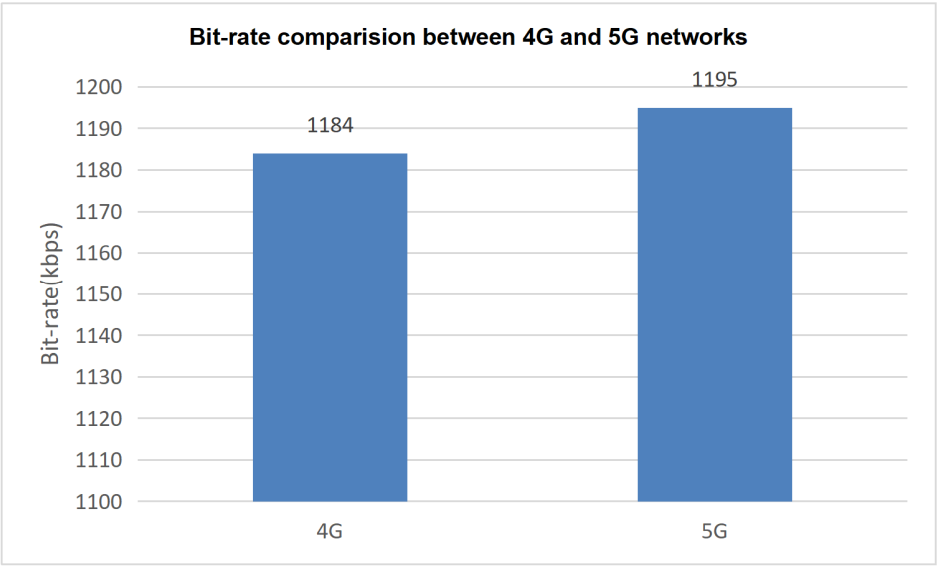


Figure 17 Bit rate comparison between 4G and 5G networks

Compared with the first screen delay of 4G network and 5G network, the experience under 5G network is significantly better than that under 4G network, and the delay under 5G network is reduced by 25%. From here, we can clearly see the huge advantages brought by 5G network. However, the average first screen delay under 5G network is about 300ms, which is still far from the 100ms extreme experience expected by users mentioned above.

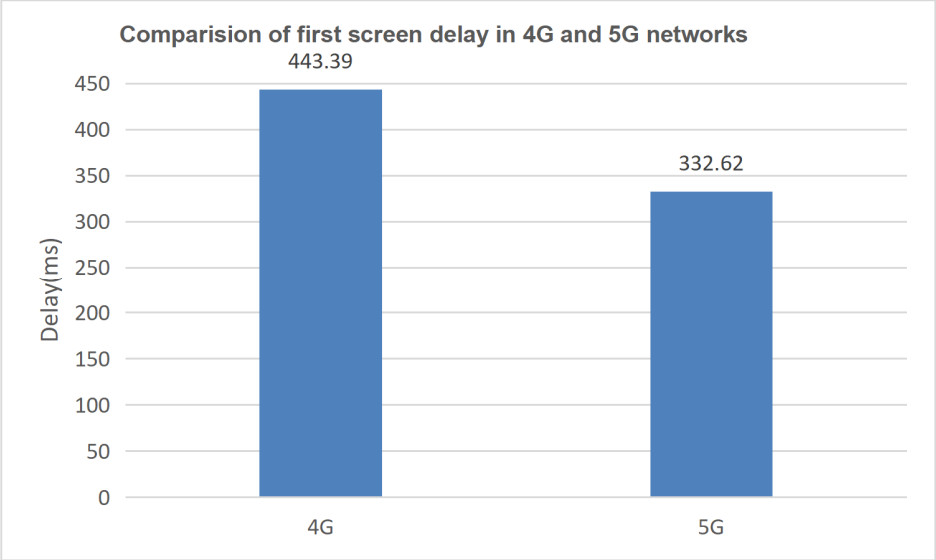


Figure 18 Comparison of first screen delay in 4G and 5G networks

The records with the first screen delay less than 1 second are extracted from the short video records, and the proportion of poor experience quality (the first screen delay exceeds 400ms) in 4G and 5G networks is obtained. The proportion of poor quality in 5G network is 24% lower than that in 4G network.

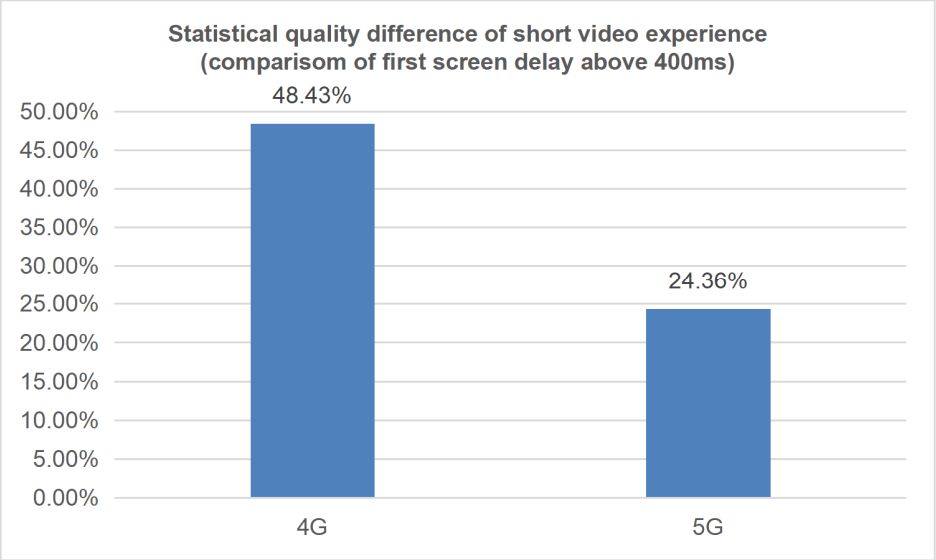


Figure 19 Comparison of 4G and 5G first screen delay

C. Comparison of 5G experience evaluation among operators

Horizontal comparison is made from the operator dimension. For the single key quality indicator (KQI) dimension, the overall first screen delay is compared. Operator B < operator A < operator C. When viewing the proportion of first screen delay of more than 400ms synchronously, operator B

