2023 15th International Conference on Quality of Multimedia Experience (QoMEX)

Impact of Quality and Distance on the Perception of Point Clouds in Mixed Reality

Minh Nguyen†, Shivi Vats†, Sam Van Damme§, Jeroen van der Hooft§, Maria Torres Vega§, Tim Wauters§, Christian Timmerer†, Hermann Hellwagner†
†Department of Information Technology, Alpen-Adria-Universität Klagenfurt, Austria
§IDLab, Department of Information Technology, Ghent University - imec, Belgium
{minh.nguyen, shivi.vats, christian.timmerer, hermann.hellwagner}@aau.at,
{sam.vandamme, jeroen.vanderhooft, maria.torresvega, tim.wauters}@ugent.be

***Abstract*—Point Cloud (PC) streaming has recently attracted research attention as it has the potential to provide six degrees of freedom (6DoF), which is essential for truly immersive media. PCs require high-bandwidth connections, and adaptive streaming is a promising solution to cope with fluctuating bandwidth conditions. Thus, understanding the impact of different factors in adaptive streaming on the Quality of Experience (QoE) becomes fundamental. Mixed Reality (MR) is a novel technology and has recently become popular. However, quality evaluations of PCs in MR environments are still limited to static images. In this paper, we perform a subjective study on four impact factors on the QoE of PC video sequences in MR conditions, including quality switches, viewing distance, and content characteristics. The experimental results show that these factors significantly impact QoE. The QoE decreases if the sequence switches to lower quality and/or is viewed at a shorter distance, and vice versa. Additionally, the end user might not distinguish the quality differences between two quality levels at a specific viewing distance. Regarding content characteristics, objects with lower contrast seem to provide better quality scores.**

***Index Terms*—Point Clouds, Quality of Experience, Subjective Tests, Mixed Reality.**

I. Introduction

A Point Cloud (PC) is a 3D representation format that allows viewers to see the details of objects without any constraint on the viewpoint. PCs can be watched on 2D screens and Virtual Reality (VR) and Augmented Reality (AR) head-mounted displays (HMDs). VR HMDs occlude the real world, so the viewer watches and interacts with PCs in a virtual environment. AR HMDs are transparent; hence, the viewer can see both the physical world and virtual objects (*i.e.*, images, videos, and text).

PCs require high-bandwidth networks for their transmission. One promising solution is adaptive streaming techniques (*i.e.*, HTTP Adaptive Streaming (HAS) [4]) combined with Point Cloud Compression (PCC) [14]. HAS adapts to the variation of the network conditions to prevent rebuffering events while providing the highest quality possible by changing the quality of PCs. PCC reduces the delivered data volume but distorts the visual quality. Therefore, understanding the impact of different factors in HAS and PCC on the QoE is of importance.

PCs have been evaluated in different viewing conditions (*i.e.*, VR HMDs and 2D screens) [2], [17]. However, research on quality assessment of PCs in MR environments is still limited. MR enhances people's perception of physical and virtual environments [3]. MR is, thus, an interesting setting for immersive telepresence applications, which we develop and assess in a research project.

Wu *et al.* [20] evaluated the quality of PCs with different quality levels. However, quality switching in the test sequences was not considered, and the device used in the subjective test was a VR device (*i.e.*, HTC Vive) rather than an AR device. The work in [17] considered the quality switching of PC videos in the context of HAS. It found that the texture of PC objects is an essential factor in the QoE. The content with fewer contrast differences can provide higher QoE. However, the PCs were displayed on a 2D screen that cannot offer a truly immersive experience. The work in [6], [16] considered different quality levels and viewing distances while watching two 3D representation formats, including PCs and Meshes. The results showed that a closer viewing distance led to lower QoE for a given quality level.

In this work, we examine the impact of different factors on QoE, including *(i)* quality switches, *(ii)* viewing distance, and *(iii)* content characteristics, while the user is watching PC videos in an MR environment through an AR device (*i.e.*, Microsoft HoloLens 2).

