Impact of Haptic Feedback in High Latency Teleoperation for Space Applications

JOE LOUCA, Bristol Robotics Laboratory, UK
KERSTIN EDER, Trustworthy Systems Laboratory, UK
JOHN VRUBLEVSKIS, Thales Alenia Space (UK), UK
ANTONIA TZEMANAKI, Bristol Robotics Laboratory, UK

Remote manipulation is a key enabler for upcoming space activities such as in-orbit servicing and manufacture (IOSM). However, due to the large distances involved, these systems encounter unavoidable signal delays which can lead to poor performance and users adopting a disjointed, 'move-and-wait' style of operation. We use a robot arm teleoperated with a haptic controller to test the impact of haptic feedback on delayed (up to 2.6 s: Earth-Moon communications) teleoperation performance for two example IOSM-style tasks.

This user study showed that increased latency reduced performance in all of metrics recorded. In real-time teleoperation, haptic feedback showed improvements in success rate, accuracy, contact force, velocity, and trust, but, of these, only the improvements to contact forces and moving velocity were also seen at higher latencies. Accuracy and trust improvements were lost, or even reversed, at higher latencies. Results varied between the two tasks, highlighting the need for further research into the range of task types to be encountered in teleoperated space activities. This study also provides a framework by which to explore how features other than haptic feedback can impact both performance and trust in delayed teleoperation.

Additional Key Words and Phrases: teleoperation, latency, haptic feedback, trust, user study

ACM Reference Format:
Joe Louca, Kerstin Eder, John Vrublevskis, and Antonia Tzemanaki. 2024. Impact of Haptic Feedback in High Latency Teleoperation for Space Applications. 1, 1 (April 2024), 22 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*This work was supported by EPSRC grant no. EP/V519650/1 and Thales-Alenia Space (UK).

1 INTRODUCTION

Long-range teleoperation capabilities will enable key developments for the next steps of the space industry, such as in-orbit servicing and manufacture (IOSM) of spacecraft, orbital debris management [1], and construction of large-scale infrastructure on the Lunar surface [16, 20, 58]. These activities would bring benefits including: lower financial costs of providing satellite services, by facilitating reusability and reducing the number of launches [24]; sustained accessibility to space, by delaying the formation of an Earth-orbiting collision belt [27]; and, improved capabilities for exploration and research [47]. Teleoperated systems are targeted for these services, over manual maintenance and manufacture by astronauts, as they avoid the high risk and cost associated with human spaceflight [24, 31]. Despite automated systems being used where possible, the variability in legacy satellite configurations [39] means that a large proportion of tasks

Authors' addresses: Joe Louca, joe.louca@bristol.ac.uk, Bristol Robotics Laboratory, University of Bristol, UK; Kerstin Eder, Trustworthy Systems Laboratory, University of Bristol, UK; John Vrublevskis, Thales Alenia Space (UK), Bristol, UK; Antonia Tzemanaki, Bristol Robotics Laboratory, University of Bristol, UK.

Manuscript submitted to ACM

would still require the adaptability of human control, and would therefore require some form of human-in-the-loop teleoperation [51].

The problem, still yet to be overcome, is how best to perform tasks with the inherently long time delays [38] – as dictated by the speed of light and the large distances between control center and robot. Round trip signal times between points of interest range from 100s ms for Earth Surface–Low Earth Orbit (LEO) to 2.6 s for Earth Surface–Moon Surface (Table 1) [32]. These communication times are further increased if travelling via a relay of communications satellites, for example, to the other side of the Earth or the Moon. At this scale of delay, telemanipulation tasks become more difficult and increase the chance of errors [23]. The cost of failure is extremely high in the space environment [44], and, hence, when considering the adoption of these technologies, the barrier for entry is also very high [11, 39]. While the concept is not new and the technology has matured to commercial usage for other applications, to date there are few implemented examples of Earth-Space telemanipulation.

Table 1. Typical 'best case' 2-way delay times between points of interest for IOS and exploration. Adapted from [32].

2-Way Delay	**Between Points**
130 ms	Deimos–Mars Surface
240 ms	Earth Surface–Geostationary Orbit (GEO)
260-360 ms	Deimos–Mars Surface (via Areostationary Orbit relay)
410 ms	Earth Surface–Earth Moon L1 *or* Earth Moon L2–Moon Surface
460-960 ms	Earth Surface–LEO via GEO relay
2,600 ms	Earth Surface–Moon Surface

The challenge, in these systems, is that once a command is sent, there can be no human intervention until the command has been acted upon, and the user will face a further return delay before they can know the outcome. As latency increases, this involves the operator taking greater risks [52], delegating more responsibility to the remote system. For this to occur smoothly, the robot must be trustworthy and the operator must trust the robot [52]. Here, we use the following definitions: 'system trustworthiness' - *assurance that a system will behave as the user expects* [3], and 'trust in the system' - *willingness to rely on the system, based on confidence that it will behave as expected without negative consequences* [9, 46]. At higher latencies, the user must predict the outcomes of their commands further in advance, which magnifies any errors as uncertainties are propagated. Additionally, users receive out-of-sync feedback, which makes it more difficult to accurately update and adjust their expectations [48]. Both of these factors, combined, lead to greater disparity between user expectations and robot behaviour.

It has been reported that, under high-latency conditions, operators will generally adopt a 'move-and-wait' strategy [49] (interrupting sequences of movements with stationary periods). This behaviour introduces stationary, predictable periods for the user to interrupt erroneous movements and re-baseline their predictions [14] in response to increased expectation-behaviour differences – i.e. it is a behavioural reflection of the decreased trust that users have in a delayed system. Using this strategy greatly increases the completion time [14], and highlights how behaviours resulting from poor trust in a system can negatively impact performance. Consequently, there is benefit in quantitatively measuring the 'demonstrated trust' in a system [52], to allow assessment and comparison of the impact of different features with fewer biases than when using self-reported measures based on subjective questionnaires commonly used in the literature [25, 46], to systematically address the issue of trust in delayed teleoperation.

1.1 Related Work

Expert operators report that the most important factor for high-latency teleoperation is understanding of the situation, i.e. information quality and quantity [56]. Providing haptic feedback could improve this situational understanding, and subsequently trust [13], by increasing the sensory information available to the user. This has been widely demonstrated to enhance performance in a range of application areas, for example in surgical [12] and nuclear robotics [50], and is also being explored for space robotics [2]. However, it is not clear whether addition of haptic feedback is advantageous with communication delays, as it could have a destabilising effect on the system [48] and, by providing users with more temporally disconnected information, this can become more cognitively demanding. Delayed bilateral haptic feedback systems must, therefore, find a balance between transparency and stability [38].

The prevailing strategies for ensuring stability in delayed haptic feedback systems are either to use Model-Mediated Teleoperation (MMT) with a time-adjustment, which allows users to interact with a time-adjusted simulation in real-time [7, 44], or to use passivity control algorithms to ensure system stability, as with a Time Domain Passivity Approach (TDPA) [41] or Wave Variable Controllers [59]. However, these strategies come at a cost. MMT is heavily reliant on having an accurate model, making it less suitable for complex, non-deterministic environments, as errors can arise when users place too much trust in an inaccurate model [44]. Passivity control, such as TDPA, results in position drift when in contact with objects, and the increased completion time as a consequence may lead astronauts to prefer no haptics or passivity control [41]. This paper examines the effect of situational awareness on trust, and therefore considers the ideal transparency condition only, where the haptic feedback provided to the operator is equal to that which was sensed by the physical remote system [38]. Instead of utilising control or modelling strategies, we focus on the often overlooked approach of using the human operator to stabilise an unstable delayed system [40].

Most delayed teleoperation research to date has assessed the impacts of ≤ 1 s delay. Negative impacts on performance as a result of delay are commonly reported in the 100 to 500 ms latency range, for systems where either only visual [28, 35, 36, 42, 57] or haptic [4, 5] feedback types were tested. However, these studies did not compare the two feedback types on the same system. Perez et al., for example, found that across four tasks in a virtual surgical robotics trainer (dv-Trainer) which used visual feedback only, completion times doubled as latency increased from 100 to 500 ms, and the number of errors in a task increased at latencies $\geq$500 ms [42]. In another virtual surgical trainer study, a similar doubling in completion times was recorded after increasing the delay from 0 to 500 ms [35].

From the literature, it is not clear at what latency the advantages of haptic feedback are lost, or become disadvantageous. Ivanova et al. found that, although telemanipulation performance (accuracy and smoothness) declined as latency was increased from 180 to 540 ms, haptic feedback was still beneficial to the user up to the largest latency tested (540 ms) in a 1-degree of freedom (DOF) tracking task [22]. However, the difference between feedback types decreased with latency, and these results do not show for how long haptic feedback maintains its advantage. Similarly, Yip, Tavakoli and Howe reported increased completion times and tip forces when latency was increased up to 500 ms in visual and haptic feedback conditions, for a 6-DOF peg-in-hole task [59]. For the delays tested ($\leq$500 ms), haptic feedback had the benefit of reducing contact forces at the cost of increasing completion time, suggesting that selecting an appropriate feedback type will depend on the task's priorities. However, the impact of adding haptic feedback at higher delays, relevant to Earth-Moon teleoperation, has not yet been assessed, and was identified by the authors as an important next step for future work [59].

There are a few examples of previous work which examined the impact of larger delays up to 4 s [21, 33, 38, 55]. These studies reported similar trends (further increases to completion time, decreased accuracy and speed), but did not successfully evaluate visual vs visual and haptic feedback systems. A comparison is required between these feedback types at higher latencies, as this would inform the design of a telemanipulation system for use in the space environment.