< operator A < operator C. There is no obvious difference in the bit rate performance of short videos.

Based on the comparison of first screen delay and bit rate, from the analysis of existing data samples, the short video service experience under operator B network is optimal.

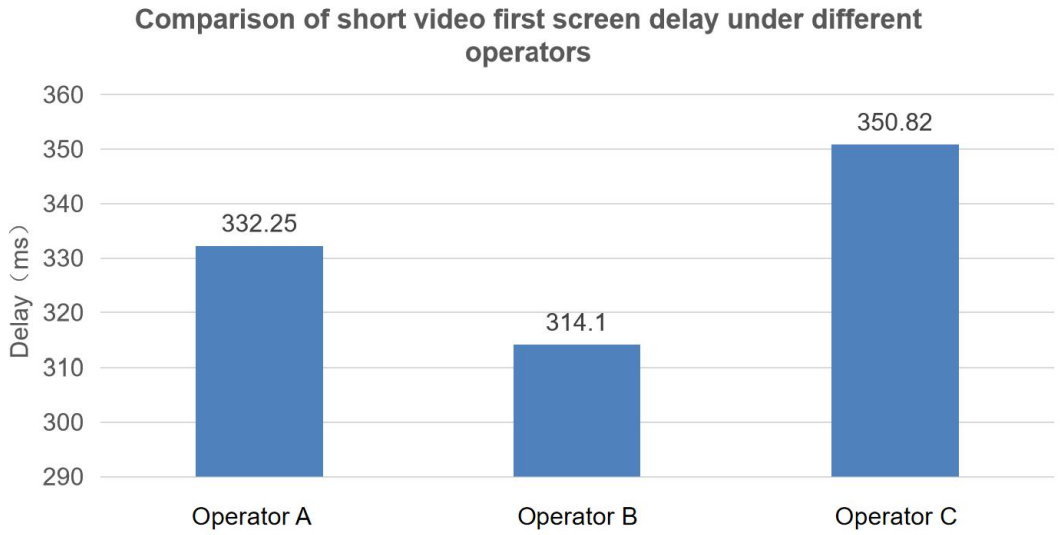


Figure 20 Comparison of first screen delay under 5G networks of different operators (first screen delay numerical optimization)

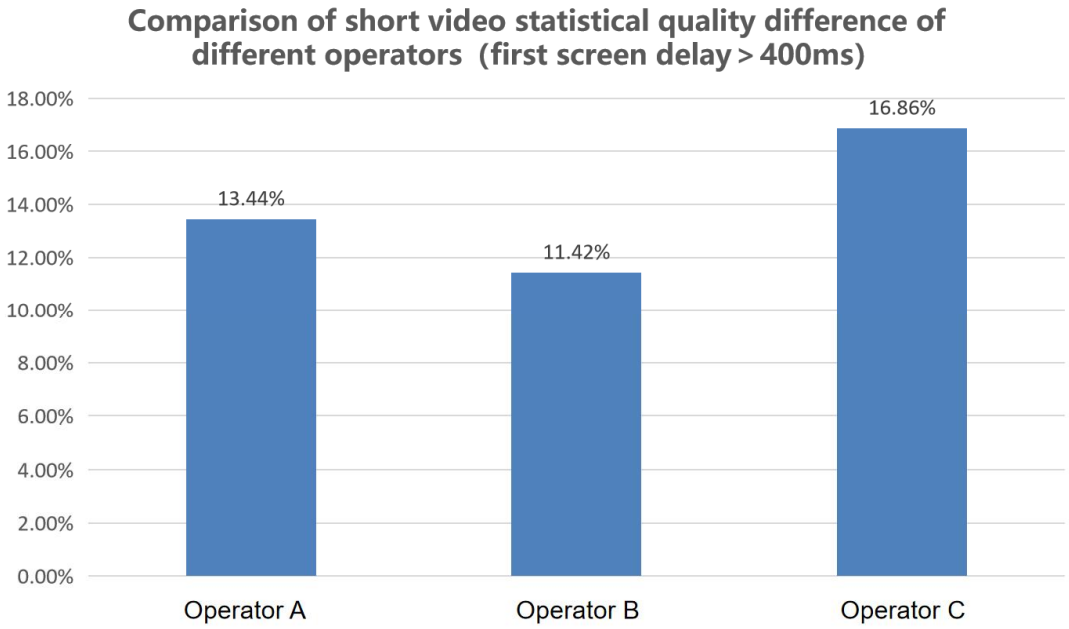


Figure 21 Comparison of proportion of first screen delay above 400ms in 5G networks of different operators