The contributions of this paper are twofold: *(i)* We provide an evaluation, based on a subjective study, of the quality of life-size digital humans in dynamic PC format in MR conditions. *(ii)* We analyze qualitative and quantitative impacts of various factors in the context of HAS on the QoE in MR environments.

II. Subjective Test for Point Cloud Assessment

This section gives an overview of the influence factors used, followed by a description of our subjective test, including dataset, equipment, environment, and tasks.

A. Influence Factors

The focus of this study is on the three main influence factors on the QoE while watching PC videos in MR environments.

Fig. 1: Raw points (left) and square shader (right) at 2.5 m.

Quality switches: In the context of HAS, the video quality can be changed due to throughput fluctuation [10], [13]. This is called a quality switch. Quality switches can be classified as *switching up* when the quality is increased and *switching down* when the quality is decreased.

Viewing distance: As 6DoF interaction allows end users to move in their space freely, the viewing distance from end users to the object can vary based on their movement. In this work, we examine the impact of viewing distance on the users' QoE.

Content characteristics: The perspective of viewers can vary depending on the sort of content [15], [17]. In this paper, four videos with different characteristics are used.

B. Dataset Preparation

We use four PC objects from the 8i Voxelized Full Bodies Database [7]: *Loot*, *LongDress*, *RedAndBlack*, and *Soldier*, captured at 30 fps for 10 seconds. The first two have lower contrast than the others [17].

We use the MPEG PCC reference software Test Model Category 2 [1] to create compressed PCs by varying quantization parameters (QPs). As there are two attributes (*i.e.*, geometry and texture) in a PC, a pair of QPs, namely geometry QP (G-QP) and texture QP (T-QP), are used in the encoding. A higher G-QP makes points deviate more from their original position. Similarly, when the T-QP increases, some color information is combined [20]. Three pairs of QPs (G-QP, T-QP) in the MPEG PCC software are selected, including *Q1:* (32, 42), *Q2:* (24, 32), and *Q3:* (16, 22). $Q3$ is thus the best quality. The bitrate of the objects is decreased with the increase of the QPs. For instance, *Loot*'s bitrates are $2.3\,\mathrm{Mbit/s}$, $5.6\,\mathrm{Mbit/s}$ and $16.7\,\mathrm{Mbit/s}$ for $Q1$, $Q2$, and $Q3$, respectively.

We develop a Unity project utilizing the Pcx Point Cloud importer [12] and the square shader from PointXR [2] to import and render the PCs, respectively. Fig. 1 compares a PC in raw points and square shader representations. Pre-tests clearly show that the square shader of [2] is visually superior to raw PCs; this format is therefore used in our test. C# scripts in [18] are utilized to control the quality and distance to design the sequences mentioned in the sequel.

C. Equipment and Environment

We use the Microsoft HoloLens 2 for the subjects to interact with our experiments. The AR HMD HoloLens 2 includes displays with 2K resolution and a diagonal field of view (FoV) of 52° [11], [19].

Following the recommendations of [8], our experiments are conducted in a room with black walls and low illumination. The tested PCs are placed in a room with life-size (1.8 m height) to obtain realistic telepresence scenarios.

TABLE I: Notation and description of the test sequences.

Notation	Description	Sequences
Qij	The video starts with quality Qi, then switches to Qj after 5 s. $i, j \in \{1, 2, 3\}$.	Loot and LongDress
Qi_Dj	The video is watched at quality Qi at distance Dj	RedAndBlack and Soldier

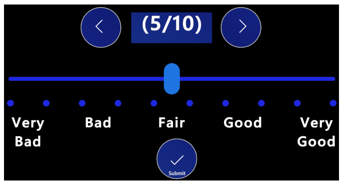


Fig. 2: Rating slider.