Further to the impacts on performance, to the authors' knowledge, there have been no prior reports of user studies investigating trust in delayed haptic telemanipulation systems. Rogers et al. [45] and Khasawneh et al. [28] explored how trust is affected by small delays (500 ms) when operating visual-only feedback mobile robots operated with a joystick. In both studies, increased delay did not reduce trust when operating a single robot. Telemanipulation, however, requires a different style of control to the teleoperation of mobile robots, so the results of these studies are not directly applicable to space manipulation tasks such as IOSM.

1.2 Contribution

Our work addresses the recommendation of Khasawneh et al. to investigate the effects of haptic feedback on trust in delayed systems [28] and, additionally, it extends it to higher latencies relevant to Earth-Moon communications.

In this paper, we investigate how haptic feedback affects performance and trust in systems with high communication delays, and identify whether there is a latency beyond which adding haptic feedback becomes detrimental. We explore this impact through two tasks which suit different aspects of haptic feedback. We also aim to assess the 'move-and-wait' behaviour as a measure of 'demonstrated trust' in teleoperation with delays. To the authors' knowledge this is the first study that assesses a delayed telemanipulation system on both the participants' perceived and demonstrated trust.

The rest of the paper is structured as follows. First, Section 2 outlines the user study, describes the telemanipulation system, the tasks, the feedback and latency conditions tested, and the metrics for assessment. Section 3 presents the results of the user study as a summary across metrics for all conditions and then in greater detail by each individual metric. The discussion in Section 4 explores the effect of increasing latency on these metrics, alongside the associated impact of haptic feedback. Finally, the conclusions of the paper are stated in Section 5, including the implications for implementing telemanipulation systems for space applications, and recommendations for further research.

2 METHODOLOGY

2.1 Population Sample

The user study was approved by the Research Ethics Committee of the University of Bristol (ID: 2022-10027-10400). Trials were carried out by 30 participants (10 female, 20 male, age range 22-55, mean age 33.17 years). Of these participants, 28 were right-handed and 2 left-handed. Each participant gave informed consent and was given an experiment overview before starting the experiment. The sample represented a broad range from people who currently, or have previously, operated robots regularly, to those who have never operated a robot. On average, this sample was accepting of technology; when asked to rate their attitude to technology on a scale from 1 to 5, 1 being extremely wary and 5 being extremely accepting, the mean value was 3.9 (σ = 1.4). Similarly, participants were quite risk tolerant; reporting a mean value of 3.5 (σ = 0.9) when asked to rate their tolerance (1: extremely risk averse, 5: extremely risk tolerant).

2.2 System description

The teleoperation system comprised two subsystems - local (user-side) and remote (task-side) (Figure 1). As information was passed between the two subsystems, a signal delay was imposed to imitate the outgoing and incoming delays which would be experienced between a control station on Earth and a manipulator in space.

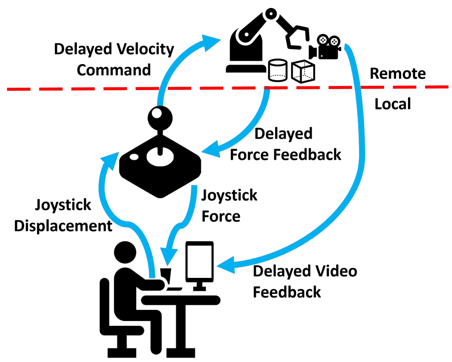

Fig. 1. Overview of the teleoperation system. Each time information crosses the red, dashed line separating local and remote subsystems, a one-way delay is added.

The remote subsystem consisted of a robot arm (KUKA LBR iiwa14 R820), fitted with a 6-axis force-torque sensor (ATI Axia80-M8) at the wrist. A marker pen was fitted as an end effector to the force torque sensor. Two cameras (Logitech C270) to capture video streams (640 x 480, 20 fps) of the task from two viewpoints.

The local subsystem consisted of a haptic arm (Haption Virtuose 6D Desktop) and a laptop (Intel i7-10850H CPU 2.70 GHz, 16GB RAM) running the ROS-based control software and displaying the two video streams of the task.

As in [17], velocity commands were sent from the haptic arm to the manipulator. In order to scale the commands appropriately for the task, velocity commands were reduced from the local to the remote subsystems at a ratio of 2.5:1. A foot pedal clutch control was included to let the user to pause control and refresh the start position. This enabled a 'click and drag'-style motion, extending the possible workspace of the haptic arm and allowing participants to reset to a more comfortable position. A direct force-reflecting bilateral control architecture was used in these experiments. The measured forces ($\leq$ 23 N) on the remote end-effector sensor were used as direct feedback forces on the local haptic device. We opted to use the high-transparency direct force-reflecting method rather than passivity controllers such as TDPA [7, 44] or Wave Variable Control [59], as those methods prioritise system stability at the cost of transparency [1]. Using the human operator to stabilise the system, on the other hand, does not reduce transparency [40]. Forces applied to the local haptic device were limited to 5 N for user safety, and scaled to 30% of the sensor reading to scale down the remote system forces to within this 0-5 N range. For these trials, angular components of velocity commands and force feedback were removed, limiting control to 3-DOF to make it simpler for the untrained participants to learn to use the system [21].

The intrinsic one-way delay of the system was measured to be 7 ms. This was below the 20 ms threshold at which delay becomes noticeable [22], so can be considered to be real-time teleoperation.

2.3 Tasks

Similar to Pryor et al. [44], our task selection was motivated by upcoming missions requiring high latency teleoperation, such as NASA's (National Aeronautics and Space Administration) On-Orbit Servicing Assembly and Maintenance 1 (OSAM-1) mission and ESA's (European Space Agency) ESPRIT Refueling Module (ERM) planned for the Lunar Gateway. The two tasks, illustrated in Figure 2, were chosen as examples of different challenges that an operator may be faced with in real-world applications, and which may benefit from the addition of haptic feedback in different ways, e.g., maintaining contact with a surface in cutting or adhesive application tasks, or aligning and mating two objects in refuelling or component replacement tasks [18]. The tasks used were inspired by the refueling task in [44] and the cutting task in [55], and were designed after discussion with experts in space industry.

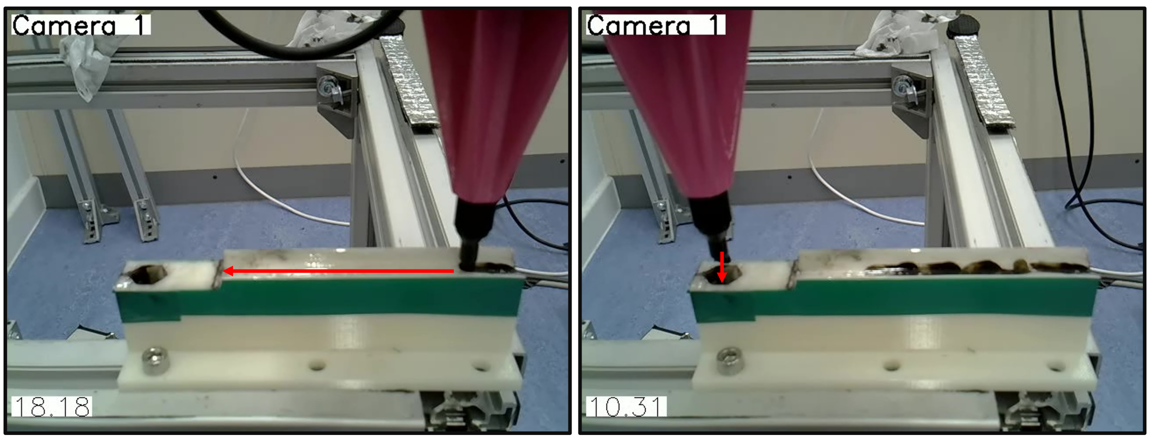

Fig. 2. Camera views of the two tasks. Left: Task A – Line Drawing, Right: Task B – Peg in Hole. Red arrows indicate direction of travel.

2.3.1 *Task A: Line Drawing.*

Participants were asked to draw a line along a 110 mm long groove, maintaining contact with, but applying as light a force as possible to, the surface of the path. This task was selected so that visual and haptic feedback both provided useful information for the user – while they could feel contact with the surface, the cameras provided visual confirmation of task success as the line was drawn. Furthermore, the soft felt tip of the marker added some compliance to the end effector, meaning a range of forces could be detected depending how hard the user pressed against the surface.

2.3.2 *Task B: Peg in Hole.*

Participants were then asked to insert the pen (11.6 mm diameter) into a hole (12.0 mm diameter, 20.0 mm deep) at the end of the path, as in typical peg in hole tasks that are commonly used to test telemanipulators [37]. This task was designed to incorporate some of the challenges faced when operating based on visual feedback alone. Firstly, the hole was placed at the end of the path where camera view is poorer (compared to the front). Additionally, the inside of the hole was coloured black to make it more difficult to visually distinguish between it and the black pen, and a pen was chosen with a filleted rim which was wider towards the base. This was to make it more difficult to visually assess progress as the pen was inserted, while reducing the clearance between the pen and the hole. This task involved more rigid object interactions than Task A, resulting in stronger, sharper forces being felt.

2.4 Conditions

2.4.1 Signal Delays.

In order to identify the signal delay limit beyond which teleoperation is not feasible, the participants were asked to repeat the task with two-way delays, t_{delay}, of: 0, 600, 1500, 2600 ms. I.e., an outbound delay of $\frac{t_{\text{delay}}}{2}$ was imposed on user commands to the remote subsystem, and an inbound delay of $\frac{t_{\text{delay}}}{2}$ was imposed on camera and force sensor data sent back to the local subsystem. This investigates delay for key points of interest (Table 1) and covers the range of delays seen up to the level of Earth-Moon Surface communications.

