Beyond Video Quality: Evaluation of Spatial Presence in 360-Degree Videos

ZOU Wenjie[1], GU Chengming[1], FAN Jiawei[1], HUANG Cheng[2,3], BAI Yaxian[2,3]

(1. School of Telecommunications Engineering, Xidian University, Xi'an 710071, China;
2. ZTE Corporation, Shenzhen 518057, China;
3. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China)

DOI: 10.12142/ZTECOM.202304012

https: //kns.cnki.net/kcms/detail/34.1294.TN.20231129.1000.002.html, published online November 29, 2023

Manuscript received: 2023-05-08

Abstract: With the rapid development of immersive multimedia technologies, 360-degree video services have quickly gained popularity and how to ensure sufficient spatial presence of end users when viewing 360-degree videos becomes a new challenge. In this regard, accurately acquiring users' sense of spatial presence is of fundamental importance for video service providers to improve their service quality. Unfortunately, there is no efficient evaluation model so far for measuring the sense of spatial presence for 360-degree videos. In this paper, we first design an assessment framework to clarify the influencing factors of spatial presence. Related parameters of 360-degree videos and head-mounted display devices are both considered in this framework. Well-designed subjective experiments are then conducted to investigate the impact of various influencing factors on the sense of presence. Based on the subjective ratings, we propose a spatial presence assessment model that can be easily deployed in 360-degree video applications. To the best of our knowledge, this is the first attempt in literature to establish a quantitative spatial presence assessment model by using technical parameters that are easily extracted. Experimental results demonstrate that the proposed model can reliably predict the sense of spatial presence.

Keywords: virtual reality; quality assessment; omnidirectional video; spatial presence

Citation (Format 1): ZOU W J, GU C M, FAN J W, et al. Beyond video quality: evaluation of spatial presence in 360-degree videos [J]. ZTE Communications, 2023, 21(4): 91 - 103. DOI: 10.12142/ZTECOM.202304012
Citation (Format 2): W. J. Zou, C. M. Gu, J. W. Fan, et al., "Beyond video quality: evaluation of spatial presence in 360-degree videos," *ZTE Communications*, vol. 21, no. 4, pp. 91 - 103, Dec. 2023. doi: 10.12142/ZTECOM.202304012.

1 Introduction

In the past decade, multimedia streaming services have had an explosive growth[1]. Among a variety of multimedia types, 360-degree videos become the major type of virtual reality (VR) content in the current stage. Major video-sharing websites such as YouTube and Facebook have already started to offer 360-degree video-on-demand and live 360-degree video streaming services.

In contrast to traditional 2D videos, 360-degree videos can provide full 360-degree scenes to end users, using the Head-Mounted Display (HMD) as a display device. With a higher degree of freedom (DoF) and wider field of view (FOV) during the viewing process, end users are provided with a stronger sense of immersion and a feeling of being in a perceptible virtual scene around the users. Different from the experience of traditional 2D videos[2 - 3], this type of feeling is usually termed as presence[4 - 7]. According to the classification of presence in Refs. [8] and [9], presence covers a broad range of aspects including spatial presence, social presence, self-presence[10], engagement, realism, and cultural presence. In the field of 360-degree video processing, researchers are more interested in spatial presence, which describes the feeling, sense, or state of "being there" in a mediated environment[4]. This feeling occurs when part or all of a person's perception fails to accurately acknowledge the role of technology that makes it appear that she/he is in a physical location and environment different from her/his actual location and environment in the physical world[11].

Over the last twenty years, a variety of work has been conducted to investigate the users' sense of presence in VR environments, especially for scenes rendered by computers[12 - 13]. These studies mainly focused on measuring specific influencing factors of the sense of presence and revealing the qualitative relationship between presence and specific human perceptual aspects in a generalized VR environment. Directly quantifying the sense of presence is, however, outside the scope of

This work is supported in part by ZTE Industry-University-Institute Cooperation Funds.

these studies. On the other hand, some researchers managed to evaluate the sense of presence using physiological signals[14 - 17]. However, this type of method requires professional equipment and the reliability of experimental results strongly relies on the accuracy of the devices.

To the best of our knowledge, most human perception research carried out for 360-degree videos only focused on the perceptual video quality instead of the spatial presence. Recently, we conducted a subjective evaluation experiment on the spatial presence of end users when watching 360-degree videos displayed on VR devices[18]. We aimed to quantitatively investigate the relationship between various impact factors and the spatial presence.

In this paper, based on the research outcomes of Ref. [18], the characteristics of the display device of 360-degree videos are considered. We propose a framework in hierarchical structure to clarify the influencing factors of the spatial presence, where both the features of 360-degree video and HMD are considered. A series of rigorous subjective experiments are designed to reveal the relationship between various influencing factors and the spatial presence. Furthermore, a quantitative evaluation model of spatial presence is built in this work. Contributions of this paper can be concluded as follows:

1) We propose the first framework to identify the components of spatial presence. This framework provides valuable input for establishing models of assessing the spatial presence of VR services.

2) We reveal the relationship between spatial presence and various related impact factors based on subjective ratings, which can be used as recommendations for further improving the quality of 360-degree video services.

3) We propose the first quantitative model to measure the spatial presence when watching 360-degree videos on the HMD. The parameters employed in the proposed model can be easily extracted, hence the model would be conveniently deployed on the network or client to assess the user's presence.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 illustrates the assessment framework and the subjective experiments. Section 4 introduces the proposed model in detail. In Section 5, the performance of the proposed model is evaluated. Conclusions are drawn in Section 6.

2 Related Work

Over the last thirty years, researchers have explained and defined the concept of presence in several different ways. For instance, LOMBARD et. al.[8] defined it as the experience of being engaged by the representations of a virtual world in 2002. Very recently, presence was defined as the feeling of being in a perceptible external world around the self[4 - 7]. The evolution of understanding and definition of the presence was summarized in Refs. [7] and [9]. As the above research is more related to psychoanalysis, straightforward solutions to the measurement of presence were outside the scope of these studies. How to measure the presence in practice is still unknown.

To acquire the subjective sense of presence, some researchers resorted to the design of subjective response questionnaires[19 - 24]. More specifically, authors in Ref. [19] designed a questionnaire, called the immersive tendencies questionnaire (ITQ), to investigate the relationship between users' sense of presence and some handcrafted influential aspects in virtual environments. Authors in Ref. [22] designed a spatial presence questionnaire, named MEC Spatial Presence Questionnaire (MSC-SPQ), to investigate the influence of possible actions, self-location, and attention allocation on users' sense of spatial presence. However, these studies only focused on revealing the qualitative relationship between specific human perceptual aspects and presence in the generalized VR environment. On the other hand, some researchers tried to evaluate the presence using physiological signals[14 - 17]. This type of measurement requires the deployment of professional equipment which is impractical for real-world applications. Therefore, designing accurate and implementation-friendly experimental methods to measure presence is of fundamental importance.

As for the human perception research specifically carried out for 360-degree videos, to our best knowledge, most studies only focused on evaluating the quality of experience aspects[25 - 33] instead of assessing the sense of presence. For instance, authors in Ref. [25] investigated how to assess the video quality of 360-degree videos corresponding to different projection approaches. A quality metric, called spherical peak signal to noise ratio (S-PSNR) was proposed to summarize the average quality over all possible viewports as the video quality. In Ref. [26], authors proposed an objective video quality assessment method using a weighted PSNR and special zero area distortion projection method for 360-degree videos. In Ref. [30], authors measured viewport PSNR values over time to assess the objective video quality of 360-degree video streaming. Recently, authors in Ref. [33] introduced visual attention in assessing the objective quality of 360-degree videos with the assumption that not all of the 360-degree scene is actually watched by users. However, as discussed above, the spatial presence of end users was not fully considered in existing research. Our recent work[18] conducted a preliminary experiment for assessing the spatial presence of end users when viewing 360-degree videos displayed on VR devices. However, modeling the spatial presence of end users is not covered. How to quantitatively evaluate users' sense of spatial presence when viewing 360-degree videos remains an open issue.

3 Subjective Evaluation Framework and Subjective Experiments

In this section, a hierarchical framework with five perception modules is first proposed to assess spatial presence. Based on this framework, five subjective experiments were designed and conducted according to each module in the framework. Results of subjective experiments are used to investi-

gate each type of human perception and facilitate the establishment of the assessment model.

3.1 Proposed Assessment Framework

As shown in Fig. 1, the proposed framework consists of three layers, namely the factor layer, the perception layer, and the presence layer from left to right. The factor layer includes several sensory cues of relevant parameters such as video, audio, VR device, and latency. These parameters can be conveniently extracted from the current VR systems. In the perception layer, users' perception is characterized into multiple dimensions including visual[34], auditory[34], and interactive perception[21 – 22, 35]. Detailed definitions of the components of perception and presence layers are discussed as follows.

1) Perceptual video quality

Perceptual video quality refers to the overall perceived quality of videos displayed on the HMD. In our previous work, three technological parameters of the video, i.e., video bitrate, video resolution and video frame rate, are extracted to assess the video quality. Two parameters corresponding to the HMD (screen resolution and refresh rate) are added in the assessment of perceptual video quality.