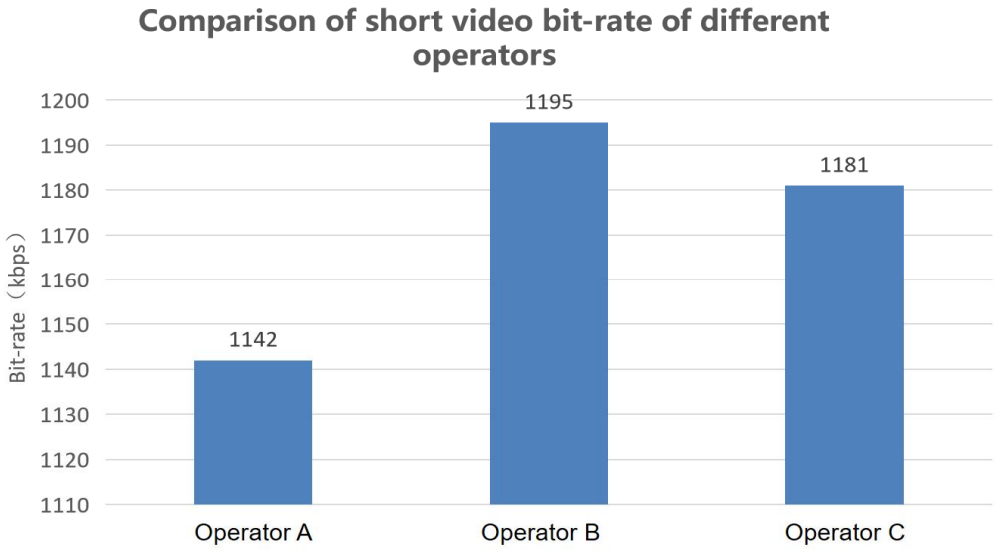


Figure 22 Comparison of Bit Rates in 5G Networks of Different Operators

3.2 Live video (viewing end)

3.2.1 Live video service process

The process of mobile live broadcast is that the host sends the generated audio and video stream to the audio and video cloud platform in real time, and the audience receives the audio and video stream from the host in real time for viewing. In the process of mobile live broadcast, real-time audio and video interaction will be frequently carried out between the anchor and audience, or the anchor will be connected to the microphone PK, which is relatively interactive and requires high end-to-end delay. The user roles in the live broadcast are divided into the host side and the audience side.

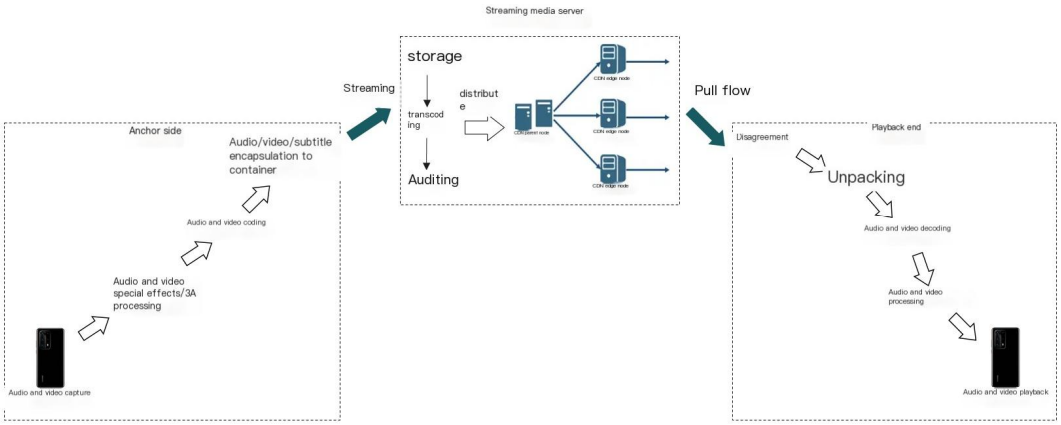


Figure 23 Live broadcast business flow chart

The service experience evaluation of live video described in this article is mainly aimed at the user experience evaluation of live audience side. The business experience data collection principle of live broadcast (viewing end) is similar to that of short video service, which is also obtained through statistics in the OTT application on the terminal side.

3.2.2 Key indicators of live broadcast (viewing end) experience evaluation

A. Experience evaluation based on single KQI indicator

In live broadcast, the audience receives the audio and video stream generated by the host in real time for viewing. The main indicators of audience experience in live broadcast are extracted from three factors: audio-visual experience quality, interactive experience quality and presentation experience quality. Audio visual experience quality includes audio experience quality and video experience quality.

a) The quality of audio-visual experience depends on the clarity, smoothness and other factors of audio and video. Here, the resolution and bit rate of live broadcast (viewing end) are taken as the main experience evaluation indicators.

b) Interactive experience quality, live broadcast (audience side) interactive experience quality also includes the jump delay of the audience side in the live broadcast room, the first screen delay of the audience entering the live broadcast room, etc. The jump delay of the audience end in the live broadcast room refers to the time required for the audience in the live broadcast room of the host A to jump from the live broadcast room of the host A to the live broadcast room of the host B in the scenario of the host connected PK (such as the host A and the host B connected PK). The first screen delay of the audience entering the live broadcast room refers to the time from the time when the audience clicks to enter the live broadcast room of the host to the time when the first frame of the live broadcast room is displayed successfully. Here we focus on the experience of audience end users, and extract the first screen delay as the main experience evaluation index.

c) The experience quality of audience presentation can be measured by such indicators as audio jam rate, video jam rate, splash screen, audio and picture synchronization delay. Here, the video jam rate of the viewing end is taken as the main experience evaluation indicator of presentation experience quality.

