Advances in Human Factors of Transportation, Vol. 186, 2025, 340–350
https://doi.org/10.54941/ahfe1006523

Operating in the Unknown – The Difference in Remote Operators' Attitudes Based on Their Knowledge of the Task at Hand

Christian Jernberg and Jan Andersson

Swedish National Road and Transport Research Institute (VTI), SE-417 55 Gothenburg, Sweden

ABSTRACT

Remotely operated vehicles are predicted to bridge the gap between conventional and autonomous vehicles. However, there are still issues to address before the technology can be fully implemented on public roads. One being how the introduction of system latency affects remote operators in different driving conditions and potentially hazardous situations. In this paper, an analysis is made of the subjective rating result of four experiments (three simulator and one remotely operated vehicle experiments) where participants acted as remote operators in varied latency conditions. A total of 114 participants in three simulator experiments drove identical scenarios. Each scenario was driven with three latency conditions in experiment one and experiment three (baseline, +100 ms and +200 ms). Two different latency conditions (baseline and +150 ms) were applied in experiment two. The latency conditions were masked for all participants, except for 26 who functioned as a control group in experiment three. Between each latency condition, rating data was collected through questionnaires regarding comfort, perceived control and realism of scenario. After the complete drive, participants were asked supplementary questions about their experience and differences between the latency conditions. In a fourth experiment, 18 participants drove a remotely operated vehicle on a track which included tasks such as line following, slalom driving, reverse parking and precision parking. Three latency conditions (baseline, +140 ms and +340 ms) were masked, and the same questionnaires were used as in the previous experiments. Rating data was compared between the experiments, as well as objective driving data collected by the simulators and the remote operation station. There seems to be a difference in attitude based on previous knowledge about the operators' current driving conditions, though the actual driving performance is similar, i.e., unaware participants adapt their behavior to the same extent as aware participants. Unaware participants contribute any sense of difference to mental alertness, the state of the mechanical controls or a learning effect from recognizing the scenario between conditions. Aware participants contribute their performance to their ability to adapt their behavior based on the level of latency. Participants who noticed the change in latency without being told seem to have an increased sense of frustration, which shows the importance of clear and correct information to the operators. However, as the performance of the participants were deemed 'good enough', the main conclusion is that there is still a need for naturalistic studies concerning the use of remote operation in a real-world scenario.

Keywords: Remote operation, Latency, Simulator experiment, Remotely operated vehicle

INTRODUCTION

Implementation of autonomous vehicles (AVs) have made some progress in the last few years, but achieving full self-driving (SAE level 5) remains a challenge. Due to unpredictable road conditions, such as bad weather, accidents, and sensor failures, a backup system needs to be in place for the foreseeable future. To facilitate for the vehicle to reach a safe state in case of an issue, a remote operator (RO) could intervene via a communication link to either assist the vehicle or take over the dynamic driving task completely.

Remote operation is still in early stages of development though, and ROs will face both technical and non-technical challenges (Beiker, 2021). On the technical side, latency is one of the main issues that needs to be considered (Neumeier et al., 2019). There have been several literary review papers on latency research (for example Zhang (2020), Amador et al. (2022) and Kamtam et al. (2024)) and it does not seem to be any clear guidelines of what levels of latency that should be allowed. Depending on the type of study, latencies of more than 170 ms are considered detrimental to RO performance, while anything up to 250 ms is still tolerable. Some studies show that anything up to 300 ms is manageable for a skilled operator, while others claim that 500 ms is ok for simpler tasks at lower speed. However, anything above 700 ms seems to be deemed impossible to manage in any kind of scenario that includes continuous driving.

There are few studies that explores the reason behind some of the non-technical issues, and some of them seem to focus on legal or organizational issues (Skogsmo et al., 2023) as opposed to human factors issues, such as changes in situational awareness, mental workload and fatigue (Blackett et al., 2022). Blackett et al. (2022) and Neumeier et al. (2019) also argues that the current literature might be too focused on artificial assignments, containing clear tasks and goals that will not correspond to a day-to-day real-life operation.