2) Perceptual audio quality

Perceptual audio quality refers to the overall perceived quality of audios offered by the VR system. The audio bitrate and audio sampling rate are extracted to assess the perceptual audio quality.

3) Visual realism

Visual realism (VRE) refers to how close the system's visual output is to real-world visual stimuli. This perception not only depends on the video quality, but also depends on how wide the FOV provided by HMD is and whether a stereoscopic vision is offered. These two additional factors have been verified to be important for the overall capability of an immersive system[36].

4) Acoustic realism

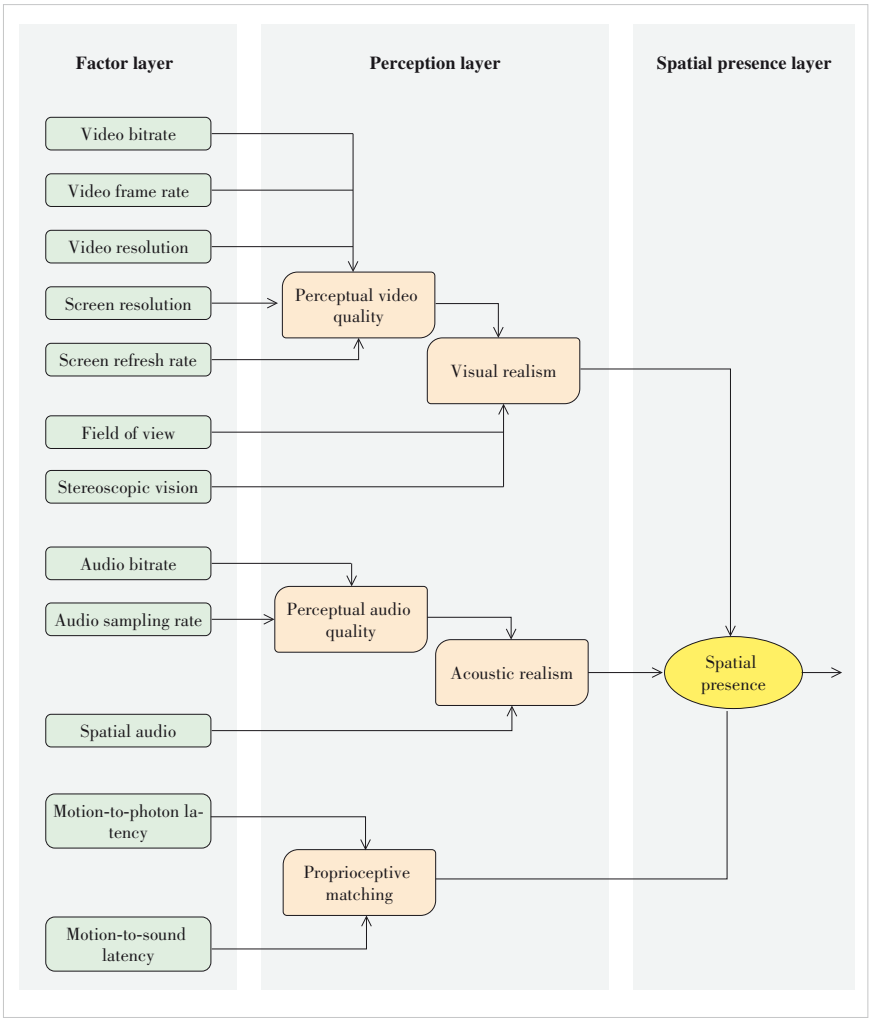
Acoustic realism (ARE) represents how close the system's aural output is to real-world aural stimuli. Perceptual audio quality is the basic experience of the audio. Moreover, spatial audio provides the capability to track sound directions and update the head movement in real time. Hence, the spatial audio and perceptual audio quality are combined to assess the overall acoustic realism.

5) Proprioceptive matching

Proprioceptive matching refers to the matching degree between the head movement and the picture/sound refresh of the HMD. As for a VR system, the tracking level is much more important in regard to the spatial presence formation[36]. Similarly, the mismatch can also occur in the spatial audio. These two mismatches, called motion-to-photon (MTP) latency and audio latency (AL)[37], are utilized to assess the capability of proprioceptive matching.

6) Spatial presence

Spatial presence refers to a user's subjective psychological response to a VR system[35]. It is correlated with VRE, ARE, and proprioceptive matching, which represents the main aspects of the experience provided by 360-degree video services.



▲Figure 1. Proposed assessment framework for assessing spatial presence

3.2 Subjective Experiments for Obtaining Spatial Presence

To explore the spatial presence, six subjective quality scoring experiments were conducted, corresponding to the five perception modules in the perception layer and one towards the spatial presence.

3.2.1 Overview of Experimental Design

A total number of 30 non-expert subjects participated in this experiment, including

16 males and 14 females aged between 22 and 33 years. All of them have normal or corrected-to-normal sight. The experiments were conducted in the test environment following ITU-T P.913[38]. A flagship HMD, i.e., HTC VIVE Pro, was employed as the display device, which has a screen with an original resolution of 2 880×1 600 pixels, a refresh rate of 90 Hz, and a horizontal FOV of 110 degrees. Moreover, a 360-degree video player with the Equirectangular projection was developed to display the videos on the HMD. The display FOV, length of the MTP latency, and audio latency can be set as desired. Our study adopted a single-stimuli scoring strategy[38].

3.2.2 Experiment 1: Obtaining Perceptual Video Quality

In this experiment, ten YUV420 original videos were employed to form a video database, including four 360-degree videos (i.e., denoted as O1 to O4) proposed by Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG[39 - 40] and six 2D videos (i.e., denoted as V1 to V6) provided by the Ultra Video Group[41], as shown in Fig. 2. The 360-degree videos have a spatial resolution of 3 840×1 920 pixels, a framerate of 30 fps and a length of 10 s. The 2D videos have a spatial resolution of 3 840×2 160 pixels, a framerate of 120 fps and a length of 5 s. The experiment was divided into 2 sub-experiments, which were designed to investigate the impact of bitrate and frame rate on the perceptual video quality. Details of the experiment settings are introduced as follows:

1) Investigating the impact of video bitrate

Four 360-degree videos, i.e., O1 to O4, were utilized to investigate the relationship between the video bitrate and the perceptual video quality. The bits per pixel (BPP) were employed to unify the coding bitrates under different resolutions. It can be calculated by

$$\mathrm{BPP} = \frac{Br}{R_H \times R_V \times f}, \tag{1}$$

where Br and f are the video bitrate and frame rate, respectively. R_H and R_V are the horizontal and vertical source resolutions. The original 360-degree videos were down-sampled and encoded using an x265 encoder according to the settings listed in Table 1.

During the experiment, video sequences were displayed in random orders using the HMD. Subjects can change their viewport by rotating their head. There was a 10-second interval between each two video sequences. Subjects could rate the perceptual video quality using the Absolute Category Rating (ACR) 5-point scale (corresponding to the perceived quality of "excellent," "good," "fair," "poor," and "bad" from 5 to 1 point) during the 10-second interval. Before the formal test, the subjects were asked to rate a few example videos to get familiar with the scoring scale and the scoring tool.

2) Investigating the impact of frame rate

To the best of our knowledge, there is no 360-degree video database containing videos with a frame rate higher than 60 fps. As the screen refresh rate of the current HMDs can reach 90 Hz, we have to use six 2D videos with a high frame rate, i.e., V1 to V6, to study the impact of frame rate. Each original video was repeated twice to generate a video of 10 s. Then, they were down-sampled to 60 fps, 30 fps, and 15 fps. These videos (including the original 120 fps) were further spatially down-sampled to 960 × 540. Videos generated from V1 to V4 were encoded with a fixed quantization parameter (QP), i.e., 22, us-

▼Table 1. Experimental setup

BPP	Bitrate/(Mbit/s)			BPP	Bitrate/(Mbit/s)
	720P (1 280×640)	1080P (1 920×960)	2K (2 160×1 080)		4K (3 840×1 920)
0.016	0.39	0.89	1.12	0.011	2.50
0.032	0.79	1.77	2.24	0.024	5.20
0.056	1.38	3.10	3.92	0.056	12.39
0.08	1.97	4.42	5.60	0.08	17.70
0.16	3.93	8.85	11.20	0.16	35.39
0.20	4.92	11.06	14.00	0.20	44.24

BPP: bits per pixel

▲Figure 2. Content of test sequences: (a) Basketball, (b) Harbor, (c) KiteFlite, (d) Gaslamp, (e) Beauty, (f) Bosphorus, (g) Honeybee, (h) Jockey, (i) ReadySetGo, and (j) YachRide

ing the x.265 encoder to generate high-quality videos. To investigate whether the QP can influence the impact of framerates on the perceptual video quality, V5 and V6 were encoded with four different QPs, i.e., 22, 32, 36, and 39, to generate four quality levels. During the experiment, video sequences were displayed in their resolution in random orders. It is noted that videos with 120 fps were displayed at 90 fps on the HMD since the refresh rate of the HMD is only 90 Hz.