2.4.2 Feedback Type.

Under each signal delay condition, the trial was repeated twice with: Visual Feedback (VF) only or Visual and Haptic Feedback (VHF). In the following Sections, the different conditions will be referred to as 'Feedback-Latency', for example, VF-0 for visual feedback only at 0 ms latency.

2.4.3 Order of Conditions.

Each participant completed the two tasks (A and B) under the eight different conditions (4 signal delays x 2 feedback types). The order of these conditions was varied systematically between participants to ensure an even distribution in order to counteract any effects of training on the system. Half of the participants completed the block of four trials for VHF followed by VF, with the other half completing VF first. Within each block of four trials, the order of delays was systematically randomised to avoid effects of adapting to latency [28]. Each of the 24 possible order combinations of the four delays was assigned a random number from 1-24 with no repetitions. This determined the delay order for each sequential block of four trials. Across the 60 blocks of trials (two each for 30 participants), half of the order combinations were repeated twice, and half were repeated three times.

2.5 Metrics

For evaluating the impact of feedback under the various task conditions, we define the following metrics.

2.5.1 Success Rate.

The percentage of participants who completed the task successfully. Success was defined as having reached the target without colliding with sufficient force to trigger the KUKA collision detection (activated if the torque at any joint exceeded 30 Nm). By defining unrecoverable conditions, we simulated the high cost of failure experienced in the space environment.

2.5.2 Completion Time.

The time taken to complete the task. Completion was defined for the two tasks as:

- Task A - Pen reached the end of the path.
- Task B - Peg fully inserted into hole.

This was measured using kinematic analysis of the robot's joints to track its end effector position over time.

2.5.3 Accuracy.

The distance between the pen tip and the target. The pen tip position was tracked using kinematics and the target position was defined for the two tasks as:

- Task A - The path along the centre of the groove. Accuracy was measured as the root mean squared error (RMSE) over the duration of the task.
- Task B - The centre of the hole when completely inserted. Accuracy was measured as the closest distance to this point.

2.5.4 Contact Forces.

Mean magnitude of forces at the end effector when in contact with an object. This was measured by the wrist force/torque sensor. Larger forces are present when the user overshoots and makes more contact than necessary with scene objects. This has real-world safety implications, e.g., it is important to avoid pressing too firmly with a blade or drill, over-tightening a bolt.

2.5.5 Perceived Trust.

Questionnaire score based on Schaefer's 14-item trust subscale. [46] This aimed to capture how much the users trusted the system. The shortened 14-item subscale was used to avoid fatigue as participants were asked to complete the questionnaire after each condition (Task A and B combined).

2.5.6 Demonstrated Trust.

We divided the 'move-and-wait' behaviour [49] into 'Move' and 'Wait' periods. 'Wait' periods were defined as any period when the magnitude of the velocity was below 0.01 mm s^{-1}. This threshold was determined by comparing the velocity profiles against qualitatively assigned 'Wait' period profiles taken from a random sample of videos of the tasks. Any non-'Wait' period was assigned as a 'Move' period.

We measured the moving behaviours: *Mean Move Duration*, the mean time between wait periods; *Mean Move Distance*, the mean distance travelled during a move period, and, *Mean Move Velocity*, the mean velocity during a move period. Waiting behaviour was also measured: *Wait Frequency*, how many times the user waited per second of moving time.

Avoiding collisions while making contact requires careful control of the end effector position within a tolerance range (contact position to collision position). The time that the user has to correct an erroneous command that moves towards a collision is:

$$t_{correction} = \textit{User Reaction Time} + \textit{Latency}$$

If: *Velocity* > *(Distance to Collision /* $t_{correction}$), the user will not be able to respond quickly enough to avoid collision. Moving at higher velocity, therefore, is a higher risk action which demonstrates greater trust that the command will be acted out as expected, and that the commands do not contain errors. Assuming a constant reaction time for each user, we would therefore expect velocity, and by extension, trust, to be inversely proportional to latency. Conversely, 'Wait' commands are safe, risk-free actions, so an increased wait frequency demonstrates reduced trust in the system.

We used these behavioural metrics to objectively measure demonstrated trust [15], which complements the subjective perceived trust scores obtained from the questionnaires (comparable to using physiological workload data to complement perceived workload scores from a questionnaire [53]). This approach avoids the potential for self-reported bias that can occur when using questionnaires [6].

2.6 Statistical Analyses

We used the MATLAB Statistics and Machine Learning Toolbox to carry out statistical tests for significant differences between samples. The success rate results were analysed using Pearson's χ^2 test. The remaining metrics were analysed using t-tests to compare the two feedback conditions at each latency level, and ANOVA tests to compare the four latencies within each feedback type.

3 RESULTS

The raw results, showing the mean and standard deviations of each metric under each test condition for the two tasks, are shown in Table 2. The results of the χ^2, t, and ANOVA statistical tests are given in Tables 3, 4 and 5, respectively.

3.0.1 *Success Rate.*

Neither latency nor feedback type had any significant effect on the overall success for the line drawing task. For Task B, however, adding haptic feedback improved the success rate from 53.3% to 90.0% at 0 ms ($\chi^2(1, N{=}58) = 9.93$, $p = 0.002$), although no significant improvements were seen at any other latencies. With increasing latency, the success rate decreased for the VHF conditions only (90% for VHF-0 to 51.7% for VHF-2600, $\chi^2(3, N{=}114) = 12.61$, $p = 0.006$), but no such trend was seen across the VF conditions.

3.0.2 *Completion Time.*

Increased latency resulted in increased task completion time for both the VF and VHF conditions ($p = 0.028$ and $p < 0.001$ respectively) in the line drawing task. Although the differences between feedback types were not significant at any latency, when haptic feedback was added, this metric increased more steeply with increasing latency than for the VF trials. Between 0 and 2600 ms, the mean completion time increased by 73.8% for VHF vs 61.7% for VF conditions.

Similarly, for the peg in hole task, increased latency led to increased completion time ($p = 0.022$) in VF conditions. With VHF, however, this was not the case. Furthermore, the addition of haptic feedback at 0 ms latency led to a significant increase in mean completion time of 40.0%, ($p = 0.026$). As latency increases, the resultant increase of the mean completion time from adding haptic feedback diminishes to 16.4% at 600 ms, 5.8% at 1500 ms, and no notable difference at 2600 ms. These differences are not statistically significant ($p = 0.092$ and 0.056, respectively), possibly due to wide distributions of values in the VF-600 and VHF-1500 conditions. This may be due to each participant attempting each task once. Repeated trials for each condition were not possible for this study due to the large number of conditions and time limitations. The mitigation steps that were taken are addressed in the discussion section.

Equivalence tests comparing the results between feedback types at each latency confirms that as the latency increases, the results for feedback types become more similar (the margin of equivalence decreases). The results of the 0, 600, 1500 and 2600 ms are statistically equivalent, when the margin of equivalence = 75%, 40%, 30% and 20%, respectively, of the mean of the VF condition.

3.0.3 *Accuracy.*

For Task A, RMSE increased with delay for both VF and VHF conditions ($p = 0.030$ and 0.023, respectively). Between the 0 ms and 2600 ms trials, RMSE increased by 25.8% for VF, and 27.9% for VHF. Using Tukey's Test, the differences were most pronounced when comparing 600 against 2600 ms delay ($p = 0.021$) for VF, and either 0 or 600 ms against 1500 ms delay ($p = 0.050$ and 0.11 respectively), for VHF. However, as with completion time, the addition of haptic feedback did not elicit any significant effects at any latency.

In Task B, increasing the latency resulted in a significant ($p = 0.003$) decrease in accuracy for the VHF condition only. The final distance to the target was 70.8% greater in VHF-2600 vs VHF-0. Accuracy in the VF conditions was largely unaffected by latency. With 0 ms delay, adding haptic feedback improved accuracy (-22.5% mean distance to target,

p = 0.038), but this was not seen in any of the higher latency conditions. Despite the weak significance (p = 0.17), there was a possible reduction in accuracy after adding haptic feedback at 2600 ms latency.

3.0.4 Contact Forces.

For Task A, increasing the latency increased the mean contact forces for VHF conditions (p = 0.037), but not for VF (Figure 3). Furthermore, providing haptic feedback reduced contact forces for all latencies (0 ms: -44.4%, p = 0.013; 600 ms: -32.7%, p = 0.054; 2600 ms: -40.2%, p = 0.002) with the exception of 1500 ms (-18.4%, p = 0.35).

Although increasing latency did not affect the mean contact forces in Task B for VF or VHF conditions, a similar trend was seen when comparing feedback types at each latency step. Adding haptic feedback reduced the mean contact forces for all latencies (0 ms: -44.2%, p = 0.023; 1500 ms: -47.6%, p < 0.001; 2600 ms: -38.4%, p = 0.012) with the exception of 600 ms (-34.6%, p = 0.15).

The insignificance seen at 1500 ms and 600 ms in Tasks A and B, respectively, may be due to each participant completing each task once.

