MSc Interaction Technology
Master Thesis

Exploring Error Recovery Strategies in Teleoperated Systems

Sabin Manuela Kerwien Lopez

Supervisors: Luc Schoot Uiterkamp, dr.ing. Gwenn Englebienne, dr.ir. Douwe Dresscher

March, 2025

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

Teleoperated robots are increasingly employed in environments requiring precision and safety. However, system errors remain inevitable; therefore, appropriate error recovery solutions are needed. This study investigates how different error recovery strategies influence operator performance and behaviour during a teleoperation task. Three strategies were tested: smooth recovery, rapid recovery, and no recovery, using a within-subject design. Performance metrics included hit scores and unclutching behaviour, while qualitative feedback provided insights into user preferences and experiences. Results indicated no significant overall effect of recovery strategy on performance, though post hoc analysis revealed differences between specific strategies. User preferences significantly impacted performance, highlighting the importance of aligning system behaviour with operator expectations. The findings underscore the need for adaptive recovery strategies tailored to user needs, enhancing the effectiveness of teleoperated systems. Future research should explore larger participant samples and incorporate physiological measures to deepen understanding of operator responses to error recovery.

Keywords: teleoperation, error recovery, user performance, human-robot interaction

Chapter 1

Introduction

1.1 The history of teleoperated robots

In 1986, the Chernobyl disaster posed challenges that humans could not safely confront. Robots became essential tools for navigating hazardous environments, showcasing their potential to protect human lives. These teleoperated robots, which are systems remotely controlled by humans to perform tasks in distant environments, have been in use for many years. The Chernobyl disaster was part of a broader evolution that began in the 1960s, when early teleoperation systems were developed for nuclear facilities [1]. In the following decade, the use of these robots expanded to space exploration [30]. They further found their use cases in industrial applications, the military, and healthcare. The shift into more sectors was mainly due to advances in the technology used, which contributed to improvements in these systems [19]. These advancements, such as improved haptic feedback systems, higher-resolution video transmission, and more intuitive user interfaces, have significantly enhanced situational awareness and control precision.

Previous literature has shown the wide-ranging benefits of using teleoperated robotics. Teleoperated systems offer the opportunity to guarantee human safety in environments that might be inhospitable or dangerous for direct human engagement. In some instances, their usage might also lead to more cost-effectiveness. Recent advancements in areas such as robotic manipulators and control algorithms have enabled teleoperated systems to achieve levels of precision that can be difficult for humans to replicate directly in some tasks [32]. Chen et al. put those advantages into context and argue that teleoperation will be of importance as a default mode for systems that work in sensitive environments [4].

While these advancements have significantly expanded the capabilities of teleoperated systems, several key challenges remain that limit their widespread adoption and effectiveness in demanding applications. Particularly, ensuring reliable and efficient control in complex or unpredictable environments. For example, issues such as the cognitive load placed on the operator and communication delay can hinder performance and even lead to critical errors. The increasing demand for teleoperated systems in safety-critical applications like telesurgery or disaster response underscores the urgent need to address these limitations. Recent developments in areas like advanced prediction algorithms offer promising avenues for improvement.

1.2 What are teleoperated systems

This thesis will set its focus on the usage of bilateral haptic teleoperation systems which are in use to tackle the aforementioned societal challenges, as they allow for tasks that require a high degree of precision or sensory awareness [12]. These systems are able to enhance human-machine interaction by acting as a bridge between the two non-co-located physical locations using the virtual domain as a medium. When talking about teleoperated systems it usually refers to five main components. The human operator that interacts with the controller system, the teleoperator or the actuator system that is influenced and influences the environment, and lastly the communication link or network (Figure 1.1). Teleoperation systems provide a bilateral or two-way communication channel for the operator and teleoperator where they are able to perceive and respond to the other components [7]. These teleoperated systems can be operated using a wide variety of control media; these can include joysticks, pedals, or devices that can be hand-held [4].

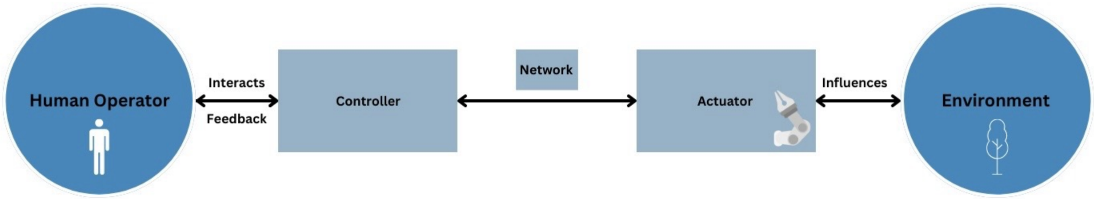


Figure 1.1: Five Components of a teleoperated system.

Through these systems, the user can receive various sensor information about the environment the robot is currently working in [8]. The sensor information of haptics is especially interesting as it includes both kinaesthetic and tactile information. When talking about kinaesthetic feedback, we talk about the force that is applied to the muscles and bones. Tactile feedback, on the other hand, refers to the force applied and felt on the skin. This information is determined by using aspects such as position, velocity, force, torque, and vibration, among other aspects. Thus, an important concept within teleoperation is force feedback. It refers to the transmission of force that is experienced by a remote robot back to the human operator's control interface. This is of special importance as it allows the user to have a sense of being present in the environment of the robot and immerse themselves in this environment. By experiencing these force cues, the operator is able to get a better understanding of the environment they are manipulating and, if needed, adjust their actions accordingly [8]. While bilateral haptic teleoperation offers significant advantages, several challenges remain that can hinder performance and limit its applicability in certain scenarios. These challenges are discussed in the following section

1.3 Current limitations of teleoperated systems

Despite these advancements, teleoperation systems face significant challenges. A particularly crucial area of concern is the occurrence of errors, which can have severe consequences in remote operations. These can arise from a multitude of sources and can be broadly categorized as human-related, environment-related, or system-related errors.

1.3.1 Human-Related Errors

The first category, human-related errors, stems from the operator's cognitive and physical limitations. In that sense, one can think about cognitive overload that occurs when the demands placed on the operator's cognitive resources exceed their capacity. This can manifest in various ways, such as difficulty attending to multiple information streams simultaneously, struggling to make timely decisions under pressure, or experiencing mental fatigue [24]. A second source of errors can be the lack of situational awareness. If the human operators have difficulties perceiving the environment they are operating in or may lack sufficient information to accurately assess the situation, the operator might make incorrect decisions [6]. Other sources of errors are stressful situations that lead to increased error rates, but also training and skill level that can affect the number of errors an operator makes during a teleoperation task.

1.3.2 Environment-Related Errors

The second category is environmental-related errors that are caused by unpredictable events or changes in the environment, thereby disrupting the teleoperation task. Complex environments, characterized by cluttered spaces, irregular terrain, or dynamic obstacles, pose significant challenges for teleoperated systems. Navigating and manipulating objects in such environments requires precise control and accurate perception, which can be difficult to achieve remotely. Lastly, a robot system might also encounter external forces, such as wind, that affect the robot's behaviour in an unpredictable and uncontrollable way.

1.3.3 System-Related Errors

Lastly, system-related errors originate from the teleoperation system itself.

