https://doi.org/10.2352/EI.2024.36.11.HVEI-237

User performance and Quality of Experience for a remote-controlled lab based moving platform

Kawthar El Ouardi[a,c], Shirin Rafiei[a,b], Chetna Singhal[d], Kjell Brunnström[a,b]
[a]RISE Research Institutes of Sweden, Kista, Sweden,
[b]Mid Sweden University, Sundsvall, Sweden,
[c]Uppsala University, Uppsala, Sweden
[d]Indian Institute of Technology Kharagpur, India

Abstract

In this study, we designed an experiment using remote-controlled lab based moving platform to evaluate the impact of resolution, latency, and field of view on the Quality of Experience, performance, user experience and depth perception. The experiment involves two tasks: driving the platform to a stop point and parking it between two boxes. Participants provided feedback through questionnaires, and their experiences were analyzed. Seven participants between 30 and 57 (average of 37) years old participated. We used Google Forms for data collection, including pre-experiment and recurring questionnaires as well as a simulator sickness questionnaire. Despite the low number of test participants leading to uncertainty in quantitative analysis, significant effects were observed, albeit with contradictory statistical outcomes. The data suggests that lower latency corresponds to better performance, with participants not always perceiving higher latency accurately. Video quality notably impacts user experience, with higher resolution being preferred.

Keywords

Remote controlled lab based moving platform, Teleoperation, Quality of Experience (QoE), User Performance, Depth perception, Resolution, Latency, Field of view.

Introduction

In 2016, the construction industry accounted for about six percent of the world's Gross Domestic Product [1]. Despite being one of the most important industries globally, it is constantly facing challenges mainly related to safety. Therefore, the use of remote-controlled platforms is becoming popular in many countries. Remotely controlled moving platforms can be used in warehouses, mines, or other use cases where remote operations are desirable. Their use is constrained by real-time streaming of intensive multimedia data using wireless networks. It is also constrained by the moving platform, controlled in real-time by a remote operator based on the rendered multimedia information.

In order to study critical perceptual parameters and the Quality of Experience (QoE) [2, 3] for the operator, a lab based remote controlled moving platform (RCMP) with video streaming capabilities was constructed [4, 5].

In comparison to the growing amount of teleoperated tools in various industries, and although it is increasing e.g., for teleoperated excavators, there is still a rather small amount of research done on the multiple kinds of platforms that exist [6]. According to Chen et al, there is an increasing interest in teleoperation although it remains a constraining task, and its limitations should be studied [7]. This research aims to investigate and select a platform and parameters that will help build an experiment. The goal is to study how some parameters affect the QoE and performance of users, by conducting a specifically designed experiment with the RCMP [4, 5], see Figure 1. We aim to test the selected parameters of the platform with test participants. It will be done through conducting a user study followed by quantitative and qualitative analysis. Conclusion will be drawn to determine how those perceptual aspects impact the remote control of the selected moving platform. This conference article is based on the M.Sc. thesis by El Ouardi [8].

This work is similar to [4] in the sense it is an experiment with the same RCMP. It differs in that it is using different tasks and to some extent different test conditions. Otherwise, there are few previous works with RCMPs as a way to investigate crucial parameters for remote control with real machines as they may be hard to get access to and need specialized drivers to operate. However, this type of experiments needs to be followed by investigation with the real machines and in their normal environment. This research can then give valuable input to such experiments.

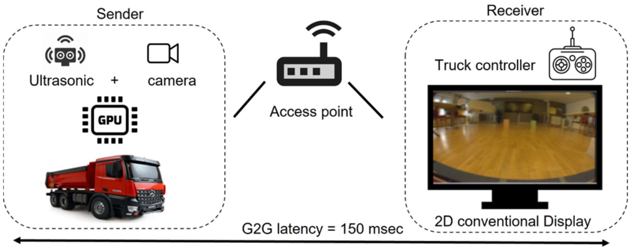

Figure 1: Remote controlled moving platform (RCMP) setup.