B. Experience evaluation based on comprehensive experience score

As with short videos, in addition to conducting experience evaluation based on a single indicator, live broadcast (viewing end) user experience evaluation can also comprehensively consider the audience's audio-visual experience quality, interactive experience quality, and presentation experience quality, which is the average subjective score (MOS) that truly reflects the mobile live

broadcast audience's user experience perception. Follow the standard T/INFOCA6-2022- Quality of Experience (QoE) of mobile live video in detail.

3.2.3 Status quo of live broadcast (viewing end) experience evaluation results

A. Overall situation of live broadcast (viewing end) business experience evaluation

For the user experience of live broadcast (viewing end), this paper focuses on the analysis of the distribution of its bit rate and first screen delay under the 5G network. The user experience data of the live broadcast (viewing end) service comes from several short video service providers of the Alliance. The data covers 22 provinces, 5 autonomous regions and 4 municipalities directly under the Central Government (two special administrative regions and Taiwan Province of China are not involved for the time being), and the overall scale of the data volume reaches 10000. Through the analysis of the data, it is found that the overall bit rate of 5G is higher than that of short video, and the 1M~2M bit rate interval accounts for the highest proportion, about 46%. However, the first screen delay is generally large, starting at 200ms.

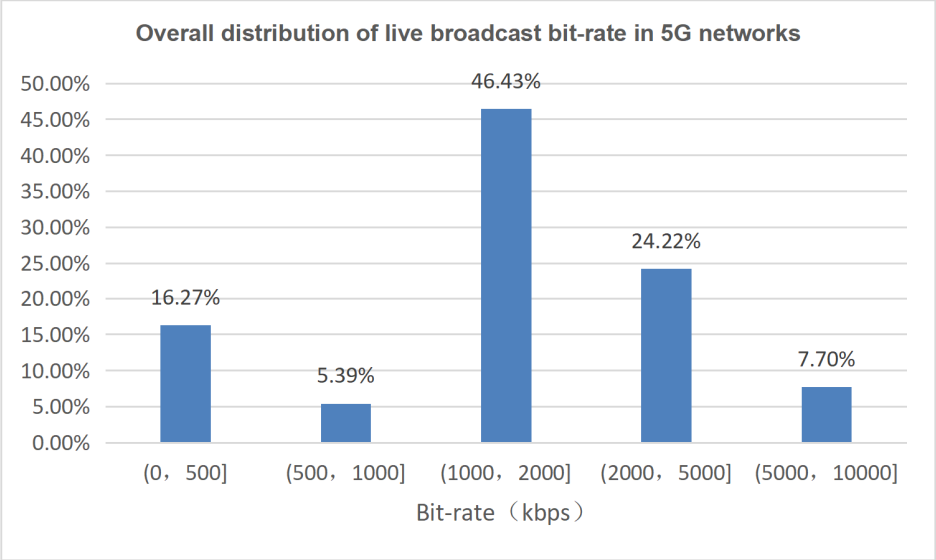


Figure 24 Overall Distribution of Live Broadcast Bit-rate

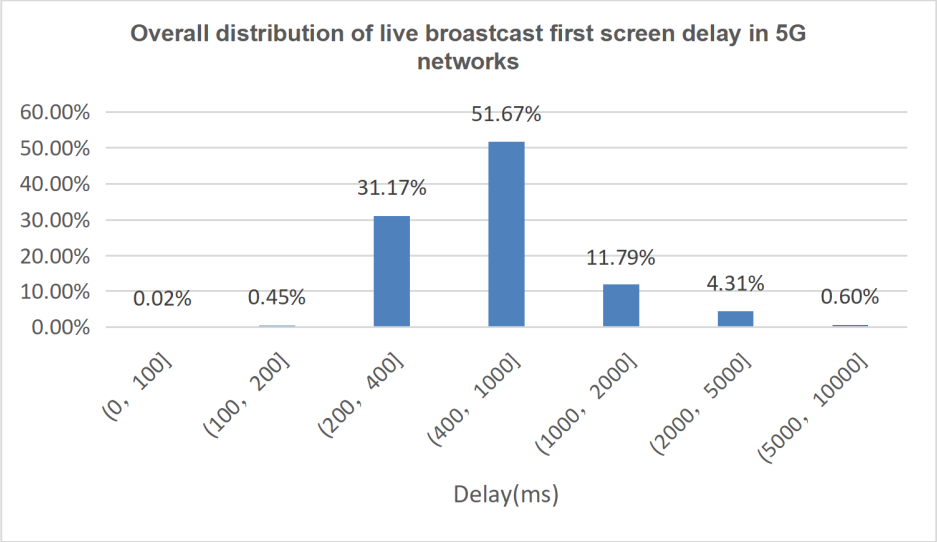


Figure 25 Overall Distribution of Live Broadcast First Screen Delay

B. 5G experience evaluation and comparison among operators

Also look at the business experience performance of live broadcast (viewing end). In terms of the first screen delay, the experience quality of each operator is poor (the first screen delay is more than 400ms). Operator A is the worst, and operator B is equivalent to C. In terms of the proportion of high bit rate above 1M, carrier B has the highest proportion. Obviously, the better network conditions of operator B enable users to experience higher bit rate live video services.

