IMPACT OF SELF-VIEW LATENCY ON QUALITY OF EXPERIENCE: ANALYSIS OF NATURAL INTERACTION IN XR ENVIRONMENTS

Carlos Cortés⋆ *Jesús Gutiérrez* ⋆ *Pablo Pérez*† *Irene Viola*‡ *Pablo César*‡ *Narciso García*⋆

⋆ Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, Spain
†Extended Reality Lab, Nokia, Spain
‡Distributed & Interactive Systems Group, Centrum Wiskunde & Informatica, Netherlands

ABSTRACT

The rise of eXtended Reality (XR) has led to multiple ways of including the user's body in interactive experiences. However, the delay limits of self-view rendering in interactive XR remain unexplored. This article presents a minimum self-view latency system and an interactive task-based experiment to study the influence of different levels of self-view delay on Quality of Experience (QoE) and task performance. During the experiment, 23 users tested 8 delay conditions (from 190 to 597 ms) while building block-based models. The results show a hard threshold in terms of involvement and overall quality around 450ms. However, the impact on adaptation and execution time was less pronounced. This indicates that although users adapted to the task in a certain way, their immersion was severely affected above a certain self-view delay value.

Index Terms— eXtended Reality, Self-View Delay, Immersive Interaction, Natural interfaces, Quality of Experience

1. INTRODUCTION

The popularization of immersive technologies has led to an increasing interest in interactive immersive applications. Specifically, eXtended Reality (XR) interaction is a growing field in the area of immersive interaction [1]. Furthermore, XR environments are intended to facilitate training that involves logistical or security complexity. Some examples are industrial [2] and medical training [3, 4]. Increasing the degree of immersion in interactive environments will improve user satisfaction and performance [5]. Thus, immersive systems should be as realistic as possible [6] and their evaluation should go through an analysis of the user experience [5].

The XR paradigm allows user interaction by blending the local and virtual realities through a self-representation of the user within the XR environment. However, how this blending is done can affect realism and, ultimately, break the user's immersion. Some examples of interfaces that can affect the way the body is introduced into the XR are: controllers, haptic gloves or grounded haptics [7]. Other methods introduce the user's body without intermediate elements, for example through image processing algorithms [8, 9]. While these methods maintain the user's immersion, they may add a significant self-view delay to the user embodiment. The self-view delay in an XR environment is defined as the difference between the time of a user's movement and the time when the user sees his/her move into the XR.

As the delay can affect the realism and immersion in the XR environment, it is very important to know its limits for interactive tasks. Consequently, self-view delay has been studied extensively in both non-immersive and immersive scenarios [10–14]. However, all studies to date emphasize non-immersive interaction or rely on very specific tasks.

In this paper, we address the impact of different levels of self-latency on an standardized interactive task while keeping the level of self-representation at a good quality. The rest of the paper is structured as follows. Section 2 describes the concept of the self-view delay in XR environments. Section 3 provides the details of the self-view delay experiment, while Section 4 presents the obtained results. Finally, Section 5 expose the conclusions of the self-view delay experiment.

2. ARTIFICIAL SELF-VIEW DELAY XR ENVIRONMENT

Our study addresses the impact of delay on Quality of Experience (QoE) and user performance in XR environments. This delay is mainly caused by the techniques for mixing local and virtual reality. Therefore, delay is a key factor in developing extended reality interaction methods. In addition, delay can harm both quality of experience and user performance. Specifically, the delay related to the self representation in XR is specially interesting as seeing yourself in the virtual reality is important to preserve the user's immersion [15, 16]. Delay in immersive environments has been addressed by other studies. However, such studies make use of intrusive devices or are based on highly artificial interactions. Moreover, these results of these studies are limited in terms of subjective QoE

This work has been partially supported by project PID2020-115132RB (SARAOS) funded by MCIN/AEI/10.13039/501100011033 of the Spanish Government.

due to the selected delay values [10, 12].

To address the limits of the self-view delay in XR we present an XR environment with minimal self-view delay along with a subjective and objective quality experiment. In addition, we adapted a standardized interaction assessment task to maximize the generalization of our results. The following subsections explain the definition of the self-view delay and the system default delay measurements.

2.1. System self-view delay

In our camera-based XR solution there are two processes that contribute to the self-view delay: capturing and rendering.

Capturing stands for the elapsed time between the instant of the user's movement and the instant the computer has available the information from the camera. This latency is the sum of the time between frames of the camera (τ_{fps}) plus the acquisition time (τ_{acq}).