METHOD

In this study, we designed an experiment using remote-controlled toy trucks to evaluate the impact of parameters like resolution, latency, and field of view on user experience and depth perception. The experiment involves two tasks: driving the truck to a stop point and parking it between two boxes, see Figures 5 and 7. Participants provided feedback through questionnaires, and their experiences were analyzed.

Apparatus

The hardware and software used for the experiment were the following:

- A toy truck – Amewi Mercedes-Benz Arocs 6x4 Kipper, 2.4 GHz, Ready-to-Run, Red
- A controller – combined with the truck
- One carton box of color white
- Two carton boxes of the same dimension and different colors (pink and green)
- Blue colored tape
- An orange line indicator of distances
- A GoPro camera – GoPro Hero 8 Black
- A Camera – Basler dart daA1600-60um (CS-Mount)
- A GPU – NVIDIA Jetson Nano Developer Kit
- A wireless network with 5 GHz

A RCMP , which can be seen in Figure 1, was selected based on its representation of a real truck. The range of the controller is 50 meters which allowed the participants to be in a separate room while manipulating it. The truck has an autonomy of 20 minutes which was deemed enough for completing half of an experiment, additional batteries were charged while one was being used. An experiment consisted of repeated tasks with twelve different setups. The speed of the truck was also considered appropriate as it was fast enough to keep the participant stimulated but slow enough to have a good perception of the view. The camera mounted on top was charged using the same battery as the one used for the truck during the experiment. It was connected to the NVIDIA Jetson Nano Developer Kit. A monitor was used to transmit the video streaming to the computer. The camera offers a maximum resolution of 1600 by 1200 pixels and a frame rate of 60 fps which can be changed using GStreamer media framework [4].

A large meeting room as shown Figure 2 to the left, was used as the environment for the experiment, which was about 12 square meters (2.4 m by 4.8 m). The tables and chair were cleared away during the experiment. Curtains and blinds were covering the windows, minimizing the influence of the light from the outside. Indoor office lighting was used. The background colors were neutral and the distances were marked using colored tape to preserve the setting through the various days of the experiment. The participants were located approximately ten meters away from the truck. They were in an open space in front of a monitor with no exterior distraction shown in Figure 2 to the right. They saw the video streaming on the monitor and gave their ratings on a separate computer.

The GoPro was used to get another viewpoint to rate the performance of the tasks. GoPro Hero 8 Black has an autonomy of 120 minutes and can be connected to a phone to record videos or stream remotely. The carton boxes were selected and built to have sizes that fit with the truck, see e.g. Figure 5. The first box represents an object that can appear in the field of view while manipulating a truck. The two other boxes are taller and thinner which makes them a good representation of poles. The orange line taped on the floor was used as a distance indicator for the participants, it represents a distance of 50 cm from the truck, see Figure 5 (left). The blue tapes represented the stop points for the truck see the Figures 5 - 7. Finally, clear tape was used to memorize the placement of the objects to keep them similar throughout the days of the experiment.

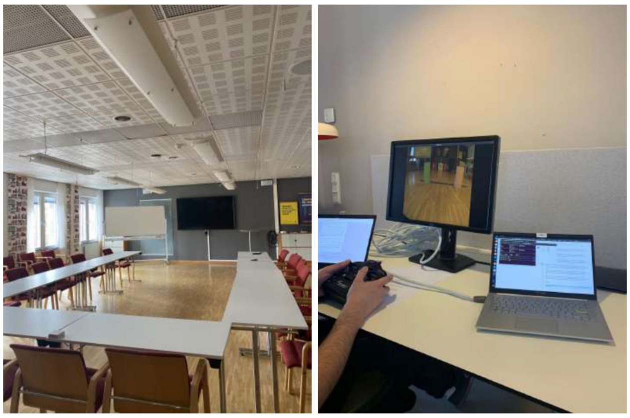

Figure 2: The experiment room to the left (the table and chairs were cleared away during the experiment) and the participants environment to the right.

In the room and in addition to the toy truck, three other boxes and a GoPro were placed. The first box of dimensions 12 cm by 16 cm by 14 cm and color white was placed approximately 150 cm in the front right of the truck. The other boxes of size 10 cm by 9.5 cm by 24.5 cm and colors light pink and light green were placed 300 cm away in front of the truck and 50 cm away from each other. The first stop point was located 130 cm from the starting point, the second one was 300 cm from the starting point. The GoPro was placed above the set-up to offer a view from above and rate the performance.