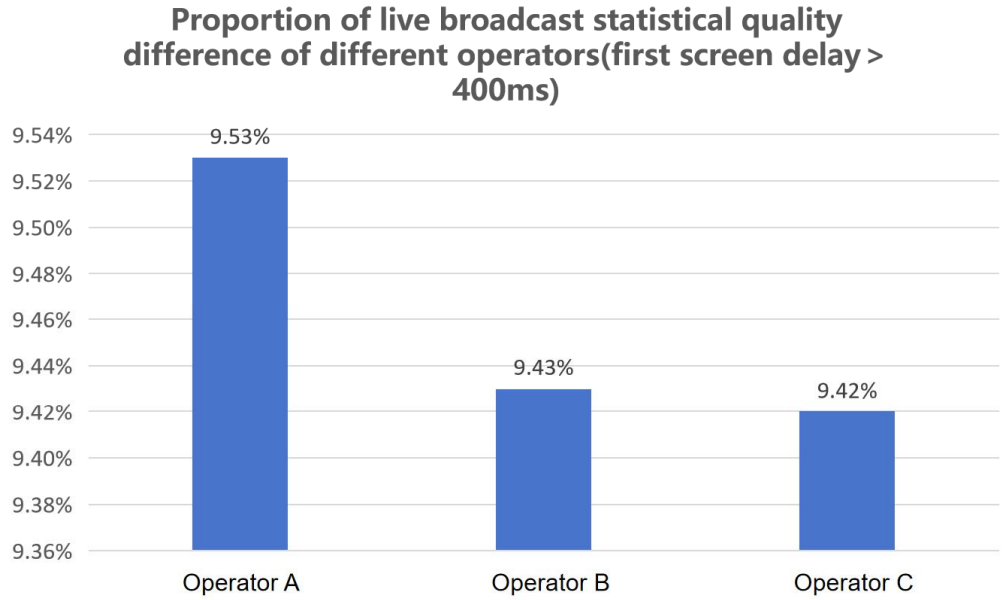


Figure 26 Comparison of first screen delay ratio of live broadcast over 400ms under 5G networks of different operators

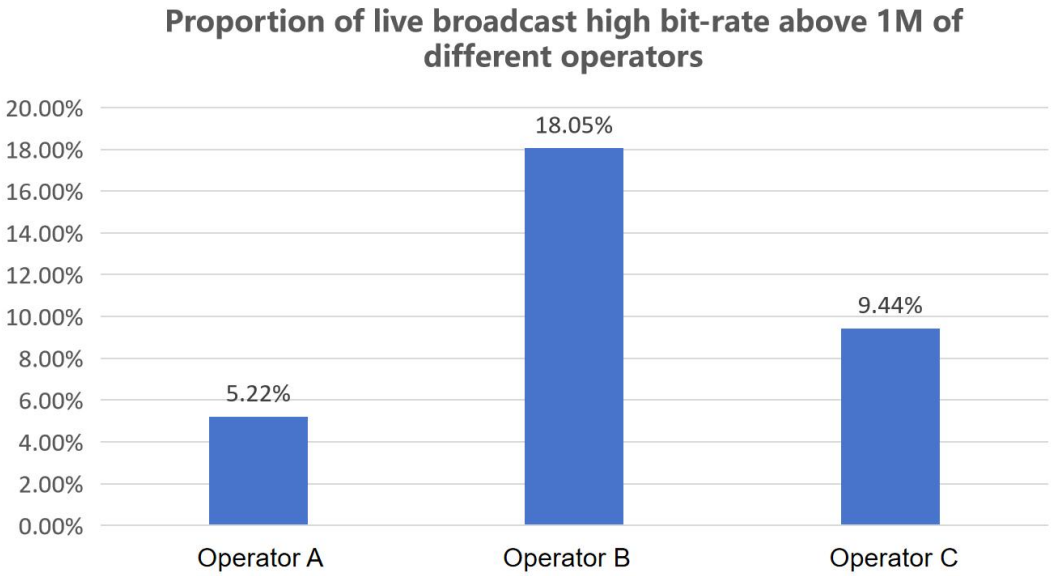


Figure 27 Proportion of High Bit Rate above 1M in 5G Networks of Different Operators

Limited by the data collection sample size of the live broadcast (viewing end) service, the experience evaluation of the live broadcast (viewing end) only includes two dimensions: the overall data description and the experience comparison between operators. The experience difference comparison between 4G and 5G networks and the experience comparison between cities are not involved.

4. Outlook

5G (the 5th generation mobile communication technology) is considered as the biggest technological breakthrough in the information field since the advent of the World Wide Web. Industry insiders believe that 5G will promote the fourth industrial revolution and change people's way of life and production. The key technologies of 5G are mainly embodied in ultra-high performance wireless transmission technology and high-density wireless network technology. As shown in the figure below, compared with 4G (the fourth generation mobile communication technology) represented by IMT Advanced, the 5G transmission rate has increased by 10 to 100 times, the peak transmission rate is 10 Gbps, the end-to-end delay can reach the order of milliseconds, the density of connected devices has increased by 10 to 100 times, and the spectral efficiency has increased by 5 to 10 times, ensuring a good user communication experience at a speed of 500 km/h. The improvement of 5G network performance not only breaks through the time and space constraints, but also realizes the interconnection between people and things, which is a leap across the times.

