THE IMPACT OF STALLING ON THE PERCEPTUAL QUALITY OF HTTP-BASED OMNIDIRECTIONAL VIDEO STREAMING

Wei Zhang, Wenjie Zou and Fuzheng Yang

School of Telecommunications Engineering
Xidian University
Xi'an, China

ABSTRACT

Recent years have witnessed increasing advances in immersive media technologies. Omnidirectional videos, which record a 360-degree spherical visual scene, have gained popularity at an accelerated speed. However, streaming this visual media is extremely bandwidth demanding when compared to traditional 2D videos, thus challenging current video streaming strategies and network conditions. Understanding how human visual system perceives the omnidirectional video becomes an urgent research topic. In this work, we focus on investigating the impact of stalling on the perceptual quality of omnidirectional video streaming. A systematic experiment with various phases was conducted to obtain the quality perception results towards omnidirectional video stalling. Experimental results showed that human visual system views the omnidirectional video in a different manner if compared with traditional 2D videos. This difference further triggers an alleviation in the annoyance level of stalling. A questionnaire focused on revealing the underlying reason regarding this difference was further designed. Findings of this work provide insights into the visual quality perception of stalling events during omnidirectional video playback. Conclusions of this paper can be used as guidelines for the design of video streaming protocols as well as network dimensioning.

Index Terms— Omnidirectional video, perceptual quality, stalling, video streaming, subjective experiment.

1. INTRODUCTION

With the development of immersive media technologies and the popularity of virtual reality (VR) display device such as Head Mounted Display (HMD), omnidirectional video service is rapidly expanding in many fields including business, entertainment, military, and education. As a representative type of immersive media, omnidirectional videos can provide users with 360-degree visual scenes which can enhance the sense of immersion and presence of end users. Due to these advantages, omnidirectional videos soon attract a large number of users. Mainstream video streaming service providers such as YouTube, Netflix all deployed omnidirectional video services in their platforms.

In recent years, HTTP-based video streaming has become a popular solution for the majority of video-sharing websites and video-on-demand services. This streaming strategy features several advantages including the reliable delivery of video streams through TCP and progressive download technology [1]. Compared with traditional UDP-based streaming, TCP-based streaming frees the video decoders from handling packet loss and prevents the video quality degradation caused by packet loss. However, due to channel throughput fluctuations and bandwidth limitations, undesirable playback interruptions, also known as stalling [2], can greatly impact the perceived quality of end users [3, 4, 5, 6]. Although some HTTP Adaptive Streaming (HAS) technologies have been proposed to mitigate the risk of playback interruptions by adapting the bitrate of the video stream to the network throughput, HTTP-based video streaming is still unable to completely avoid rebuffering during playback [7]. Therefore, it becomes crucial for the research community and industrial to evaluate the quality perception behaviour of HTTP-based video streaming services.

In [4, 5], authors considered the impact of stalling frequency on the perceived quality of traditional 2D videos. They concluded that a single stalling is a better solution than repeated stalling events. In [6], authors compared the influence of initial rebuffering and stalling during watching on the quality perception of 2D videos. Experimental results showed that users are extremely sensitive to stalling events if compared with the initial rebuffering. In [8], a subjective experiment was conducted to investigate the effect of stalling on the perceived quality of mobile video watching. Results showed that the subjective quality of users were highly correlated with stalling duration and the number of stalling events. In general, the results of existing psychophysical studies revealed that the stalling events perceived during the video

This research was supported by the National Nature Science Foundation of China (Grant No. 61801364) and the Fundamental Research Funds for the Central Universities (Grant No. JB180105)

Fig. 1. Illustration (5 out of 20) of the omnidirectional videos used in the experiment.

streaming can strongly degrade the quality perception results of end users. It may provide inspirations for researchers regarding the quality perception of stalling events in omnidirectional videos. However, compared with traditional 2D videos, omnidirectional videos provide end users a larger field of vision (FOV) and a higher degree of freedom (DOF) in selecting desired viewpoints. This intrinsic characteristic of omnidirectional videos may alter the viewing behaviour of end users and thus impact their quality perception results when stalling occurs. To this end, it is crucial to investigate how stalling plays a role in the quality perception of omnidirectional videos.

To our best knowledge, so far there exists few study in the literature towards the effect of stalling on the perceived quality of omnidirectional videos, especially for highlighting the quality perception difference between 2D viewing condition using flat monitors and immersive condition using HMDs. A recent research in [9] investigated the impact of stalling in the context of omnidirectional video streaming using HMD. Experimental results showed that stalling events seriously impact the perceived quality of omnidirectional video streaming in a very similar way as they do in the context of traditional viewing conditions. It should be noted that this research focused on investigating the stalling pattern with different numbers of stalling, and the duration of each stalling was fixed. As the duration of stalling is also a crucial factor to influence the perceived quality of end users, it is therefore necessary to investigate the effect of stalling in different durations.

