The Effects of Predictive Displays on Performance in Driving Tasks with Multi-Second Latency: Aiding Tele-Operation of Lunar Rovers

Adrian Matheson[1], Birsen Donmez[1], Faizan Rehmatullah[1],
Piotr Jasiobedzki[2], Ho-Kong Ng[2], Vivek Panwar[2], Mufan Li[1].
[1]University of Toronto, Toronto, ON, Canada; [2]MacDonald Dettwiler Space and Advanced Robotics Ltd., Brampton, ON, Canada

Tele-operation of a Lunar rover from a control station on Earth involves a latency of several seconds due primarily to the finite speed (light-speed) of command and sensor signals, and this latency creates a difficult control task for the human operator. Two predictive displays, which seek to aid viewer perception of *present* events, were designed and evaluated for the specific task of driving a rover with multi-second latency. These displays provided visual information to the human operator on the rover's real-time locomotion, as predicted from control inputs executed by the operator. A human-subject experiment with 12 participants was conducted in which the participants navigated an actual rover through obstacle courses. There were four experimental conditions repeated by each participant: (1) delayed video feed only, (2, 3) two predictive displays based on delayed video feed, and (4) a reference condition of video feed with no delay. Inferential statistics show that both predictive displays significantly improved performance in terms of time taken to complete the courses, and one of the displays facilitated performance approaching that with no delay. No trends were observed in terms of collisions with or encroachments near obstacles.

INTRODUCTION

Uncrewed rovers are currently the main platforms for surface-based exploration of extraterrestrial bodies such as solid planets and moons. The great distances between these rovers and any rover control station mean that they must operate with some combination of autonomy and remote operation, where the latter has a latency of several seconds (Earth's moon), tens of minutes (Mars), or more. Earth's moon is close enough, at just over one light-second away from Earth, to make full tele-operation of a Lunar rover very nearly practical. However, command signals from a rover station on Earth take time to reach the rover on the moon, and sensor signals (which are needed to inform operators of which commands to execute next) take an equal time to pass from the rover to the station. Adding in the time required for processing of both information streams at their source and sink, the total latency involved is approximately 3 seconds. Even with a latency of only several seconds, tele-operation in a time-delayed environment, in particular for the task of basic driving, is a difficult and highly stressful task for humans (Sheridan, 1993; Wright, 2007). It results in high levels of cognitive workload (Lovi et al., 2010). Further, extended exposure to such an environment can also lead to mental fatigue due to cognitive overload (Lim et al., 2010).

Control with latency can increase operator cognitive workload specifically due to a lack of clear correspondence between input and output. "[A] principle to reduce cognitive workload is to maintain a correlation with commands issued by the operator and the expected result of those commands as observed by the movement of the robot and changes in the interface (Nielsen, Goodrich, & Ricks, 2007, p. 936)." Lunar tele-operation, here defined as the operation of an instrument on Earth's moon from a control station on Earth, has this basic problem of lack of (quick) input-output correspondence, and it requires the user to make a predictive map of the outcomes of the input commands.

Previous research shows that the use of predictive displays in time-delayed situations can greatly reduce the time to complete tasks and improve accuracy. Bejcyz, Kim, and Venemo (1990) implemented such a display using high-fidelity graphics to generate a "phantom robot" controllable in real time. Preliminary experiments showed a significant performance improvement. Similarly, with use of predictive displays on simple manipulation tasks, a time improvement of over 70% was reported by Noyes and Sheridan (1984).

Given a 3-second delay, by the time a rover operator reacts to what appears to be a nearby hazard, the rover may already have encountered the hazard. To prevent this and mitigate the need for cumbersome rover operation procedures, MacDonald-Dettwiler and Associates Ltd. (MDA) is working toward the design of multiple predictive displays for the forthcoming Lunar Exploration Light Rover (LELR) (McCoubrey et al., 2012). MDA had previously designed visual aids for rover operation, and through experiments performed in co-operation with human factors experts, determined them to be valuable (Langley, Nimelman, Mukherji, L'Archeveque, & Milgram, 2010). These previous displays primarily aided navigation, however, whereas the aim for the newer predictive displays, which are evaluated in this paper, is to increase operator performance for manoeuvres on a smaller time scale (operation versus navigation).