Participants

Seven participants between 30 and 57 (average of 37) years old, were recruited through the test persons' recruitment site Accindi (https://www.accindi.se/). The majority of the participants were between 30 and 45 years old. The experiment took place over a few weeks. Amongst the participants, 5 out of 7 participants previously had experience with controllers. Three of them had previously experience with toy trucks, and the same number of people were using eye corrections.

Procedure

The first task focused on depth perception, while the second task assessed comfort during remote parking. The parameters were varied to create different setups, see Table 1.

Table 1: Test conditions used in the experiment. W and h in resolution referred to width and height.

Test conditions	Resolution	Glass-to-Glass Latency	Field of view
1	720$_w$ x 640$_h$	150 ms	120 degrees
2	320$_w$ x 240$_h$	150 ms	120 degrees
3	720$_w$ x 640$_h$	150 ms	108 degrees
4	320$_w$ x 240$_h$	150 ms	108 degrees
5	720$_w$ x 640$_h$	500 ms	120 degrees
6	320$_w$ x 240$_h$	500 ms	120 degrees
7	720$_w$ x 640$_h$	500 ms	108 degrees
8	320$_w$ x 240$_h$	500 ms	108 degrees
9	720$_w$ x 640$_h$	850 ms	120 degrees
10	320$_w$ x 240$_h$	850 ms	120 degrees
11	720$_w$ x 640$_h$	850 ms	108 degrees
12	320$_w$ x 240$_h$	850 ms	108 degrees

The first task consisted of driving the toy truck from the start position, see Figure 5, to the beginning of a blue line taped on the floor which was defined as a stop point, see Figure 6. The aim of this task was to assess the depth perception using various boxes. During and before the experiment, the participants answered questions about their confidence in their depth perception, the distances with objects shown on the screen, and how comfortable they felt with the interface. They rated the distances in cm between themselves and the three boxes. When the participants thought they reached the line, they were asked to stop.

The second task consisted of parking between two boxes which represented poles, see Figure 7. Again, a line was taped on the floor to give a stopping point for the participant. The aim of this task was to assess the depth perception, performance, and comfort of the participant while remotely parking the toy truck, see also [4, 9]. During and after the experiment, the participants answered questions about how comfortable they felt with the interface, and how was their overall experience. Their performance (time and accuracy) while performing the task was also recorded. When they thought they were parked correctly/reached the beginning of the second line, the participants were asked to stop and answer the last questionnaire.

Participants' performance was recorded using a GoPro camera. The total time of the experiment was 80 min per participant. The collected data was then analyzed to draw conclusions about the impact of the experimental parameters (Table 1) on the user.

Performance was based on stopping closest to the stop points, and distance from it were calculated using the following scale:

- 1 = Very Below (-30 cm compared to original distance or stop point),
- 2 = Slightly Below (between -30 cm and -10 cm),
- 3 = Accurate (between -10 cm and +10 cm),
- 4 = Slightly Above (between +10 cm and +30 cm),
- 5 = Very Above (+30 cm).

Additionally, the reasons for the difficulty in the task were encoded in the following way:

- D = the perceived Delay,
- V = the Video quality,
- B = Both,
- O = Other.

Before the start of the experiment, the participants read an instruction form and signed a consent form. To retrieve the results from the participants, questionnaires were built using Google Forms. The initial questionnaire was used to retrieve personal data regarding the participant: age, profession, vision problems, and experience with controllers and toy trucks. They also answered a Simulator Sickness Questionnaire (SSQ) [10]. The main questionnaire for the experiment was divided into three parts: before completing the first task, after completing the first task and after completing the second task. For each of these parts, an emphasis was put on the parameter that was the most relevant. For example, after completing the first task, the participants were asked to assess their depth perception as they were the furthest to all the objects. After completing the second task, they were asked how comfortable they felt maneuvering the truck as they had to park between two objects. The questionnaire was repeated for each set of parameters. In the end, the user answered an SSQ again and a feedback questionnaire. The questionnaires were implemented in Google Forms.