During communication of operator and actuator systems that are geographically distant, the main problems that arise are time delay and packet loss [12]. The delay can range from milliseconds to seconds and is dependent on the distance and communication. It is important to notice that delays above 1800 ms up to seconds are considered excessive and might not even qualify as teleoperation anymore [18]. Humans are able to detect delays from about 10 to 20 ms on [4]. Previous research emphasized that feedback delays when operating with teleoperated systems negatively influence the performance of this system. This becomes apparent in the user's perception and ability to manipulate the environment. These feedback delays can result in the actuator making contact with the environment before the leader is able to display any force [27]. Alternatively, a so-called bounce can also occur that materializes when an operator stops executing force on the actuator as it waits for feedback [29]. The stiffness of the arm decreases, which is then not accurate to the delayed feedback that the operator received. The end result can be described as an unexpectedly high feedback force in the opposite direction of the movement. This pushes on the operator, making them move and in turn the actuator causing a bounce [29].

A study highlighted this further and showed that delay causes significant problems in the transparency of such a system [4]. This can be explained by the decoupling effects that an operator experiences while interacting with a teleoperated system. The effect is experienced as the natural abstract processing is decoupled from the actual physical environment [4]. To counteract this, two approaches were proposed. First, the idea emerged to create robots that are able to fulfil the task independently [18]. Adding automation can help with the challenges that are met such as latency or bandwidth limitations that, in turn, lead to delays [3]. Secondly, the approach to have a system with different levels of control emerged. In that sense, the level of automation added to such a system can be seen as a spectrum, with certain researchers focusing on different types of automation [3]. Basañez et al. [3] explain the spectrum in more detail through the degree of automation on a scale from one, "The computer offers no assistance; human must do it all" to ten, "The computer decides everything and acts autonomously, ignoring the human". The levels used by Ferrel and Sheridan [7], on the other hand, are described as either direct control or supervision by the operator, with sub-goals that are set in the course of completing the task. This latter idea is focused on cooperation between the robot and the human operator. Experiments with this type of control showed that higher degrees of accuracy and reliability could be achieved when operators performed the task in sequences and waited for feedback. While these approaches offer some mitigation of the effects of delay, they do not fundamentally address the core issue of communication latency. In teleoperation, there is no choice in the delay that a system experiences. Thus, a novel approach to directly tackle the problems of delay, instability, and loss of transparency is needed.

1.4 Model-mediated teleoperation (MMT)

The idea that emerged to mitigate these detrimental effects of delays and the negative consequences of the interaction between humans and machines was to use model-mediated teleoperation [32]. Different types of models can be used in model-mediated teleoperation. For this specific research, the focus will be on a two-sided model. Experimental results showed that using such a model has benefits as operators were able to operate the robot in a controlled manner [2].

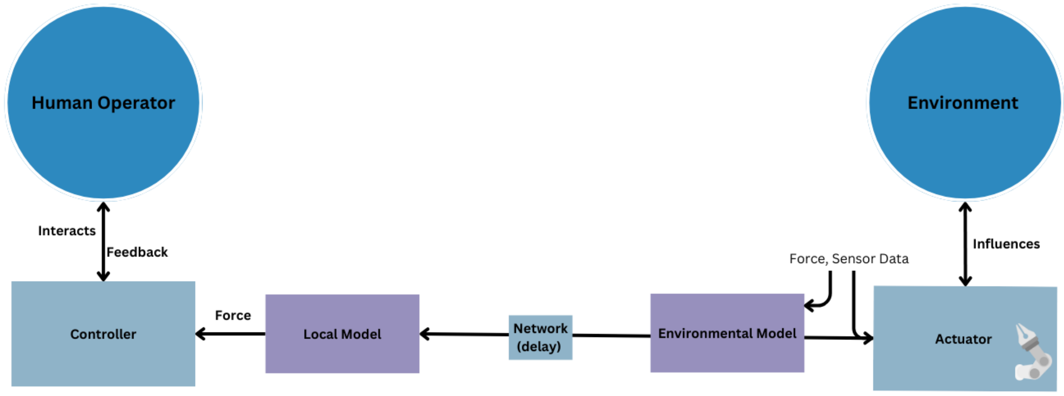


Figure 1.2: A simplified architecture of MMT.

The underlying concept is to predict from one side, either the operator or actuator side, what the other side is going to do. In order to do so, an object model is used on the side of the controller that approximates the environment of the actuator [25]. The model is created using geometric properties, for example by creating a 3D model including obstacles and targets, as well as contact dynamic properties, namely the physical properties of the model such as object mass or friction. This way a virtual world is created that can be used by the operator to interact with the environment. The actuator then follows the commands of the operator and moves in this environment. At the same time, it gathers sensor data such as force, position, and images about the environment to improve the model of the environment. The model is able to provide real-life feedback and is not affected by delays. Estimating this model is more straightforward than the model of the operator since there are many physical properties of the environment that can be used to develop the model.

On the side of the actuator, a second local model is constructed based on certain parameters relevant to modeling the operator; these can include aspects such as affordances, the task, human limitations, and many more. These specific aspects can be rather difficult to model because of the variability and complexity of human behaviour, resulting in errors.

This second model manifests itself in predictions of what the controller is going to do next [25]. A simplified architecture is depicted in Figure 1.2 that highlights the differences to the previously described teleoperated system.

1.4.1 Limitations

Although Model-mediated teleoperation (MMT) is a promising approach that addresses the critical challenges in teleoperated robotics, particularly the one arising from communication delays, MMT also has challenges.

One of the main limitations of MMT is the accuracy of the environmental models. The performance of an MMT system heavily relies on the fidelity of these models, which are used to predict and simulate the remote environment. Real-world environments are often dynamic and complex, involving factors such as moving objects and unforeseen obstacles. Updating the models in real time to reflect changes in the environment is critical for maintaining accuracy but is computationally demanding. This creates a trade-off between the model's complexity and the system's responsiveness. For instance, studies have shown that delays in updating the model can result in a loss of situational awareness for the operator [14]. Moreover, inaccurate or outdated models can lead to misrepresentations of the environment, causing errors in operator decisions and system actions [5].

In addition to modelling the environment, MMT systems also attempt to predict the operator's behaviour and intentions. However, human behaviour is inherently variable and context-dependent, making it difficult to create reliable predictive models. Errors in these operator models can lead to misaligned system responses, confusion, and decreased task performance [33]. For example, predictive errors might cause the robot to act in ways that do not align with the operator's intentions, leading to frustration and reduced trust in the system.

Errors in teleoperation systems, including those employing MMT, are inevitable due to the complexity of real-world environments and the inherent variability in human behaviour [27]. These errors may arise from unexpected situations, inaccuracies in predictive models, or limitations in the communication link between the operator and the robot. Given their inevitability, it is crucial to focus on how to recover from these errors effectively rather than solely attempting to avoid them. Research has shown that errors, when handled correctly, can provide valuable feedback to improve system performance and operator experience [9]. Moreover, focusing on recovery strategies ensures that systems remain functional and reliable even under adverse conditions, which is critical for applications in high-stakes environments such as healthcare, disaster response, and industrial automation. This thesis emphasizes the importance of developing and evaluating robust error recovery strategies to enhance the resilience and usability of teleoperated systems.

1.5 Error Recovery Strategies

Previous research has mainly focused on overcoming modelling errors or errors made by the operator [15]. Research on operator errors has shown the impact such errors can have, as they tend to lead to an increased task load; operators tend to experience frustration, and they overall affect the performance of the teleoperated system negatively [22]. There is, however, a gap in the literature, namely taking the perspective of researching the effect that system errors have on the operator. To expand on this, it is also important to research what the best recovery strategy is to overcome these errors and how they affect the operator and their performance. This is because they are, as of now, inevitable. The research should therefore not focus on how to avoid such errors but more on how to correctly recover from them.