In this paper, we hypothesize that the stalling event has different impact on the perceived quality between omnidirectional videos and traditional videos. To validate our hypothesis, we first designed a systematic experiment with various phases towards the subjective quality of omnidirectional video streaming. A questionnaire was also designed to further reveal the viewing behaviour difference between the two viewing conditions.

2. EXPERIMENTAL DESIGN

2.1. Stimuli

A total number of twenty omnidirectional videos were selected as test stimuli. These videos features a resolution of 4K and a duration ranging from 10 to 30 seconds. To ensure the stimulus selection is not restrictive in terms of the diversity of video content, we further screened the spatial perceptual information (SI) and temporal perceptual information (TI) for each content. Experimental results showed that the distributions of SI and TI cover a wide range of space indicating a sufficient stimulus diversity. To generate stimuli with different perceived quality levels, different durations of stalling were simulated as shown in Table 1. More specifically, after a certain period of initial playback (e.g., 8 seconds in our setup), the video sequences will suffer from a stalling event with 13 different durations from 1 second to 50 seconds. To notify the viewers that the stalling is finished, we empirically set a two-second playback as with [10] before the viewers giving their opinion scores.

Table 1. The playback and stalling duration (/seconds) setting in our experiment.

Initial playback	stalling	playback
8	1	2
8	2	2
8	4	2
8	5	2
8	6	2
8	8	2
8	10	2
8	11	2
8	13	2
8	15	2
8	20	2
8	30	2
8	50	2

2.2. Procedure

To compare the effect of stalling between traditional 2D viewing condition and the immersive HMD viewing condition, the experiment was divided into three phases:

2.2.1. Phase 1: quality rating in 2D viewing condition

In this experiment, subjects were asked to view the omnidirectional videos shown on a flat screen. A fixed viewpoint (i.e., the initial viewpoint) per stimulus was shown on the screen to subjects without giving freedom of changing viewpoint. A 32-inch LCD monitor with a native resolution of 3840×2160 pixels was utilized as the display device. In total,

Table 2. Questions used for acquiring the overall user experience after Phase 2.

	questions	rating scale
Q1	I will explore the 360-degree visual scene when stalling occurs	fully disagree — fully agree
Q2	The exploration of scenes during stalling alleviates its impact on perceptual quality	not at all — very much
Q3	The annoyance level rises up along with the extension of stalling duration	fully disagree — full agree

40 subjects participated in Phase 1, including 21 males and 19 females. All of them reported normal or correct-to-normal sight. Since standardised protocol for conducting subjective experiment on omnidirectional videos do not exist, we thus followed the suggestion of ITU-T P.913 [11] as the experiment protocol. In our setup, each video has 13 derivatives with different durations of stalling. This may involve massive stimulus repetition which leads to undesired carry-over effects such as learning effect [12]. To reduce this effect, we followed a rigorously validated methodology as used in [12] based on the between-subjects design. We randomly divided subjects into two groups with 20 subjects each. For each video content, one group has only to view 6 derivatives and the other view the rest 7 derivatives. A five-point ACR rating scale (corresponding to the perceived quality of excellent, good, fair, poor and bad) was used to rate the perceived quality. The mean opinion score (MOS) representing the average quality perception results is then derived according to the data processing method in Rec. ITU-T P.913 [11].

2.2.2. *Phase 2:quality rating in immersive viewing condition*

In Phase 2, subjects were asked to rate the perceived quality of the same omnidirectional videos with stalling using an HMD. Note, a two-day interval was adopted as a "washout" period between Phase 1 and Phase 2 to further reduce the memory effect due to content repetition. All 40 subjects participated in the experiment and were asked to view the videos "as they normally would". Again, the between-subjects design was utilised and the same grouping method was followed. The HMD has a screen resolution of 2160×1200 pixels, a refresh rate of 90 Hz and a horizontal FOV of 110 degrees. A scoring interface was also designed for subjects that allows them to rate the perceptual quality without taking off the HMD.

2.2.3. *Phase 3: Post-questionnaire*

To investigate the viewing behaviour of users during the period of stalling in Phase 2, a questionnaire including three items was designed. Each subjects should answer all the questions after they scoring all the video sequences using an HMD. Details regarding the questionnaire are shown in Table 2. A 5-point Likert scale is used to rate individual items as included in the questionnaire. We managed to understand the viewing behaviour when stalling occurs during the HMD viewing condition. More specifically, we focused on evaluating the change of viewpoint during stalling event.