The rating scales were based on the five graded absolute category rating scale [11, 12] as shown in Figure 3. On this scale the participants rated: video quality, depth perception, perception of delay, comfort of interface and difficulty. For the perception of delay the degradation category rating scale (see Figure 4) [11, 12], was used.

First part questions

The orange line on the floor on the left indicates a distance of 50 cm from the truck to the end of the line. It serves as an indicator for distances and can be seen in e.g. Figure 7Figure 5.

- How many colorful boxes do you see? _
- How far do you estimate the distance to the white box on the right? (in cm) _
- How far do you estimate the distance to the green box? (in cm)
- How far do you estimate the distance to the pink box? (in cm)
- How confident do you feel in your distance estimations? Ratings given on the scale in Figure 3.

Figure 3: Rating scale used with the categories Bad, Poor, Fair, Good and Excellent.

Second part questions

The questions to be answered after driving the truck to the first line.

- Please rate the video quality. (see Figure 3)
- Please rate your depth perception while performing the task. (see Figure 3)
- Please rate how perceptible is the delay between the video and the movement of the controller. (see Figure 4)

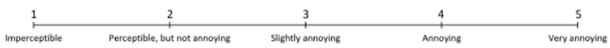

Figure 4: Rating scale used with the categories Imperceptible, Perceptible, but not annoying, Slightly annoying, Annoying and Very annoying.

Last part questions

The questions to be answered after parking the truck and completing all the tasks.

- How comfortable do you experience the interface? (see Figure 3)
- How difficult was it to perform the task? (see Figure 3)
- Would you say the difficulty is related to the video quality or to the delay you experience?
  - Delay
  - Video quality
  - Both
  - Other

  If other, please specify. _
- How would you rate your overall experience? (see Figure 3)

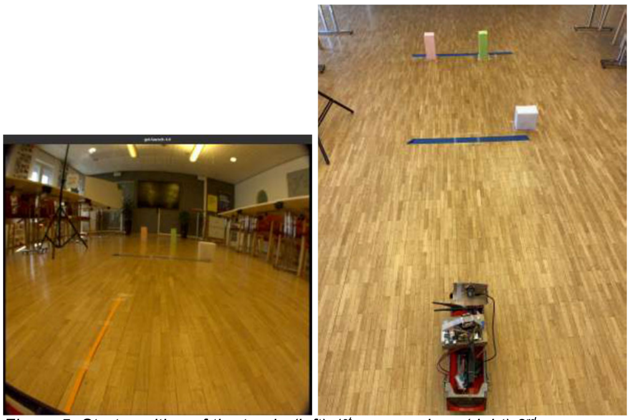

Figure 5: Start position of the truck. (left) 1st person view. (right) 3rd person view

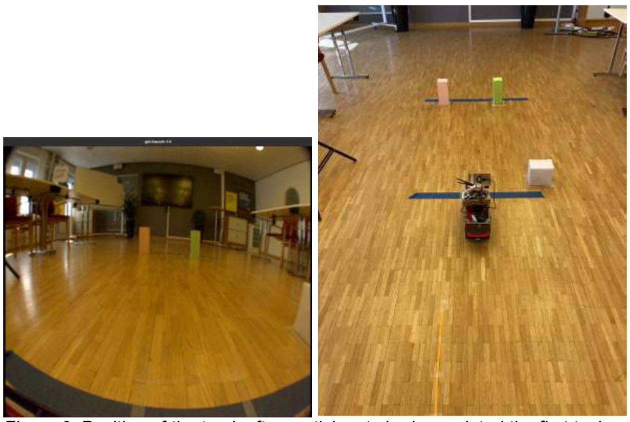

Figure 6: Position of the truck after participants had completed the first task. (left) 1st person view. (right) 3rd person view