Predictive Displays

The two predictive displays developed by MDA for potential use in the LELR utilize images from a two-dimensional, forward-facing "drive camera". Both rely on recording driving commands issued by the operator and predicting the rover position based on this information, as shown in Figure 1. Both assume these commands will be properly followed by the rover, and the accuracy of the predictions also depends on the accuracy of rover model (kinematic and dynamic) and of the expected traction in Lunar soil. Crucially, these early models also assume a horizontal planar land surface. This modelling and prediction is required only for approximately 3 s of travel at speeds up to 20 km/h, however. Also, since the predictive displays are merely modifications of the most recent video, any errors due to modelling and prediction are not cumulative.

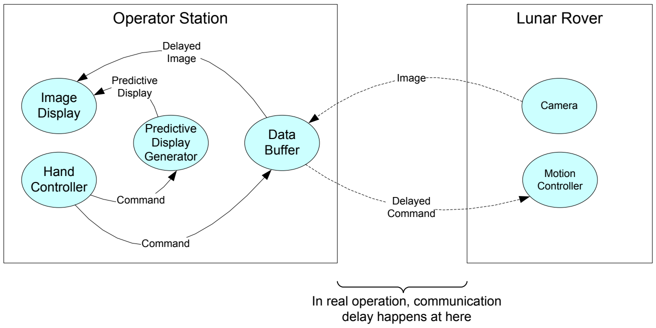

Figure 1 – Overview of the Lunar rover system including predictive displays.

Projected display: an immersed display simulating 0-second delay video. This display is obtained by estimating the current rover position within the delayed drive camera image, finding the current field of view edges given the rover's location and orientation, and manipulating the delayed image through cropping and projection, to approximate the view from the current rover location (Figure 2). This is a very good approximation while driving ahead, but suffers from perspective distortions during and immediately following turns—especially sharp turns. In the latter situation, slowing down (typical behaviour before a sharp turn) allows the view to "catch up". Figure 3 (Left) explains this display further.

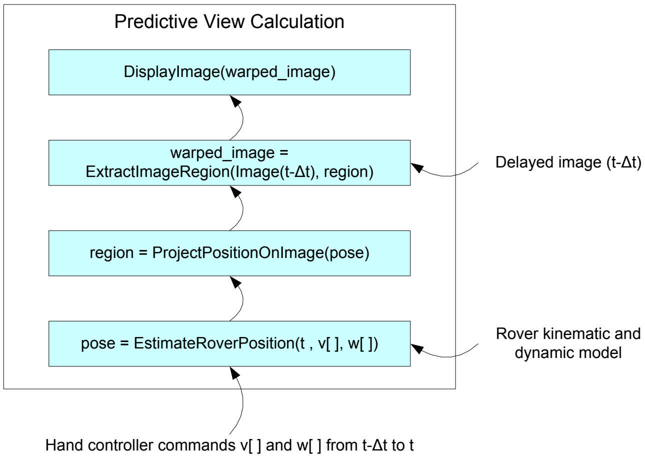

Figure 2 – Logic for the projected display.

Avatar display: a tethered display with 3-second delay video and a 0-second delay rover position overlay. This display involves the delayed drive camera image overlaid with symbols indicating the estimated current rover position and several recent positions (from 0 to 3 s in the past), as shown in Figure 4. The current position symbol (the "avatar") is a green quadrangle which depicts the rover footprint and responds to controls in real-time. The symbols for recent positions are white quadrangles which are chronologically evenly spaced and serve to outline the recent path of the rover as well as to convey speed. Figure 3 (Right) shows a sample of this display.

Figure 3 – Projected (left) and Avatar (right) views for the same current rover position and 3-second history.
(Left) Projected display. The large panel is an illustrative view of the raw video (3-s delay) overlaid with a red border marking the edges of current rover field of view. The small panel is the view seen by the operator. It is a projection of the red-bordered area into a full screen.
(Right) Avatar display. In this case the operator does see the full delayed video area pictured, and does see the overlay pictured. The green quadrangle and white quadrangles represent the present location and past locations of the rover, respectively.