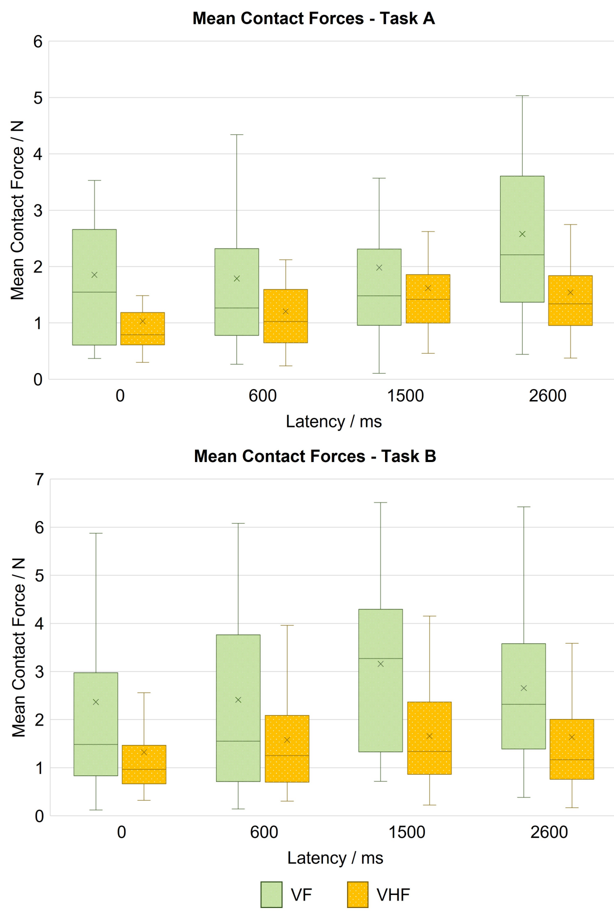

Fig. 3. Mean end-effector contact force for VF and VHF conditions over increasing latency for Task A (left) and B (right).

3.0.5 *Move - Wait Behaviour.*

During Task A, increasing latency from 0 to 2600 ms decreased the mean move distance, duration and velocity for both VF and VHF conditions (VF: distance -64.4%, $p < 0.001$; duration -54.0%, $p < 0.001$; velocity -37.0%, $p < 0.010$. VHF: distance -67.8%, $p < 0.001$; duration -59.3%, $p < 0.001$; velocity -50.7%, $p < 0.001$). It is worth noting that these metrics are all interlinked. Addition of haptic feedback resulted in increased move velocity and move distances for all latencies. (Velocity for 0, 600, 1500 and 2600 ms: +70.1%, +40.3%, +26.2% and 33.3%; $p = 0.026$, 0.004, 0.006 and 0.007. Distance: +47.3%, +56.8%, +49.1%, +33.4%; $p = 0.068$, < 0.001, 0.004, and 0.080, respectively). While increasing latency decreased the mean duration of move periods, this was not affected by adding haptic feedback. Instead, VHF resulted in higher mean moving velocities over similar durations, which, in turn, resulted in moving farther during each move period, compared with the VF conditions. However, the velocity improvements provided by VHF diminished as latency increased.

For Task B, increasing latency also decreased move distance, duration and velocity (VF: distance -67.0%, $p < 0.001$; duration -52.5%, $p < 0.001$; velocity -44.7%, $p = 0.003$. VHF: distance -69.5%, $p < 0.001$; duration -55.6%, $p < 0.001$; velocity -43.4% $p = 0.041$). However, adding haptic feedback did not offer any notable velocity improvements at any latency, except for 1500 ms (+38.0%, $p = 0.027$).

The mean moving velocities were similar between the two tasks, comparing VF trials at each latency step. Participants almost exclusively approached Task B by, firstly, aligning the end effector above the hole without making contact, and then inserting it into the hole. This essentially converted the task into a VF task for both feedback conditions up until inserting the pen. As this part comprised the majority of the task duration, the lack of difference between feedback conditions at low latencies for Task B may be explained, even though it was so pronounced for Task A.

In Task A, for both feedback types, increased latency increased the wait frequency (VF: +162.1%, $p < 0.001$; VHF: +107.7%, $p < 0.001$). Task B showed similar effects on wait behaviour. Wait frequency increased with increasing latency (VF: +121.8%, $p < 0.001$; VHF: +127.3%, $p < 0.001$). However, adding haptic feedback did not have any significant effect on the wait frequency at any latency.

3.0.6 *Perceived Trust.*

Across the two tasks, increased latency was found to reduce trust for both feedback types, although this decrease was steeper for VHF trials. Between 0 and 2600 ms, mean trust score percentages decreased by 16.5% and 24.6% points for VF and VHF conditions ($p = 0.012$ and $p < 0.001$). While there was a possible increase in trust as a result of haptic feedback at 0 ms (+5.7%, $p = 0.15$), this gain was lost as latency increased. No significant difference was seen at 600 ms, while a significant decrease of 5.6% was seen at 1500 ms ($p = 0.046$). At 2600 ms, no effect was seen by adding haptic feedback.

As this metric relied on individual responses to a questionnaire, we also evaluated each participant's scores in relation to their other responses (Table 6). For each participant, we calculated the change in trust score after increasing the latency by one step, or after adding haptic feedback. From this, we report the average change in trust score across participants, as well as the proportion of participants who reported reduced trust following the change of conditions.

The majority of participants reported decreasing trust scores between each increasing latency step. For VHF, the initial rate of decrease was fast, but this slowed for higher latencies. There were particularly large decreases from 0-600 (-11.5%, $p < 0.001$) and 600-1500 ms (-8.6%, $p = 0.006$), but a smaller, insignificant decrease from 1500-2600 ms (-3.9%, $p = 0.21$). Conversely, for VF, trust scores decreased more steadily, but this rate became faster and more significant at higher latencies, -4.5% ($p = 0.15$) from 0-600 ms up to -7.2% ($p = 0.012$) from 1500-2600 ms.

Table 2. Mean (*M*) and standard deviations (*SD*) of the eight metrics for both Tasks A and B. For perceived trust, combined values are given for Tasks A and B together.

		VF								VHF							
		0 ms		600 ms		1500 ms		2600 ms		0 ms		600 ms		1500 ms		2600 ms	
Task	**Metric**	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
	Success Rate / %	80.0	-	76.7	-	82.8	-	69.0	-	93.3	-	90.0	-	86.2	-	86.2	-
	Completion Time / s	75.00	40.84	92.98	51.65	106.76	64.00	121.28	48.84	79.70	70.38	88.66	39.93	122.74	56.20	138.53	59.99
	RMSE / mm	3.87	1.55	3.51	1.16	4.17	1.74	4.87	2.47	3.49	1.52	3.64	1.61	4.70	2.24	4.46	1.71
	Contact Force / N	1.85	1.57	1.79	1.45	1.98	1.77	2.58	1.59	1.03	0.72	1.20	0.84	1.62	0.97	1.54	0.98
A	**Move Vel. / mm s^{-1}**	11.46	5.11	9.06	4.08	9.17	5.42	7.21	4.21	19.49	21.33	12.72	7.96	11.57	6.71	9.61	7.01
	Move Dist. / mm	48.87	36.97	30.16	27.77	22.46	13.67	17.38	12.03	71.97	73.28	47.29	36.22	33.49	21.7	23.18	21.33
	Move Dur. / s	3.08	1.79	2.18	1.47	1.59	0.81	1.42	0.87	3.49	2.61	2.42	1.30	1.80	0.94	1.41	0.73
	Wait Freq. / waits s^{-1}	0.36	0.20	0.63	0.49	0.75	0.36	0.94	0.55	0.40	0.34	0.52	0.31	0.69	0.36	0.83	0.36
	Success Rate / %	53.3	-	73.3	-	65.5	-	65.5	-	90.0	-	76.7	-	58.6	-	51.7	-
	Completion Time / s	80.21	51.16	117.17	63.73	136.34	77.25	150.36	75.35	112.28	75.13	136.34	67.31	144.28	80.44	151.98	56.50
	Accuracy / mm	6.24	3.38	5.86	3.41	6.39	3.56	6.84	3.86	4.84	2.40	5.77	3.14	6.32	4.18	8.26	4.05
	Contact Force / N	2.37	2.44	2.41	2.43	3.16	1.75	2.65	1.64	1.32	1.19	1.58	1.27	1.66	1.08	1.63	1.43
B	**Move Vel. / mm s^{-1}**	13.51	8.40	9.63	5.62	8.06	5.63	7.47	6.46	15.42	9.01	12.08	9.92	11.12	8.86	8.72	6.57
	Move Dist. / mm	37.90	32.54	24.70	18.16	13.72	9.35	12.52	12.24	45.27	31.06	30.33	31.86	21.51	17.93	13.80	8.77
	Move Dur. / s	2.04	1.50	1.60	0.71	0.95	0.37	0.97	0.47	2.13	1.14	1.52	0.86	1.20	0.67	0.95	0.44
	Wait Freq. / waits s^{-1}	0.56	0.30	0.70	0.38	1.18	0.50	1.24	0.55	0.53	0.34	0.79	0.42	0.99	0.44	1.21	0.52
A & B	**Trust Score / %**	77.60	17.50	73.10	18.41	67.83	21.38	61.13	21.52	83.26	16.75	71.71	21.29	62.24	22.25	58.62	25.50

Table 3. Output of Pearson's χ^2 tests, performed on success rate data for Tasks A and B.

Task	**0 ms**		**600 ms**		**1500 ms**		**2600 ms**		**VF**		**VHF**	
	$\chi^2(1, N{=}58)$	p	$\chi^2(1, N{=}58)$	p	$\chi^2(1, N{=}56)$	p	$\chi^2(1, N{=}56)$	p	$\chi^2(3, N{=}114)$	p	$\chi^2(3, N{=}114)$	p
A	2.31	0.13	1.92	0.17	0.13	0.72	2.48	0.12	1.76	0.62	1.07	0.76
B	9.93	0.002	0.09	0.77	0.29	0.59	1.14	0.29	2.68	0.44	12.61	0.006

Table 4. Outputs of *t*-Tests, comparing the differences between VF and VHF conditions at each latency level.

