Subjective evaluation of dynamic point clouds: Impact of compression and exploration behavior

Jesús Gutiérrez[1], Gulzhanat Dandyyeva[2], Matteo Dal Magro[2], Carlos Cortés[1], Michele Brizzi[3],
Marco Carli[3], Federica Battisti[2],
[1]*Grupo de Tratamiento de Imágenes, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, Spain*
[2]*Department of Information Engineering, Università di Padova, Italy*
[3]*Department of Industrial, Electronic and Mechanical Engineering, Università Roma Tre, Italy*
Email: {jesus.gutierrez, carlos.cs}@upm.es, {gulzhanat.dandyyeva, matteo.dalmagro}@studenti.unipd.it,
{michele.brizzi, marco.carli}@uniroma3.it, federica.battisti@unipd.it

***Abstract*—The objective of this work was to explore the subjective quality assessment of dynamic point clouds with compression artifacts and to analyze the exploration behavior of the users while visualizing them with a Head-Mounted Display with 6 Degrees-of-Freedom. Thus, this paper presents a subjective study on dynamic point clouds using the Absolute Category Rating methodology and considering different compression rates using the MPEG standard Video-based Point Cloud Compression. Firstly, results on the impact of compression artifacts on the perceived Quality of Experience of the users are reported, showing the validity of Absolute Category Rating, although more than 20 observers may be needed to obtain robust conclusions. Results on users' exploration behavior show no significant differences when visualizing point clouds with different qualities, no changes in the behavior during the test session, and no correlation between exploration activity and quality assessments. Further research will be conducted to help identify appropriate methodologies for the subjective assessment of point clouds and for understanding users' exploration behavior.**

***Index Terms*—Point cloud, Quality of Experience, compression, Virtual Reality exploration**

I. INTRODUCTION

In recent years, we have witnessed an increase in multimedia systems that attempt to give the feeling of immersiveness to a user. The availability of such systems has required significant advances in both hardware (acquisition and rendering systems) and software. In particular, for rendering volumetric data, promising results have been obtained with Point Clouds (PCs). A PC is a discrete set of data points in space, each point can be characterized by attributes such as color, surface normal, reflectivity, etc. It is generally produced by 3D scanners or camera arrays. Due to the huge amount of data generated during the acquisition process, several compression algorithms have been developed. For instance, in [1] a detailed description and comparison of two MPEG codec algorithms Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC) is carried out.

While achieving data compression, these algorithms can introduce perceptible artifacts. Thus, in order to optimize these algorithms and to provide the highest possible Quality of Experience (QoE) to the end users, various assessment studies have been performed and even a Grand Challenge has been recently organized[1]. Some works in the state-of-the-art have investigated the impact on the perceived quality of typical degradations that affect PCs [2], [3] using traditional double-stimulus methodologies to provide an explicit reference to observers who may be unfamiliar with this type of content and using 2D and 3D displays showing videos with pre-defined trajectories around the PCs [2]–[6]. Some studies of the QoE have been performed using Head Mounted Displays (HMDs) to visualize the PCs with 6 Degrees-of-Freedom (6DoF), which are more representative of real use cases, using both double [7] and single stimulus methodologies [8], [9]. In these cases, understanding user exploration behavior plays an important role in optimizing the processing pipeline of PCs [10], especially when considering PCs representing humans, since one of the leading applications of this technology is social communication using extended reality (XR) [11], [12].

As with other immersive media technologies [13], further studies are needed to identify and standardize appropriate methodologies for subjective assessment of PCs and to better understand how users interact with them. Therefore, the main objectives of this work were twofold: 1) to explore the validity of a simple and well-established methodology, originally designed for 2D content, for assessing the quality of dynamic PCs with 6DoF; and 2) to analyze users' exploration behavior in this context and its possible influence on the subjective assessment. The following research questions were considered:

- RQ1: Is the traditional Absolute Category Rating (ACR) appropriate for quality assessment of dynamic PCs?
- RQ2: Does the subjective test induce simulator sickness?
- RQ3: Are there differences in the way people explore different PCs?
- RQ4: Are there differences in the way people explore PCs with different qualities?
- RQ5: Does the exploration behavior of the observers change during the test session?

This work was supported by the European Union's HORIZON Research and Innovation Programme through Project "XRECO: XR mEdia eCOsystem" under Grant 101070250 and through Project PID2020-115132RB (SARAOS) funded by MCIN/AEI/10.13039/501100011033 of the Spanish Government.