3.2.3 Experiment 2: Obtaining Visual Realism

Three high-quality stereoscopic videos (3 840×3 840 resolution) were downloaded. Note that the audio tracks were not used in this experiment. These videos last for 20 s and have a frame rate of 30 fps. The projection mode is equirectangular. They were firstly separated into two monoscopic videos, namely the left and right videos, separately. To investigate the impact of stereoscopic vision, the left videos and stereoscopic videos were utilized as the test materials that were further encoded into three quality levels: 1 Mbit/s, 5 Mbit/s and 14 Mbit/s for monoscopic videos and 2 Mbit/s, 8 Mbit/s and 18 Mbit/s for stereoscopic videos. The FOV was set to be 60 degrees, 90 degrees and 110 degrees, respectively. The ACR 5-point scale was also used in this experiment to record the evaluation scores for the perceptual video quality and visual realism. To obtain visual realism, the subjects were asked a question: "To what extent are your visual experiences in the virtual environment consistent with that in the real world?".

3.2.4 Experiment 3: Obtaining Perceptual Audio Quality

The audio tracks from the perceptual evaluation of audio quality (PEAQ) conformance test listed in ITU-R BS.1387[42] were employed as the reference. More specifically, six samples, four music pieces and two speeches, were used, as summarized in Table 2. The sampling frequency of all audio files is 48 kHz. Stereo (two-channel) audio files were used for the test. They were encoded using the Advanced Audio Codec (AAC) encoder with a bit rate of 8 kbit/s, 16 kbit/s, 32 kbit/s, 64 kbit/s, 128 kbit/s, 256 kbit/s, and 320 kbit/s, respectively and a sampling rate of 48 kHz. The generated audio sequences were displayed to subjects on a high-fidelity headphone in a random order. After each display, the subjects were asked to rate the quality levels of audio files in ACR 5-point scales.

3.2.5 Experiment 4: Obtaining Acoustic Realism

The left videos in Experiment 2 encoded with 14 Mbit/s and corresponding audio files were used in this experiment to investigate the influence of the audio quality and spatial audio on acoustic realism. The audio component of these videos was in eight channels with each representing the sound from one direction. The original audio files were encoded using the AAC codec with a bit rate of 128 kbit/s and a sampling rate of 44.1 kHz. The sound from front-left and front-right was firstly mixed into the stereo audio. Then, the stereo audio files and original spatial audio files were encoded with 16 kbit/s, 32 kbit/s, 64 kbit/s, and 128 kbit/s to generate four quality levels. After the display of each audiovisual sequence, two questions were asked: "How do you rate the quality of the audio you just heard?" and "To what extent are your acoustic experiences in the virtual environment consistent with that in the real world?". Then, the subjects used the ACR 5-point scale to score the audio quality and acoustic realism of the test sequences separately.

3.2.6 Experiment 5: Obtaining Proprioceptive Matching

In this experiment, the influence of the MTP latency and AL on proprioceptive matching was investigated. First, three left vision videos in Experiment 2 with "excellent" video quality were displayed with seven lengths of MTP latency, i.e., 0 ms, 20 ms, 60 ms, 100 ms, 200 ms, 300 ms, and 500 ms, in a random order. Their audio files (high quality, 128 kbit/s) were displayed with no audio latency. Then, these videos were displayed with no MTP latency while the corresponding spatial audio files (high quality, 128 kbit/s) were displayed with eight different lengths of audio latency, i.e., 0 ms, 20 ms, 60 ms, 150 ms, 300 ms, 500 ms, 1 000 ms, and 2 000 ms, respectively. The subjects were asked to score the degree of proprioceptive matching for the test sequences with the ACR 5-point scale.

3.2.7 Experiment 6: Obtaining Spatial Presence

As listed in Table 3, the original stereoscopic videos (i.e., denoted as S1 to S3) and corresponding stereo audio files in Experiment 2 were first encoded and displayed on the HMD with no MTP latency and AL. Then, the original audiovisual files were encoded with high quality and displayed with six MTP latencies, i.e., 0 ms, 20 ms, 80 ms, 150 ms, 300 ms, and

▼Table 2. Experimental setup

File Name	Signal Type	File Name	Signal Type
FCODSB1.WAV	music	LCODPIP.WAV	music
GCODCLA.WAV	music	NCODSFE.WAV	speech
LCODHRP.WAV	music	KREFSME.WAV	speech

▼Table 3. Experimental setup

No.	Video / (Mbit/s)	Audio / (kbit/s)	No.	Video / (Mbit/s)	Audio / (kbit/s)
S1	2	16	S2	8	16
S1	8	32	S2	18	32
S1	18	64	S2	2	64
S1	18	16	S3	2	32
S1	2	64	S3	8	64
S1	4	128	S3	18	128
S2	2	16	S3	8	16
S2	8	64	S3	18	32
S2	18	128	S3	2	128

500 ms, respectively. We adopted the 5-point spatial presence scale proposed in Ref. [43] where a point from 5 to 1 indicates the degree of being there from "very strong" to "not at all". The question designed in the experiment was "To what extent did you feel like you were really inside the virtual environment?".

After the subjective tests, the reliability of the subjective results in each experiment was checked using the Pearson Linear Correlation Coefficient (PLCC) adopted by ITU-T Recommendation P.913[38]. According to the suggested threshold of 0.75[38], only the results from two subjects were discarded.

4 Spatial Presence Assessment Model

In the previous section, we construct several test scenarios under different impact factor settings and launched subjective experiments to obtain users' rating scores. These scores are the ground truth of spatial presence under different impact factor settings. In this section, the characteristic of users' perception in each module is analyzed based on the preliminary observation of the experiment results. The weight of each impact factor is determined using the linear regression method.

4.1 Perceptual Video Quality Assessment Module

As studied in Refs. [44] and [45], the impact of frame rate and quantization is separable. We follow this conclusion and hypothesize that the perceptual video quality can be predicted as follows:

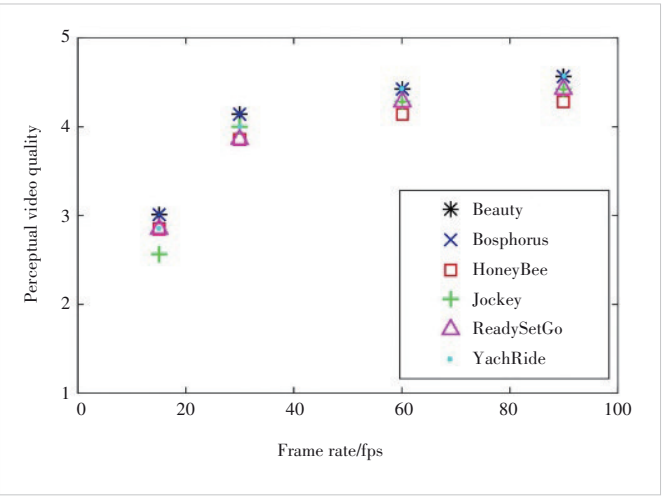
$$\mathrm{PVQ}(\mathrm{BPP}, f) = \mathrm{SQF}(\mathrm{BPP}) \cdot \mathrm{TCF}(f), \tag{2}$$

where f represents the frame rate and BPP is the bits per pixel. SQF and TCF are the spatial quality factor and temporal correction factor, respectively. The first term SQF (BPP) measures the quality of encoded frames without considering the impact of frame rate. The second term models how the Mean Opinion Score (MOS) varies with the change of frame rate.

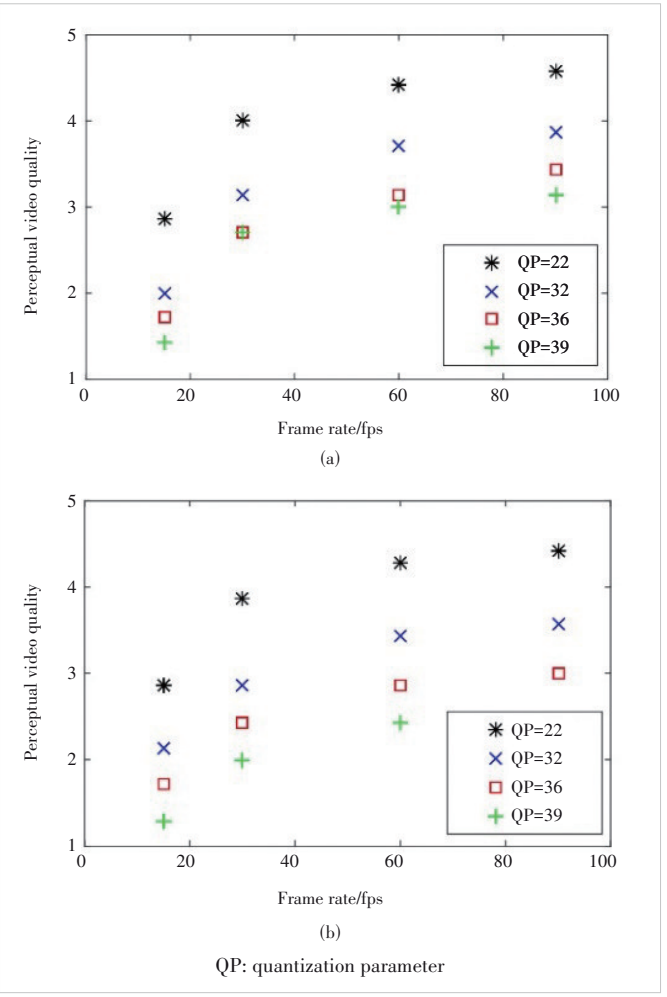
4.1.1 Temporal Correction Factor

Fig. 3 shows the relationship between the frame rate and the perceptual video quality. We can see that the perceptual video quality increases along with the rise of frame rate. Fig. 4 presents the experimental results of the two videos encoded with four different QPs. It can be found that no matter what the QP level is, MOS reduces consistently as the frame rate decreases. In order to examine whether the decreasing trend of MOS against the frame rate is independent of the QP, the MOS scores were normalized and shown in Fig. 5, where the normalized MOS (NMOS) is the ratio of the MOS with the MOS at 30 fps. More specifically, the NMOS is calculated as

$$\mathrm{NMOS}(\mathrm{QP}, f) = \frac{\mathrm{MOS}(\mathrm{QP}, f)}{\mathrm{MOS}(\mathrm{QP}, 30)}. \tag{3}$$