To examine how a remote driver of a vehicle is affected by latency both objectively and subjectively, a series of experiments were conducted with similar setups (see Method). One of the main research questions was to understand how noticeable a change of latency actually was for a participant. Is the willingness to accept an increase in latency connected to the explicit knowledge about the worsened condition? And if ROs do not notice the change in conditions, are they still affected by it in ways that confirm previous latency studies?

METHOD

The data collected and analyzed in this paper was gathered in four experiments, performed with the help of two different simulators, and a remotely operated vehicle (ROV). The data used is presented in Table 1 below.

The structures of the experiments were similar: the participants performed the dynamic driving task (DDT) on a predetermined course while being exposed to different latency conditions. The current latency was masked for the participants, who at the end of the experiment were told about the change of latency between the laps and asked, amongst other things, if they could determinate in which order they had been driving with the different conditions (except for one group in experiment 3 who were told before). The participants answers were noted in the test protocol and the experiment leader also noted if they were correct, partly correct or completely wrong. After all the experimental conditions and the debriefing about the latency differences, participants were asked once again to rate how much they trusted an autonomous vehicle (same question as before the experiment), and if they thought that they could drive a vehicle remotely with or without passengers.

The questions with free text answers (written down by the experiment leader in the protocols) are the empirical base for the analysis in this paper. However, the results from the ratings performed will be used to contextualize the free text answers, i.e., the free text answers might be more difficult or easier to interpret if all participants were correct in their ratings about latency or if they all were incorrect. The free text answers were given in a specific context, and the context is therefore described below.

The Simulator Experiments

In the simulators, the course consisted of a low-speed *urban* environment and a high-speed *rural* environment, with simulated traffic to increase realism. In each environment, five hazards, such as a child running out from behind a bus, slow-moving bicyclists and a cars entering the ego lane just in front of the participant in three separate cases (starting from parking pocket next to road, not stopping at intersection and making a left turn while meeting in opposing lanes), were introduced to force the participants into some kind of action. For each hazard, objective measurements (such as reaction time, collision and speed variation, etc.) were recorded by the simulator for analysis and publication in separate papers (Jernberg et al. (2024), Jernberg et al. (under review), and Jernberg and Andersson (in preparation) respectively). For a more substantial description of the simulator, the scenario and the objective ratings recorded, see Jernberg et al. (2024). Subjective ratings were gathered with questionnaires before, during and after the drive (see below).

Experiment one and experiment two were performed in the same mid-fidelity simulator (see Figure 1a). Each experiment had 31 participants filling out all questionnaires successfully. Participants belonged to one of two user groups, Experienced Drivers (EDR) or Experienced Gamers (EGA). In experiment one, three latency conditions were used (89 ms, 189 ms, and 289 ms), i.e., the participant drove the course three times. In experiment two, two latency conditions were used (89 ms, and 239 ms) as well as two view conditions ("normal view" with the camera height equal to the viewpoint of a driver in a passenger car, and "roof view", with the camera height equal to the viewpoint of a driver in a truck). Each participant drove the course in each latency condition for both view conditions, i.e., a total of four laps.

Experiment three was performed in an updated simulator with improved hardware and graphics (see Figure 1b). As in experiment one, each of the 52 participants drove the course in three latency conditions (74 ms, 189 ms, and 289 ms). However, in experiment three, 26 participants drove with the current latency masked, while 26 participants drove with the explicit knowledge of exactly how much latency they were operating with.