---

[1] https://sites.google.com/view/icip2023-pcvqa-grand-challenge/


(a) Loot (b) Red&Black (c) Soldier (d) Longdress

Fig. 1: Screenshots of the SRC point clouds.

- RQ6: Are there different types of observers in terms of exploration activity and, if yes, do they rate differently the quality?

In order to answer these questions, a subjective experiment, described in Section II, was performed. In this test, the ACR methodology was used to evaluate the quality of dynamic PCs compressed with different rates using the MPEG standard V-PCC. The observers visualized these PCs using a Head-Mounted Display (HMD) that allowed the tracking of their positions and head movements. The hypotheses made to these research questions are checked with the analysis of the results in section III. Finally, the main conclusion and future objectives are exposed in section IV.

II. SUBJECTIVE EXPERIMENT

A. Stimuli

Four dynamic PCs representing humans, depicted in Fig. 1, were used as source (SRC) contents in this experiment, specified in the MPEG Common Test Conditions [14] and published in [15]. Human PCs were selected given our future interests on studying QoE in social XR scenarios, which include photo-realistic representations of the users. All of them contain 300 frames (at 30 fps) and 1024x1024x1024 (RGB) points. To generate the test stimuli, these PCs were encoded using MPEG V-PCC (TMC2v15) [16] with five rate points (defining the quality of the texture, the geometry, and the precision of the occupancy map), as shown in Table I, and the provided configurations for all-intra encoding described in [14].

TABLE I: V-PCC Rate settings for the test stimuli.

Rate	Geometry QP	Texture QP	Occupancy Map Precision
R01	32	42	4
R02	28	37	4
R03	24	32	4
R04	20	27	4
R05	16	22	2

B. Equipment and Environment

The tests were performed at the *Universidad Politécnica de Madrid* (Spain), in a test room where the observers could move comfortably. Point clouds were visualized using Pico Neo 3, which is an untethered device. An application was developed with Unity3D to reproduce dynamic PCs in a virtual environment based on an empty room with medium gray walls. The PCs were displayed approximately in real (human) size and they were placed at a distance of 1 meter from the starting position of the observer. The rendering shape of the points was a circle of 0.05 units, so discontinuities were not noticeable in the shapes of the PCs from the initial position. Also, the application displayed an interface to rate the quality of the displayed PCs. In addition to these ratings, the head position and rotation data were stored for each participant while visualizing each PC.

Fig. 2: Test session structure.

C. Methodology

The test protocol followed the general guidelines of ITU recommendations for subjective quality assessment experiments [17], [18]. In particular, ACR was used to evaluate the quality of the test PCs, while the Simulator Sickness Questionnaire (SSQ) was used to measure cybersickness [19]. The PCs were shown to the participants for 10 seconds and they could examine them by freely moving around them. Then, participants rated the perceptual quality within the virtual reality environment and started to visualize the following PC after pressing a button to continue. The sequence of PCs shown to each participant was randomized. Concerning the SSQ, the participants were asked to fill it in three different moments during the tests (details in subsection II-D) to assess the evolution of the symptoms along the experiment.

D. Test Session

The structure of the whole test session performed with each participant was divided into seven parts, as depicted in Fig. 2.

First of all, the conditions and procedures of the experiment were explained to the participants. The welcome session also involved a vision test and the signing of the informed consent by the participant for processing his/her data according to the General Data Protection Regulation (GDPR) of the European Union. Afterward, the form with demographic data and SSQ were filled. Subsequently, a training session was conducted to make the participants familiar with the equipment, the interaction area, the rating methodology, etc., and to provide examples of the test stimuli using two dynamic PCs with the lowest and highest quality levels. Then, a first test session, which lasted approximately 10 minutes, was conducted by visualizing and evaluating a first set of dynamic PCs. Once it finished, there was a small break of 5 minutes for the participants to rest and fill again the SSQ. After this break, the rest of the test stimuli were displayed and evaluated. In the end, the observers filled out the last SSQ and provided their feedback about the tests. Finally, they were remunerated for their participation in this study.

E. Observers

Twenty participants (10 women and 10 men), aged 19-29 years (mean of 22.7 and standard deviation of 2.6), took part in the tests. Among them, 47% of the participants were international students. All observers were assessed on (corrected-to-)normal vision. Also, participants were requested to fill out a questionnaire about their experience in using VR headsets.