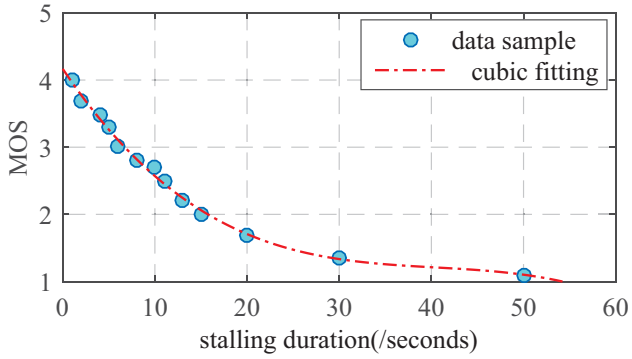

Fig. 2. The relationship between perceived quality and the stalling events in different durations for traditional 2D viewing condition.

3. EXPERIMENTAL RESULTS

3.1. Data validation

To make use of questionnaire rating scores as a solid "ground truth", various tests were conducted toward the quesitonnaire. To check the reliability of the questionnaire, Cronbach's α test [13] was performed. The Cronbach's α value ranges from 0 to 1 with a higher value indicating a higher reliability of the questionnaire. Following the rule of thumb, a value greater than 0.8 indicates a high reliability. In our case, the Cronbach's α value is 0.833, showing that the questionnaire is reliable. To also check the validity of the questionnaire, Kaiser-Meyer-Olkin (KMO) test [14] was performed. The KMO value ranges from 0 to 1 with a higher value indicating a higher sampling adequacy. Experimental results showed that the KMO value of the questionnaire is 0.697, demonstrating that the questionnaire holds sufficient validity for conducting further data analysis.

3.2. The effect of stalling on two viewing conditions

As illustrated in Fig. 2, the perceived quality of subjects gradually decreases along with the increase of stalling duration. In contrast, as shown in Fig. 3, the observed tendency can be divided into three stages. The first stage corresponds to the stalling duration from 1 second to 10 seconds where the perceptual quality decreases gradually. The second stage corresponds to the stalling duration from 10 seconds to 30 seconds where the perceptual quality of end users maintains sta-

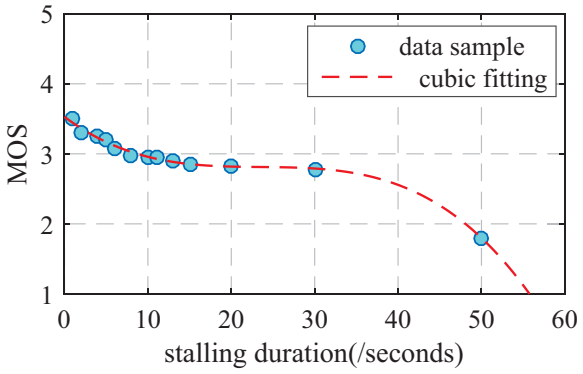

Fig. 3. The relationship between perceived quality and the stalling events in different durations for immersive HMD viewing condition.

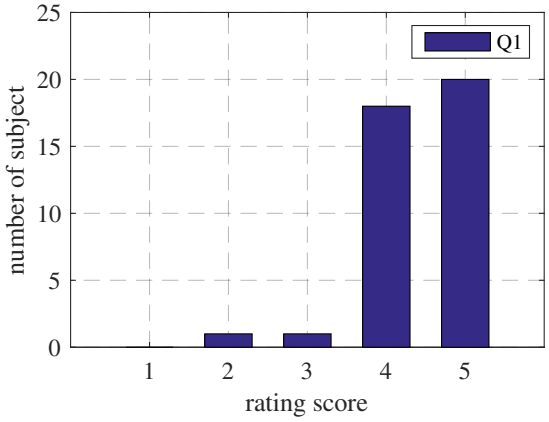

Fig. 4. The histgram of the rating score distribution corresponding to Q1 in the questionnaire.

ble. The third stage corresponds to the stalling duration from 30 seconds to 50 seconds where the subjective quality of end users decreases rapidly. The comparison between these two plots clearly evidenced that stalling is relatively less annoying when streaming omnidirectional videos.