Kontogiannis [9] argues that teleoperated systems, in error detection, can use both a forward and backward strategy that can be employed to re-enter the plan of the system. A backward recovery means that the system will be set back to the state it was in before the error occurred. Forward recovery, on the other hand, means that the system will be brought to an intermediate state to find a more suitable solution later on [9]. Stein et al. [26] further describe a forward recovery as re-entering the system at a later stage, or later task plan. Moreover, there is also the approach of compensatory recovery where the robot will be brought to the intended goal by, for example, using extra resources. Lastly, there is also the option to opt for no recovery strategy and give the operator full control to recover the error. If the error is too great and compensation from the operator is needed, this might be achieved by unclutching, namely virtually disconnecting the input, moving the input to the correct position, and clutching it back in [13]. The decision to choose a certain strategy or combination of strategies depends on the type of error [9]. Literature argues that it is common to move the robot back to the last commanded position. Additionally, the forward recovery strategy has the disadvantage of potentially making the user repeat certain actions that would have happened but were skipped after the error occurred, therefore making the process more time-consuming. It was therefore decided to use backward recovery as the main error recovery strategy that this research would focus on.

Chapter 2

Problem Definition

2.1 Research Question

Teleoperated robotics can work within the spectrum of automation, as previously mentioned systems can be fully controlled, some might be operated under supervised autonomy, and the rest lie in between these two extremes [3]. Depending on the task, the cost, the resources, or other aspects, and the therefore resulting teleoperating system, it can be placed along that spectrum.

In the specific case of this research, the teleoperated robot can be placed somewhere in the middle of this spectrum, as it is not fully automated and also not controlled. This is because it uses predictions to estimate where the operator is going to move next, based on the model of the operator. Those predictions, however, can contain errors and therefore the robot's movement might not coincide with the movement the operator originally intended for the robot. The system might then try to correct or recover the error. It would be of interest to research how people react to these types of errors experienced when operating a teleoperated robot. Thus, there is one research question to be answered:

RQ1: How do different error recovery strategies influence performance?

This question is of interest since errors tend to change over time, and errors will most likely never be eradicated. Hence, it is of interest to see how a system can move from the current error to the next prediction while minimizing the error that happened before. In this study, the following conditions will be explored: 1) an attempt at recovery by moving back to the current known predicted position in a smooth manner, 2) an attempt at recovery by moving back to the current known predicted position in a rapid manner, and lastly 3) no error recovery. For the scope of this research, "performance" is measured by the number of correct buttons that were hit per minute as well as the amount of time users unclutched. Both of these measures will be measured in the time span of an error as well as in a baseline measurement by taking the average over the time span excluding the time span of the three errors.

In human-computer interaction, user preferences are known to significantly influence outcomes. Studies have shown that when interacting with a system whose behaviour aligns with the individual's expectations and choices, users tend to perform better [20]. In that sense, if a particular behaviour matches the personal mental model, this can lead to stronger perceived control and even reduce cognitive workload. This can be explained as such behaviour may be perceived as more intuitive or effective and, in turn, enhance the performance of the user. The idea is, therefore, that in this study, participants will, if they have a preference for a particular recovery strategy, showcase better performance in that strategy. Therefore, the following hypothesis was developed.

> *H1: There is a correlation between recovery strategy preference and baseline performance.*

Zheng and Daneshmend [34] suggest that in teleoperated systems, removing the dependency of an autonomous system on a human operator should be achieved, as it allows the human to solely focus on the task instead of additionally recovering from errors. Based on this idea, the following hypothesis was developed.

> *H2: Participants will showcase a better performance if there is an active recovery strategy*

In this specific case, an active recovery refers to the recovery strategies smooth and rapid, where the system attempts a recovery. The third Recovery strategy, namely no strategy, is seen as a passive recovery strategy since the operator has to recover instead of the system doing it.

Chapter 3

Methods

3.1 Participants

A convenience sample was used, where participants were primarily recruited within the circle of acquaintances of the researchers. Participants needed to be over the age of 18 but did not need any other previous requirements. Apart from being asked for their demographic information, they were also asked about their previous knowledge of teleoperated robotics.

A total of 21 participants participated in the study, with eleven female and ten male participants. The age ranged from 19 to 55 years. Nineteen participants indicated to be on a beginner level with regards to teleoperated robots while two indicated that they were on a novice level. In the end, two of the participants' data recordings were faulty and could therefore not be used. Thus, the quantitative data analysis will be based on 19 participants.

3.2 Materials and Measures

3.2.1 Informed Consent

Participants of the study were asked to read and actively sign the informed consent. The informed consent contained information regarding the study, the use of information gathered in the study, as well as contact information in case questions arose.

3.2.2 Quantitative Measurements

A closer look was taken at unclutching to see how participants react after an error occurs. Moreover, how many correct buttons were pressed during the whack-a-mole task was observed.

3.2.3 Interview

Lastly, an interview was conducted with the participants once the experiment was concluded. Here, questions about the interaction of the user with the robot and the experience were explored further. The questions are meant to allow the researcher to engage in a semi-structured interview and to follow a certain scheme that dives into the topics of the user experience and attitude, impact on task, error understanding, and error recovery. The questions were developed based on the study done by Weiss et al. [31]. The authors propose four methodological points of view, namely the user experience while interacting with a robot, the perceived usability, and attitude towards such a robot. Mirnig et al. [17] propose to start with introductory questions and then move on to questions regarding the perceived error in the interaction. Thus, the questions about error understanding and recovery make up the last part of the interview.

3.3 Procedure

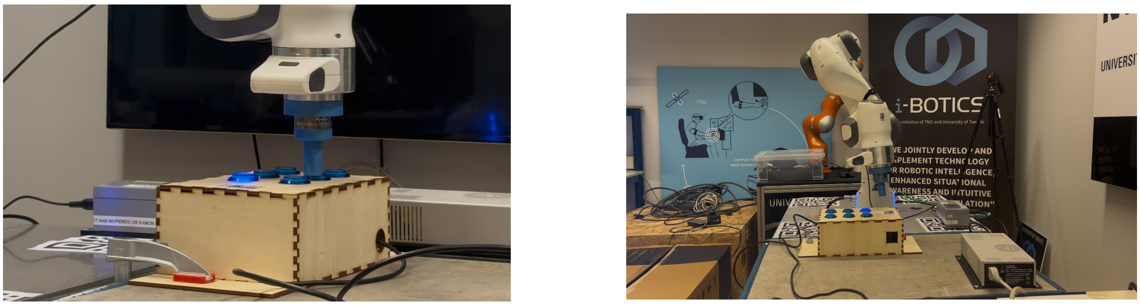
The general setup of the experiment uses the Franka Emika Panda arm as well as the corresponding code to activate and manage the robot.

Figure 3.1: Set up of the experiment.

The robot can be controlled remotely by using a Force Dimension Omega 7. The Omega 7 system can provide haptic feedback because it is a 7 DOF control unit. A complete technical setup is described in the Appendix (Appendix B). Participants were

seated in front of the robot and could look directly at the whack-a-mole game. They were then asked to complete a whole game of the whack-a-mole task.
The whack-a-mole task consists of targets (buttons) that light up and need to be hit by the participant to score. In this experiment, the buttons change every four seconds if the participant is too slow and does not hit it in time. The game requires quick reactions and precise movements, which are key aspects of teleoperation tasks. Moreover, the random appearance of moles mimics unpredictable scenarios in teleoperation, such as sudden environmental changes. The simplicity of the game mechanics (hit the mole when it appears) ensures that participants can focus on the teleoperation process and the effects of errors and recovery strategies without being distracted by complex task requirements. It therefore is a fitting context to evaluate the research question in.