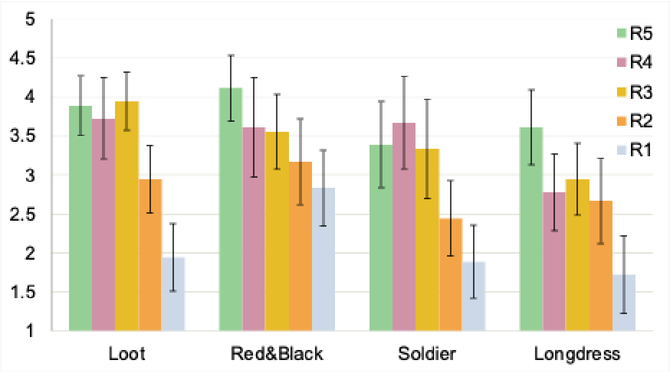


Fig. 3: Quality results.

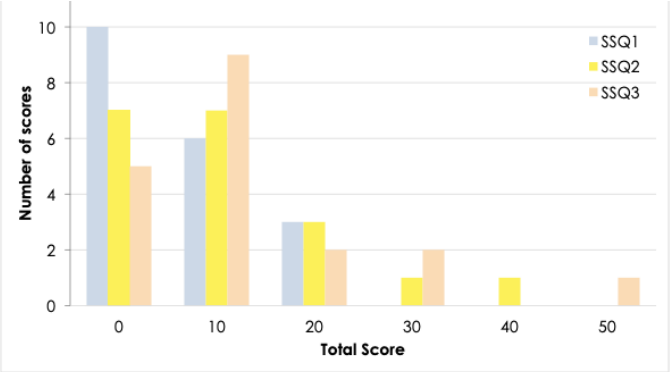


Fig. 4: SSQ results.

According to the results, 74% of the participant were using it for the first time, 10% of them had used it less than 5 times, and 16% had used it more than 20 times. After all tests, one participant's data were discarded due to hardware problems during the session, and the quality ratings from another one were not considered due to errors in data collection.

III. RESULTS

A. Quality

The Mean Opinion Scores (MOSs) obtained from the quality assessments provided by the participants for the test PCs is shown in Fig. 3, together with the 95% confidence intervals. In general, the expected trend of obtaining worse MOSs for more severe compression rates is shown, although similar results were obtained in various cases for R5, R4, and R3. In order to check statistically significant differences among the tested conditions, a Kolmogorov-Smirnov test was first performed, which showed normality of the data. So, paired t-tests were performed with Bonferroni corrections for multiple comparisons. The results from these tests are in accordance with the statistical significance shown by the confidence intervals in Fig. 3. Thus, statistical significance can be assumed for those conditions where these intervals do not overlap. Firstly, it is worth noting that even the best compression rate do not provide MOSs higher than 4, which can be due to the inexperience of the participants in watching this type of content that presents holes and discontinuities that are more visible when getting too close to the PCs. Secondly, compression artifacts have a different impact depending on the SRC point cloud, as shown by low MOSs obtained for R4 and R3 in Longdress, which is a more dynamic PC than, for example, Red&Black. It is worth mentioning that there were a few undesirable freezes with Loot, but they do not seem to have impacted the main results. With respect to RQ1, these results show that, as hypothesized, ACR can be a suitable methodology for the quality assessment of dynamic PCs with compression artifacts, although more test participants may be required to obtain more robust and significant results.

A similar trend of the results can be observed in [8], [9]. Also, in comparison with [4], the MOSs obtained for the corresponding PCs in both tests present a high Pearson correlation (0.818), even though in that test the PCs were visualized in a 2D screen. Obtaining similar assessments using HMDs and 2D screens was also demonstrated in [9].

B. Simulator Sickness

As aforementioned, the SSQ was used to evaluate the simulator sickness and to address RQ2. Our hypothesis to this research question was that, given the structure of the test session (see Fig. 2 and the limited time in which the participants were using the HMD, the simulator sickness symptoms would be mild. Figure 4 shows the histogram distribution of the Total Score (obtained from the ratings of the individual symptoms, according to [19]) for the three times that the participants answered the questionnaire along the whole session (i.e., at the beginning of the test, during the break between the two test sessions, and at the end of the test). Although the results show that simulator sickness may increase along the session, the obtained scores are low enough (in comparison with other validated experiments [13]) to guarantee that the procedure followed in this experiment is appropriate in terms of participants' physiological discomfort.