Rendering represents the elapsed time between the moment when the computer receives the frames from the camera and the moment when the computer displays the processed information to the user. This latency is the sum of the image processing (τ_{proc}) plus the display rendering delay (τ_{disp}). So, the overall self-view delay is defined as:

$$\textit{self-view delay} = \tau_{fps} + \tau_{acq} + \tau_{proc} + \tau_{disp} \tag{1}$$

2.2. System default self-view delay measurement

The scope of this experiment is to measure the impact of the self-view delay in interactive XR environments. In this context, it is necessary to know the intrinsic delay of the system before adding the artificial ones. We used a method inspired in the "numerical latency measurement" [11] to measure the intrinsic delay. Our implementation uses a regular display instead of a hardware one for the reference clock as is illustrated in the Figure 1. The difference between the reference clock and the captured one in the high framerate video gives us an estimation of the intrinsic self-view delay of the XR environment. After 70 measurements, the mean estimated intrinsic delay was $190ms \pm 9ms$. This result is in line with other HTC Vive Pro pass-though delay measurements [1] [2]. We also developed a tunable artificial latency adder by buffering the camera frames before displaying them into the virtual environment. Finally, we used the same numerical latency method together with software to measure the added delay for each buffered frame. After several iterations, we concluded that each frame adds 37 ms.

3. SELF-DELAY EXPERIMENT

To measure the impact on the QoE of the self-view delay in interactive XR environments, we designed a new task inspired

[1] https://stereolabs.com/blog/vive-pro-ar-zed-mini/

[2] https://softserveinc.com/en-us/blog/passthrough-ar-headset-comparison

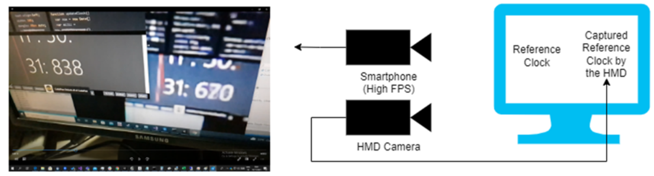
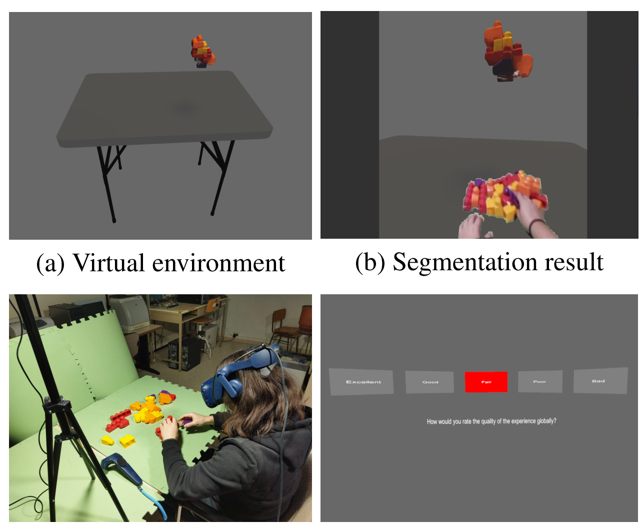


Fig. 1. Example and diagram of the offset latency measurement system.

in a standardized one for measuring the impact of audiovisual degradation in interactive communications [17]. The original task consisted of building a model using Lego-style blocks with the help of another user through a regular teleconference system. An example of implementation of this task in its original form is [18]. However, as we wanted to measure the impact of self-view delay, we developed a single user experience including a realistic 3D model of each designed shape. This model was inserted within the virtual environment so the user could see the 3D model. The Figure 2 shows a user building a model while watching the reference 3D model using their own hands and real blocks.



(a) Virtual environment (b) Segmentation result (c) Local reality setup (d) Votation application

Fig. 2. XR environment setup.

3.1. XR environment setup & apparatus

XR needs coordinated information from the real world embedded in the virtual one. The design of the virtual scenario has been made with the intention of not distracting the user. Thus, virtual scene is composed by a grey room with a simple gray table. Figure 2(a) shows a distant view of the scene.

Building the XR environment requires the integration of the local reality. In our implementation, local reality is cap-

Table 1. Questionnaire used in the experiment.