Participants first got some minutes to get used to the robot and to test the functions out; once they felt comfortable, they let the researcher know. The researcher started a recording of all the data. Additionally, a timer of five minutes was set and the participants were instructed to try and hit as many moles as possible (Figure 3.2).



Participant hitting mole

Robot executing error

Figure 3.2: Game play

Then the robot was manipulated by introducing an artificial position error, by doing a nudge once the operator pressed a key (Figure 3.2). The robot, thus, moves away from the path the user is following in addition to the user input. During those five minutes, there were three different conditions the participant was in.

1. **Condition "Smooth":** Error strategy where the robot moves back to the current known position in a smooth manner.

2. **Condition "Rapid":** The robot moves back to the current known position in a rapid manner.

3. **Condition "No":** The robot does not intervene and maintains the error position and requires the user to compensate. Here, no repair is attempted.

In the smooth recovery strategy, the robot arm moved with a speed of 0.1 m/s, whereas in the rapid recovery strategy, the robot moved with a speed of 0.5 m/s. The maximum distance the robot could move in each recovery condition was 0.1 m, after which it would automatically stop.

These conditions were randomly introduced to the user; thus, every participant experienced all three conditions and recovery strategies, making it a within-participant set-up.

Once the five minutes were over, the researcher started a voice recording and the interview. The interview was concluded with a debrief during which the research aim of the experiment was explained. The participants were informed that during their game an error and three different error recovery strategies were artificially added by the researcher. Participants were asked for further input on these errors and recovery strategies.

3.4 Data Analysis

In order to analyse the data and to answer the research question and hypothesis, statistical tests were chosen. First, both a Shapiro-Wilk as well as a Kolmogorov-Smirnov Test were used to test for normality of the data. Based on the results of this test, either a repeated measures ANOVA or a Friedman's Test was conducted to research the effect of the error recovery strategies on user performance. These tests were run a second time using normalized data to account for the varying baseline. In order to do so, an average of the baseline was taken and subtracted from each score. This was followed up with a Post Hoc analysis where pairwise comparisons using multiple Bonferroni corrections were run.

To test the first hypothesis, namely that participants with a strong preference for a strategy will perform better, a third ANOVA or Friedman's Test was run. To understand better which conditions differ, paired t-tests, with a Bonferroni correction, were run. Participants were divided into participants with a strong preference, weak or divided preference, and no preference.

To research individual differences, two-tailed paired sample t-tests were run, as well as the nonparametric alternative, the Wilcoxon Signed Rank Test. These tests were also used to answer the second research question, namely that participants will perform better if there is an active recovery strategy.

Chapter 4

Results

This section presents the findings of the study, that focused on the three error recovery strategies: 1) a smooth recovery strategy, 2) a rapid recovery strategy, and 3) no error recovery strategy. Key outcomes related to task performance and user reactions are reported. Both quantitative and qualitative results are included to provide a comprehensive understanding.

4.1 Hit Scores

4.1.1 Normality Test

Results of the Shapiro-Wilk and a Kolmogorov-Smirnov Test suggest that the data is normally distributed (**W=0.96, p=0.64** and **D=0.10, p=0.97**). Lastly, the results of the Quantile-Quantile (Q-Q) Plot are shown in Figure 4.1, visualizing the distribution of the data. Based on the results, parametric statistical tests were used.

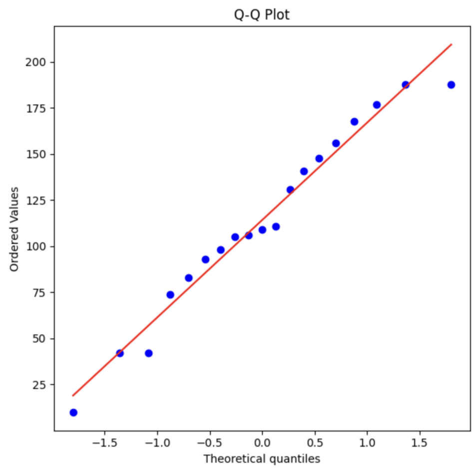


Figure 4.1: Q-Q Plot.

4.1.2 Descriptive Statistics

Overall, the highest score of hits was 188 total hits or 37 hits per minute, while the lowest score was a total of 15 hits or 3 hits per minute. On average, the no recovery strategy showed the highest number of hits per minute m_{hits}=3.9, followed by the no smooth condition m_{hits}=3.7, and lastly the rapid recovery strategy m_{hits}=2.7. Figure 4.2 shows that the no recovery condition has the highest variability in the data, whereas the rapid condition shows the lowest variability.

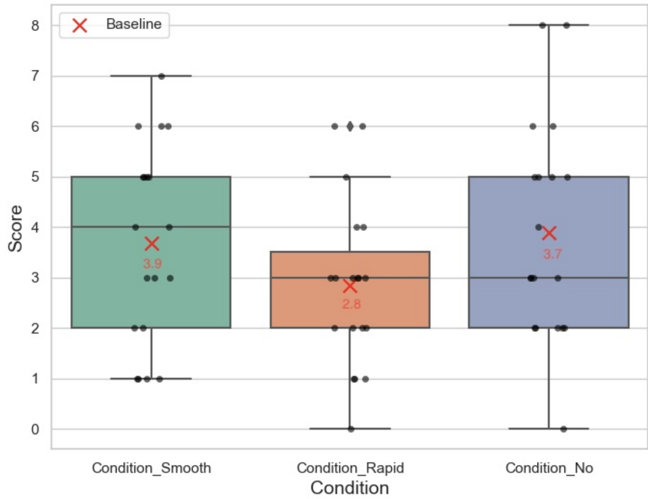


Figure 4.2: Boxplot for Hits across conditions with the baseline values.

4.1.3 Influence of recovery strategies on hit scores

ANOVA

The repeated measures ANOVA revealed a non-significant main effect of the condition, the recovery strategy, on user performance, **F=2.50, p=0.121**. Thus, the different conditions did not have a detectable impact on the dependent variable, the performance of the participants. To account for a varying baseline a second repeated measures ANOVA was run using normalized scores. The results of this second repeated measure ANOVA showed no significant result **F= 2.5, p= 0.09**. A significant effect was found on individual differences among participants with regards to their results, **F = 2.62, p = 0.007**. Hence, there is a significant variability between the participants.

Post Hoc Analysis

Next, t-tests with Bonferroni corrections were run. Three paired sample t-tests were conducted to compare the hits in the three conditions. There was a significant difference between the smooth recovery strategy (m_{hits}=3.7, SD=1.9) and the rapid one (m_{hits} =2.8, SD=1.6), **t=2.8,p=0.01**. A significant difference can also be seen between recovery strategies two and three (m_{hits}=3.9, SD=2.2), **t=-2.1, p= 0.04**. No significant difference was found between recovery strategies one and three, **t=-0.33, p=0.75**. As shown in Figure 4.2, the rapid recovery strategy showed overall the lowest scores with the lowest distribution, while both the smooth and the no recovery conditions show a fairly similar distribution.

4.1.4 Performance Differences Based on Recovery Strategy Preference

The first hypothesis stated that there is a correlation between recovery strategy preference and baseline performance.