C. Exploration behavior

Fig. 5 shows the heat maps of the most visited locations (on the floor) by the observers for each SRC point cloud (aggregated for all compression rates). As it can be observed, the participants mainly explored the PCs from a position that allowed them to not only see the front part but also around them. These results address RQ3, to which we hypothesized that the exploration behavior would be similar for the four considered PCs since they are all human representations. The results support this hypothesis since no significant differences can be observed in Fig. 5 on the way people explore the different PCs. Possibly, a slightly higher exploration activity can be observed with Longdress, which would be in line with the findings in [10]. In this study, more dispersion in exploratory movements was found for more dynamic PCs, which is the case of Longdress since it moves forwards and does not stay around a fixed point like the other PCs. In addition, Fig. 6 shows the distribution of the viewing directions in elevation for each SRC point cloud. We focus on the elevation (i.e., pitch) since we observed that the participants mainly looked straight ahead to the PCs with minimal rotation

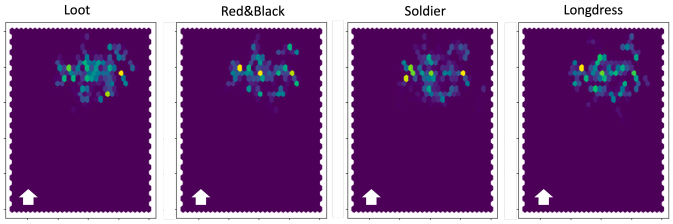


Fig. 5: Heat maps (aggregated per SRC) of the distribution of the observers' position while exploring the PCs (white arrow with the PC's orientation).

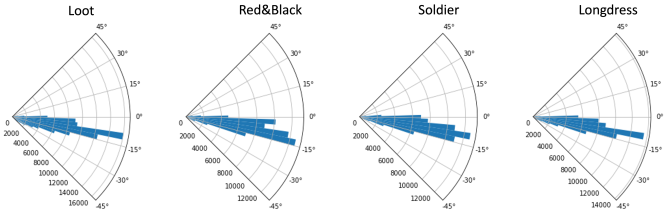


Fig. 6: Distribution of the viewing direction in elevation of the observers while exploring the PC's (aggregated per SRC).

in the yaw axis, which was also observed in [10]. As can be noticed, there are no significant differences among the different PCs, which supports our previous statements. It can be observed that the participants had the tendency to look slightly down, which may be because they probably tend to direct their heads a bit downwards to fit the whole PC in the viewport and be able to spot and notice imperfections in any part of the PC. In general, participants (average height to the HMD of 1.589 meters) watched the PCs at a distance of 4.273 meters, so, looking straight to the PCs at this distance they do not fit in the visible viewport.

Similarly, Fig. 7 and Fig. 8 do not show differences among the exploration behaviors for different compression rates. These results contradict our hypothesis for RQ4, since we expected more activity with PCs in the high-quality range, where artifacts may be less noticeable, so the observers may search more actively to identify them for their assessment. Nevertheless, similar conclusions were obtained in [10].

To answer RQ5 and check whether the exploration behavior of the participants changed throughout the whole session, we analyze the distribution of their positions while watching the PCs (aggregating for all PCs) in the first and the second test session (i.e., before and after the break). In this sense, we hypothesized that, since the observers watched each SRC point cloud several times during the test, they would explore less after visualizing them the first time. The results shown in Fig. 9(a) contradict this hypothesis, since users seem to move more in the second session. This behavior could be explained by the inexperience of most of the participants in visualizing this type of content and in using HMDs, so in the first session users tend to be more cautious in exploring and moving, while in the second session, they start to get used to it and try to experience more. This is also supported by the average distance traveled by the participants in both test sessions, which resulted in 0.656 meters for session 1 and 0.852 meters for session 2. In addition, Fig. 9(b) shows the distribution of the viewing angles in elevation for both sessions. As can be seen, in the first session users looked higher and lower, extending their viewing angles between 0 and -20º, while in the second session, they focused on a narrow range between -5º and -15º. Probably, during the first session, users learned that the best way to evaluate the quality of PCs is to look within a range of viewing angles, so that, as aforementioned, the whole PC falls within the viewport.

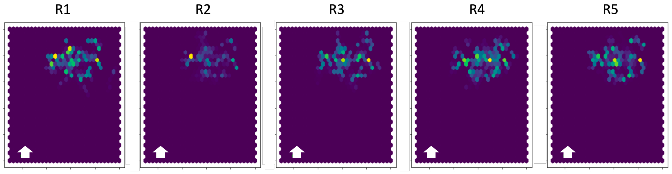


Fig. 7: Heat maps (aggregated per rates) of the distribution of the observers' position while exploring the PCs (white arrow with the PC's orientation).