D. Experiment Tasks

We design two tasks for each participant to watch 36 sequences, 10 s each. Table I describes the sequences.

1) Task 1: Impact of video quality switches: The participant watches nine sequences for each of the two objects, including three sequences with static quality and six sequences with a quality switch in the middle of the sequences. The objects are 5 m from the participant such that the whole body can be watched. *Loot* and *LongDress* are used in this task as they have different contrasts.

2) Task 2: Impact of viewing distance: The participant watches static-quality sequences of the other two objects at quality levels $Q1$, $Q2$, and $Q3$ at three distances: *(D1)* 1.25 m (only face and shoulder in FoV), *(D2)* 2.5 m (only upper body in FoV), and *(D3)* 5 m (full body in FoV).

The order of tasks and sequences of each task is randomized. After watching each sequence, the participant is asked to rate the perceptual quality (*i.e.*, 1, 2 – very bad, ..., 9, 10 – very good) through the immersive slider shown in Fig. 2.

Before the experiment, the participants are asked to provide some background information, including age, gender, eyesight, and experience in viewing VR, AR, and MR contents. The total length of a single experiment is around 25 minutes.

III. Results and Discussions

A. Participants

A total of 36 participants attended the subjective test, including 22 (61%) males, 13 (36%) females, and 1 (3%) non-binary. 3 (8%) were in the age group from 18 to 24 years, 18 (50%) were between 25 and 34, 12 (33%) between 34 and 44, 2 (6%) between 45 and 54, and 1 (3%) between 55 and 64. The color vision of the participants is evaluated through the Ishihara test [5]. Four participants failed this test, so their ratings are excluded.

B. Task 1 – Impact of Quality Switches

Fig. 3 describes the ratings of participants for different quality switches. There is no remarkable improvement in the quality scores when the video starts at quality $Q1$ (*i.e.*, $Q11$, $Q12$, and $Q13$). ANOVA analysis indicates no significant difference ($p > 0.05$) among the quality scores for both *Loot* ($p = 0.07$) and *LongDress* ($p = 0.13$). This can be attributed

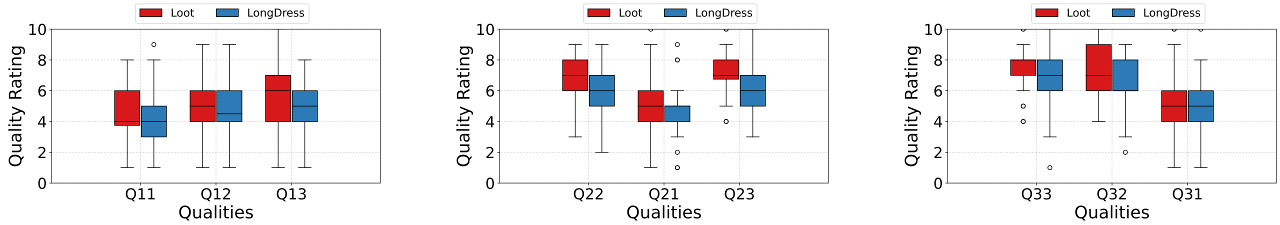
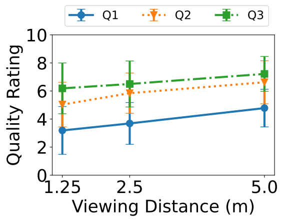
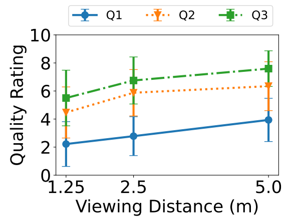


Fig. 3: Quality ratings for different quality levels and quality switches.

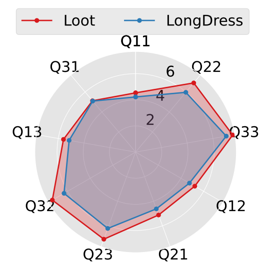


(a) RedAndBlack.