Factor	Question
Global QoE	How would you rate the quality of the experience globally?
Sens.Haptic	How well could you move or manipulate objects in the virtual environment?
Sickness	Did you feel any sickness or discomfort during the experience? Please rate it
Involv.	How natural did your interactions with the environment seem?
Adapt.	How quickly did you adjust to the virtual environment experience?
Adapt.	How well could you concentrate on the assigned tasks rather than on the mechanisms used to perform them?
Involv.	How compelling was your sense of objects moving through space?
Involv.	How much did your experiences in the virtual environment seem consistent with your real world ones?

tured from the HTC Vive Pro cameras and segmented by the XR engine shaders. System's segmentation uses a chroma-key algorithm like in [19]. These methods for integrating the local reality keep the delay at the minimum while preserving the user's immersion [8, 19]. Figure 2 (b) shows the result of the chroma segmentation during the experiment task.

3.2. Methodology

Before starting the experiment, the user had to sit in front of a table, adjust the headset, and adjust his or her position according to the table. (see Figure 2(c)). The experiment consisted of two regular sessions and a training one. During the training session, the subject got used to the XR environment and the procedure of the task. In addition, the training session included the best and the worst conditions (190 and 597 ms), so the user knew the delay range in advance. This is a common practise in QoE assessment [20, 21]. The 3D model used during the training session was not included in the regular ones. During the regular sessions, each subject had to indicate the beginning of the experience before each construction. After the confirmation, the experimenter started the session. Then, the subject had to complete the building task reproducing the 3D model with a fixed self-view delay value. After that, they had to vote within the virtual environment using the MIRO360 app [22]. Then, the process started again with another model and delay value. After four iterations, the user rested for 10 minutes before starting the second session [23].

3.3. Questionnaire

After finishing each condition (model plus delay), users had to fill a questionnaire evaluating important aspects of the QoE interaction in XR: global quality, involvement, adaption, haptic sensation and simulator sickness. The questionnaire included eight questions (see Table 1). This questionnaire is a subsampling of Presence Questionnaire of Witmer and Singer, which was validated in [24] for interactive immersive environments. All items were evaluated in a Likert-like 5-level scale.

3.4. Stimuli

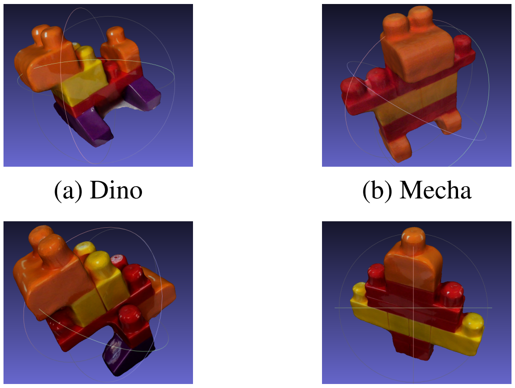
The users had to perform the task eight times with different delay values in two separate sessions. For selecting the delays, we performed a pre-test where 6 users tried from 0 to 12 buffered frames, which means a delay range from 190 to 634 ms. During the pre-test, each user played for approximately 5 minutes in total. For each delay, they had to vote verbally from 1 (bad) to 5 (excellent) how well they felt about interacting with the blocks. We found that the opinion scores fell from 3 to 2 between 375-486 ms. Moreover, these results are in line with the literature about delay impact on the self perception and task-based delay adaption [10, 12, 13, 25]. Consequently, we decided to use more delay stimulus around those values. The selected delays were [190, 264, 338, 375, 412, 449, 523, 597] ms. These eight values were separated in two sets for each regular session. Set A contained [190, 338, 412, 523] ms while set B contained [264, 375, 449, 597] ms. In addition, we balanced the starting set (12 users started with A, and 11 started with B), and the order of each delay value in each set was randomized during the experience. The available set of individual blocks for building each model was the same for all models. During regular sessions, the subjects had to build four different models (see Figure 3). The model shapes were also randomized, except for model Rocket which was assigned to delay values 264 ms and 523 ms to have anchor values in each set.

3.5. Subjects

We conducted a lab trial with 23 subjects (7 female and 16 male; ages between 21 and 34). Each one had to complete the task eight times mixing four models and eight delay values separated in two sessions.

4. RESULTS

During the experiment, we collected the scores for each QoE factor and the task elapsed time. Before aggregating all the scores, we analyzed the influence of the model



(a) Dino (b) Mecha

(c) T-Rex (b) Rocket

Fig. 3. 3D models to reproduce during the task.