		0 ms		**600 ms**		**1500 ms**		**2600 ms**	
Task	**Metric**	***t*(29)**	***p***	***t*(29)**	***p***	***t*(28)**	***p***	***t*(28)**	***p***
A	**Completion Time**	-1.11	0.28	0.45	0.66	-1.48	0.16	-1.29	0.21
	RMSE	1.07	0.29	-0.47	0.64	-1.36	0.19	0.97	0.34
	Contact Force	2.64	0.013	2.01	0.054	0.95	0.35	3.33	0.002
	Move Velocity	-2.35	0.026	-3.10	0.004	-2.96	0.006	-2.93	0.007
	Move Distance	-1.90	0.068	-3.73	<0.001	-3.18	0.004	-1.82	0.080
	Move Duration	-0.86	0.40	-1.10	0.28	-1.17	0.25	0.02	0.98
	Wait Frequency	-0.75	0.46	1.72	0.10	0.25	0.81	1.36	0.18
B	**Completion Time**	-2.52	0.026	-1.79	0.092	-2.19	0.056	-1.44	0.19
	Accuracy	2.17	0.038	0.12	0.90	0.07	0.94	-1.41	0.17
	Contact Force	2.41	0.023	1.48	0.15	4.32	<0.001	2.71	0.012
	Move Velocity	-1.13	0.27	-1.51	0.14	-2.33	0.027	-1.85	0.075
	Move Distance	-1.21	0.24	-1.09	0.29	-2.47	0.020	-0.69	0.49
	Move Duration	-0.02	0.98	0.59	0.56	-1.97	0.059	0.21	0.83
	Wait Frequency	0.13	0.90	-1.43	0.16	1.88	0.070	0.24	0.81
A & B	**Trust Score**	-1.47	0.15	0.41	0.69	2.09	0.046	0.81	0.42

Table 5. Output of ANOVA tests, comparing the differences between latencies for each of the feedback conditions.

		VF		**VHF**	
Task	**Metric**	***F*-statistic**	***p***	***F*-statistic**	***p***
A	**Completion Time**	$F(3, 87) = 3.18$	0.028	$F(3, 101) = 6.08$	<0.001
	RMSE	$F(3, 114) = 3.08$	0.030	$F(3, 114) = 3.31$	0.023
	Contact Force	$F(3, 113) = 1.50$	0.22	$F(3, 114) = 2.92$	0.037
	Move Velocity	$F(3, 114) = 3.97$	<0.010	$F(3, 113) = 3.50$	0.018
	Move Distance	$F(3, 114) = 9.04$	<0.001	$F(3, 113) = 6.78$	<0.001
	Move Duration	$F(3, 114) = 9.75$	<0.001	$F(3, 113) = 9.58$	<0.001
	Wait Frequency	$F(3, 114) = 10.00$	<0.001	$F(3, 113) = 9.34$	<0.001
B	**Completion Time**	$F(3, 72) = 3.42$	0.022	$F(3, 78) = 1.28$	0.29
	Accuracy	$F(3, 114) = 0.39$	0.76	$F(3, 114) = 5.01$	0.003
	Contact Force	$F(3, 108) = 0.84$	0.47	$F(3, 112) = 0.45$	0.72
	Move Velocity	$F(3, 114) = 4.97$	0.003	$F(3, 111) = 2.84$	0.041
	Move Distance	$F(3, 114) = 10.04$	<0.001	$F(3, 111) = 8.59$	<0.001
	Move Duration	$F(3, 114) = 10.44$	<0.001	$F(3, 111) = 11.02$	<0.001
	Wait Frequency	$F(3, 114) = 17.65$	<0.001	$F(3, 111) = 12.36$	<0.001
A & B	**Trust Score**	$F(3, 114) = 3.79$	0.012	$F(3, 114) = 7.67$	<0.001

Although VHF scored more highly at 0 ms (VHF: 83.3% vs VF: 77.6%), by 2600 ms the scores for each feedback were similar, with VF actually scoring higher (VHF: 58.6% vs VF: 61.1%). This was because trust declined more steeply in the VHF conditions, mainly occurring between 0-1500 ms. For VF, on the other hand, the decline was more gradual, with a greater impact between the two highest latencies. This is further supported by comparing the VF and VHF conditions at each latency. At 0 ms, the majority of participants reported higher trust scores for VHF, and on average these scores were larger than for VF. The strongest difference between VF and VHF occurred at 1500 ms (-5.6%), where 69.0% of participants reported a lower score for VHF than VF.

Table 6. Comparison of each individual participant's trust scores between latency steps and feedback types.

Comparing Conditions [A-B] Feedback	Latency / ms	Mean (SD) change in trust score	*t*-score	*p*	% of participants reporting lower trust in B vs A
VF	**0-600**	-4.5% (0.17)	$t(29) = -1.48$	0.15	55.7%
	600-1500	-4.5% (0.12)	$t(28) = -2.07$	0.047	58.6%
	1500-2600	-6.7% (0.13)	$t(28) = -2.69$	0.012	65.5%
VHF	**0-600**	-11.5% (0.15)	$t(29) = -4.13$	<0.001	80.0%
	600-1500	-8.6% (0.15)	$t(28) = -2.99$	0.006	69.0%
	1500-2600	-3.6% (0.15)	$t(28) = -1.28$	0.21	72.4%
VF-VHF	**0**	+5.7% (0.21)	$t(29) = 1.47$	0.15	33.3%
	600	-1.4% (0.19)	$t(29) = -0.41$	0.69	43.3%
	1500	-5.6% (0.14)	$t(28) = -2.09$	0.046	69.0%
	2600	-2.5% (0.17)	$t(28) = -0.81$	0.42	48.3%

4 DISCUSSION

Table 7. Summary of resultant changes in different metrics due to either increasing the latency or adding haptic feedback for the two tasks. Arrows indicate an increase or decrease in the given metric. Green indicates practical advantages; Red indicates disadvantages; Yellow indicates a mixed response.

	Task A		Task B	
Metric	**Latency ↑**	**+ Haptic**	**Latency ↑**	**+ Haptic**
Success	–	–	$\downarrow^{VHF}$	$\uparrow^{0}$
Time	↑	–	$\uparrow^{VF}$	$\uparrow^{0*}$
Accuracy	↓	–	$\downarrow^{VHF}$	$\uparrow^{0}$
Force	$\uparrow^{VHF}$	$\downarrow^{0,\ 2600}$	–	↓
Velocity	↓	↑	↓	$\uparrow^{1500}$
Wait Freq.	↑	–	↑	–
Trust	↓	$\downarrow^{1500}$	↓	$\downarrow^{1500}$

$^{VF\ /\ VHF\ /\ \#\#}$ Directional change was only seen between the VF / VHF / ## ms latency conditions.

* Difference between VF and VHF decreased as latency increased.

4.1 Impact of Latency

Firstly, we consider the effect of increased latency. By summarising the results from the previous section in Table 7, it is clear that, for both tasks, increasing the latency negatively impacted the metrics assessed. This highlights the challenges presented by high-latency teleoperation, and is comparable to the effects reported in the literature, although the magnitude of the effects are different [4, 5, 22, 33, 35, 42, 59].

Considering completion time, for example, the literature reports two-fold increases as delays are increased from 0-500 ms for 1-DOF [42] and 6-DOF [59] tasks. However, our results showed a more moderate increase in completion time from 0-600 ms delay (VF: 22% and 46%, VHF: 11% and 21% for Tasks A and B, respectively). Furthermore, in another study, a nine-fold increase was reported under a 3 s delay for a 6-DOF manipulation and cutting task [33], which was far greater than our results (VF: 61% and 74%, VHF: 87% and 35%).

There were some unexpected exceptions to these trends. When considering success rate, no effect was seen after increasing latency in either the VF or VHF conditions for Task A. This was likely due to the task being easy to complete, but difficult to complete 'well'. Although several other performance metrics had deteriorated, participants were successful in the 2600 ms delay trials 69% and 86% of the time for VF and VHF conditions, respectively (Table 2). This result was contrary to those reported by Perez et al. [42], where the researches recorded a significant increase in errors above 500 ms delay, and a delays ≥700 ms made it very difficult for participants to complete the tasks successfully, particularly for more complex tasks.

For Task B, success rate was only impacted by latency in the VHF conditions, with the VF conditions all seeing similarly poor results across all latencies. Surprisingly, the worst of these, by far, was VF-0 with 53%, which also saw the greatest improvement by adding haptic feedback (increasing to 90%). This improvement highlights the benefits of VHF that are frequently reported in the literature, particularly for tasks where visual data alone is insufficient to reliably complete the task. We suspect that the unusually high failure rate at VF-0 was because users were over-confident in what they perceived to be an 'easier' condition, without realising how difficult the task was. Task B had much smaller margins for error when aligning the pen than Task A. Additionally, Task B involved contact between more rigid materials, meaning the collision threshold was passed more quickly.

In VHF conditions for Task B, although not statistically significant, the overall trend was still towards greater completion times as latency increases. It appears, however, that large distribution of values for the VHF-1500 condition may have reduced the significance of the data. As the participants were asked to complete the two tasks under a total of eight different conditions, time limitations of the experiment meant that we were only able to ask them to complete one trial for each condition. In order to counteract the increased variance as a result of this, while staying within time limits for individual participants, we systematically randomised the task order and used a sufficiently large population sample size. However, considering the many metrics and conditions assessed in this user study, we were unable to avoid large distributions in every situation.