The Remotely Operated Vehicle Experiment

Experiment four (Jernberg and Zhou (in preparation)) were conducted at the ITRL Laboratory connected to the KTH Royal Institute of Technology, Sweden, and at Arlanda test track, 36 kilometers north of the laboratory. A ROV (see Figure 1c) was placed at the test track while the participants performed the DDT from a remote operation station (see Figure 1d). The course for the ROV consisted of four parts: a circular track where the participants were asked to a) follow a path that alternated between the left and right lane markings of the track, b) a slalom course, c) a simulated loading dock that required reverse driving, and d) a parking space. Measures such as speed, position, steering wheel reversal rate, collisions, and total task time were gathered by the ROV for analysis. Participants drove the course once for each of the three latency conditions (155 ms, 295 ms, and 495 ms) as practice, and then two laps for each latency, i.e., six laps were measured by the ROV in total.

Figure 1: The apparatus used for the different experiments. The simulator used in experiments one and two (a). The simulator used in experiment three (b). The ROV (c) and the remote operation station (d) used in experiment four.

The Questionnaires

During all four experiments, the same questionnaires were used. Table 1 shows the questions asked with the corresponding rating scale, as well as when the question was asked during the experiment. Participants got a form to fill out before the driving session to gather demographic data, such as age, gender and current type of driving license, as well as the *Before the experiment* questions. After each change of conditions (latency and/or view), the participants got to fill out a questionnaire while remaining in the driving seat, i.e., the "*After each condition*" questions. After the completed driving session, the participants were interviewed and debriefed simultaneously, while asked the "*After the experiment*" and "*After the debriefing*" questions by the test leader who noted the answers and asked potential follow up questions. Each question with a rating scale was analyzed and presented in a paper corresponding to respective experiment.

Thematic Analysis

The replies to the open questions (see Table 1) were compiled and sorted by experiment and by latency condition. The replies were printed out and coded in order to find patterns and common themes which were later analyzed.

Table 1: Questions asked before, during and after the participants drove their test laps, with corresponding rating scales where applicable. The last question was only for the control group (experiment three) which knew about the current latency condition during the drive.

Question	Rating Scale	Timing
How much do you trust a vehicle operated by automation or any other non-conventional way?	*Not at all* to *completely* (1-5)	Before/after the experiment
How well do you think you could operate such a vehicle?	*Not at all* to *completely* (1-5)	Before/after the experiment
How comfortable do you experience the driving session?	*Not at all* to *completely* (1-5)	After each condition
How do you experience the control of the vehicle?	*Not at all* to *completely* (1-5)	After each condition
How do realistic were the situations in the scenarios?	*Not at all* to *completely* (1-5)	After each condition
How much simulator sickness do you experience?	*None* to *a very high degree* (1-7)	After each condition
How much did the simulator sickness affect your driving?	*None* to *a very high degree* (1-7)	After each condition
NASA TLX ratings of Mental demands, Physical demands, Temporal demands, Performance, Effort, and Frustration	*Very low* to *Very high* (1-20)	After each condition
What affected you the most during your drive?	*Open question*	After each condition

Continued

Table 1: Continued

Question	Rating Scale	Timing
Did you notice any difference between the different laps?	*Open question to masked group*	After the experiment
How did you experience the drive with regard to the latency?	*Open question to control group*	After the debriefing
In which order do you think you drove with the different latency conditions?	*Open question to masked group*	After the debriefing
How much of a difference was it to drive in the different latency conditions?	*Open question to control group*	After the debriefing
Do you think these limits should be allowed in actual traffic?	*Open question to control group*	After the debriefing

RESULTS AND ANALYSIS

Starting with the simulator experiments, there were two major groups of themes. In the first group, participants did not express any difference in their own performance due to perceived changes in the simulation or their own status, but rather by their own choices. This group, the Notices Not group, contains three main themes: i) Nothing Particular, ii) Learning Effects, and iii) Other Road Users. In the other group, participants noticed that something affected them, but they were unsure of what. This group, the Something Is Different group, contains two main themes: iv) Apparatus Issues and v) Mental State.

In the simulator experiments, these themes exist both in the masked groups as well as in the control group. In the masked groups a total of three participants explicitly mentioned that they felt or believed that there was some kind of delay difference (latency) in the experiment. Out of these, two reported that they were competing in e-racing.