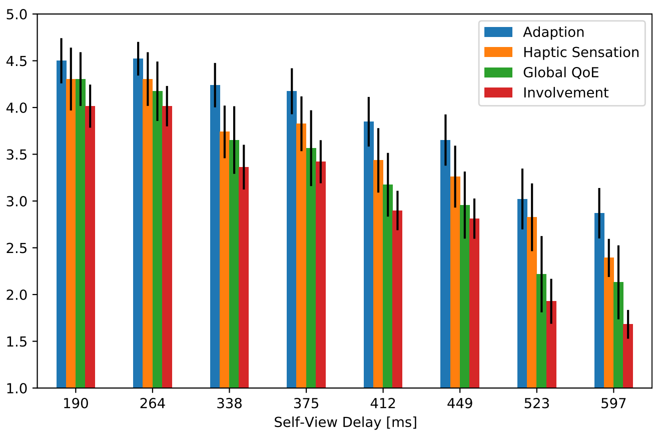


Fig. 4. Mean scores of the different QoE factors per delay.

shape on the subjective scores. Firstly, we ensured that the scores for each factor conformed to a normal distribution ($kurt < |2|, skewness < |2|$ for all factors votations) [26]. The result of the ANOVA allowed us to discard the influence of content (model shape) on voting for all the QoE factors ($\rho > 0.05$). The scores for sickness have not been included in the analysis as the participants did not felt sick at all for any delay value. Consdering the static nature of the XR environment, the absence of cybersickness is in line with the results of previous studies [14, 23]. Taking this into account, Figure 4 shows the Mean Opinion Scores (MOSs) and 95% confidence intervals obtained for the considered QoE factors.

From the results of the $global\ quality\ factor$ (GQOE), we can observe no statistical difference from the reference delay (190 ms) until the 375 ms score. After that, the GQOE maintains at an acceptable level (above 3) until 449 ms. For the 523 and the 597 ms delay the GQOE score decreases until the range of (2, 2.5) what denotes a strong QoE disruption.

$Involvement$ stands for the average score of the three involvement questions in the Table 1. Here, we can observe a similar behavior to the global quality results. However, for the 523 and 597 ms, we can observe that the mean values are even worse for this factor even though the starting score (190 ms) is around 0.5 points lower than the GQOE.

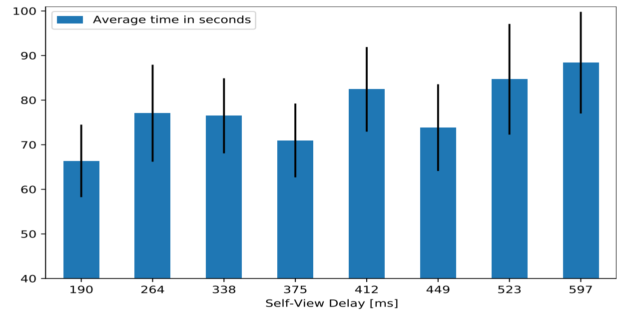


Fig. 5. Average time of accomplishment per delay.

$Adaption$ factor is constructed averaging all the scores of the two adaption questions. The scores follow the same trend that the previous factors. However, for longer delays, adaptation scores remained at acceptable levels (around 3). This idea that the people adapts somehow to the delays is supported by some previous studies [10, 25].

The results for $haptic\ sensation$ follow a similar trend. However, the values reached for the larger delay values are close to acceptable levels (3) as is the case for adaptation.

$Time\ to\ accomplish$ We can observe in Figure 5 that there is a clear upward relationship between experiment performance and added delay. After an ANOVA analysis, we observed a significant influence of the delay on the time to build each model ($\rho < 0.05$). However, Tukey post-hoc analysis indicated that we could only find significant differences for the extreme values (190 vs 597) ms. Moreover, we can observe in Figures 4 and 5 that the impact of delay on involvement or global quality is much more pronounced than for execution time. This is in line with the results of the adaptation factor. That is, people felt less immersed as delay increased, but, in contrast, immersion disruption did not have such effect on their ability to perform and adapt to the task.

5. CONCLUSIONS AND FUTURE WORK

This paper presents a study on how the self-view delay affects the QoE and the performance in interactive immersive XR environments. The task selected for the study is inspired by a widely validated ITU-T interactive task [17]. The results show that there is a threshold around 450 ms for the QoE factors (global QoE, involvement, and haptic sensation) where the QoE falls to levels of non-acceptance. However, the time to accomplish and the adaption factor show that the users can adapt to these delay scenarios.

Future studies should investigate whether these delay values hold in environments where the task requires the user to communicate with others (SocialVR) and/or to move around the environment (6DoF).

6. REFERENCES

[1] “Enterprise XR is Coming of Age - Tech Trends.” [Online]. Available: https://techtrends.tech/tech-trends/enterprise-xr-is-coming- of-age/

[2] F. Cassola and et al., “A novel tool for immersive authoring of experiential learning in virtual reality,” *Proc. IEEE Conf. Virtual Reality and 3D User Interfaces, VRW 2020*, pp. 44–49, 2021.