Although the means trended upwards, there was no significant change in mean contact forces with increasing latency for VF conditions in Task A or B, and for VHF conditions in Task B. Therefore, this does not appear to be one of the major challenges presented by high latency. The distribution of forces was much narrower for all VHF conditions vs their VF counterparts (Figure 3), and this wider variance for VF conditions may account for the low significance between samples. While the lower end of the ranges is comparable between VF and VHF conditions, the upper quartile in VHF conditions was limited to around 2 N in both tasks, compared with 3-4 N in VF conditions. This demonstrates that the feedback provided in both conditions was sufficient for users to carry out both tasks with gentle contact forces.

However, VHF enabled more of the population to achieve this higher performance. Although the upper quartiles of contact forces were larger for Task B than Task A in VF conditions, these were similar across VHF conditions.

4.2 Impact of Haptic Feedback

At 0 ms delay, adding haptic feedback improved:

- Success Rate (Task B only)
- Accuracy (Task B only)
- Contact Forces (Tasks A and B)
- Moving Velocity (Tasks A and B)
- Perceived Trust (Combined Task A and B)

In previous studies exploring the benefits of real-time haptic teleoperation, similar improvements have been demonstrated in success rate, accuracy and reduced contact forces [12].

If only considering completion time, adding haptic feedback with no delay was disadvantageous for Task B only, similar to reports in the literature [59]. The literature reports, both, faster [10] and slower [54] moving velocities as a result of adding haptic feedback, indicating that this may be task dependent. The tasks used here resulted in faster moving velocities with VHF, but the magnitude of the differences between VF and VHF velocities varied between the tasks, indicating that this is affected by the various individual constraints and challenges presented by each specific task.

Next, we consider whether the addition of haptic feedback mitigated any of the negative impacts of increased latency. Out of the above metrics, only the mean contact forces and mean move velocities were improved by adding haptic feedback at higher latencies, for both tasks. The combination of these two benefits demonstrate how haptic feedback enables the users to move ‘confidently’ (high velocity, low force), as opposed to ‘over-confidently’ (high velocity, high force) or ‘over-cautiously’ (low velocity, low force), resulting in a faster but safer operation. The gap between VF and VHF conditions decreases at higher latencies, however, reducing the magnitude of this benefit.

Similarly to our study, Yip Tavakoli and Howe [59] reported increased tip forces when latency increased up to 500 ms in VF and VHF conditions in a peg-in-hole task, while the tip forces were also lower for VHF than VF at each latency. Unlike our setup, their system used a passivity based wave variable controller for stability, but we still demonstrated the same benefits in the absence of a passivity controller.

Of the benefits seen when adding haptic feedback in real-time operation, success rate, accuracy and trust improvements were not maintained at higher latencies. In fact, haptic feedback possibly even reduced accuracy at 2600 ms latency for Task B. This expands upon the results of Ivanova et al. [22], where a reduction in VHF performance (accuracy and smoothness) benefits was seen at delays up to 540 ms. In bilateral haptic telemanipulators such as this, the user becomes the damper in the system. If they are not ready to absorb the force feedback, it will move the local haptic arm and send a new command. By absorbing the feedback force, users maintain a steady position and constant contact force. At low latencies, users can predict when the robot is about to make contact and are, hence, ready to damp the feedback. At higher latencies, however, the user needs to predict when the haptic feedback will arrive some seconds in advance to be ready to absorb it. If they react too late, the feedback will make the local haptic controller move, and therefore the remote robot too, and if they react too early, they may send an unwanted command. This makes unexpected movements more likely with haptic feedback at high latency, and accounts for haptic feedback improving accuracy at 0 ms, but reducing accuracy at 2600 ms delay.

Although most participants reported greater trust in the VHF than VF systems at 0 ms delay, the results in Section 3.0.6 show that perceived trust degrades more rapidly in systems with VHF than in those with VF, up to 1500 ms. From 1500-2600 ms delay, however, trust declined more steeply for VF, leading to similar trust scores for both feedback conditions at the highest latency. These delays are relevant for systems in space (Table 1). Our results show that, at 1500 ms delay VHF leads to reduced trust compared to the VF system. This is within the range of communication delays from Earth to LEO via GEO relay, so has direct implications for applications in IOSM. Operators may be more willing to rely on a VF system during these medium delays. This preference was also reported by expert operators of telemanipulators from the underwater and ordnance disposal sectors, in our previous work [34]. Our results build on those of Rogers et al. [45] and Khasawneh et al. [28], who found no significant reduction in trust when increasing the latency from 0-500 ms. These studies used VF mobile robots operated with a joystick and, hence, differ from our work which involves a physical manipulator using a haptic device. Despite this, in both VF mobile robot studies, increased latency did not reduce trust when operating a single robot, which is also supported by our results at lower latencies (VF-0 - VF-600). It should also be noted that they used between-subjects designs, which make comparing subjective scores between participants more difficult than the within-subjects design that we used in our work, albeit at the cost of having fewer repetitions for each condition.

VHF provides more information on the task to the user, which improves transparency [4, 19]. However, this information becomes more temporally disconnected as latency increases, which can be more cognitively demanding. Human response times to visual and haptic stimuli vary due to differing neural pathways; it takes approximately 100 ms for meaningful fingertip reaction to a haptic stimulus, whereas this increases to >200 ms for a visual stimulus [26]. This difference may be the cause of the more rapid decline in trust seen in VHF conditions as latency increased. The destabilising effect of haptic feedback in the delayed teleoperation systems [48] may have also negatively impacted trust more in the VHF rather than the VF conditions as latency increased. Without any prior knowledge of the system, it was not obvious to participants why the haptic arm was "pushing them" off the path seconds after they had moved. Participants also reported that the robot was "unhelpful", "disruptive", "interfering" and "counter-productive" in the high latency VHF conditions, but not in the VF equivalent.

It may seem intuitive that trust would degrade with increased latency, but it is not yet clear how significant the problem is, or how to solve it. These results quantify the problem in order to test the effectiveness of one possible solution, haptic feedback. Furthermore, these results identify that adding haptic feedback would range from improving trust (with no delay), to reducing trust (with high delay). This corroborates the findings of our previous work, where expert operators reported mixed opinions on whether haptic feedback would build their trust in a delayed system [34].

Although move velocity increased and there was no change in the wait frequency or duration, adding haptic feedback led to an increase in the completion time at all latencies for Task B. This was due to an increase in the number of movements which did not progress the task; for example, spending more time finalising the peg alignment before fully inserting it into the hole. Completion times for VF and VHF conditions became more similar as latency increased.

Of the two tasks explored here, Task B better demonstrates the benefits of haptic feedback, particularly in real-time. However, it was also susceptible to the negative aspects of haptic feedback, e.g., increased completion time and, hence, there is a trade-off to be made depending on the desired application. In Task A, haptic feedback offered fewer benefits, but did not hinder performance. The tasks used here represent a subset of the possible actions which might be required of telemanipulators, and are analogous to some of the tasks to be encountered in space applications. Comparing just these two examples highlights how the impact of haptic feedback can vary between tasks. Both tasks were carried out in 3-DOF using haptic feedback to support visual feedback, although the camera views for Task B were partially occluded. Even for these relatively simple examples, we found that participants approached the task using a variety of strategies, which increased the variance within our results. Prescribing a strategy for participants to use may be an effective method of decreasing this variance when assessing more complex tasks.

There will be other telemanipulation tasks encountered in space with different requirements and constraints, and will, therefore, be differently suited to haptic feedback devices. For larger movements, a simple GUI based control is often preferable to a complex haptic controller [4]. A similar view was held by expert NASA robot operators [43]. Direct telemanipulation will likely be most relevant for less prescriptive tasks, where it is difficult to model the environment or automate control and, thus, the adaptability of human decision making is required [51]. More research is therefore needed into high latency systems for the variety of potential tasks.

4.3 Limitations and Future Work

One of the aims of this work was to assess the use of 'move-and-wait' behaviour as a measurement of demonstrated trust. We have defined trust as willingness to rely on a system, based on confidence that it will behave as expected without negative consequences. If 'unwillingness' is demonstrated by not taking any actions that might have negative consequences, we can say that 'willingness' is demonstrated by taking more actions with risks of negative consequences. Waiting 100% of the time would demonstrate complete 'unwillingness', whereas waiting less, or moving more (e.g., with greater velocity) would demonstrate greater 'willingness'. 'Willingness' does not in itself demonstrate trust, however, as it does not take into account the user's understanding of the action's associated risks, nor does it consider their attitude towards the risks. We define four possible states for a 'willing' user: cautious (unaware of the risk, concerned with the consequences), oblivious (unaware of the risk, unconcerned of the consequences), reckless (aware of the risk, unconcerned with the consequences), or trusting (aware of the risk, concerned with the consequences). Our results show that as latency is increased, participants use reduced moving velocities and increased wait frequencies, demonstrating a reduced 'willingness' to use the system. However, without identifying their understanding of the system, we cannot be sure that we were not detecting cautious, oblivious or reckless behaviour. It is plausible that untrained users may be cautious or oblivious, or that participants in a safe and approved laboratory experiment may be unconcerned with the consequences [15], and were therefore reckless, particularly when compared with a real-world teleoperation application in space. Therefore, future work is required to assess this measure further, with a particular emphasis on using trained participants who understand the risks of using their systems, and ensuring the participants are suitably concerned with avoiding any negative consequences of errors.

Although our sample size was comparable to (or greater than) similar studies [4, 30, 35, 36, 42, 45, 55, 57, 59], one limitation of this work was the use of only naive operators. This introduced a large variance within the sample, making the differences between some conditions less explicit. In the space industry, operators are experts, trained to adapt to delay conditions ($\leq$ 600 ms) [28, 58]. As trust is developed over time [8, 19], experience and training will have an impact. We have previously identified this as a key requirement for developing trust in teleoperated systems [34]. Future work will look into how training impacts performance and trust at higher latencies. Training users in the same way should reduce sample variance, meaning smaller samples should show significant results (overcoming the practical constraints of conducting larger user studies).