Group One – Notices Not

In all the simulator experiments it was clear that some people felt largely unaffected while driving. Sometimes, however, participants would use the phrase "nothing special this time" after already mentioning other issues during previous laps. This could indicate that some participants tried to find some novel effect for each lap rather than repeating an answer. As it would be reasonable to assume that the shift from a conventional car to a simulated environment would have some effect, the statement most likely shows that the apparatus matches the participants' expectations of how a simulator should act and behave to such a degree that it was not seen as a distraction. Especially when taken into account the number of participants who, when asked "*did you notice any difference between the laps?*", claimed that they did not. The participants could simply expect that a simulator would have some kind of lag and delay as well as an in-built instability and this expectation made the experience equal for each lap.

More commonly though, participants felt that knowing the track increased their performance and affected their driving. As they kept driving, they became more aware of the hazards that could occur, and they learned to look for clues as to where they were. This is of course a limitation of the experimental design. The participants were told to drive as normal as possible, and a few participants explicitly told the test leader that they tried to ignore the previous laps to be able to act as realistic as possible, but that it was difficult. Others claimed that they felt scared from previous laps and kept slowing down before each location that potentially could include some kind of hazard.

> "*I felt very invisible, so during the last lap I hit the brakes everywhere I knew a car could come. Therefore, I was not as affected or frustrated.*"

In the control group (Exp. 3), where participants knew the current latency, some explicitly said that the learning effect affected them more than the added latency or that the knowledge balanced out the worsened driving condition. Others found that it was beneficial to have driven the lap before, as they got the opportunity to prepare themselves and account for the slower reaction of the brake pedal when the latency increased.

These comments show a weakness in most simulator studies with repeated tasks. There still is a demand for more naturalistic remote studies though (Neumeier et al., 2019; Blackett et al., 2022), and a simulator is the only way to keep full control of the experiment while ensuring the participants safety in the automotive industry. Therefore, the loss of novelty effects as the experiment progresses is a necessary compromise.

The Other Road Users theme also highlights the importance of more naturalistic studies. If a driver is supposed to operate a vehicle remotely, it will be necessary for the RO to interact with other road users, and sometimes situations will occur that could be unpredictable and cause bodily or material harm. In these experiments, this was mentioned frequently. The amount of erratic road users was clearly distracting and the child running out from behind a bus caused distress in some participants that would cause them to miss the following event. Several participants in the control group also mentioned that the other road users affected them most during the drive. The bad behavior of the simulated drivers distracted them to the point where the latency was not an issue anymore. As one participant put it:

> "*[I was mostly affected by the fact that] I very quickly realized that the other drivers were lunatics.*"

Even though the ROs are not in physical danger themselves, our mid-fidelity simulator managed to create immersion well enough to leave participants with unease for a period of time after the hazardous events. One could speculate that this feeling would increase in a real-world scenario, for example if a real child ran out in front of a ROV. In all the hazards during our experiments, the participants were never the cause of the situation. However, if there were to be a collision, regardless of who caused it, a RO would be physically detached from the scene in such a way that they would have video feed of the event, but no way to physically interact or help. Interacting with other road users is a field of study that still is in need of more research when it comes to remote operations, and these experiments highlight this further.

Group Two – Something is Different

The state of the operator environment, as well as the state of the operator, could also have a larger effect than the latency itself. As an example, there were several comments about the brakes being too stiff or the steering wheel too sensitive. However, there were also several comments that expressed a sensation of changed difficulties, which could indicate that the latency changes indeed were noticed, but not on an explicit level. Participants instead contributed this either to getting better at controlling the simulator or to the simulator getting worse.