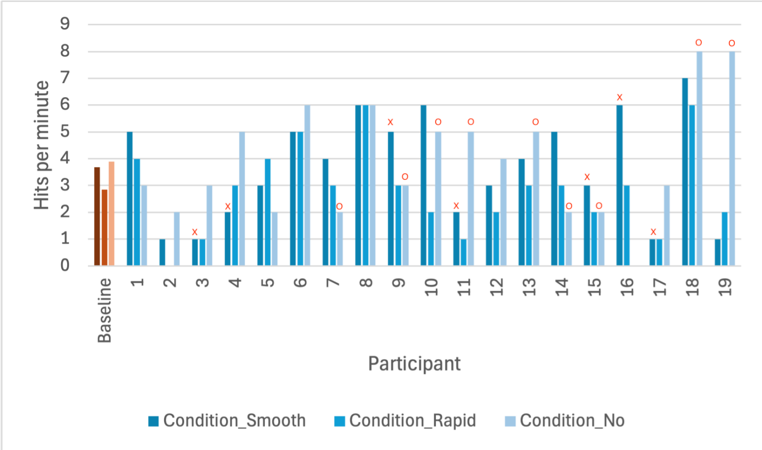


Figure 4.3: Bargraph for Hits across conditions with preferences for a strategy .

Figure 4.3 illustrates the distribution of hit scores for each participant across three recovery strategy conditions. Each bar represents the number of hits per participant under each condition, as well as participants' stated preferences, with marks above the preferred condition. While some participants performed best under their preferred condition (e.g., Participant 16 preferred Smooth and scored highest there), others demonstrated incongruence between preference and performance (e.g., Participant 7 preferred "No" but performed better in the "Smooth" condition). Several participants, such as Participant 9 and Participant 15, exhibited a divided preference between two strategies.

To test this first hypothesis and that if participants have a strong preference they will showcase a better performance in that preferred condition, an ANOVA was run. The results were significant **F=10.31, p=0.005** showing that there is a significant difference; at least one of the strategies leads to different performance compared to the others. Results of the follow-up t-tests show that there was a significant result (**p = 0.0152**), for the performance of the participant under the condition with a stronger preference (C_smooth) compared to the weaker preference condition (C_rapid). There was also a significant difference between weaker (C_rapid) and no preference (C_no)(**p = 0.0065**), suggesting that participants may perform worse in conditions of equal/no preference (C_rapid) compared to no preference (C_no). Lastly, there was no significant difference (**p = 0.4054**), between having a strong preference or no preference.

4.1.5 Individual differences

A closer look can be taken at the scores of the participants across the conditions. The line graph in Figure 4.4 shows the scores of each participant in each condition.

It becomes apparent that some participants score higher in certain conditions and lower in others. From the interviews, it became apparent that some participants had a strong preference for either of the conditions. In that sense, participants expressed that "the reposition yourself (no recovery condition) was the worst," while others "did not like it (the robot) to do anything that I am not doing" and therefore expressed a preference for no recovery strategy.

Overall, participants who expressed a strong preference for the smooth recovery condition showed a lower average score of hits per minute during their baseline recordings (m_{hits}=17.8), whereas participants with a preference for no recovery condition had a higher overall score (m_{hits}=22.4). The results of the t-test showed that for participants who had

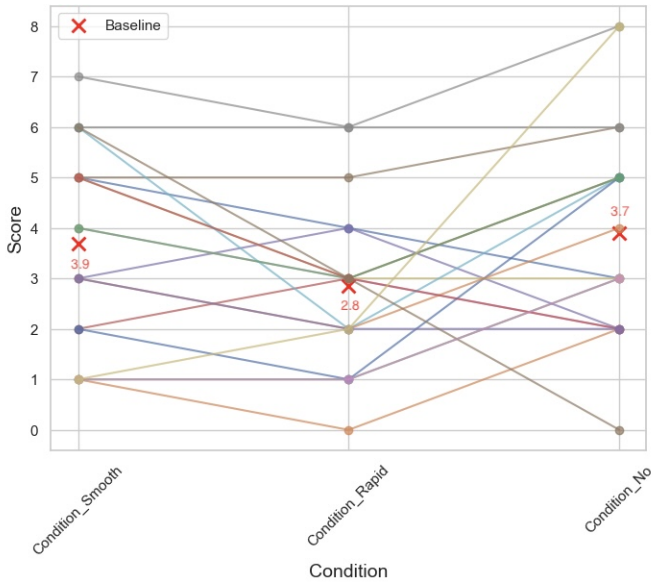


Figure 4.4: Linegraph for Hits across conditions with the baseline values.

a strong preference for the smooth recovery strategy, there was no significant difference between the scores in this condition in comparison to the baseline, $t_{18} = 0.58, p = 0.604$. For participants that chose no recovery strategy as their preference, a significant result could be seen $t_{18} = 2.14, p = 0.049$; meaning that there was a significant difference between the scores in their preferred condition and in the baseline condition.

To test if the baseline performance of participants with a preference for the smooth strategy or no strategy differed statistically, a second t-test was run. The results show that there is a significant difference between the groups $t = 2.718, p = 0.024$. The positive result of the t-test suggests that the group preferring the no recovery strategy has a significantly higher hit score than participants that prefer the smooth condition.

4.1.6 Active versus passive recovery strategy

To answer one of the hypotheses, namely that participants perform better if there is a recovery strategy over no recovery strategy, a paired t-test was conducted. Results show that there is no significant difference between these two groups, **t=1.14, p=0.27**.

4.2 Unclutching behaviour

4.2.1 Normality

The results of the Shapiro-Wilk and Kolmogorov-Smirnov Test suggest that the data is not normally distributed (**W=0.877, p=0.02 and D=0.200, p=0.35**). Lastly, the results of the Histogram as well as the Quantile-Quantile (Q-Q) Plot are shown in Figure 4.5, visualising the distribution of the data. Based on the results, researchers decided to continue using nonparametric tests.

4.2.2 Descriptive Statistics

Overall, both recovery strategies 1 and 2 showed the lowest medians with a Mdn=2. The rapid condition had a median of Mdn=3, as well as the lowest variability.

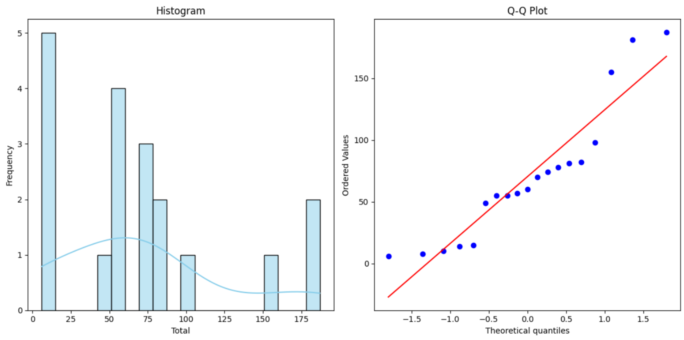


Figure 4.5: Histogram and Q-Q Plot.

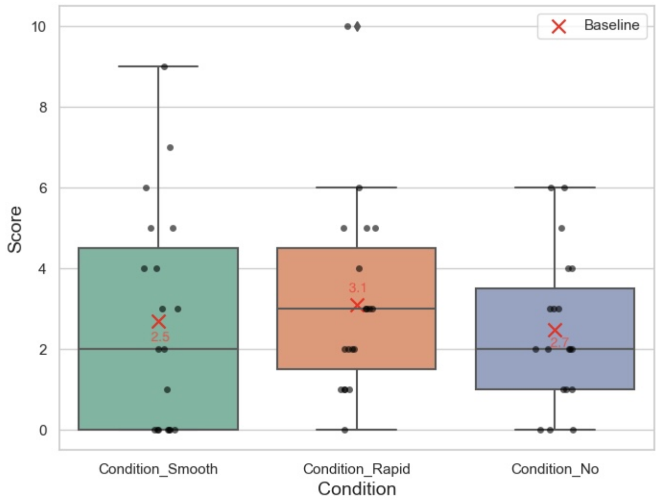


Figure 4.6: Boxplot of unclutching across conditions with the baseline values.