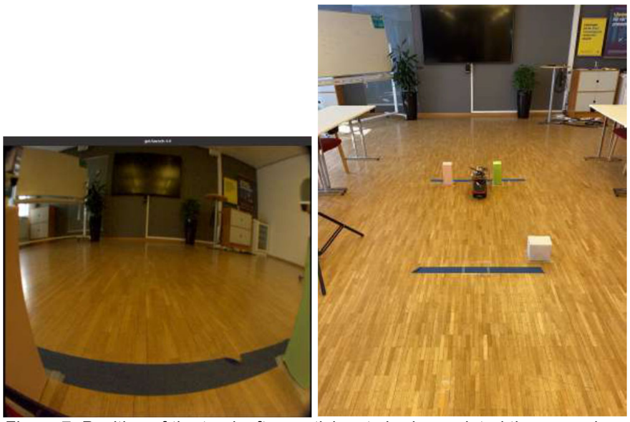

Figure 7: Position of the truck after participants had completed the second task. (left) 1st person view. (right) 3rd person view

Results

In this paper we present part of the gathered results, the full comprehensive description of the results is given in [8]. A couple of significant effects were found, which will be described below.

The results of the depth estimation performance are presented in Figures 8 – 11 In a repeated measures Analysis of Variance (ANOVA), the main effect of field of view was significant for user performance ($F(1, 4) = 9.53$, $p = 0.037$ and partial eta-squared 0.7). A post-hoc test by Tukey Honestly Significance Difference (HSD), shows that there is a significant difference between the performance for the small FOV (108º) combined with the lowest latency (150 ms) compared to the larger FOV (120º) combined with the largest latency (800 ms), see Figure 11.

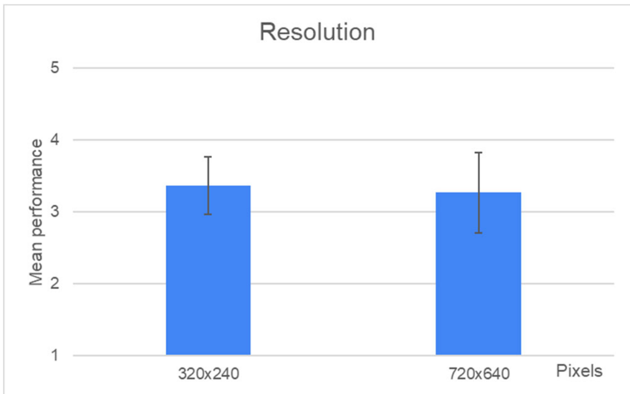

Figure 8: Overall impact on resolution on Video Quality. Close to 3 here means better performance. Error bars represent 95% confidence intervals.

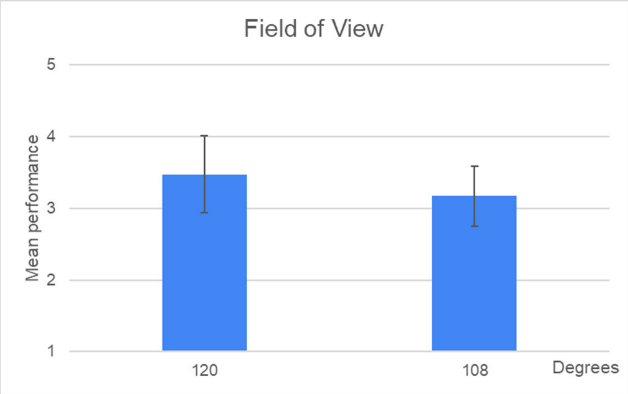

Figure 9: Overall performance impact by Field of View. Close to 3 here means better performance. Error bars represent 95% confidence intervals.

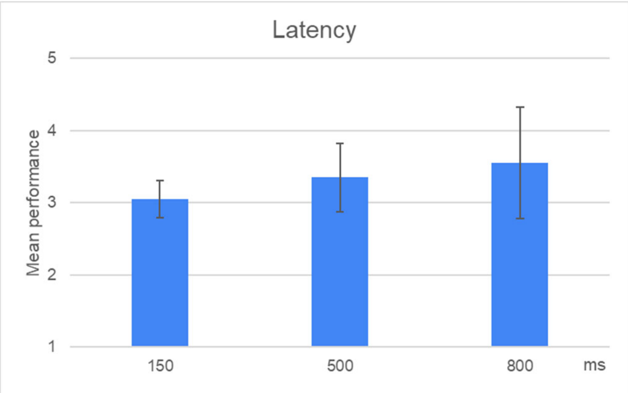

Figure 10: Overall performance impact by Latency. Close to 3 here means better performance. Error bars represent 95% confidence intervals.