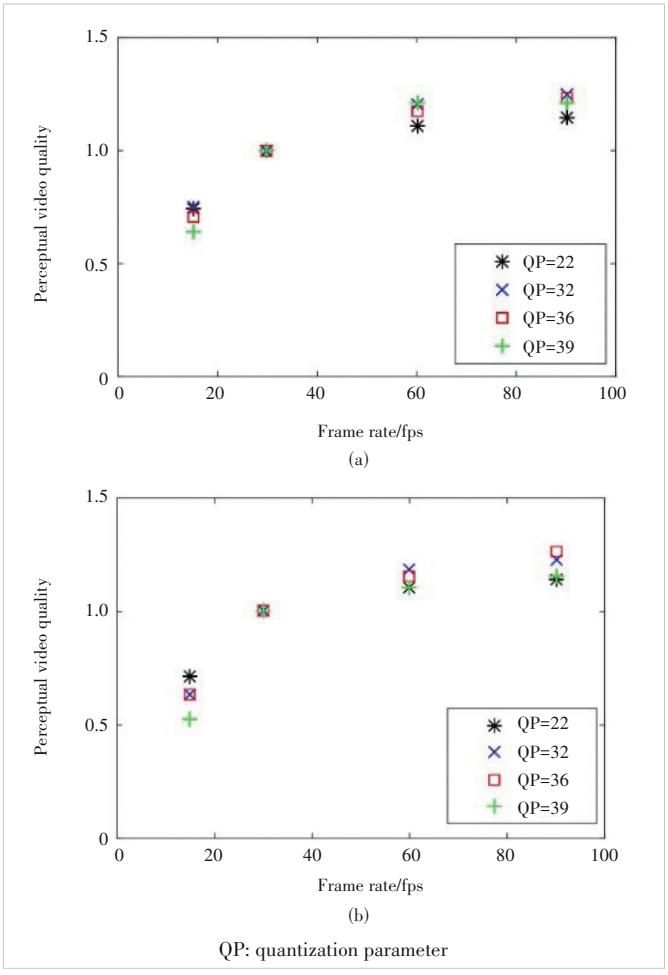


▲ Figure 3. Relationship between the frame rate and perceptual video quality



▲ Figure 4. Experimental results of (a) ReadySetGo and (b) YachRide encoded with four different QPs

As can be seen in Fig. 5, these NMOS scores corresponding to different QPs almost overlap with each other, indicating



▲ **Figure 5. Relationship between the frame rate and normalized Mean Opinion Score (NMOS): (a) ReadySetGo and (b) YachRide**

that the decrease of MOS with the frame rate is independent of the QP. This observation follows the conclusions drawn in Refs. [44] and [45] and confirms our hypothesis. The trend in Fig. 5 can be fitted using the function as

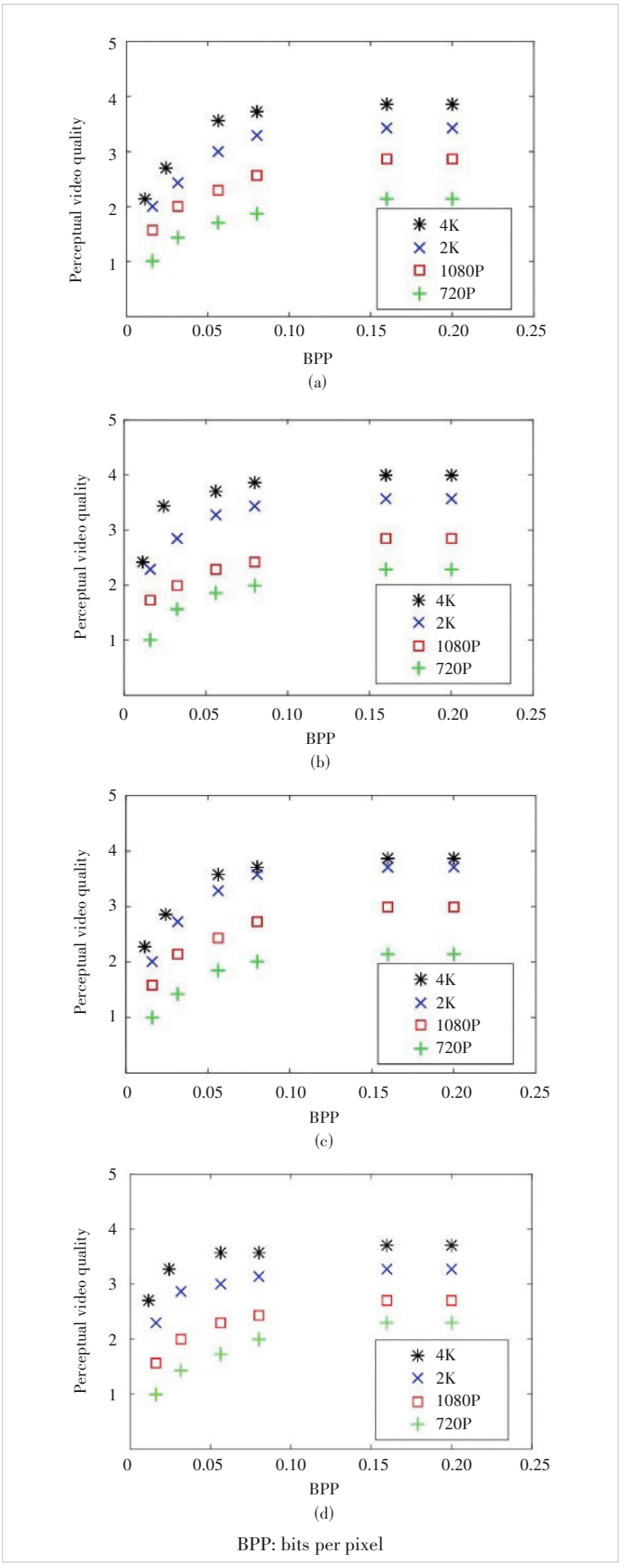
$$\mathrm{TCF}(Fr) = v_1 \cdot \exp(v_2 \cdot f) + v_3, \tag{4}$$

where v_1, v_2, and v_3 are −1.672, −0.095 31 and 1.112, respectively, which were obtained by regression.

4.1.2 Spatial Quality Factor

In this subsection, we investigate and modeled the spatial quality, which is mainly influenced by the bitrate, video resolution, and screen resolution. Fig. 6 shows the relationship between the BPP and the perceptual video quality of four 360-degree videos. It can be seen that the perceptual video quality increases with the rise of BPP. However, as for videos at different resolutions, the increasing trends of the perceptual video quality are different. This trend can be represented as

$$\mathrm{SQF}(\mathrm{BPP}) = v_4 \cdot \ln(v_5 \cdot \mathrm{BPP} \cdot 1\,000 + 1), \tag{5}$$



▲ **Figure 6. Relationship between BPP and perceptual video quality: (a) Basketball, (b) Harbor, (c) KiteFlite, and (d) Gaslamp**

where v_4 and v_5 are the model coefficients that can be obtained by regression. The values of v_4 and v_5 are listed in Table 4. It can be seen that the values of v_4 are very close to each other while that of v_5 are quite distinct for different video resolutions. Hence, the average value of v_4 is used as a fixed coefficient. The value of v_5 is then regressed again.