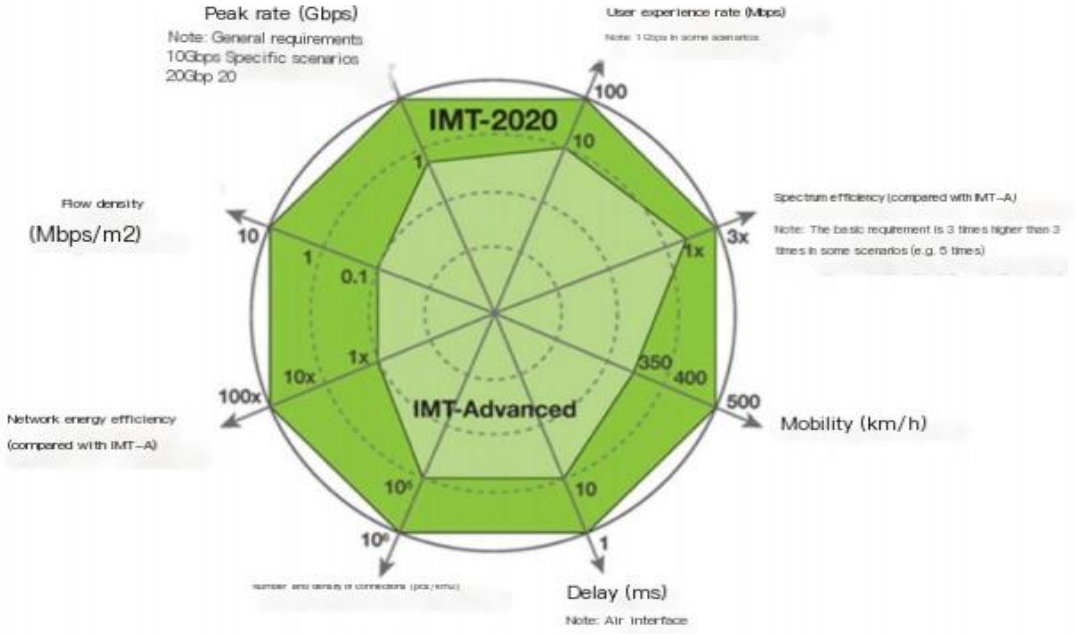


Photo source: White Paper on 5G Economic and Social Impact

Figure 28 4/5G Key Performance Comparison

4.1 Basic trend of video technology in 5G era

5G, with a peak rate of at least ten times higher than 4G, transmission delay of millisecond level and connectivity of 100 billion level, deeply integrates advanced technologies such as artificial intelligence, virtual and augmented reality, becomes a basic productivity, subverts and reconstructs all links of network transmission, content production and terminal reception, empowers video, and brings more extreme experience of information transmission and interaction. In the 5G era, video will be more real-time, high-definition, immersive and interactive.

4.1.1 Ultra HD video streaming

5G network can be targeted to solve the pain point of large data transmission of UHD video. At present, the 4G network bandwidth is limited, and 4K/8K video transmission and online playback cannot achieve the desired effect. When users watch videos, they are often affected by playback stutters, blurred screens, or automatic reduction of video resolution. The faster transmission rate and larger network capacity of 5G will completely solve this problem. Users can get a better experience when downloading and watching. Ultra HD video may become the standard configuration of 5G video. In addition to the sharp increase in video picture definition, 5G can provide a stable real-time transmission capability for UHD video, which is more flexible than traditional cable transmission, free from space constraints, and can meet the more flexible UHD video transmission needs. That is, 5G's advantages of high bandwidth and low latency will enable both the mobility and real-time playback of UHD video, and the live broadcast of UHD video will usher in a broad application prospect. On January 6, 2022, the Beijing Winter Olympics Train and

High speed Railway 5G Ultra HD Studio will be launched at Qinghe Station of Beijing Zhangjiakou High speed Railway. With 5G technology, the live broadcast of the Olympic Games will be realized anytime, anywhere and in HD.

4.1.2 Virtual reality video and interactive experience

XR (Extended Reality) is the general name of VR (Virtual Reality), AR (Augmented Reality) and MR (Hybrid Reality). It will release its full potential with the support of 5G technology. High quality XR experience is extremely demanding for network transmission capability and network delay. In the 4G environment, insufficient rendering ability, weak interactive experience, poor terminal mobility, etc., made it quiet after a short period of noise in 2015 and 2016, and failed to achieve large-scale commercial use. 5G is expected to make up for the application shortcomings of XR. First, the characteristics of high bandwidth and low latency mean that the channel transmission capacity has been greatly improved, which can improve the interaction efficiency between the virtual world and the real world, and ensure the service experience of XR. Second, edge computing and cloud computing technology supported by 5G can solve the problem that XR clients rely on high-performance computing hardware devices, ensure that people can enjoy XR content on portable and high-performance terminal devices, realize cordless and lightweight, and interact with the virtual world in a new way in life and work. The combination of 5G and augmented reality or virtual reality technology brings many new possibilities for video experience. 5G+360 ° VR real-time panoramic video can open the "God's perspective", so that each user can independently choose the viewing angle, picture and specific content, such as locking their favorite characters to watch in sports games or live events. In addition, 5G can enable holographic technology, and 5G+VR live video synchronized in different places, so that users in different places can be exposed to the same space-time environment, which can be widely used in many scenes such as remote interactive teaching, conferences, and medical treatment. From April 15, 2020, the Central Video 5G new media platform of China National Radio and Television Station, in conjunction with China Telecom, will launch the "Everest Twelve Hour" series of slow live broadcast. Relying on China Telecom's "5G+Cloud Network" technology, through 4K high-definition images and VR perspective, the 24-hour real-time landscape changes of Everest will be presented to the national audience for the first time in a 360 degree panorama. Thousands of Internet users will be able to enjoy the magnificent and steep Everest in a "immersive" way without leaving their homes. Other research shows that 5G will be conducive to the development and use of haptic feedback equipment, and will add a new perception dimension to AR/VR/MR video, further upgrading the entertainment experience. Ericsson's consumer survey report points out that in the 5G era, consumers will watch videos outside their home every week or spend three hours more on mobile devices, and one third of the time will use VR/AR glasses. And globally, 50% of consumers believe that by 2025, all people will wear AR glasses.