Some participants noticed that it was harder to maintain speed or to steer as the experiment progressed and felt that there was a change in the resistance of the controls. This effect was not consistent with the latency conditions, i.e., it does not seem to be a relation between increased latency and the increase in control difficulties. Some reported that the lap with the maximum latency was their best, both in the masked group as well as in the control group. One participant expands:

> "*I felt rather unaffected by the latency. I forgot that I had latency at +200 which was the second lap, and that lap still felt like the best. Maybe because I was used to the situations, but I wasn't tired and feeling sick yet.*"

As this happened in both the masked group, and the control group, it shows that a latency to up to 289 ms is manageable after a short learning period when the task is simply to drive from A to B, even if unexpected events occur. The medium latency conditions (189 ms or 239 ms) were barely noticeable for the masked groups and deemed as similar to the baseline (89 ms or 74 ms) by the control group.

A more consistent effect though, is that participants with the increased latency conditions at the end of the experiment felt that their own mental state affected them most, namely that they felt that they underperformed due to tiredness. Other participants also expressed that the experiment had a tiring effect, but that it was manageable. Participants with consistently decreasing latency conditions were also more likely to contribute the increased performance to their own ability to learn how to control the vehicle.

The ROV Experiment

Due to the difference in procedure, there were some differences between the ROV experiment and the simulator experiments, but also some similarities. As the participants in experiment four did operate an actual vehicle remotely, there was no point in trying to mask that part of the experiment. Therefore, the expectations were different, and the participants assumed that there was going to be a certain amount of latency. The latency conditions were also different as the lowest baseline that was technically possible to achieve was 155 ms roundtrip. Therefore, the medium latency was comparable to the maximum latency in the simulator experiments (295 ms compared to 289 ms), and the maximum latency was 495 ms. However, even so, there still was a participant who claimed to be unaffected by the latency conditions, and could neither notice a difference between the laps, nor guess in which order the conditions were driven.

The main difference for the thematic analysis though, was the lack of other road users to consider. Otherwise, the main themes still apply to the ROV experiment as well. Participants felt that the steering became more sensitive in the added latency conditions, and they felt more tired and fatigued. However, as most participants noticed that something changed in the maximum latency condition, there was a suspicion that the other laps also contained a difference in latency, except for the participants who ended the experiment in this condition.

The other main difference was that the comments made by each individual participant usually covered more than one of the main themes. It is unclear whether this is due to the novelty factor, i.e., the technology is new and exciting, or whether there were simply more things that affected the participants, making it more difficult to decide which one affected them the most. For example, it was not unusual that participants claimed to be affected most by the stiff pedals, the sensitive steering, a feeling of nausea and the suspicion that this is all due to a change in the simulator.

The third difference was the introduction of an actual physical environment. Some participants assumed that the increased difficulty in controlling the vehicle was due to an increase in wind where the ROV was located.

CONCLUSION

Claims that latencies above 170 ms would be detrimental to operators performance, above 250 ms would be intolerable and that 300 ms possibly could be manageable for a trained operator seems to be considered something of a ground truth (Zhang, 2020; Amador et al., 2022; Kamtam et al., 2024). Our studies show that if the main task is to drive as naturally as possible in a naturalistic environment, it is very common that latencies of up to 289 ms are not explicitly noticeable. Even so, participants rated their ability to control the vehicle significantly lower in the maximum latency condition for experiment one (Jernberg et al., 2024) and three (Jernberg and Andersson, in preparation), but this was not coherent with the objective data recorded by the simulator, i.e., there was not much of a difference in actual performance between neither groups, nor conditions.

Blackett et al. (2022) writes that *"Known performance effects of teleoperations on human operators can include fatigue and high cognitive load"* and *"It is reasonable to assume that the addition of signal latency will further compound the likelihood and detrimental effects of fatigue and* *cognitive workload, since it means that tasks will take longer to complete, not to mention that it probably also creates frustration and uncertainty with the task*" (Blackett et al., 2022, p. 6). Our experiments show that participants that are unaware of any latency changes will not consistently rate their cognitive workload or frustration higher for latencies up to 300 ms, even when they rate their degree of control lower. However, comments affirm that introducing latency could increase the feeling of fatigue. Also, when introducing latencies of 500 ms, preliminary results show a significant increase in self-rated mental demand, temporal demand, effort and a decrease in performance ratings (Jernberg and Zhao, in preparation).