[3] Y. Wang and et al., “Research on the Application of Medical Teaching Based on XR and VR,” *Proc. 2nd Int. Conf. on Big Data and Informatization Education, ICBDIE 2021*, pp. 688–691, 2021.

[4] J. Sánchez-Margallo and et al., “Application of Mixed Reality in Medical Training and Surgical Planning Focused on Minimally Invasive Surgery,” *Frontiers in Virtual Reality*, vol. 2, 2021.

[5] B. Lee and et al., “Immersive analytics,” *IEEE Comp. Graphics and Appl.*, vol. 39, no. 3, pp. 16–18, 2019.

[6] R. Costa and et al., “Tactful Networking: Humans in the Communication Loop,” *IEEE Trans. Emerging Topics in Computational Intel.*, vol. 5, no. 1, pp. 92–107, 2021.

[7] U. Radhakrishnan and et al., “A systematic review of immersive virtual reality for industrial skills training,” *Behaviour and Information Technology*, vol. 40, no. 12, pp. 1310–1339, 2021.

[8] M. McGill and et al., “A dose of reality: Overcoming usability challenges in vr head-mounted displays,” in *Proc. 33rd Annual ACM Conf. Human Factors in Computing Systems, CHI ’15*, 2015, p. 2143–2152.

[9] E. González-Sosa and et al., “Subjective evaluation of egocentric human segmentation for mixed reality,” in *2021 IEEE Int. Conf. Artificial Intelligence and Virtual Reality (AIVR)*, 2021, pp. 232–236.

[10] K. Brunnström and et al., “Latency impact on Quality of Experience in a virtual reality simulator for remote control of machines,” *Signal Processing: Image Communication*, vol. 89, p. 116005, 2020.

[11] R. Gruen and et al., “Measuring system visual latency through cognitive latency on video see-through ar devices,” in *2020 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 791–799.

[12] P. Caserman and et al., “Effects of End-to-end Latency on User Experience and Performance in Immersive Virtual Reality Applications,” *Lecture Notes in Computer Science*, vol. 11863, pp. 57–69, 2019.

[13] V. Dam and et al., “Effects of prolonged exposure to feedback delay on the qualitative subjective experience of virtual reality,” *PLoS ONE*, vol. 13, no. 10, 2018.

[14] C. Cortes and et al., “Influence of Video Delay on Quality, Presence, and Sickness in Viewport Adaptive Immersive Streaming,” in *2020 12th Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–4.

[15] A. Doumanoglou and et al., “Subjective quality assessment of textured human full-body 3D-reconstructions,” in *2018 Tenth Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–6.

[16] S. Jung and et al., “Over My Hand: Using a Personalized Hand in VR to Improve Object Size Estimation, Body Ownership, and Presence Sungchul,” *Proc. Symposium on Spatial User Interaction - SUI ’18*, pp. 60–68, 2018.

[17] Rec. ITU-T P.920, “Interactive test methods for audiovisual communications,” 2000.

[18] M. Schmitt and et al., “Towards Individual QoE for Multiparty Videoconferencing,” *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1781–1795, 2018.

[19] A. Villegas and et al., “Realistic training in VR using physical manipulation,” *Proc. IEEE Conf. Virtual Reality and 3D User Interfaces, VRW 2020*, pp. 109–118, 2020.

[20] Rec. ITU-T P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications,” 2008.

[21] Rec. ITU-R BT.500-13, “Methodology for the Subjective Assessment of the Quality of Television Pictures,” 2012.

[22] C. Cortes, P. Perez, and N. Garcia, “Unity3D-based app for 360VR subjective quality assessment with customizable questionnaires,” in *IEEE 9th Int. Conf. on Consumer Electronics (ICCE-Berlin)*, 2019, pp. 281–282.

[23] J. Gutiérrez, P. Pérez, M. Orduna, A. Singla, C. Cortés, and et al., “Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T Rec. P.919,” *IEEE Trans. Multimedia (Early Access)*, 2021.

[24] P. Pérez and et al., “Ecological validity through gamification: an experiment with a mixed reality escape room,” in *2021 IEEE Int. Conf. Artificial Intelligence and Virtual Reality (AIVR)*, 2021, pp. 179–183.

[25] A. Farshchiansadegh and et al., “Adaptation to visual feedback delay in a redundant motor task,” *Journal of Neurophysiology*, vol. 113, no. 2, pp. 426–433, 2015.

[26] D. George, *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e*. Pearson Education India, 2011.