To reflect the impact of video resolution and screen resolution on perceived video quality, we employ the integrated assessment parameter that we proposed in the previous work[46], i.e., the number of effective video pixels per degree (ED-PPD) displayed on the screen of HMD. The effective pixels do not include the pixels interpolated by the up-sampling process. This parameter is calculated as

$$\mathrm{ED-PPD}=\begin{cases}\dfrac{R_H}{360}, & R_H \le R_{SH}\cdot\dfrac{360}{\mathrm{FOV}}\\ \dfrac{R_{SH}}{\mathrm{FOV}}, & R_H > R_{SH}\cdot\dfrac{360}{\mathrm{FOV}}\end{cases}, \tag{6}$$

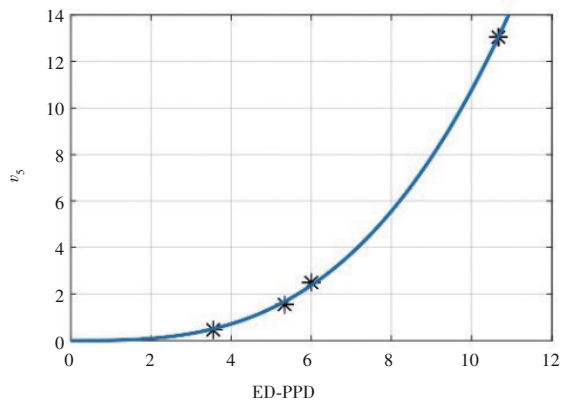
where R_H and R_{SH} are the horizontal resolution of 360-degree video and screen, respectively. When the horizontal pixels of the video displayed on the screen are more than the horizontal pixels on the screen, the ED-PPD will be saturated.

Fig. 7 shows the relationship between the ED-PPD and v_5. It can be seen that the values of v_5 and ED-PPD are in accordance with the power function relationship, which can be expressed as

$$v_5 = v_6 \cdot \mathrm{ED-PPD}^{v_7}, \tag{7}$$

▼Table 4. Values of v_4 and v_5

Video Resolution	v_4	v_5
720P	0.497 4	0.525 7
1080P	0.529 2	1.369 0
2K	0.537 1	4.584 0
4K	0.497 4	16.580 0



ED-PPD: effective video pixels per degree

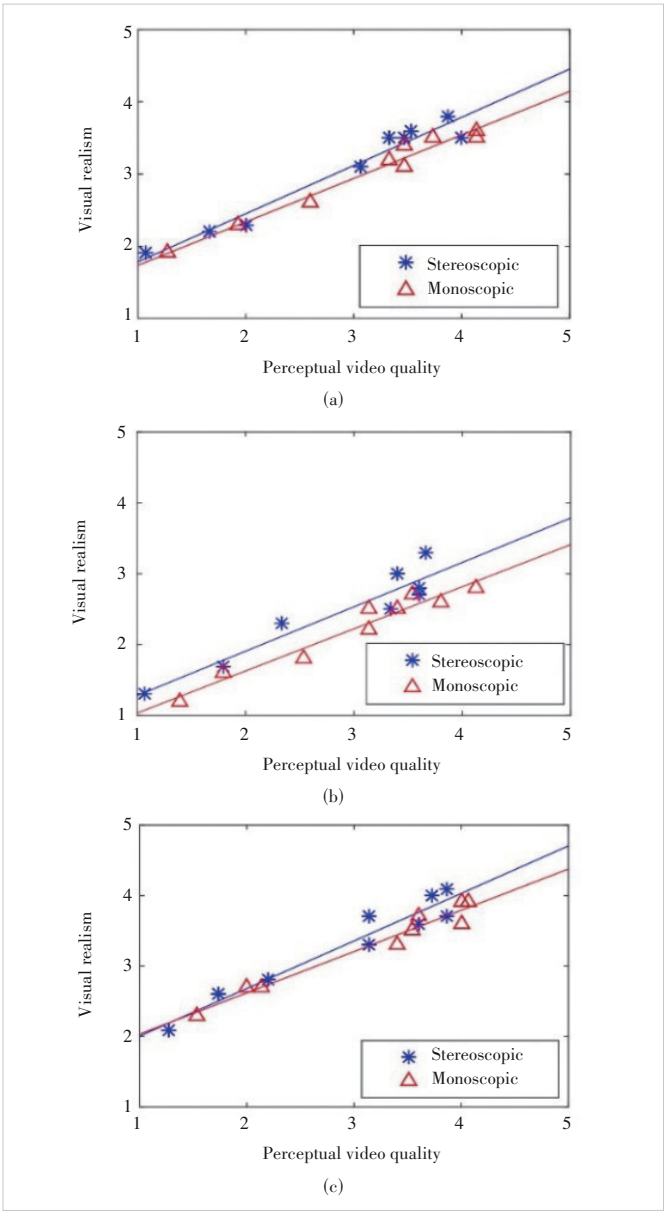
▲Figure 7. Relationship between ED-PPD and v_5

where v_6 and v_7 are equal to 0.011 7 and 2.962, respectively.

By substituting Eqs. (4), (5) and (7) into Eq. (2), the perceptual video quality of 360-degree videos can be modeled.

4.2 Visual Realism Assessment

According to the results of Experiment 2, Fig. 8 shows the relationship between perceptual video quality and visual realism. It can be seen that there is a strong correlation between the perceptual video quality and visual realism. For the influence of FOV, it can be observed that a higher FOV leads to a higher visual realism. The Kruskal-Wallis H test showed that there is a significant effect of FOV on visual realism, with p = 0.001 for monoscopic videos and p = 0.039 for stereoscopic



▲Figure 8. Relationship between the perceptual video quality and visual realism: (a) 60 FOV (field of view), (b) 90 FOV, and (c) 110 FOV

videos. A one-way analysis of variance (ANOVA) test indicates that there is no significant effect of the type of vision on visual realism. Based on the results above, the video quality and FOV appear to have a more significant impact on visual realism than the type of vision. Thus, the relationship of perceptual video quality, FOV, and visual realism can be calculated by

$$\mathrm{VRE}(\mathrm{PVQ},\mathrm{FOV}) = \max\left(\min\left(v_8\mathrm{PVQ} + v_9\mathrm{FoV} + v_{10}, 5\right), 1\right), \tag{8}$$

where v_8, v_9 and v_{10} are equal to 0.595, 0.02 and −0.735, respectively.

4.3 Perceptual Audio Quality and Acoustic Realism Assessment

We first model the perceptual audio quality using the experimental results of Experiment 3. Fig. 9 shows the logarithmic relationship between the audio bitrate and the perceptual audio quality. This relationship can be represented as

$$\mathrm{PAQ}(\mathrm{ABr}) = 1 + v_{11} - \frac{v_{11}}{1 + \left(\frac{\mathrm{ABr}}{v_{12}}\right)^{v_{13}}}, \tag{9}$$

where v_{11}, v_{12} and v_{13} are equal to 4.103, 42.36 and 1.251, respectively.

As for AR, Fig. 10 shows the relationship between the perceptual audio quality and acoustic realism. It can be found that there is a significant linear relationship between the audio quality and acoustic realism for stereo audio (R^2 = 0.881, F = 213.251, and p = 0.000 < 0.05) and for spatial audio (R^2 = 0.955, F = 73.791, and p = 0.000 < 0.05). The relationship in Fig.10 can be expressed as

$$\mathrm{AR}(\mathrm{PAQ}) = v_{14}\mathrm{PAQ} + v_{15}, \tag{10}$$

where v_{14} and v_{15} are equal to 0.733 and 0.634 for the stereo audio, and equal to 0.682 and 1.167 for the spatial audio.

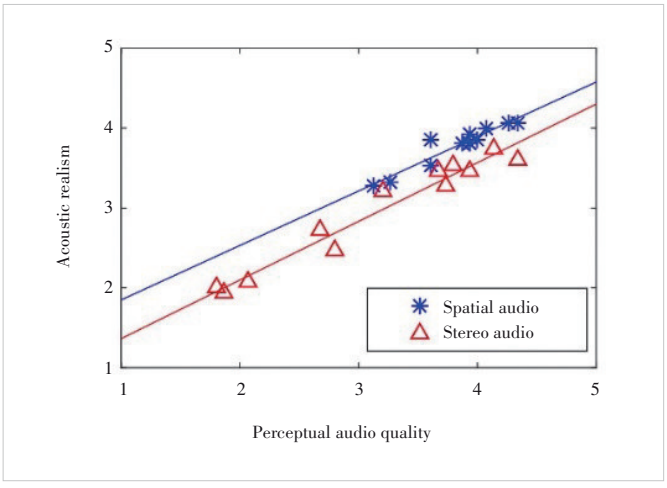
4.4 Proprioceptive Matching Assessment

Fig. 11 shows the relationship between the two types of delay and proprioceptive matching. It can be seen that the proprioceptive matching decreases with the increase of both the MTP latency and AL. Here, the degradations of proprioceptive matching caused by the MTP latency and AL are calculated by