4.1.3 Content ecology "super video

5G, the Internet of Things, artificial intelligence and other new technologies will jointly drive an era of pan media and pan video. The advantages of 5G technology are strongly changing the transmission ecology and content pattern of video. Video may become the main social expression and occupy the main position of content transmission. The most important evolution of 5G compared with 4G is to push the world into the era of interconnection of everything. Based on 5G core technologies such as Multi User Shared Access (MUSA), low-cost and low-power 5G mass connections can be realized. The density of 5G connections reaches 1 million/km2. The high-intensity Internet of Everything enables people to use various devices as the Internet interface to connect with the virtual world, so that everything can be transmitted. It is also expected that video will go deep into vertical industries, such as smart cities, smart medicine, smart agriculture, etc., and become a new carrier for information recording, transmission and interaction, so that the "Internet of Everything" will come into reality and enrich video usage scenarios. In addition, 5G is integrated with 4K/8K, AR/VR, AI and other technologies, and new video technologies such as motion capture, real-time interaction, and virtual characters accelerate iterative penetration. The boundary between the virtual world and the real world is further blurred, which will lead to more diverse forms of video content. Experts predict that 5G will affect the development of all mobile Internet application services towards "video streaming", including virtual reality and other types of "super video", greatly enriching people's work and entertainment experience.

4.2 New features of content oriented mobile short video service experience

At present, mobile video for public customers is mainly video content (mainly vertical screen viewing) that is played on various new media platforms (such as dithering, fast hand, etc.), suitable for watching in the mobile state and short-term leisure state, and is pushed with high frequency (mainly vertical screen viewing), and the duration varies from a few seconds to a few minutes.

The process of on-demand video includes starting the terminal/player, selecting the viewing content, and performing on-demand, viewing and other operations. The factors that affect the user experience of on-demand video come from the terminal's software and hardware, transmission protocol, codec, network, platform, content source quality and other aspects; The key factors to be considered can be summarized into three aspects: audio-visual experience quality, user interaction experience quality and viewing experience quality.

The process of real-time audio and video interactive live broadcast is that the host sends the generated audio and video stream to the real-time audio and video cloud platform in real time, and the audience receives the audio and video stream of the host in real time for viewing. In real-time

interactive live broadcast of audio and video, the host and audience need frequent audio and video interaction, or the host and audience need to conduct a link game PK, which is highly interactive and requires high end-to-end delay. User roles in live broadcast are divided into anchor and audience.

The process of live broadcast on the host side is that the host side sends the generated audio and video stream in real time. The main indicators to measure the quality of the host side include audio and video quality, interaction quality, and the transmission quality of the host side.

In live broadcast, the audience receives the audio and video stream generated by the host in real time for viewing. The main indicators to measure the client experience in live broadcast include audio experience quality, video experience quality, interactive experience quality, and audience presentation experience quality.

Content is king". The experience of users watching mobile videos increasingly reflects the trend of content dominance, that is, how to improve the attractiveness of video content to users. This will involve the following aspects: First, users' evaluation of the value of video content, including practical guidance and perceptual experience; Second, users' recognition and loyalty to the anchor; Third, users' interest preferences. For these new trends of content led video experience, it is necessary to dig new acquisition parameters, such as the number of times users watch the same program source, the same theme video, the same anchor, and whether there is speed doubling operation in watching to comprehensively evaluate the user's comprehensive experience.

In addition, with the evolution of video services to new directions such as XR and Meta Universe, it is necessary to collect more key parameters in terms of device interaction, user behavior, etc., to further expand the user experience evaluation mode.

5. Appendix

5.1 Explanation of terms

The following terms and definitions are applicable to this document:

- Mobile Internet

Mobile Internet emphasizes the use of cellular mobile communication networks (such as 4/5G) to access the Internet, so it often refers to mobile terminals using mobile communication networks to access the Internet and use Internet services; Wireless Internet emphasizes wireless access, which includes various wireless access technologies (such as Wi Fi) in addition to cellular network

- Mobile video service

Service providers provide various video playback services through various mobile terminals (mobile phones, PADs), including Wi Fi and mobile network (4G/5G) access and other different forms.

- Short video

Short video refers to the video content played on various new media platforms, suitable for watching in the mobile state and short-term leisure state, and pushed frequently, ranging from a few seconds to a few minutes. The content integrates skills sharing, humor, fashion trends, social hot spots, street interviews, public education, advertising creativity, business customization and other topics. Due to its short content, it can be made into a single film or a series of columns.

- Live video

Live video can be divided into two types. One is to provide TV signal watching on the Internet, such as live broadcast of various sports games and artistic activities. This kind of live broadcast principle is to collect TV (analog) signals, convert them into digital signals, input them into computers, and upload them to websites for people to watch in real time, which is equivalent to "network TV"; The other is "webcast": set up independent signal acquisition equipment (audio+video) on the site, upload it to the server through wired or wireless network, and publish it to the website for viewing. The audience can also interact with the host in real time. Unless otherwise specified in special circumstances, this white paper refers to the latter.