To clarify the underlying reason for the difference in the tendency, we analysed the questionnaire results over individual items. Figure. 4 shows the rating score distribution toward Q1 in the questionnaire. It can be seen that most subjects (38 out of 40) strongly agree that they will explore the 360-degree visual scene when stalling event occur. It indicates that the viewing strategy of subjects are dramatically changed when compared to viewing traditional 2D videos where no surrounding visual scene is provided when stalling occurs. Figure. 5 shows the distribution of rating scores of Q2. It shows that most subjects (37 out of 40) felt that exploring the visual scene during the stalling can alleviate its annoyance level. It thus explains why the perceived quality maintains approximately the same when the duration of stalling extends. Moreover, as shown in Fig. 6, not all subjects agreed

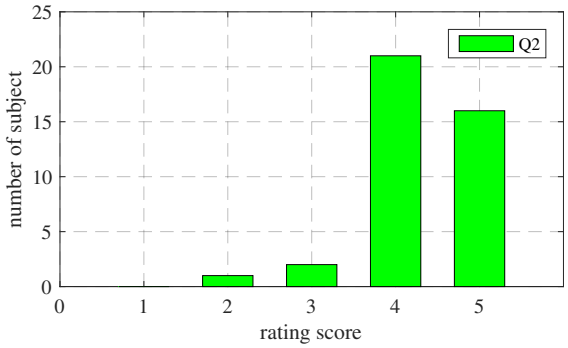

Fig. 5. The histgram of the rating score distribution corresponding to Q2 in the questionnaire.

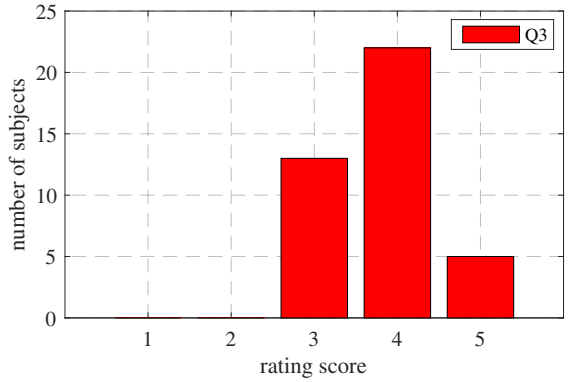

Fig. 6. The histgram of the rating score distribution corresponding to Q3 in the questionnaire.

that the annoyance level rises up along with the extension of stalling duration. Instead, 13 out of 40 subjects thought that the stalling has less impact on their quality judgement by giving a score of 3, which is necessary to be taken into account in the design of streaming strategies.

4. CONCLUSION

In this work, we focused on investigating the impact of stalling events on the perceptual quality of omnidirectional video streaming. A systematic experiment with various phases was designed to clarify the knowledge. Experimental results showed that human visual system views the omnidirectional video in a different manner if compared with traditional 2D viewing condition. This difference further triggers an alleviation in the annoyance level of stalling. A questionnaire focused on revealing the underlying reason regarding this phenomena was further designed. Conclusions of this paper can be used as guidelines for the design of video streaming strategies as well as network dimensioning.

5. REFERENCES

[1] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of http adaptive streaming," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, Firstquarter 2015.

[2] P. Ameigeiras, A. Azcona-Rivas, J. Navarro-Ortiz, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A simple model for predicting the number and duration of rebuffering events for youtube flows," *IEEE Communications Letters*, vol. 16, no. 2, pp. 278–280, February 2012.

[3] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, Dec 2013.

[4] X. Tan, J. Gustafsson, and G. Heikkilä, "Perceived video streaming quality under initial buffering and rebuffering degradations," in *MESAQIN Conference*, June 2006, vol. 90.

[5] J. Gustafsson, G. Heikkila, and M. Pettersson, "Measuring multimedia quality in mobile networks with an objective parametric model," in *Proc. of the 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 405–408.

[6] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *Proc. of the 4th International Workshop on Quality of Multimedia Experience*, July 2012, pp. 1–6.

[7] J. Yao, S. S. Kanhere, I. Hossain, and M. Hassan, "Empirical evaluation of http adaptive streaming under vehicular mobility," in *NETWORKING*, Jordi Domingo-Pascual, Pietro Manzoni, Sergio Palazzo, Ana Pont, and Caterina Scoglio, Eds., Berlin, Heidelberg, 2011, pp. 92–105.

[8] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, March 2013.

[9] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *Proc. of the 9th International Conference on Quality of Multimedia Experience*, May 2017, pp. 1–6.

[10] W. Zou, F. Yang, J. Song, S. Wan, W. Zhang, and H. R. Wu, "Event-based perceptual quality assessment for http-based video streaming with playback interruption," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1475–1488, June 2018.

[11] ITU-T, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," *Rec. ITU-T P.913*, Mar. 2016.

[12] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2424–2437, May 2017.

[13] J. M. Bland and D. G. Altman, "Statistics notes: Cronbach's alpha," *British Medical Journal*, vol. 314, no. 7080, pp. 572, 1997.

[14] B. Williams, A. Onsman, and T. Brown, "Exploratory factor analysis: A five-step guide for novices," *Australasian Journal of Paramedicine*, vol. 8, no. 3, 2010.