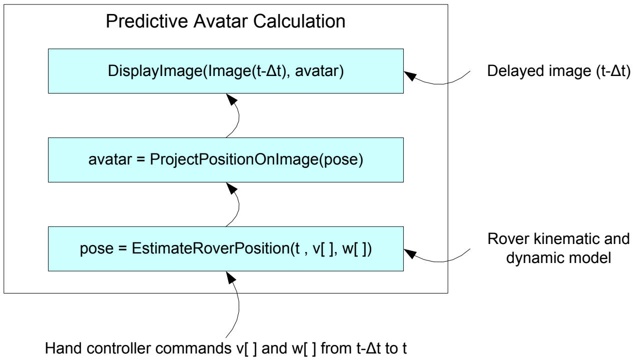

Figure 4 – Logic for the avatar display.

METHOD

To evaluate the predictive displays, a human subject experiment was conducted at the MDA facility in Brampton, Ontario, where participants tele-operated a rover through an obstacle course. Three of the study conditions involved artificial delay of the video stream by 3 s, to approximate the effect of Lunar tele-operation of a rover. In one of these conditions subjects received only the delayed video, and in each of the remaining conditions subjects used one of the predictive displays. The study also included a fourth condition in which subjects received video with no delay.

Participants. 12 participants were recruited from the University of Toronto Engineering and Computer Science student bodies. The average age of the participants was 22 years with a standard deviation of 3.6 years. The participants had no known prior experience with rover tele-operation or the types of predictive displays under study, but were trained prior to data collection, as detailed in following sections.

Experimental design. The experiment was a within-subject design with each participant completing four experimental conditions. A 3-s delay was introduced to the forward video feed for the following three conditions: raw delayed video (no assistance), projected display, and avatar display. The fourth condition provided a baseline of driving performance and style (e.g., speed-accuracy trade-off) as well as a benchmark time for completing the task. It involved presenting the raw video in real time and was administered last for each participant. The three 3-s delay conditions were presented to the participants in a counterbalanced order to control for possible learning effects.

Apparatus and experimental design. An iRobot ATRV Jr. served as the rover for this experiment. The rover was equipped with an AXIS 212 PTZ network camera (resolution 640 by 480 pixels), mounted forward-facing, to supply the rover raw forward video. This video was transferred from the rover to the control computer via a wireless router, and for conditions with time delays the delay was produced by buffering the video at the control computer and delaying its processing. That is, rather than delaying control and video signals each by 1.5 s, the video was simply delayed by 3 s.

Participants were situated in an indoor control room and viewed the forward video through a standard computer monitor at a normal viewing distance while seated or standing, according to preference. Participants controlled the rover using a MicroSoft Xbox 360 controller with the left analog (i.e., high-resolution multi-step) control stick dedicated to longitudinal translation (forward-backward motion), and the right stick dedicated to yaw (azimuth adjustment, turning).

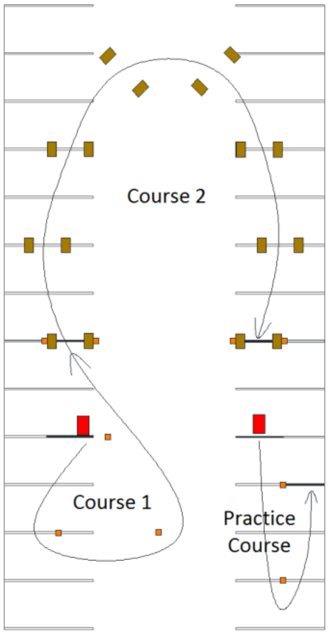

Figure 5 – Obstacle course (single continuous course through "Course 1" and "Course 2", and a practice course; red figures are rover starting positions; orange and brown figures are obstacles).

The task for the experiment was to drive the rover through the obstacle course depicted in Figure 5. The obstacles were positioned with relatively little room to manoeuvre in order to provide a challenging driving task. The obstacles were placed relatively regularly to avoid presenting a challenging navigation task and the associated cognitive load and learning effects. It was expected that a challenging navigation task would have caused variance in the data not closely related to the experimental manipulation of interest. Before each condition, participants were also clearly explained the layout of the course so as to, again, minimize the cognitive workload. Participants were, however, not allowed to view the obstacle course directly at any time, and only ever saw it via camera during driving.