4.2.3 Influence of recovery strategies on unlcutching

Friedman's Test

The result of the Friedman's Test, **F=3.39, p=0.18** indicates that there is no significant difference in the number of unclutching between the three recovery strategies. Next, to account for a varying baseline, a second Friedman's Test was run using normalized scores. The results of this statistical test showed no significant results **F=0.13, p=0.08**.

Post Hoc Analysis

The results showed a significant difference between conditions "smooth" and "rapid" (p=0.037, Bonferroni corrected). There was no significant difference between conditions "smooth" and "no recovery strategy" (**p=1.00**) and conditions "rapid" and "no recovery strategy" (**p=0.129**).

4.2.4 Performance Differences Based on Recovery Strategy Preference

The first hypothesis states that there is a correlation between recovery strategy preference and baseline performance.

This bar chart in Figure 4.7 illustrates the unclutching scores of each participant under the three recovery strategy conditions. There is variability in unclutching behaviour across

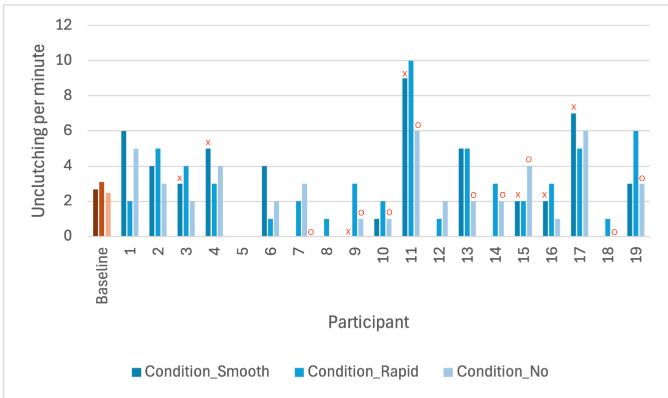


Figure 4.7: Bargraph for unclutching scores per minute across conditions with preferences for a strategy .

participants and conditions. For example, Participant 11 shows a particularly high number of unclutching actions under the "No" condition, whereas Participant 5 and Participant 18 showed very few unclutching actions overall. Some participants' preferred conditions correlate with fewer unclutching actions (e.g., Participant 7 shows fewer unclutchings under "No", which they also preferred). In contrast, others display higher unclutching frequencies in their preferred condition (e.g., Participant 4 preferred Smooth despite unclutching most in that condition).

To test the first hypothesis, and that people with strong preferences will perform better in their preferred condition, a Friedman's Test was conducted. The results were significant (**F=12.15, p=0.0042**), indicating that at least one of the strategies produces different performance results than the others. The results of the Wilcoxon Signed Rank Test suggest that there was a significant (**p = 0.0148**) difference in participant performance between the condition with a stronger preference (C_Smooth) and the weaker preference condition (C_Rapid). Additionally, there was a significant difference between weaker preference (C_Rapid) and no preference (C_No) (**p = 0.0073**). Finally, there was no significant difference ((**p = 0.4126**) between having a strong preference versus no preference.

4.2.5 Individual Differences

The graph in Figure 4.8 illustrates the variability in performance across participants and conditions. Each line represents a single participant's score in each condition, highlighting the patterns observed. The red 'X' markers indicate baseline performance, which also varies across conditions.

As mentioned in the previous section, some participants expressed a strong preference for one of the three recovery strategies. It was therefore decided to take a closer look at the unclutching behaviour of participants with such a preference. Overall, the no recovery strategy condition showed the lowest amount of unclutching per minute ($m_{unclutching}$=2.5), followed by the smooth condition ($m_{unclutching}$=2.7) and lastly the rapid condition with on average the highest amount of unclutching ($m_{unclutching}$=3.1).

The results of the Wilcoxon Signed-Rank test showed that there was no significant difference between the smooth condition and the baseline **W=2.1, p=0.32**. For the no recovery condition and the baseline condition, no significant difference could be found **W=2.5, p=0.79**.

To test if the baseline performance of participants with a preference for the smooth

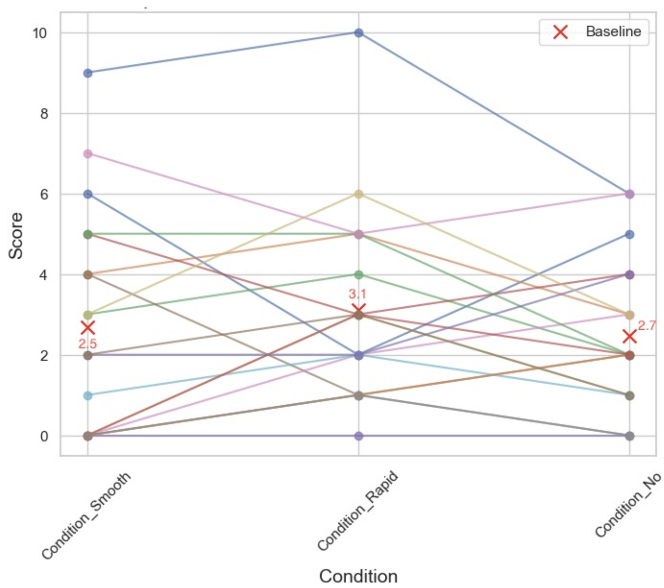


Figure 4.8: Lingraph for unclutching across conditions with the baseline values.

strategy or no strategy differed statistically, a second Wilcoxon Signed-Rank test was run. The results show that the difference in the unclutching scores between participants who prefer the smooth strategy and those who prefer no recovery strategy is not significant **W=16.5, p=0.38**.

4.2.6 Active versus passive recovery strategy

To answer the second hypothesis, a Wilcoxon Signed-Rank Test was used. The results show that there is no significant difference between the active versus passive recovery strategy, **W=47.5, p=0.28**.

Chapter 5

Discussion

5.1 Recapitulation and Implications of the present study

This study examined the impact of different error recovery strategies on operator performance and behaviour in a teleoperation task. The aim was to answer the following research question: **How do different error recovery strategies influence performance?** The strategies included a smooth recovery strategy, a rapid recovery strategy, and no error recovery strategy. Both quantitative and qualitative data were collected to provide a comprehensive view of how these strategies affected the operators' performance and reactions.

5.1.1 Summary of Findings

Research Question: How do different error recovery strategies influence performance?

The results showed no significant overall effect of recovery strategy on hit performance, as revealed by repeated measures ANOVA. However, post hoc analysis suggested significant differences between specific recovery strategies. Notably, as revealed by the descriptive statistics, recovery strategy one (smooth recovery) outperformed strategy two (rapid recovery). In strategy three (no error recovery) a better performance than in the rapid strategy was observed. Lastly, no significant difference was found between the smooth strategy and no recovery strategy. These findings suggest that while participants' performance can vary across strategies, the relationship between strategy type and performance is nuanced. This aligns with previous research highlighting the importance of individual adaptability and preference in error recovery contexts [23]. Unclutching behaviour did not significantly differ across recovery strategies, as indicated by the Friedman test. Post hoc analysis revealed a significant difference between the smooth strategy and the rapid strategy, but no significant differences between other pairs of strategies. These findings suggest that while recovery strategies may influence unclutching behaviour to some extent, the effects are not consistent or robust.

Hypothesis 1: There is a correlation between recovery strategy preference and baseline performance.