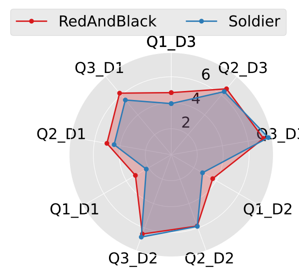


(b) Soldier.

Fig. 4: Average quality ratings for different distances.



(a) Task 1.



(b) Task 2.

Fig. 5: Average quality ratings of participants. It should be noted that the sequence Qii in Task 1 is equivalent to Qi_D3 ($i \in \{1, 2, 3\}$) in Task 2 as they are encoded at quality Qi and viewed at 5 m.

to the severe distortion of $Q1$ in the initial 5 s that impairs the QoE of watching the full 10 s video.

Regarding *switching down*, when the quality changes from $Q2$ or $Q3$ to $Q1$ (*i.e.*, $Q21$ or $Q31$), the quality ratings are remarkably reduced, compared to constant quality ($Q22$ and $Q33$). However, there are no significant differences when the quality changes between $Q2$ and $Q3$. We conducted paired samples t test [9] to validate this observation further. It shows non-significant p-values between $Q22$ and $Q23$ (*e.g.*, $p = 0.61$ for *Loot*) as well as between $Q33$ and $Q32$ (*e.g.*, $p = 0.12$ for *Loot*). Combined with the results in the previous section, we claim that the end user hardly recognizes the quality differences between $Q2$ and $Q3$. That comes to a recommendation that it is unnecessary to change the quality from $Q2$ to $Q3$ when the object is viewed at a 5 m distance.

Additionally, the QoE is affected by the encoding parameters, QPs. The lower the QPs are, the higher the QoE is. For example, the medians of $Q11$, $Q22$, and $Q33$ are 4, 7, and 8, respectively, for *Loot*.

C. Task 2 – Impact of Viewing Distance

Fig. 4 shows the quality ratings of the test objects at different viewing distances. It is noticeable that the distance significantly impacts the visual quality of the objects: the higher the viewing distance, the higher the quality scores. The reason is that, at a higher distance, the viewers do not recognize some quality distortion; thus, they give higher quality scores.

In addition, we observe that to achieve the same visual quality, the object should be encoded with lower QPs (*i.e.*, more data) if viewed closer. For example, *RedAndBlack* at quality $Q1$ is rated on average 4.8 at 5 m, and this object has to be encoded at $Q2$ to gain a similar score (*i.e.*, 4.9) if it is viewed at 1.25 m ($p = 0.5$ in a paired t test).

D. Impact of Content Characteristics

In this paper, we also evaluate the impact of content characteristics on the visual perception of participants for both tasks, as shown in Fig. 5. *Loot* and *RedAndBlack* achieve higher quality ratings in most cases. For example, the quality scores of *Loot* and *RedAndBlack* with quality $Q1$ viewed at distance $D3$ (*i.e.*, 5 m) are 4.5 and 4.8, respectively. Under the same conditions, these figures for *LongDress* and *Soldier* are 4.2 and 3.9, respectively. This can be explained by the fact that the participants are less sensitive to the quality distortion/changes for the content with fewer contrast differences. This extends the results in the work [17] on 2D screens to an MR environment with AR HMDs, in which the texture of the objects is a crucial factor to the viewers.

IV. Conclusions

In this paper, we investigate the impact of different factors on the QoE of PC videos in MR environments, including quality switches, viewing distance, and content characteristics. The experimental results show that the QoE will be decreased if the sequence switches to lower quality and/or is viewed at a shorter distance, and vice versa. In addition, lower-contrast contents may provide higher QoE. We also suggest the sequence should be encoded at lower QPs (*i.e.*, better quality) to maintain the quality score when viewed at a shorter distance.