- On demand load duration

When the user uses the on-demand service to watch the video content without the title advertisement, the time interval from the user triggering the play operation to the terminal playing the first screen.

- Channel switching duration

When the user uses the live TV service, the time interval between switching from the existing live TV channel to another live TV channel and triggering the switching operation through the remote control and other control devices to complete the appearance of the new live TV channel screen.

- Playback duration

When the user is watching the video, the sum of the user's viewing time/the total number of playing times shall be the statistical average value of all test samples.

- Calton duration

When the user is watching the video, the sum of the picture stuck duration/the total number of stuck times shall be the statistical average value of all test samples.

- Stuck times

In the process of watching the video, the total number of stuck times/total number of play times shall be the statistical average value of all test samples.

- Proportion of stuck duration

When the user watches the video, the total stuck duration/total playback duration shall be the statistical average of all test samples.

- Carton rate

The total number of play times/total number of play times when the user is watching the video, and the statistical average value of all test samples is used.

- No stuck rate (zero stuck rate)

The total number of play times without jamming/total number of play times shall be the statistical average value of all test samples.

- UVES score

According to the calculation method specified in the uVES standard, the video source quality (resolution, frame rate, encoding method, code rate, picture complexity), interactive experience (on-demand loading duration, channel switching duration), and broadcast experience (on-demand stuck rate, live broadcast splash rate) data obtained from the test terminal are comprehensively calculated to obtain a score; It can be used to measure the subjective experience of users watching videos.

- Average download rate

The "media download data amount/download duration" within a certain time period counted from the video player terminal is usually measured in bits per second (bit/s or bps), and the download duration does not include the spare time without actual download data.

- Quality difference rate

In a statistical period, if the user's experience quality of watching video is lower than the threshold value (v) of poor quality for more than the specified time, it is defined as a poor quality user. Poor quality rate=number of poor quality users/total number of users

- Business scope

The video services covered by this white paper currently include short video, long video and live broadcast services accessed through mobile terminals (mobile phones, pads) under the mobile cellular network.

- data sources

The data related to the development of the domestic industry used in this white paper comes from the data officially released by the Ministry of Industry and Information Technology, the data provided by members of the Video Experience Alliance, and the data reported publicly by the industry (marked where cited).

The relevant data of mobile video service used in this white paper is from the sampling statistics of the existing network provided by the video service providers in the alliance members, and the relevant data of radio and television video service/IPTV service/OTTTV service is from the acquisition data of the sampling spot test of the existing network of the alliance members.

- Sampling statistical method of user experience quality

The video service providers in the alliance members count the data of mobile terminal online services on their respective platforms for a continuous period of time, and then calculate the weighted average of the data from multiple video service providers.

- Organization and implementation of sampling point measurement

In order to verify the effectiveness of the statistical data provided by the alliance units, the member units of the alliance organization implemented the sampling spot test of user experience quality. Shanghai Jiaotong University, Dekeshi Communication (Shanghai) Co., Ltd. and Shanghai Aice Communication Technology Co., Ltd. are mainly responsible for defining the test process and providing data analysis tools. The State Key Laboratory of Network and Exchange of Beijing University of Posts and Telecommunications assists in data sorting and analysis, and has completed the collection, data analysis and validation of point test data in the past two months.

5.2 Reference

[1] The 50th Statistical Report on Internet Development in China, http://www.cnnic.net.cn/NMediaFile/2022/1020/MAIN16662586615125EJOL1VKDF.pdf

[2] The 49th Statistical Report on the Development of Internet in China,

https://www.cauc.edu.cn/jsjxy/upfiles/202203/20220318171634656.pdf

[3] Ericsson Mobile Market Report in June 2022, https://www.ericsson.com/zh-cn/reports-and-papers/mobility-report/reports/june-2022

[4] Research Report on the Development of Online Audiovisual in China in 2021, http://www.cnsa.cn/attach/0/2112271351275360.pdf

[5] QuestMobile2021 China Mobile Internet Annual Report, https://www.questmobile.com.cn/research/report-new/222

[6] Application Scenarios and Development Trends of Short Video and Long Video in 5G Era

[7] Ministry of Industry and Information Technology, Statistical Bulletin of Communication Industry in 2021, https://www.miit.gov.cn/gxsj/tjfx/txy/art/2022/art_e8b64ba8f29d4ce18a1003c4f4d88234.html

[8] The Ministry of Industry and Information Technology, the Ministry of Education, the Ministry of Culture and Tourism, the State Administration of Radio and Television, and the State General Administration of Sport. Action Plan for the Integration and Development of Virtual Reality and Industrial Applications (2022-2026), October 28, 2022, https://www.miit.gov.cn/zwgk/zcwj/wjfb/tz/art/2022/art_775aaa3f77264817a5b41421a8b2ce22.html

[9] Interpretation of the Action Plan for the Integration and Development of Virtual Reality and Industrial Applications (2022-2026) by the Electronic Information Department of the Ministry of Industry and Information Technology, https://www.miit.gov.cn/zwgk/zcjd/art/2022/art_acb9482eac994486913b592cebf71b81.html

UHD World Association
世界超高清视频产业联盟

联系我们:
UWA联盟邮箱: support@theuwa.com
UWA联盟官网: www.theuwa.com