The hypothesis that participants would perform better in their preferred strategy was partially supported. The bar graphs depicting individual performance patterns across recovery strategies highlighted the variability in how participants responded to each condition. For instance, while some participants performed best under their stated preferred condition, others showed a mismatch between preference and performance (e.g., Participant 7 preferred the no recovery strategy but performed better during the smooth condition).

In addition to that, the tests revealed a significant effect of preference. Post hoc analysis indicated that there was a significant difference between strong preference and a divided preference. There was, additionally, a significant difference between divided preference and no preference. Lastly, there was no significant difference between strong preference and no preference. Participants performed worse in conditions of equal preference, reinforcing the idea that preference can play a role in performance but may depend on the specific strategies. This could be seen as preference for the smooth strategy meaning on average a lower baseline hit score. On the other hand, those favouring no recovery strategy achieved better scores relative to baseline conditions. Participants with no preference may have approached the task with a more flexible or neutral mindset, focusing on completing the task regardless of the system's recovery behaviour. This could have allowed them to adapt more readily to each recovery strategy without the distraction of comparing or critiquing the system's actions. These results were also statistically significant. This finding suggests that the recovery strategy preference might have a differential impact on task performance, specifically in terms of hitting accuracy or success. Thus, suggesting that alignment between user preferences and system behaviour can positively impact task efficiency [11].

Hypothesis 2: Participants will showcase a better performance if there is an active recovery strategy

The results indicate that there was no significant difference in participant performance between having an active recovery strategy and no recovery strategy. Additionally, individual recovery strategies varied in their effectiveness, with the rapid strategy underperforming compared to the smooth strategy and no recovery, but none of these differences collectively indicated a broad advantage of active recovery strategies over no recovery strategy.

5.2 Interpretation of Findings

It is important to note that none of the participants could recall the errors when asked during the interview. In that sense, some participants, after having been told that errors were introduced, mentioned that they "do not remember noticing the recovery (as) it was so subtle that it did not affect the game play". Most of the participants either could not explain the error, "I did not understand it (the behaviour of the robot) completely" or assumed that the error was made by them "I feel like I was the one making the errors".

Role of Recovery Strategy Type

The findings of hypothesis two suggest that while certain recovery strategies may enhance performance for specific individuals or contexts, the presence of an active recovery strategy alone does not universally lead to better performance. Instead, the effectiveness of a recovery strategy may depend on its alignment with user preferences and its design characteristics [16]. The mixed results regarding recovery strategies suggest that strategy two (fast recovery) may not align well with user expectations or task demands, leading to poorer performance. This, for example, manifests itself in the higher unclutching frequency in this strategy which could signal frustration or disengagement with this recovery approach. One participant explained that that condition "swung very much to the right side and after that, it was hard to find the control back". A second participant explained further that this strategy interrupted the task performance since they first had "the complete range of motion without having to have (to) unclutch", and then afterwards they lost that range of motion and they "immediately tried to push against it and got out of the flow" and had to get back into it again. In contrast to that, the similar performance levels between the smooth strategy and no recovery strategy may indicate that participants value predictability and more control over the system's actions [28].

Impact of Preference

The findings support the hypothesis that preference influences performance. Participants with strong preferences likely feel more confident and engaged, which can positively affect task performance. The lack of significant differences between strong preference and no preference, however, suggests that preferences must align with task demands to enhance performance. Strong preferences for an ineffective strategy may not yield benefits. A strong preference does not always guarantee a better performance in all conditions, as having a strong preference versus not having a strong preference was not significant. This lack of significant difference between these two conditions suggests that other factors, such as individual participant variability, may also influence performance.

5.3 Practical Implications

Recovery strategies should prioritize user preference to enhance user satisfaction and task performance. Kontogiannis [10] suggests that experiencing errors leads to an increased workload. In these abnormal situations, when errors are encountered, the goal should be to reduce complexity and manage information overload as this is crucial for effective error recovery. Furthermore, it could be beneficial to avoid strategies that are overly fast or disruptive, as these can hinder performance and increase unclutching behaviour [21]. A one-size-fits-all recovery approach may not suit the diverse range of user behaviors and preferences. The results further indicated that user performance can diverge even from their expressed preferences, suggesting that adaptive systems that can learn and respond to individual user patterns could further optimize performance and user satisfaction. Incorporating mechanisms that allow users to customize or select recovery strategies based on their preferences could, further, improve task outcomes [20]. This is particularly relevant in systems where users have diverse needs or skill levels. The observed variability in the data underscores the need for flexible teleoperation systems. Especially for users who demonstrate indecisiveness or fluctuating preferences, the system could provide recommendations or gradually tune recovery behaviours to match observed user tendencies. Based on feedback from certain participants, additional feedback when errors occurred could be beneficial. One participant expressed that "I wasn't always sure if I did that or the robot as you don't get much feedback." Stein et al. [26] support this idea and express that there should be multiple overlapping ways of feedback for better error recovery.

5.4 Limitations and Future Research

This study presented certain limitations. First, the study categorized preference levels but did not account for how these preferences were formed. Future research could explore the factors influencing user preferences, such as prior experience or task familiarity. Next, the results were based on the task performance, measured by hits and unclutching behaviour, as well as reports. For future research, it could be of interest to incorporate measures like eye tracking or heart rate variability to gain deeper insights into user engagement and stress levels during error recovery. The idea of working with teleoperated systems that incorporate user modelling techniques or machine learning algorithms that dynamically adjust recovery behaviours based on observed performance and user input over time can also be explored. Lastly, the small sample size may limit the generalizability of the findings. Future studies should include a larger participant pool to validate the results.

5.5 Conclusion

This study emphasizes the need for tailoring error recovery procedures based on user preferences and task requirements. While preferences and active methods can have an impact on performance, their effectiveness is determined by how well they correspond with user wants and expectations. Teleoperation systems can improve their performance and satisfaction by developing adaptive and user-centered recovery methods.

Bibliography

[1] Jeffrey Abouaf. Trial by fire: teleoperated robot targets chernobyl. *IEEE Computer Graphics and Applications*, 18(4):10–14, 1998.

[2] Andreas Achhammer, Carolina Weber, Angelika Peer, and Martin Buss. Improvement of model-mediated teleoperation using a new hybrid environment estimation technique. In *2010 IEEE International Conference on Robotics and Automation*. IEEE, May 2010.

[3] Luis Basañez, Emmanuel Nuño, and Carlos I Aldana. Teleoperation and level of automation. In *Springer Handbook of Automation*, pages 457–482. Springer, 2023.

[4] Jessie Y C Chen, Ellen C Haas, and Michael J Barnes. Human performance issues and user interface design for teleoperated robots. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, 37(6):1231–1245, November 2007.

[5] Zheng Chen, Fanghao Huang, Chunning Yang, and Bin Yao. Adaptive fuzzy backstepping control for stable nonlinear bilateral teleoperation manipulators with enhanced transparency performance. *IEEE transactions on industrial electronics*, 67(1):746–756, 2019.

[6] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.

[7] William R Ferrell and Thomas B Sheridan. Supervisory control of remote manipulation. *IEEE spectrum*, 4(10):81–88, 1967.

[8] B Hannaford. A design framework for teleoperators with kinesthetic feedback. *IEEE Trans. Rob. Autom.*, 5(4):426–434, 1989.

[9] Tom Kontogiannis. User strategies in recovering from errors in man–machine systems. *Saf. Sci.*, 32(1):49–68, April 1999.

[10] Tom Kontogiannis. A systems perspective of managing error recovery and tactical re-planning of operating teams in safety critical domains. *Journal of safety research*, 42(2):73–85, 2011.