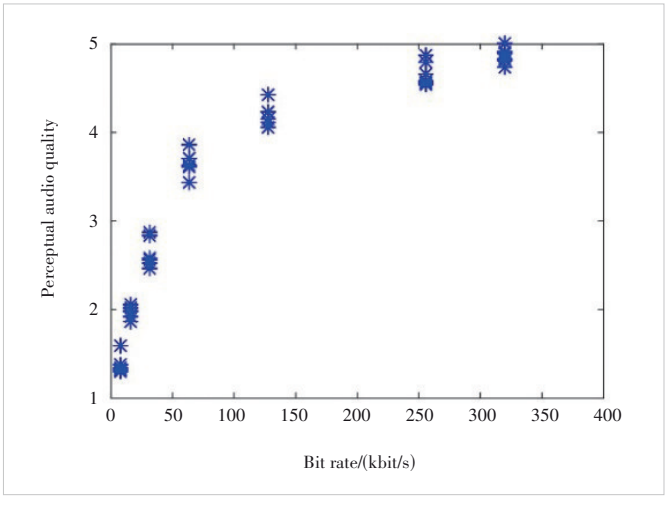
$$\mathrm{DMOS}(\mathrm{MTP}) = 5 - \mathrm{MOS}(\mathrm{MTP}), \tag{11}$$

$$\mathrm{DMOS}(\mathrm{AL}) = 5 - \mathrm{MOS}(\mathrm{AL}). \tag{12}$$

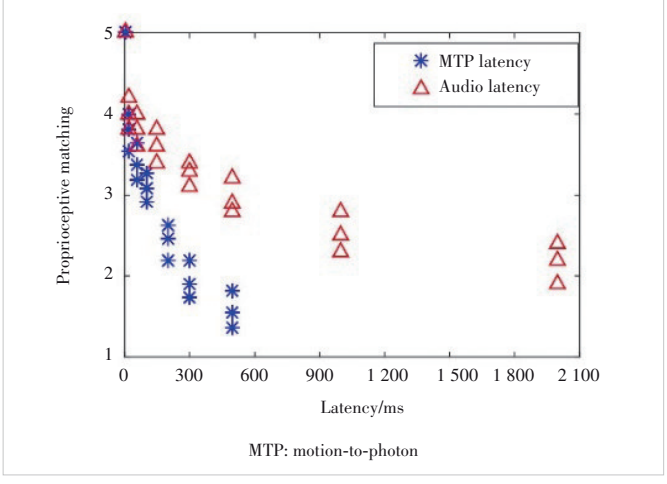
Fig. 12 shows the relationship between the two types of delay and the degradation of proprioceptive matching. This rela-



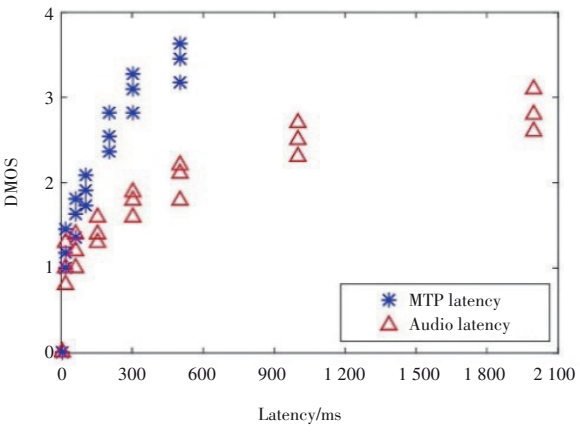
▲ **Figure 10. Relationship between the perceptual audio quality and acoustic realism rated on Head-Mounted Display (HMD)**



▲ **Figure 9. Relationships between the audio bit rate and perceptual audio quality**



▲ **Figure 11. Relationship between the two types of delay and the proprioceptive matching**



DMOS: difference mean opinion score MTP: motion-to-photon

▲Figure 12. Relationship between the two types of latency and the degradation of proprioceptive matching

tionship can be represented by

$$\mathrm{DMOS(MTP)} = \max\left(\min\left(\ln\left(v_{16}\mathrm{MTP} + 1\right), 4\right), 0\right), \quad (13)$$

$$\mathrm{DMOS(AL)} = \max\left(\min\left(v_{17}\ln\left(v_{18}\mathrm{AL} + 1\right), 4\right), 0\right), \quad (14)$$

where v_{16}, v_{17} and v_{18} are equal to 0.065 46, 0.428 9 and 0.275 4, respectively. We modeled the proprioceptive matching as

$$\mathrm{PM(MTP,AL)} = \max\left(\min\left(5 - \mathrm{DMOS(MTP)} - \mathrm{DMOS(AL)}, 5\right), 1\right). \quad (15)$$

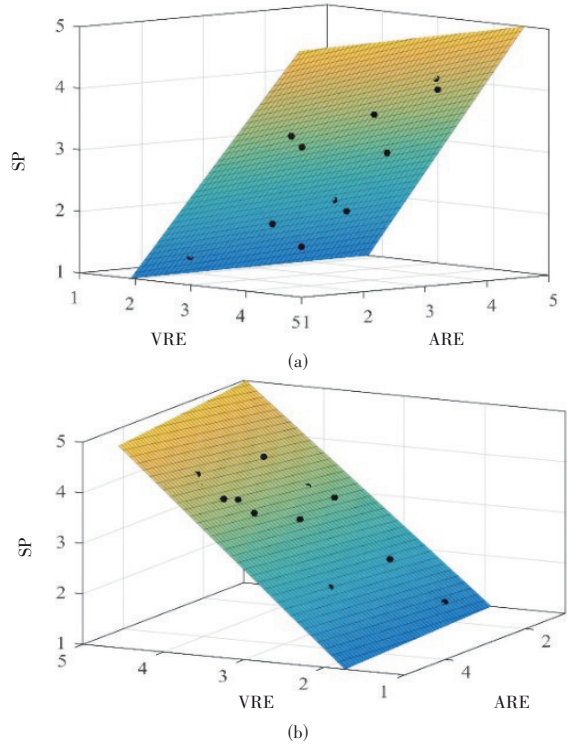
4.5 Spatial Presence Assessment

First, the relationship between the visual/acoustic realism and the spatial presence is modeled. As shown in Fig. 13, the spatial presence increases with the rise of VRE and ARE. This phenomenon confirms the conclusion drawn in our previous work[18]. The relationship shown in Fig. 13 can be calculated as

$$\mathrm{SPAV(VRE, ARE)} = \min\left(\max\left(v_{19}\mathrm{VRE} + v_{20}\mathrm{ARE} + v_{21}\mathrm{VRE}\times\mathrm{ARE} + v_{22}, 1\right), 5\right), \quad (16)$$

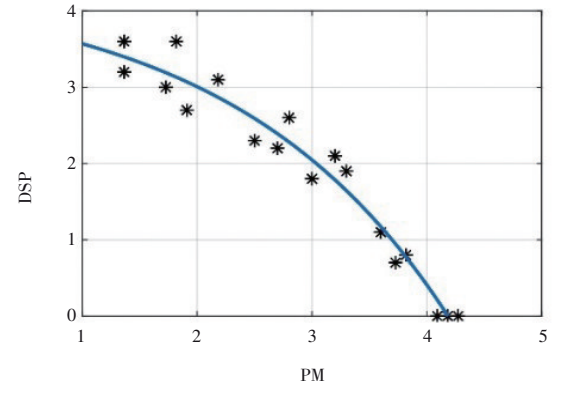
where v_{19}, v_{20}, v_{21}, and v_{22} are equal to 1.285, 0.01, 0.027 4, and −1.529, respectively; SPAV represents the spatial presence provided by the visual and acoustic experience.

Second, the impact of proprioceptive matching is investigated. Fig. 14 shows the relationship between the proprioceptive matching and the degradation of spatial presence. We can find that the degradation of spatial presence decreases with the increase of proprioceptive matching. The relationship in



ARE: acoustic realism SP: spatial presence VRE: visual realism

▲Figure 13. Relationships between the two types of realism and the spatial presence



DSP: degradation of spatial presence PM: proprioceptive matching

▲ Figure 14. Relationships between the proprioceptive matching and degradation of spatial presence

Fig. 14 can be modeled as

$$\mathrm{DSP(PM)} = v_{23}\cdot\exp\left(v_{24}\cdot\mathrm{PM}\right) + v_{25}, \quad (17)$$

where v_{23}, v_{24} and v_{25} are equal to −0.467 9, 0.533 8 and 4.367, respectively. Hence, the spatial presence can be calculated by

$$\mathrm{SP(SPAV, DSP)} = \min\left(\max\left(\mathrm{SPAV} - \mathrm{DSP}, 1\right), 5\right). \quad (18)$$

By utilizing the proposed model, the spatial presence of

360-degree video can be assessed based on the corresponding technical parameters extracted from the VR system.