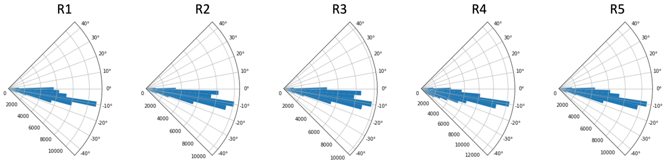
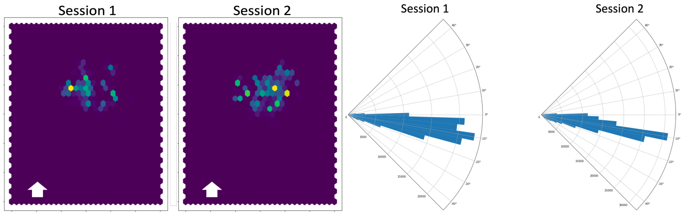


Fig. 8: Distribution of the viewing direction in elevation of the observers while exploring the PC's (aggregated per rates).

Finally, to address RQ6 and analyze if there are different types of users in terms of their exploration behavior, we analyzed their activity by computing the average distance traveled by each participant in both test sessions. The results are shown in Fig. 10. As we hypothesized, some observers tend to move more (e.g., participants 1, 8, etc.), while others stay almost static while observing the PCs (e.g., participants 6, 9, etc.). To investigate if there is any relationship between the way that the participants assessed the quality of the PCs and their exploration activity, Fig. 11 depicts the voting patterns of each observer for all the test PCs. While it can be seen that some participants were more positive (e.g., user 8) or negative (e.g., user 1) with their scores, no clear relationship can be found in this sense between users that explored more and those that moved less. It is worth noting that no outlier removal was applied since given the novelty of quality assessment of immersive media, traditional methods (e.g., recommended in ITU-T BT.500 [17] and ITU-T P.913 [20]) are not suitable and further research is required [13].

IV. CONCLUSION

This paper explored the subjective quality assessment of dynamic PCs with compression artifacts using ACR methodology and analyzed the exploration behavior of users while visualizing them with an HMD. The results showed that ACR can be a valid methodology, but more than 20 observers may be needed for significant results. The analysis of the



(a) Heat maps of the position. (b) Viewing directions in elevation.

Fig. 9: Distributions of the positions and viewing directions of the observers while exploring the PCs in the two test sessions.

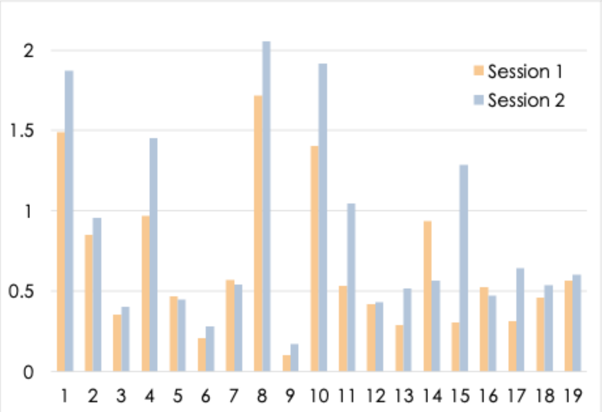


Fig. 10: Average distance in meters traveled by each user while exploring the PCs in the two sessions.

exploration behavior of the users did not show significant differences in exploration activity between PCs of different qualities, changes in behavior over the test session, or correlation between exploration activity and quality assessments. Future work will focus on validating the methodology for the evaluation of transmission errors and on investigating eye-tracking data to further understand how users watch PCs. Also, the resulting datasets and tools will be made publicly available to support the research on this topic.

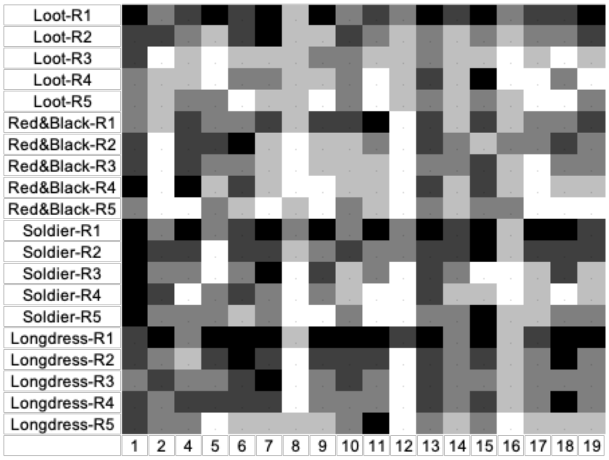


Fig. 11: Diagram of the quality scores provided by each user (black: 1, white: 5).