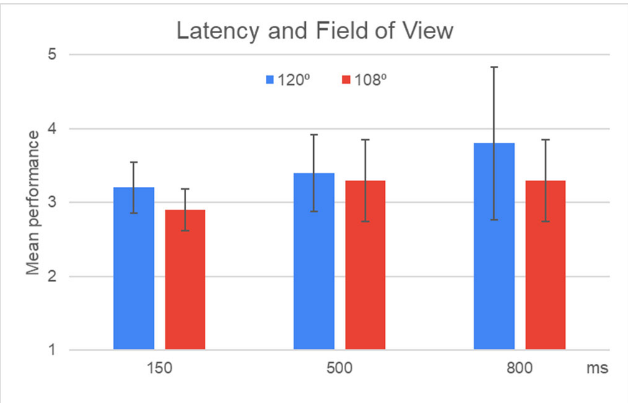

Figure 11: Interaction between Latency and FOV for the performance. Close to 3 here means better performance. There is a significant effect (p = 0.01) between low latency (150 ms) with small FOV and the largest latency (800 ms) and the larger FOV. Error bars represent 95% confidence intervals.

The impact of resolution on the Mean Opinion Scores (MOS) rated for video quality, are shown in Figures 11 and 12. In a repeated measures ANOVA the main effect of resolution was statistically significant ($F(1, 5) = 13.35$, $p = 0.015$ and partial eta-squared 0.7). The results also show that the video quality was mostly experienced as lower than good. There was no significant interaction effect. A post-hoc test by Tukey HSD, shows that there were significant differences between the MOS for all the lower resolutions with all the higher resolutions at all the latency levels ($p < 0.05$), see Figure 12, which is inline with a significant main effect of resolution.

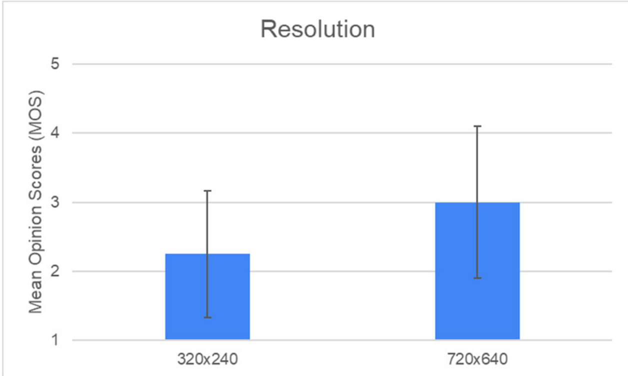

Figure 12: The MOS of video quality for the two resolutions, which was significantly different from each other (p = 0.015). Error bars represent 95% confidence intervals.

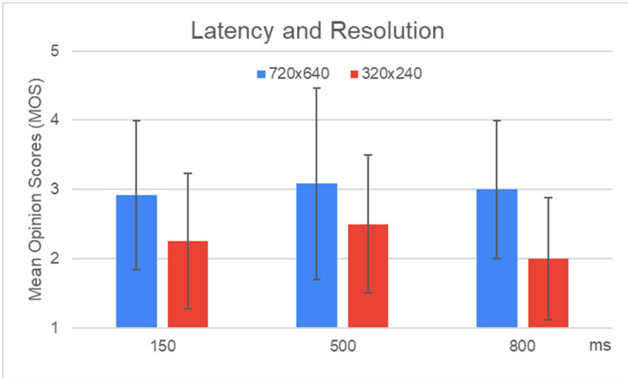

Figure 13: The interaction between the MOS of video quality for the two resolutions and the three latencies. There was no significant interaction effect, but there were significant differences between the MOS for all the lower resolutions with all the higher resolutions at all the latency levels (p < 0.05). Error bars represent 95% confidence intervals.