5 Performance Evaluation

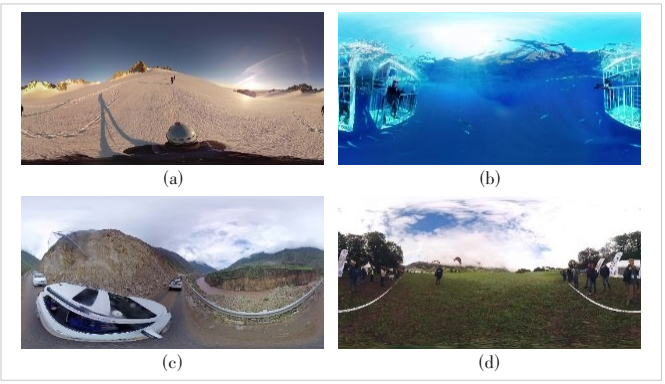
The performance of the proposed model was evaluated on a test set consisting of another four YUV420 360-degree video sequences that had a video resolution of 3 840×1 920 and a video framerate of 30 fps. Screenshots of the video content are shown in Fig. 15. Four lossless audio files (PCM, 48 kHz) containing two channels were utilized as the background sound of these 360-degree videos. The 360-degree videos were firstly down-sampled to 2K resolution and encoded with a BPP of 0.02, 0.06, 0.14, and 0.19 using the x.265 encoder. The audio files were encoded with 16 kbit/s, 64 kbit/s, 128 kbit/s, and 256 kbit/s using the AAC codec. We conducted two experiments to verify the performance of the proposed model by changing the video bitrate, audio bitrate, and MTP latency. In the first experiment, audiovisual files were displayed without MTP latency. The display FOV was set to be 90 degrees and 110 degrees, respectively. The details of the setting are shown in Table 5. In the second experiment, audiovisual files with 4K resolution were displayed with three MTP latencies, i.e., 40 ms, 120 ms, and 260 ms, respectively. The display FOV was set to be 110 degrees. The details of the setting are shown in Table 6. A total number of 30 subjects participated in these two experiments. After each display, the subjects provided their ratings on the spatial presence on a five-point scale.

Since there is no model evaluating the spatial presence that can be used as a comparison, we only show the performance of the proposed model. The performance is evaluated in two ways: 1) comparing predicted scores of the spatial presence with the subjective MOS, and 2) comparing the predicted scores with the subjective scores rated by individuals.

5.1 Predicted Scores vs MOS

Three commonly used performance criteria are employed to measure the performance of the proposed model: Pearson Correlation Coefficient (PCC), Root-Mean-Squared Error (RMSE), and Spearman Rank Order Correlation Coefficient (SROCC).

The model performance is given in Table 7. It can be found that reliable prediction performance is obtained when using the proposed spatial presence evaluation model.



▲ **Figure 15. Content of test sequences: (a) Driving, (b) Shark, (c) Glacier, and (d) Paramotor**

▼**Table 5. Setup for the video**

Video (BPP)	Audio/(kbit/s)
0.02	16, 64, 128, 256
0.06	16, 64, 128, 256
0.14	16, 64, 128, 256
0.19	16, 64, 128, 256

BPP: bits per pixel

▼**Table 6. Setup for the audio**

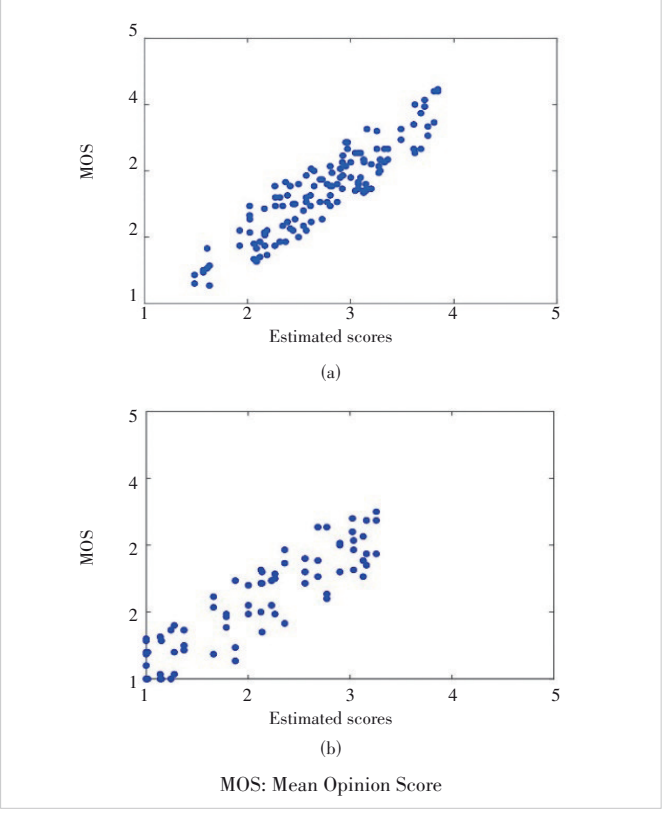
Video (BPP)	Audio/(kbit/s)
0.06	16, 64, 256
0.14	16, 64, 256
0.19	16, 64, 256

BPP: bits per pixel

▼**Table 7. Experimental results**

Experiment	PCC	SROCC	RMSE
1	0.910	0.894	0.277
2	0.908	0.900	0.335

PCC: Pearson Correlation Coefficient
RMSE: Root-Mean-Squared Error
SROCC: Spearman Rank Order Correlation Coefficient



▲ **Figure 16. Scatter plots of the subjective spatial presence versus predicted objective scores: (a) result of Experiment 1 and (b) result of Experiment 2**

To visualize the performance, Fig. 16 shows the scatter plots of objective scores predicted by the proposed model against the subjective MOSs. This figure clearly shows that the proposed model exhibits good convergence and monotonicity performance.

5.2 Predicted Scores vs Individual Ratings

To also check the accuracy of the proposed model, we evaluated the performance of the model against the individual ratings of subjects. Again, PCC, SROCC, and RMSE were calculated. For Experiment 1, we found that the PCC, SROCC, and RMSE ranged from 0.882 to 0.926, 0.878 to 0.922, and 0.443 to 0.227, respectively. For Experiment 2, we found that the PCC, SROCC, and RMSE ranged between 0.886 to 0.924, 0.881 to 0.918, and 0.462 to 0.214. Among the 30 subjects, the lowest, medium and highest prediction results are shown in Table 8. It can be found that a relatively good prediction performance is always guaranteed using the proposed model.

We also calculated the percentage that the predicted scores match the subjective scores to better verify the accuracy of the proposed model. A match is found if a predicted score (after the rounding process) is the same as the subjective score rated by the participants. The results show that the proposed model matches the subjective ratings with an accuracy of 83.7% and 82.4% for Experiments 1 and 2, respectively. It can be concluded that the proposed model manifests itself as a reliable spatial presence indicator that can be directly used in current 360-degree video applications.

▼Table 8. Model performance

Experiment	Subject No. 1			Subject No. 2			Subject No. 3		
	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
1	0.882	0.878	0.443	0.926	0.922	0.227	0.908	0.902	0.282
2	0.886	0.881	0.462	0.924	0.918	0.214	0.904	0.898	0.343

PCC: Pearson Correlation Coefficient
RMSE: Root-Mean-Squared Error
SROCC: Spearman Rank Order Correlation Coefficient

6 Conclusions

In this paper, we propose a spatial presence assessment framework for measuring users' sense of spatial presence in 360-degree video services. Well-designed subjective experiments are conducted to obtain accurate subjective ratings of spatial presence. An objective spatial presence prediction model is further proposed. Experimental results show that the proposed model can achieve good prediction accuracy in terms of PCC, SROCC, and RMSE. The proposed scheme serves as guidelines for the research community to better understand the spatial presence perception. It also provides valuable recommendations for the industry to further improve its quality of service.

References

[1] DELOITTE. Digital democracy survey: a multi-generational view of consumer technology, media and telecom trends, digital democracy survey 9th edition [R]. 2022
[2] ZHU W H, ZHAI G T, TAO M X, et al. Quality of experience estimation of ultra-high definition content [J]. ZTE technology journal, 2021, 27(1): 37 – 43. DOI: 10.12142/ZTETJ.202101009
[3] LI J L, ZHAO X, YANG Y. A review of interactive video quality assessment methods [J]. ZTE technology journal, 2021, 27(1): 44 – 47. DOI: 10.12142/ZTETJ.202101010
[4] LOMBARD M, JONES M T. Defining presence [M]//Immersed in media. Cham: Springer International Publishing, 2015: 13 – 34. DOI: 10.1007/978-3-319-10190-3_2
[5] SCHUEMIE M J, VAN DER STRAATEN P, KRIJN M, et al. Research on presence in virtual reality: a survey [J]. CyberPsychology & behavior, 2001, 4(2): 183 – 201. DOI: 10.1089/109493101300117884
[6] SEO Y, KIM M, JUNG Y, et al. Avatar face recognition and self-presence [J]. Computers in human behavior, 2017, 69: 120 – 127. DOI: 10.1016/j.chb.2016.12.020
[7] FELTON W M, JACKSON R E. Presence: a review [J]. International journal of human-computer interaction, 2022, 38(1): 1 – 18. DOI: 10.1080/10447318.2021.1921368
[8] LOMBARD M, DITTON T. At the heart of it all: the concept of presence [J]. Journal of computer-mediated communication, 2006, 3(2). DOI: 10.1111/j.1083-6101.1997.tb00072.x
[9] NORTH M M, NORTH S M. A comparative study of sense of presence of virtual reality and immersive environments [J]. Australasian journal of information systems, 2016, 20. DOI: 10.3127/ajis.v20i0.1168
[10] GONÇALVES G, MELO M, BARBOSA L, et al. Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive VR [J]. Virtual reality, 2022, 26(1): 1 – 14. DOI: 10.1007/s10055-021-00530-5
[11] SKARBEZ R, BROOKS Jr F P, WHITTON M C. A survey of presence and related concepts [J]. ACM computing surveys (CSUR), 2017, 50(6): 1 – 39. DOI: 10.1145/3134301
[12] LAARNI J, RAVAJA N, SAARI T, et al. Ways to measure spatial presence: review and future directions [M]//Immersed in media. Cham: Springer International Publishing, 2015: 139 – 185. DOI: 10.1007/978-3-319-10190-3_8
[13] AL-JUNDI H A, TANBOUR E Y. A framework for fidelity evaluation of immersive virtual reality systems [J].Virtual reality, 2022, 26(3): 1103 – 1122. DOI: 10.1007/s10055-021-00618-y
[14] EGAN D, BRENNAN S, BARRETT J, et al. An evaluation of heart rate and electrodermal activity as an objective QoE evaluation method for immersive virtual reality environments [C]//Proc. 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2016: 1 – 6. DOI: 10.1109/QoMEX.2016.7498964
[15] CHESSA M, MAIELLO G, BORSARI A, et al. The perceptual quality of the oculus rift for immersive virtual reality [J]. Human-computer interaction, 2019, 34(1): 51 – 82. DOI: 10.1080/07370024.2016.1243478
[16] TERKILDSEN T, MAKRANSKY G. Measuring presence in video games: an investigation of the potential use of physiological measures as indicators of presence [J]. International journal of human-computer studies, 2019, 126: 64 – 80. DOI: 10.1016/j.ijhcs.2019.02.006
[17] GRASSINI S, LAUMANN K. Questionnaire measures and physiological correlates of presence: a systematic review [J]. Frontiers in psychology, 2020, 11: 349. DOI: 10.3389/fpsyg.2020.00349
[18] ZOU W J, YANG F Z, ZHANG W, et al. A framework for assessing spatial presence of omnidirectional video on virtual reality device [J]. IEEE access, 2018, 6: 44676 – 44684. DOI: 10.1109/ACCESS.2018.2864872
[19] WITMER B G, SINGER M J. Measuring presence in virtual environments: a presence questionnaire [J]. Presence: teleoperators and virtual environments, 1998, 7(3): 225 – 240. DOI: 10.1162/105474698565686
[20] LESSITER J, FREEMAN J, KEOGH E, et al. A cross-media presence questionnaire: the ITC-sense of presence inventory [J]. Presence: teleoperators and virtual environments, 2001, 10(3): 282 – 297. DOI: 10.1162/105474601300343612