[11] Christopher Korte, Sudhesh Sudhakaran Nair, Vasile Nistor, Thomas P Low, Charles R Doarn, and Grant Schaffner. Determining the threshold of time-delay for teleoperation accuracy and efficiency in relation to telesurgery. *Telemedicine and e-Health*, 20(12):1078–1086, 2014.

[12] D A Lawrence. Stability and transparency in bilateral teleoperation. *IEEE Trans. Rob. Autom.*, 9(5):624–637, 1993.

[13] Jae-Young Lee and Shahram Payandeh. Forward error correction for reliable teleoperation systems based on haptic data digitization. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, November 2013.

[14] Jin Sol Lee, Youngjib Ham, Hangue Park, and Jeonghee Kim. Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation. *Automation in Construction*, 135:104119, 2022.

[15] Jiann-Fuh Liaw and Pei-Hsuan Tsai. Target prediction to improve human errors in robot teleoperation system. In *2017 International Conference on Applied System Innovation (ICASI)*. IEEE, May 2017.

[16] Weijane Lin, Hong-Chun Chen, and Hsiu-Ping Yueh. Using different error handling strategies to facilitate older users' interaction with chatbots in learning information and communication technologies. *Frontiers in Psychology*, 12:785815, 2021.

[17] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. To err is robot: How humans assess and act toward an erroneous social robot. *Front. Robot. AI*, 4, May 2017.

[18] Probal Mitra and Günter Niemeyer. Model-mediated telemanipulation. *Int. J. Rob. Res.*, 27(2):253–262, February 2008.

[19] G Niemeyer and J-J E Slotine. Stable adaptive teleoperation. *IEEE J. Ocean. Eng.*, 16(1):152–162, January 1991.

[20] Kent L Norman and Jurek Kirakowski. The wiley handbook of human computer interaction volume. 2018.

[21] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297, 2000.

[22] Thomas Piercy, Guido Herrmann, Angelo Cangelosi, Ioannis Dimitrios Zoulias, and Erwin Lopez. Using skeletal position to estimate human error rates in telemanipulator operators. *Frontiers in Robotics and AI*, 10:1287417, 2024.

[23] Syed Hammad Hussain Shah, Abeer Badawy, Faisal Jamil, Tariq Alsboui, and Ibrahim A Hameed. Toward intelligent telepresence robotics for enhancing elderly healthcare in smart care home. In *International Conference on Human-Computer Interaction*, pages 180–195. Springer, 2024.

[24] TB Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, 1992.

[25] Jingzhou Song, Yukun Ding, Zhihao Shang, and Ji Liang. Model-mediated teleoperation with improved stability. *Int. J. Adv. Robot. Syst.*, 15(2):172988141876113, March 2018.

[26] M R Stein, C P Sayers, and R P Paul. The recovery from task execution errors during time delayed teleoperation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94)*. IEEE, 2002.

[27] Matthew R Stein, Craig P Sayers, and Richard P Paul. The recovery from task execution errors during time delayed teleoperation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'94)*, volume 1, pages 547–554. IEEE, 1994.

[28] JOHN A STOOP and ERIC A VAN KLEEF. Reliable or resilient: Recovery from the unanticipated. *International Journal of Performability Engineering*, 11(2):169, 2015.

[29] Luc Schoot Uiterkamp, Francesco Porcini, Gwenn Englebienne, Antonio Frisoli, and Douwe Dresscher. EMG-based feedback modulation for increased transparency in teleoperation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, October 2022.

[30] Jean Vertut. *Teleoperation and robotics: Applications and technology*. Springer Science & Business Media, March 2013.

[31] Astrid Weiss, Regina Bernhaupt, Manfred Tscheligi, and Eiichi Yoshida. Addressing user experience and societal impact in a user study with a humanoid robot. In *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction (Edinburgh, 8-9 April 2009), SSAISB*, pages 150–157, 2009.

[32] Xiao Xu, Burak Cizmeci, Clemens Schuwerk, and Eckehard Steinbach. Model-mediated teleoperation: Toward stable and transparent teleoperation systems. *IEEE Access*, 4:425–449, 2016.

[33] Tianhao Zhang, Zoe McCarthy, Owen Jow, Dennis Lee, Xi Chen, Ken Goldberg, and Pieter Abbeel. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 5628–5635. IEEE, 2018.

[34] Y Zheng and L K Daneshmend. Learning error-recovery strategies in telerobotic systems. In *Proceedings. 1991 IEEE International Conference on Robotics and Automation*. IEEE Comput. Soc. Press, 2002.

Appendix A

Interview Questions

1. **Considering the interaction with the robot overall, how would you rate your experience?**

   When you have control of the robot and you get used to it, quite easy to use.

   It was pretty responsive.

   It was quite or relative Intuitive. Sometimes it had a little hick-up, which was kind of annoying.

   I think it kind of feels the same as just a normal computer mouse, but it takes some getting used to.

   At times it was very smooth and it just like it did exactly what I did. And sometimes I hit it too hard and like start and when stacking up like in the end, it was kind of doing its own thing.

2. **What were your initial thoughts and feelings when you first started using the teleoperated robot?**

   It reminds me of when you kind of stand in front of a mirror and you try to do like a movement.

   I had to get used to the force, that's for sure, because I was using way too much force and sometimes it was a bit difficult to estimate the distance.

   I knew how to move it, but I didn't know how sensitive it was, you know, so I didn't want to go too hard

3. **How did the interaction influence your mood throughout the session?**

   I just was focused on doing it as efficient as possible

   When it was stuck, I was kind of frustrated, But in general I had a very stable mood and it went well.

   So it started in a more cramped way and then decided. To relax and then I feel my performance, might change towards the end.

4. **Have you experienced any issues while interacting with the robot? / Have you encountered any malfunctions or technical issues?**

   Aside from the When it was stuck again, it just continued. Normally it wasn't acting weird or having issues.

   Yes, that it moved it's position without me doing anything. Movement wasn't as smooth as I thought it was going to be.

Yes! Sometimes it got stuck on the button and then I think you had to reset it or something.

Apart from the jerking and the tilting no.

No, there were a couple of situations where the robots suddenly flew off to the right, which I didn't understand entirely.

5. **How did the system recover from these issues (errors)?**

   I solved them myself.

   I feel like it was mainly me, but on the other hand I don't think I actually like let go and recentered (no unclutching). Might be both.

6. **How did these issues (errors) affect your gameplay experience and task performance?**

   But they gave me, I think a few seconds delay every time to get it back.

   Something that I encountered at some point is that it like drifted a little bit. Like at the start I set up a little area which could perfectly hit all the buttons, but then sometimes it drifted a little bit and then that area was skewed, so I had to unclutch.

7. **Debrief**

   I was mostly bothered by the fast one because you thought you were doing a mistake and then you immediately tried to push against it and got out of the flow.

   I like the first one best. I like that it recovers itself, he has his own life and should also behave.

   It seems like it recovered itself. I don't remember noticing the recovery so it was so subtle that it did not affect the game play.

Appendix B

Setup

To conduct the experiment, the following setup was used. The Franka Research 3 robot arm was further expanded by adding a force torque sensor installed in the final ‘limb’ of the robot. It was further completed by a 3D printed tip that the participants used to hit the targets. The force torque sensor is used to measure force and torque when they are applied.

Figure B.1: Technical Setup

In front of the robot a second table was placed where on top the whack-a-mole box was secured using clamps. Then came the table with the participant seated where also the Omega 7 system was found (Figure B.2).

Figure B.2: Omega 7 system.