The test participants were also asked to rate the perception of delay between the movement of the hand controller and the video on an impairment scale: Imperceptible, Perceptible, but not annoying, Slightly annoying, Annoying and Very annoying. The results are shown in Figure 14. There were no significant effects here, but interesting is that the perception was no different for the different latencies at least not on average, which about slightly below three i.e. slightly annoying.

The SSQ response from the participants was obtained before and after each experiment. A difference was noted for some of the parameters (Difficulty concentrating, Difficulty focusing, and Eyestrain) for most of the participants, while one participant felt motion sickness after the experiment. Around a third of the symptoms (Increased salivation, Dizziness, Vertigo, Stomach awareness, Burping) were never felt by any participants before or after the experiment.

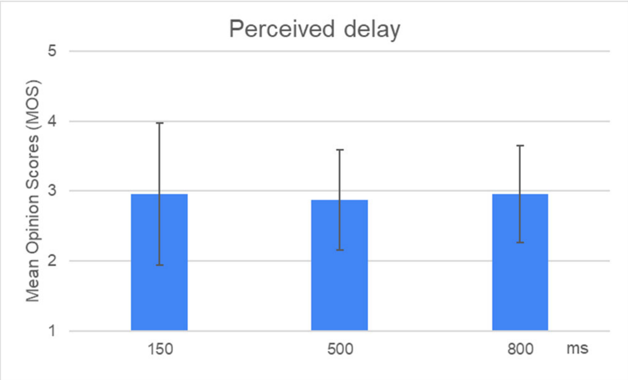

Figure 14: The perception of delay between the hand controller and the video for different latencies, rated on the impairment scale: Imperceptible, Perceptible, but not annoying, Slightly annoying, Annoying and Very annoying. Error bars represent 95% confidence intervals.

Discussion

There is a great uncertainty in the quantitative analysis due to the low number of test participants. A couple of significant effect were found, although the statistical analysis was contradictive as the calculated confidence intervals were not indicating statistical significance, but statistical testing was.

Overall, the performance was slightly better when the latency was the lowest and became gradually worse when the latency was increased, see Figure 10. It is interesting to note that the latency was not necessarily rated as more perceptible by the participants when it was actually higher. In fact, the latency was on average rated the same (2.9 points out of 5) for the three different ones which is shown in Figure 14. It can be mentioned that the standard deviation for the perceived delay was about 1 on average indicating that participants were not always certain of their assertion. In the post questionnaire, the participants mostly wrote "Delay" as the main reason for difficulty when performing the tasks. Although not fully perceived, latency impacts the performance of the user as well as how hard they perceive a task to be.

The impact of the video quality was the most noticeable as all the categories were rated in favor of the better higher resolution (760x640) which is shown in Figure 12. A few participants suggested a higher resolution as an improvement. This is inline with the quantitative results that the MOS for video quality shows that the participants did not think the video quality was good or better.

Other technical issues revolving around network connections were also reported, enforcing the fact that stable communication is necessary to use teleoperation tools. Overall, a better resolution and low latency lead to a better user experience. This enables the users to work with a remote platform for a longer period of time without feeling symptoms such as difficulty concentrating or focusing. The depth perception can also be severely affected by poor video quality although not directly proven through this experiment. Proper training and different angles can be needed depending on the experience of the user with controllers as well as the tasks that need to be performed.

Conclusions

In conclusion, despite the low number of test participants leading to uncertainty in quantitative analysis, significant effects were observed, albeit with contradictory statistical outcomes. The data suggests that lower latency corresponds to better performance, with participants not always perceiving higher latency accurately. Video quality notably impacts user experience, with higher resolution being preferred. Technical issues with network stability underscore the importance of reliable communication in teleoperation. Overall, better resolution and lower latency contribute to improved user experience, potentially mitigating symptoms like difficulty concentrating or focusing. However, the impact on depth perception and the need for tailored training and angles remain areas for further exploration.