REFERENCES

[1] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, “An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC),” *APSIPA Trans. on Signal and Information Processing*, vol. 9, Apr. 2020.

[2] E. Alexiou, M. V. Bernardo, L. A. Da Silva Cruz, L. G. Dmitrovic, C. Duarte, E. Dumic, T. Ebrahimi, D. Matkovics, M. Pereira, A. Pinheiro, and A. Skodras, “Point cloud subjective evaluation methodology based on 2D rendering,” in *International Conference on Quality of Multimedia Experience*, May 2018.

[3] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “Point cloud rendering after coding: Impacts on subjective and objective quality,” *IEEE Transactions on Multimedia*, vol. 23, pp. 4049–4064, Nov. 2021.

[4] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, “Textured mesh vs coloured point cloud: A subjective study for volumetric video compression,” in *Int. Conf. on Quality of Multimedia Experience*, May 2020.

[5] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, “Objective and subjective qoe evaluation for adaptive point cloud streaming,” in *International Conference on Quality of Multimedia Experience*, May 2020.

[6] E. Dumic, F. Battisti, M. Carli, and L. A. da Silva Cruz, “Point cloud visualization methods: a study on subjective preferences,” in *European Signal Processing Conference*, Jan. 2021, pp. 595–599.

[7] E. Alexiou, N. Yang, and T. Ebrahimi, “PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality,” in *Int. Conference on Quality of Multimedia Experience*, May 2020.

[8] S. Subramanyam, I. Viola, J. Jansen, E. Alexiou, A. Hanjalic, and P. Cesar, “Subjective qoe evaluation of user-centered adaptive streaming of dynamic point clouds,” in *International Conference on Quality of Multimedia Experience*, Sep. 2022.

[9] I. Viola, S. Subramanyam, J. Li, and P. Cesar, “On the impact of vr assessment on the quality of experience of highly realistic digital humans: A volumetric video case study,” *Quality and User Experience*, vol. 7, no. 1, p. 3, Dec. 2022.

[10] S. Rossi, I. Viola, and P. Cesar, “Behavioural analysis in a 6-DoF VR system: Influence of content, quality and user disposition,” in *Workshop on Interactive EXtended Reality*, Oct. 2022.

[11] R. Mekuria, K. Blom, and P. Cesar, “Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, Mar. 2017.

[12] C. Cortés, J. Gutiérrez, P. Pérez, I. Viola, P. César, and N. Garćıa, “Impact of self-view latency on Quality of Experience: Analysis of natural interaction in XR environments,” in *IEEE International Conference on Image Processing*, Oct. 2022.

[13] J. Gutiérrez, P. Pérez, M. Orduna, A. Singla, C. Cortés, P. Mazumdar, I. Viola, K. Brunnström, F. Battisti, N. Cieplińska, D. Juszka, L. Janowski, M. Leszczuk, A. Adeyemi-Ejeye, Y. Hu, Z. Chen, G. V. Wallendael, P. Lambert, C. Díaz, J. Hedlund, O. Hamsis, S. Fremerey, F. Hofmeyer, A. Raake, P. César, M. Carli, and N. García, “Subjective evaluation of visual quality and simulator sickness of short 360° videos: ITU-T Rec. P.919,” *IEEE Transactions on Multimedia*, vol. 24, pp. 3087–3100, Jul. 2021.

[14] ISO/IEC JTC1/SC29/WG11, “Common test conditions for point cloud compression,” Moving Picture Experts Group Meeting, Output doc. N18474, Geneva, Switzerland, Mar. 2019.

[15] E. d’Eon, T. M. Bob Harrison, and P. A. Chou, “8i Voxelized Full Bodies - A voxelized point cloud dataset,” SO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Input document WG11M40059/WG1M74006, Geneva, Switzerland, Jan. 2017.

[16] MPEG, “MPEG-PCC-TMC2,” https://github.com/MPEGGroup/mpeg-pcc-tmc2, 2022.

[17] ITU-R Recommendation BT.500-14, “Methodology for the Subjective Assessment of the Quality of Television Pictures,” Oct. 2019.

[18] ITU-T, “Subjective test methodologies for 360º video on head-mounted displays,” Recommendation P.919, Oct. 2020.

[19] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness,” *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, Nov. 1993.

[20] 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,” Recommendation P.913, Mar. 2016.