Acknowledgements

This work has been funded by the European Union (SPIRIT project, Grant Agreement 101070672, https://www.spirit-project.eu/). Sam Van Damme and Jeroen van der Hooft are funded by the Research Foundation Flanders (FWO), grant numbers 1SB1822N and 1281021N, respectively.

References

[1] Video Point Cloud Compression - VPCC - MPEG-PCC-TMC2 Test Model Candidate Software. https://github.com/MPEGGroup/mpeg-pcc-tmc2, 2022. Accessed: 27 March 2023.

[2] E. Alexiou, N. Yang, and T. Ebrahimi. PointXR: A Toolbox for Visualization and Subjective Evaluation of Point Clouds in Virtual Reality. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.

[3] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2):1–36, 2018.

[4] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann. A survey on bitrate adaptation schemes for streaming media over http. *IEEE Communications Surveys Tutorials*, 21(1):562–585, 2019.

[5] J. Birch. Efficiency of the Ishihara Test for Identifying Red-Green Colour Deficiency. *Ophthalmic and Physiological Optics*, 17(5):403–408, 1997.

[6] K. Cao, Y. Xu, and P. Cosman. Visual Quality of Compressed Mesh and Point Cloud Sequences. *IEEE Access*, 8:171203–171217, 2020.

[7] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou. 8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset. In *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Input Document M40059/M74006*, 2017.

[8] ITU. Methodologies for the Subjective Assessment of the Quality of Television Images, Recommendation ITU-R BT. 500-14 (10/2019), 2020. Geneva, Switzerland.

[9] T. K. Kim. T Test as a Parametric Statistic. *Korean Journal of Anesthesiology*, 68(6):540–546, 2015.

[10] M. Nguyen, C. Timmerer, and H. Hellwagner. H2BR: An HTTP/2-based Retransmission Technique to Improve the QoE of Adaptive Video Streaming. In *Proceedings of the 25th ACM Workshop on Packet Video*, pages 1–7, 2020.

[11] A. Pose-Díez-de-la Lastra, R. Moreta-Martinez, M. García-Sevilla, D. García-Mato, J. A. Calvo-Haro, L. Mediavilla-Santos, R. Pérez-Mañanes, F. von Haxthausen, and J. Pascau. HoloLens 1 vs. HoloLens 2: Improvements in the New Model for Orthopedic Oncological Interventions. *Sensors*, 22(13):4915, 2022.

[12] K. Takahashi. Pcx - Point Cloud Importer/Renderer for Unity. https://github.com/keijiro/Pcx, 2021. Accessed: 27 March 2023.

[13] B. Taraghi, M. Nguyen, H. Amirpour, and C. Timmerer. INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming. *IEEE Access*, 9:118087–118098, 2021.

[14] C. Timmerer and A. C. Begen. A Journey Towards Fully Immersive Media Access. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2703–2705, 2019.

[15] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang. A Subjective Study on QoE of 360 Video for VR Communication. In *2017 IEEE 19th international workshop on multimedia signal processing (MMSP)*, pages 1–6. IEEE, 2017.

[16] S. Van Damme, I. Mahdi, H. Kumar Ravuri, J. van der Hooft, F. De Turck, and M. Torres Vega. Immersive and Interactive Subjective Quality Assessment of Dynamic Volumetric Meshes. In *15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2023.

[17] J. van der Hooft, M. Torres Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz. Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.

[18] S. Vats, M. Nguyen, S. Van Damme, J. van der Hooft, M. Torres Vega, T. Wauters, C. Timmerer, and H. Hellwagner. A Platform for Subjective Quality Assessment in Mixed Reality Environments. In *15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2023.

[19] VRCompare. Microsoft Hololens 2. [Online] Available: https://vr-compare.com/headset/microsofthololens2. Accessed: 10 February 2023.

[20] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong. Subjective Quality Database and Objective Study of Compressed Point Clouds with 6DoF Head-Mounted Display. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12):4630–4644, 2021.