Acknowledgement

This research has been funded by Sweden´s Innovation Agency (VINNOVA, dnr. 2021-02107) through the Celtic-Next project IMMINENCE (C2020/2-2), and the VINNOVA project CONTROL (dnr. 2022-02670) as well as the Swedish foundation for strategic research (SSF dnr. FID18-0030).

References

[1]. Gerbert, P., *et al.* (2016). *The Transformative Power of Building Information Modeling: Digital in Engineering and Construction.* Available from: https://www.bcg.com/publications/2016/engineered-products-infrastructure-digital-transformative-power-building-information-modeling, Access Date: 26 Sept 2023.

[2]. Le Callet, P., *et al.* (2012). *Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)* (Version 1.2 (http://www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf)): Lausanne, Switzerland.

[3]. ITU-T. (2017). *Vocabulary for performance, quality of service and quality of experience* (ITU-T Rec. P.10/G.100). International Telecommunication Union (ITU): Place des Nations, CH-1211 Geneva 20.

[4]. Rafiei, S., *et al.* (2023). *Human Interaction in Industrial Tele-Operated Driving: Laboratory Investigation*. in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. 2023.

[5]. Singhal, C., *et al.* (2023). *Real-time Live-Video Streaming in Delay-Critical Application: Remote-Controlled Moving Platform*. in *2023 98th Vehicular Technology Conference: (VTC-Fall)*. Hong Kong: IEEE Xplore; To appear.

[6]. Lee, J.S., *et al.*, (2022). *Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation.* Automation in Construction. **135**, p. 104119. doi:https://doi.org/10.1016/j.autcon.2021.104119.

[7]. Chen, J.Y.C., *et al.*, (2007). *Human Performance Issues and User Interface Design for Teleoperated Robots.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). **37**, p. 1231-1245.

[8]. El Ouardi, K. *Investigating perceptual aspects of real-time video streaming based remote moving platform control*. (Doc nr: 23040), IT, Student thesis. 2023

[9]. Brunnström, K., *et al.*, (2020). *Latency impact on Quality of Experience in a virtual reality simulator for remote control of machines.* Signal Processing: Image Communication. **89**, p. 116005. doi:https://doi.org/10.1016/j.image.2020.116005.

[10]. Kennedy, R.S., *et al.*, (1993). *Simulator Sickness Questionnaire: An Enhanced Method of Quantifying Simulator Sickness.* The International Journal of Aviation Psychology. **3**(3), p. 203-220.
[11]. ITU-T. (2008). *Subjective video quality assessment methods for multimedia applications* (ITU-T Rec. P.910). International Telecommunication Union, Telecommunication standardization sector.
[12]. ITU-R. (2023). *Methodology for the subjective assessment of the quality of television pictures* (ITU-R Rec. BT.500-15). International Telecommunication Union (ITU).

Author Biography

Kawthar El Ouardi *has received a Master of Science degree in Computer Science at Uppsala University (2023). She has since been working as a developer at Apotea.*

Shirin Rafiei *received her B.Sc. and M.Sc. degrees in Electronics and Telecommunications in 2009 and 2014, respectively. Since 2020, she has been a researcher and Ph.D. student at RISE Research Institutes of Sweden AB, and Mid Sweden University. Her research focuses on Quality and User Experience and Augmented Telepresence.*

Chetna Singhal *received her B.Eng. in Electronics and Telecommunications from University of Pune (2008), M.Tech in Computer Technology (2010), and Ph.D. from Indian Institute of Technology (IIT) Delhi (2015). Since then she has worked as a faculty in the Department of Electronics and Communication Engineering, IIT Kharagpur. She has also worked as ERCIM Fellow (visiting researcher) at RI.SE, Stockholm, for a year during 2022-2023. Her work has focused on Multimedia communications and wireless networks.*

Kjell Brunnström *is a Senior Scientist at RISE Research Institutes of Sweden AB and Adjunct Professor at Mid Sweden University. He is leading development for video quality assessment as Co-chair of the Video Quality Experts Group (VQEG). His research interests are in Qualtiy of Experience for visual media especially immersive media. He is area editor of the Elsevier Journal of Signal Processing: Image Communication and has co-authored > 100 peer-reviewed scientific articles including conference papers.*