[21] JENNETT C, COX A L, CAIRNS P, et al. Measuring and defining the experience of immersion in games [J]. International journal of human-computer studies, 2008, 66(9): 641 – 661. DOI: 10.1016/j.ijhcs.2008.04.004
[22] VORDERER P, WIRTH W, GOUVEIA F R, et al. MEC spatial presence questionnaire (MEC-SPQ): short documentation and instructions for application [R]. 2004
[23] CUMMINGS J J, WERTZ E E. Capturing social presence: Concept explication through an empirical analysis of social presence measures [J]. Journal of computer-mediated communication, 2022, 28(1): zmac027. DOI: 10.1093/jcmc/zmac027
[24] LIN J J W, DUH H B L, PARKER D E, et al. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment [C]//Proc. IEEE Virtual Reality. IEEE, 2002: 164 – 171. DOI: 10.1109/VR.2002.996519
[25] YU M, LAKSHMAN H, GIROD B. A framework to evaluate omnidirectional video coding schemes [C]//Proc. 2015 IEEE International Symposium on Mixed and Augmented Reality. IEEE, 2015: 31 – 36. DOI: 10.1109/ISMAR.2015.12
[26] ZAKHARCHENKO V, CHOI K P, PARK J H. Quality metric for spherical panoramic video [C]//Proc. SPIE Optics and Photonics for Information Processing X. SPIE, 2016. DOI: 10.1117/12.2235885
[27] XU M, LI C, LIU Y F, et al. A subjective visual quality assessment method of panoramic videos [C]//Proc. 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2017: 517 – 522. DOI: 10.1109/ICME.2017.8019351
[28] UPENIK E, ŘEŘÁBEK M, EBRAHIMI T. Testbed for subjective evaluation of omnidirectional visual content [C]//Proc. 2016 Picture Coding Symposium (PCS). IEEE, 2017: 1 – 5. DOI: 10.1109/PCS.2016.7906378
[29] SCHATZ R, SACKL A, TIMMERER C, et al. Towards subjective quality of experience assessment for omnidirectional video streaming [C]//Proc. Ninth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2017: 1 – 6. DOI: 10.1109/QoMEX.2017.7965657
[30] OZCINAR C, CABRERA J, SMOLIC A. Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality [J]. IEEE journal on emerging and selected topics in circuits and systems, 2019, 9(1): 217 – 230. DOI: 10.1109/JETCAS.2019.2895096
[31] GHAZNAVI-YOUVALARI R, ZARE A, AMINLOU A, et al. Shared coded picture technique for tile-based viewport-adaptive streaming of omnidirectional video [J]. IEEE transactions on circuits and systems for video technology, 2019, 29(10): 3106 – 3120. DOI: 10.1109/TCSVT.2018.2874179
[32] DUAN H Y, ZHAI G T, MIN X K, et al. Perceptual quality assessment of omnidirectional images [C]//IEEE International Symposium on Circuits and Systems (ISCAS). 2018: 1 – 5. DOI: 10.1109/ISCAS.2018.8351786
[33] XU M, LI C, CHEN Z Z, et al. Assessing visual quality of omnidirectional videos [J]. IEEE transactions on circuits and systems for video technology, 2019, 29(12): 3516 – 3530. DOI: 10.1109/TCSVT.2018.2886277
[34] ERMI L, MÄYRÄ F. Fundamental components of the gameplay experience: analysing immersion [C]//Digital Games Research Conference 2005, Changing Views: Worlds in Play. DBLP, 2005: 37 – 53
[35] SLATER M, WILBUR S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments [J]. Presence: teleoperators and virtual environments, 1997, 6(6): 603 – 616. DOI: 10.1162/pres.1997.6.6.603
[36] CUMMINGS J J, BAILENSON J N. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence [J]. Media psychology, 2016, 19(2): 272 – 309. DOI: 10.1080/15213269.2015.1015740
[37] ZHAO J B, ALLISON R S, VINNIKOV M, et al. Estimating the motion-to-photon latency in head mounted displays [C]//Proc. 2017 IEEE Virtual Reality (VR). IEEE, 2017: 313 – 314. DOI: 10.1109/VR.2017.7892302
[38] ITU. Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment: ITU-T recommendation P.913 [S]. 2021
[39] ASBUN E, HE Y, HE Y, et al. AHG8: interdigital test sequences for virtual reality video coding [R]. 2016
[40] SUN W, GUO R. Test sequences for virtual reality video coding from letinVR [R]. 2016
[41] MERCAT A, VIITANEN M, VANNE J. UVG dataset: 50/120 fps 4K sequences for video codec analysis and development [C]//Proc. ACM Multimedia Systems Conference. ACM, 2020: 297 – 302. DOI: 10.1145/3339825.3394937
[42] ITU. Methods for objective measurements of perceived audio quality: ITU-R recommendation BS 13871 [S]. 2001
[43] BAILENSON J N, SWINTH K, HOYT C, et al. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments [J]. Presence: teleoperators and virtual environments, 2005, 14(4): 379 – 393. DOI: 10.1162/105474605774785235
[44] OU Y F, MA Z, LIU T, et al. Perceptual quality assessment of video considering both frame rate and quantization artifacts [J]. IEEE transactions on circuits and systems for video technology, 2011, 21(3): 286 – 298. DOI: 10.1109/TCSVT.2010.2087833
[45] OU Y F, LIU T, ZHAO Z, et al. Modeling the impact of frame rate on perceptual quality of video [C]//Proc. 15th IEEE International Conference on Image Processing. IEEE, 2008: 689 – 692. DOI: 10.1109/ICIP.2008.4711848
[46] ZOU W J, YANG F Z, WAN S. Perceptual video quality metric for compression artefacts: from two-dimensional to omnidirectional [J]. IET image processing, 2018, 12(3): 374 – 381. DOI: 10.1049/iet-ipr.2017.0826

Biographies

ZOU Wenjie received his BS and PhD degrees from Xidian University, China in 2009 and 2017, respectively. He is currently a lecturer with the Multimedia Communication Laboratory, Xidian University. His research interests include QoE, video quality assessment, and multimedia compression.

GU Chengming received his BE degree in communication engineering from Xidian University, China in 2021. He is currently working toward an ME degree in information and communication engineering with Xidian University. His research interests include video coding and processing.

FAN Jiawei received his BE degree in telecommunications engineering from Xidian University, China in 2022. He is currently working toward his ME degree in electronic information with Xidian University. His research interests include video coding and processing.

HUANG Cheng (huang.cheng5@zte.com.cn) received his MS degree from the School of Computer Science and Engineering, Southeast University, China. He is currently a senior system architect and project manager of video technology research at ZTE Corporation. His research interests include visual coding, storage, transport, and multimedia systems.

BAI Yaxian received his MS degree in communication engineering from Wuhan University of Technology, China. She is currently a senior engineer at ZTE Corporation. Her research interests include video coding and processing and point cloud compression and transmission.