Finally, the masked groups and the control group had similar driving performances, but with different presuppositions. The masked groups only noticed the difference in latency by proxy, i.e., they noticed mental fatigue, worsened vehicle controls, or the sense of getting better by learning the machine and adjusted their driving accordingly. The control group on the other hand found coping strategies and managed to balance the worsened condition by driving more defensively. Neither group suffered more major accidents in the simulator, and both could be considered good enough in performing their driving task. They also responded positively to allowing these levels of latency in real-life situations with some limitations on speed and/or geographical locations.

However, in the ROV experiment, where participants noticed that the latency changed between the laps by themselves, the attitudes were more negative and, in some cases, almost hostile, showing the importance of transparent latency information to the operator.

The main conclusion is that before we can allow for widespread implementation of remote operation, there is still a need for further studies in a naturalistic environment as the effect of latency seems to be task dependent and non-linear.

ACKNOWLEDGMENT

The authors would like to acknowledge Vinnova, who financed these studies in the REDO and REDO2 projects. We would also like to thank Maytheewat Aramrattana and Smith Sapira, VTI, for supporting with the simulation hardware and software making experiment 1–3 possible, as well as Lin Zhao and Mikael Nybacka, KTH, for supporting with the ROV and the remote operation station. Finally, thanks to Caroline Andersers, Johan Egeskog, Johanna Larsson, Kristina Larsson, Magnus Nylin and Gunilla Sörensen for supporting with the data collection.

REFERENCES

Amador, O., Aramrattana, M., & Vinel, A. (2022). A Survey on Remote Operation of Road Vehicles. IEEE Access, 10, 130135–130154. https://doi.org/10.1109/ACCESS.2022.3229168

Beiker, S. (2021). Unsettled Issues in Remote Operation for On-road Driving Automation. SAE International. https://doi.org/10.4271/EPR2021028

Blackett, C., Fernandes, A., Teigen, E., & Thoresen, T. (2022). Effects of Signal Latency on Human Performance in Teleoperations. In T. Ahram & R. Taiar (Eds.), Human Interaction, Emerging Technologies and Future Systems V (Vol. 319, pp. 386–393). Springer International Publishing. https://doi.org/10.1007/978-3-030-85540-6_50

Jernberg, C., Sandin, J., Ziemke, T., & Andersson, J. (2024). The effect of latency, speed and task on remote operation of vehicles. Transportation Research Interdisciplinary Perspectives, 26, 101152. https://doi.org/10.1016/j.trip.2024.101152

Kamtam, S., Lu, Q., Bouali, F., Haas, O., & Birrell, S. (2024). Network Latency in Teleoperation of Connected and Autonomous Vehicles: A Review of Trends, Challenges, and Mitigation Strategies. Sensors, 24, 3957. https://doi.org/10.3390/s24123957

Neumeier, S., Wintersberger, P., Frison, A.-K., Becher, A., Facchi, C., & Riener, A. (2019). Teleoperation: The Holy Grail to Solve Problems of Automated Driving? Sure, but Latency Matters. Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 186–197. https://doi.org/10.1145/3342197.3344534

Skogsmo, I., Andersson, J., Jernberg, C., & Aramrattana, M. (2023). One2Many: Remote operation of multiple vehicles. Statens väg- och transportforskningsinstitut. https://urn.kb.se/resolve?urn=urn: nbn: se: vti: diva-19710

Zhang, T. (2020). Toward Automated Vehicle Teleoperation: Vision, Opportunities, and Challenges. IEEE Internet of Things Journal, 7(12), 11347–11354. IEEE Internet of Things Journal. https://doi.org/10.1109/JIOT.2020.3028766