The tasks tested in this work were simplified versions based on examples from IOSM. More complex and realistic versions of these tasks could also be explored in future work to determine whether the same trends are seen. For example, a modified Task A could be to follow an irregular, undulating surface using 6-DOF control, or Task B could be to mate fragile electrical connectors without any visual feedback. Other tasks could be designed in the future to explore other capabilities in examples relevant to the space environment such as peeling multi-layer insulation [29] or component inspection without visual feedback. Future work should map the expected tasks for IOSM, or other space applications, against a categorised list of capabilities required for those tasks, for example, 'maintain contact with a surface', 'grasp an object', or 'manipulate held objects'. From there, research can determine how appropriate haptic feedback is for each task, given the latency and the required capabilities. This framework could also be extended to other feedback modalities in the future, for example, vibrotactile feedback may be more compatible with latency [37].

While haptic feedback does improve trust in real-time systems, this is not the case at higher latencies. In order to facilitate wider adoption of high latency telemanipulation systems, further research is required to identify which features can improve trust in the presence of large delays. Additional work is required to experimentally assess technical features such as control [41, 59] or model-based [7, 44] approaches, as well as non-technical, user-centric approaches such as training [34], for their effectiveness in developing trust.

5 CONCLUSIONS

Increasing latency negatively affects several performance metrics, with most detrimental impact occurring between 600 and 1500 ms. This is within the range of best-case transmission times for Earth–LEO communications via a GEO relay and, hence, it would be impactful for IOSM tasks. This limit may also be exceeded in non-optimised communication systems between Earth Surface–Earth Moon L1 and Earth Moon L2–Moon Surface, extending these implications, beyond IOSM, to activities to and from the Lunar Gateway.

Haptic feedback undoubtedly has some advantages in telemanipulator systems, particularly in real-time applications. At higher latencies, however, haptic feedback can sometimes become a hinderance, rather than a help. Where, and to what degree, haptic feedback will be beneficial in delayed systems will be task dependent. For example, if the top priority for a peg in hole task is a fast completion time with little importance given to accuracy (e.g., connecting compliant connectors), adding haptic feedback would not necessarily be the best choice. It is our recommendation that, when considering which features to include in these systems, system developers should first consider the type of task to be conducted and what attributes would be most valuable.

REFERENCES

[1] Jordi Artigas, Ribin Balachandran, Marco De Stefano, Michael Panzirsch, Roberto Lampariello, Alin Albu-Schaeffer, Jan Harder, and Juergen Letschnik. 2016. Teleoperation for on-orbit servicing missions through the ASTRA geostationary satellite. *IEEE Aerospace Conference Proceedings* 2016-June (2016). https://doi.org/10.1109/AERO.2016.7500785

[2] Jordi Artigas, Ribin Balachandran, Cornelia Riecke, Martin Stelzer, Bernhard Weber, Jee Hwan Ryu, and Alin Albu-Schaeffer. 2016. KONTUR-2: Force-feedback teleoperation from the international space station. In *Proceedings - 2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden, 1166–1173. https://doi.org/10.1109/ICRA.2016.7487246

[3] A Avizienis, J Laprie, and B Randell. 2001. Fundamental Concepts of Dependability. 010028 (2001). http://citeseer.ist.psu.edu/avizienis00fundamental.html

[4] Keyhan Kouhkiloui Babarahmati, Carlo Tiseo, Quentin Rouxel, Zhibin Li, and Michael Mistry. 2020. Robust High-Transparency Haptic Exploration for Dexterous Telemanipulation. *arXiv preprint arXiv:2003.01463* (2020). http://arxiv.org/abs/2003.01463

[5] Ribin Balachandran, Jee Hwan Ryu, Mikael Jorda, Christian Ott, and Alin Albu-Schaeffer. 2020. Closing the Force Loop to Enhance Transparency in Time-delayed Teleoperation. *Proceedings - IEEE International Conference on Robotics and Automation* (2020), 10198–10204. https://doi.org/10.1109/ICRA40945.2020.9197420

[6] Sebastian Bauhoff. 2014. Self-Report Bias in Estimating Cross-Sectional and Treatment Effects. In *Encyclopedia of Quality of Life and Well-Being Research*, Alex C Michalos (Ed.). Springer Netherlands, Dordrecht, 5798–5800. https://doi.org/10.1007/978-94-007-0753-5{_}4046

[7] Hadi Beik-Mohammadi, Matthias Kerzel, Benedikt Pleintinger, Thomas Hulin, Philipp Reisich, Annika Schmidt, Aaron Pereira, Stefan Wermter, and Neal Y. Lii. 2020. Model Mediated Teleoperation with a Hand-Arm Exoskeleton in Long Time Delays Using Reinforcement Learning. *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020* September (2020), 713–720. https://doi.org/10.1109/RO-MAN47096.2020.9223477

[8] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with Trust for Human-Robot Collaboration. *ACM/IEEE International Conference on Human-Robot Interaction* (2018), 307–315. https://doi.org/10.1145/3171221.3171264

[9] K. Chopra and W. A. Wallace. 2003. Trust in electronic environments. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003* (2003). https://doi.org/10.1109/HICSS.2003.1174902

[10] Mauricio Correa, Daniel Cárdenas, Diego Carvajal, and Javier Ruiz-del-solar. 2022. Haptic Teleoperation of Impact Hammers in Underground Mining. *Applied Sciences (Switzerland)* 12, 3 (2022). https://doi.org/10.3390/app12031428

[11] Peter B. de Selding. 2010. MDA may scrap in-orbit satellite servicing plan. https://spacenews.com/mda-may-scrap-orbit-satellite-servicing-plan/

[12] Issam El Rassi and Jean Michel El Rassi. 2020. A review of haptic feedback in tele-operated robotic surgery. *Journal of Medical Engineering and Technology* 44, 5 (2020), 247–254. https://doi.org/10.1080/03091902.2020.1772391

[13] Akuadasuo Ezenyilimba, Margaret Wong, Alexander Hehr, Mustafa Demir, Alexandra Wolff, Erin Chiou, and Nancy Cooke. 2023. Impact of transparency and explanations on trust and situation awareness in human–robot teams. *Journal of cognitive engineering and decision making* 17, 1 (2023), 75–93.

[14] William R. Ferrell. 1965. Remote manipulation with transmission delay. *IEEE Transactions on Human Factors in Electronics* HFE-6, 1 (1965), 24–32. https://doi.org/10.1109/thfe.1965.6591253

[15] Rebecca Flook, Anas Shrinah, Luc Wijnen, Kerstin Eder, Chris Melhuish, and Séverin Lemaignan. 2019. On the impact of different types of errors on trust in human-robot interaction: Are laboratory-based HRI experiments trustworthy? *Interaction Studies* 20, 3 (2019), 455–486.

[16] Werner Grandl. 2007. Lunar Base 2015 Stage 1. *Acta Astronautica* 60, 4-7 SPEC. ISS. (2007), 554–560. https://doi.org/10.1016/j.actaastro.2006.09.031

[17] Blake Hannaford and Jee Hwan Ryu. 2002. Time-domain passivity control of haptic interfaces. *IEEE Transactions on Robotics and Automation* 18, 1 (2002), 1–10. https://doi.org/10.1109/70.988969

[18] Blake Hannaford, Laurie Wood, Douglas A McAffee, and Haya Zak. 1991. Performance evaluation of a six-axis generalized force-reflecting teleoperator. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 620–633.

[19] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434. https://doi.org/10.1177/0018720814547570

[20] A Scott Howe and Silvano Colombano. 2010. The Challenge of Space Infrastructure Construction. In *AIAA Space 2010 Conference & Exposition*. 8619.

[21] Huan Hu, Camilo Perez, Han Xu Sun, and Martin Jagersand. 2017. Performance of predictive display teleoperation under different delays with different degree of freedoms. *Proceedings - 2016 International Conference on Information System and Artificial Intelligence, ISAI 2016* (2017), 380–384. https://doi.org/10.1109/ISAI.2016.0087

[22] E Ivanova, J Eden, S Zhu, G Carboni, A Yurkewich, and E Burdet. 2021. Short Time Delay Does Not Hinder Haptic Communication Benefits. *IEEE Transactions on Haptics* 14, 2 (2021), 322–327. https://doi.org/10.1109/TOH.2021.3079227

[23] Caroline Jay, Mashhuda Glencross, and Roger Hubbold. 2007. Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment. *ACM Transactions on Computer-Human Interaction* 14, 2 (8 2007). https://doi.org/10.1145/1275511.1275514

[24] Sharon A. Jefferies, Carey M. McCleskey, Brian Nufer, Roger A. Lepsch, Raymond G. Merrill, David North, John Martin, and D. R. Komar. 2015. Viability of a reusable in-space transportation system. *AIAA SPACE 2015 Conference and Exposition* (2015), 1–22. https://doi.org/10.2514/6.2015-4580

[25] Jiun-Yin Jian, Ann Bisantz, and Colin Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4, 1 (3 2000), 53–71. https://doi.org/10.1207/S15327566IJCE0401{_}04

[26] R. S. Johansson and J. R. Flanagan. 2008. Tactile Sensory Control of Object Manipulation in Humans. In *The Senses: A comprehensive reference*, Esther Gardner and Jon H. Kaas (Eds.). Vol. 6. Elsevier Inc., Chapter 6.05, 67–86.

[27] Donald J Kessler, Burton G Cour-palais, and Burton G Cour-Palais. 1978. Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research: Space Physics* 83, A6 (1978), 2637–2646.

[28] Amro Khasawneh, Hunter Rogers, Jeffery Bertrand, Kapil Chalil Madathil, and Anand Gramopadhye. 2019. Human adaptation to latency in teleoperated multi-robot human-agent search and rescue teams. *Automation in Construction* 99, January 2018 (2019), 265–277. https://doi.org/10.1016/j.autcon.2018.12.012

[29] Thierry Laliberte ́ and Cle ́ ment M Gosselin. 2006. Development of a blanket manipulation tool for satellite servicing. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 42568. 843–850.

[30] J. Corde Lane, Craig R. Carignan, Brook R. Sullivan, David L. Akin, Teresa Hunt, and Rob Cohen. 2002. Effects of time delay on telerobotic control of neutral buoyancy vehicles. *Proceedings - IEEE International Conference on Robotics and Automation* 3, May (2002), 2874–2879. https://doi.org/10.1109/ROBOT.2002.1013668

[31] Dan Lester. 2013. Achieving human presence in space exploration. *Presence: Teleoperators and Virtual Environments* 22, 4 (2013), 345–349. https://doi.org/10.1162/PRES{_}a{_}00160

[32] Dan Lester and Harley Thronson. 2011. Human space exploration and human spaceflight: Latency and the cognitive scale of the universe. *Space Policy* 27, 2 (2011), 89–93. https://doi.org/10.1016/j.spacepol.2011.02.002

[33] Andrew Lewis, David Drajeske, John Raiti, Angelique Berens, Jacob Rosen, and Blake Hannaford. 2020. RAVEN-S : Design and Simulation of a Robot for Teleoperated Microgravity Rodent Dissection Under Time Delay. (2020), 11332–11337.

[34] Joe Louca, John Vrublevskis, Kerstin Eder, and Antonia Tzemanaki. 2023. Elicitation of trustworthiness requirements for highly dexterous teleoperation systems with signal latency. *Frontiers in Neurorobotics* 17 (2023), 1187264.

[35] Mitchell J H Lum, Jacob Rosen, Hawkeye King, Diana C W Friedman, Thomas S Lendvay, Andrew S Wright, Mika N Sinanan, and Blake Hannaford. 2009. Teleoperation in surgical robotics–network latency effects on surgical performance. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 6860–6863.

[36] Ryan D. Madder, Stacie VanOosterhout, Abbey Mulder, Jared Bush, Samuel Martin, Adam J. Rash, Jose Mariano Tan, Jessica L. Parker, Andrew Kalafut, Yao Li, Nicholas Kottenstette, Per Bergman, and Brent Nowak. 2020. Network latency and long-distance robotic telestenting: Exploring the potential impact of network delays on telestenting performance. *Catheterization and Cardiovascular Interventions* 95, 5 (2020), 914–919. https://doi.org/10.1002/ccd.28425

[37] M.J. Massimino and T.B. Sheridan. 1992. Sensory Substitution for Force Feedback in Teleoperation. *IFAC Proceedings Volumes* 25, 9 (1992), 109–114. https://doi.org/10.1016/s1474-6670(17)50178-0

[38] Syeda Nadiah Fatima Nahri, Shengzhi Du, and Barend Jacobus Van Wyk. 2022. A Review on Haptic Bilateral Teleoperation Systems. *Journal of Intelligent and Robotic Systems: Theory and Applications* 104, 1 (2022). https://doi.org/10.1007/s10846-021-01523-x

[39] NASA. 2010. *On-Orbit Satellite Servicing Study Project Report.* Technical Report. National Aeronautics and Space Administration, Goddard Space Flight Centre.

[40] Reut Nomberg and Ilana Nisky. 2023. Human-in-the-Loop Stability Analysis of Haptic Rendering With Time Delay - The Effect of Delayed Feedback Control in the Operator Model. *IEEE Robotics and Automation Letters* 8, 10 (2023), 6379–6386. https://doi.org/10.1109/LRA.2023.3301773

[41] Michael Panzirsch, Harsimran Singh, Thomas Kruger, Christian Ott, and Alin Albu-Schaffer. 2020. Safe Interactions and Kinesthetic Feedback in High Performance Earth-To-Moon Teleoperation. *IEEE Aerospace Conference Proceedings* March (2020). https://doi.org/10.1109/AERO47225.2020.9172665

[42] Manuela Perez, Song Xu, Sanket Chauhan, Alyssa Tanaka, Khara Simpson, Haidar Abdul-Muhsin, and Roger Smith. 2016. Impact of delay on telesurgical performance: study on the robotic simulator dV-Trainer. *International Journal of Computer Assisted Radiology and Surgery* 11, 4 (4 2016), 581–587. https://doi.org/10.1007/s11548-015-1306-y

[43] Will Pryor, Balazs P. Vagvolgyi, William J. Gallagher, Anton Deguet, Simon Leonard, Louis L. Whitcomb, and Peter Kazanzides. 2019. Experimental evaluation of teleoperation interfaces for cutting of satellite insulation. *Proceedings - IEEE International Conference on Robotics and Automation* 2019-May (2019), 4775–4781. https://doi.org/10.1109/ICRA.2019.8793968

[44] Will Pryor, Balazs P. Vagvolgyi, William J. Gallagher, Anton Deguet, Simon Leonard, Louis L. Whitcomb, and Peter Kazanzides. 2020. Interactive Planning and Supervised Execution for High-Risk, High-Latency Teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 4775–4781. https://doi.org/10.1109/ICRA.2019.8793968

[45] Hunter Rogers, Amro Khasawneh, Jeffery Bertrand, and Kapil Chalil Madathil. 2017. An investigation of the effect of latency on the operator's trust and performance for manual multi-robot teleoperated tasks. *Proceedings of the Human Factors and Ergonomics Society* 2017-Octob (2017), 390–394. https://doi.org/10.1177/1541931213601579

[46] Kristin E. Schaefer. 2013. *The Perception And Measurement Of Human-robot Trust.* Ph. D. Dissertation. University of Central Florida. https://doi.org/10.1016/0022-3999(73)90100-1

[47] George R. Schmidt, Geoffrey A. Landis, and Steven R. Oleson. 2011. HERRO missions to mars and venus using telerobotic surface exploration from orbit. *AIAA SPACE Conference and Exposition 2011* January 2016 (2011). https://doi.org/10.2514/6.2011-7343

[48] Thomas B Sheridan. 1993. Space Teleoperation Through Time Delay: Review and Prognosis. *IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION* 9, 5 (1993).

[49] T. B. Sheridan and W. R. Ferrell. 1963. Remote Manipulative Control with Transmission Delay. *IEEE Transactions on Human Factors in Electronics* HFE-4, 1 (1963), 25–29. https://doi.org/10.1109/THFE.1963.231283

[50] Robert Skilton, Nathan Hamilton, Robert Howell, Chris Lamb, and Jorge Rodriguez. 2018. MASCOT 6: Achieving high dexterity tele-manipulation with a modern architectural design for fusion remote maintenance. *Fusion Engineering and Design* 136, March (2018), 575–578. https://doi.org/10.1016/j.fusengdes.2018.03.026

[51] Jan Smisek. 2017. *Systematic Framework for Teleoperation with Haptic Shared Control.* Ph. D. Dissertation. https://doi.org/10.4233/uuid:e7d8b0d5-8d7a-4d68-b106-c37f30d455a4

[52] Rachel E. Stuck, Brianna J. Tomlinson, and Bruce N. Walker. 2022. The importance of incorporating risk into human-automation trust. *Theoretical Issues in Ergonomics Science* 23, 4 (2022), 500–516. https://doi.org/10.1080/1463922X.2021.1975170

[53] Andrew J Tattersall and Penelope S Foord. 1996. An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics* 39, 5 (1996), 740–748.

[54] Antonia Tzemanaki, Gorkem Anil Al, Chris Melhuish, and Sanja Dogramadzi. 2018. Design of a wearable fingertip haptic device for remote palpation: Characterisation and interface with a virtual environment. *Frontiers in Robotics and AI* 5 (2018), 62.

[55] Steve Vozar, Simon Leonard, Peter Kazanzides, and Louis L. Whitcomb. 2015. Experimental evaluation of force control for virtual-fixture-assisted teleoperation for on-orbit manipulation of satellite thermal blanket insulation. *Proceedings - IEEE International Conference on Robotics and Automation* 2015-June, June (2015), 4424–4431. https://doi.org/10.1109/ICRA.2015.7139811

[56] Janis Wojtusch, Daniela Taubert, Thorsten Graber, and Kim Nergaard. 2019. Evaluation of Human Factors for Assessing Human-Robot Interaction in Delayed Teleoperation. *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018* (2019), 3787–3792. https://doi.org/10.1109/SMC.2018.00641

[57] Song Xu, Manuela Perez, Kun Yang, Cyril Perrenot, Jacques Felblinger, and Jacques Hubert. 2015. Effect of latency training on surgical performance in simulated robotic telesurgery procedures. *The International Journal of Medical Robotics and Computer Assisted Surgery* 11, 3 (9 2015), 290–295.

https://doi.org/10.1002/rcs.1623

[58] Zhihui Xue, Jinguo Liu, Chenchen Wu, and Yuchuang Tong. 2021. Review of in-space assembly technologies. *Chinese Journal of Aeronautics* 34, 11 (2021), 21–47. https://doi.org/10.1016/j.cja.2020.09.043

[59] Michael C. Yip, Mahdi Tavakoli, and Robert D. Howe. 2012. Performance analysis of a haptic telemanipulation task under time delay. *Advanced Robotics* 25, 5 (2012), 651–673. https://doi.org/10.1163/016918611X558216

Received 1 October 2022; revised 26 October 2023; accepted 8 February 2024