The obstacle course was a continuous two-part course with different designs for the two parts. "Course 1", the first part of the main obstacle course, included acute angle turns and obstacles on only one side of the rover. This matched an expected rover requirement to make large changes in direction, as would be required to follow high-level navigation at waypoints along a planned exploration route. It also provided a simple obstacle avoidance task in which the driver could use as much separation distance as desired between the rover and each obstacle. "Course 2", the second part of the main obstacle course, included mild turns but also pairs of closely-spaced obstacles ("gates") through which to drive the rover. This matched an expected Lunar rover requirement for low-level manoeuvring, as for passing through areas with large rocks.

Before completing the main course for each condition, participants were brought up to a rough minimum baseline of manoeuvring proficiency with the corresponding display through completing a practice course. This course comprised very sharp turns as such turns were expected to allow participants the greatest range of feedback from the interfaces.

Procedure. Prior to each experimental condition, a short explanation of the display was given. An overview of the controls and the objective of the driving task were also provided each time. The objective was to complete each run as quickly as possible with zero collisions with obstacles. After each of their runs of the practice and main courses, participants also completed forms on the Task Load Index developed by NASA (Hart & Staveland, 1988).

Data recording. Separate run times for Courses 1 and 2 were recorded. For Course 2, the closest lateral (with respect to the gate) separation distances between the rover and the nearer obstacle in each gate were also recorded, with collisions recorded as negative values matching the horizontal distance by which the rover displaced the (moveable) obstacle.

A GoPro Hero camera (resolution 1280 by 720 pixels) mounted downward-facing on the rover enabled the measuring of these separation distances. A grid aligned longitudinally and laterally with the rover was overlaid onto the video during post processing of the data to aid this measurement. Figure 6 is a snapshot from the video with the associated grid. The recording of top-down video for the real-time condition for one study participant failed. Obstacle separation distance performance analyses were thus completed without these data.

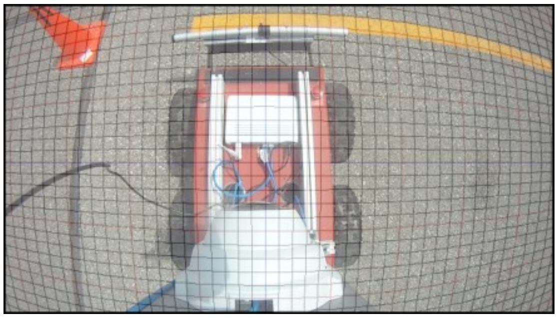

Figure 6 – Downward-facing camera view with grid overlay.

RESULTS

Display condition comparison overview. Comparisons between outcomes for the two predictive displays and the raw delay display are the most valid as experiment subjects were exposed to these conditions in counterbalanced order. Comparisons of these outcomes with those for the real-time display may be less reliable as the real-time display was always the last condition for each participant, but they are informative and included in the analysis. Any potential learning effect bias is in favour of the real-time condition.

Course completion times. Figure 7 depicts the times taken by the participants to complete Course 1. Figure 8 depicts the same for Course 2.

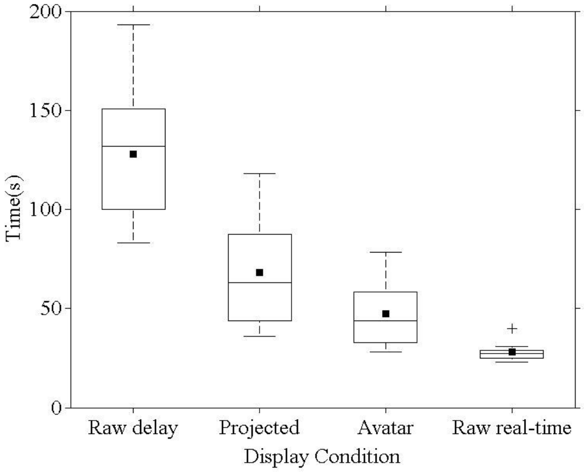

Figure 7 – Time taken to complete Course 1 (obstacles on single side).

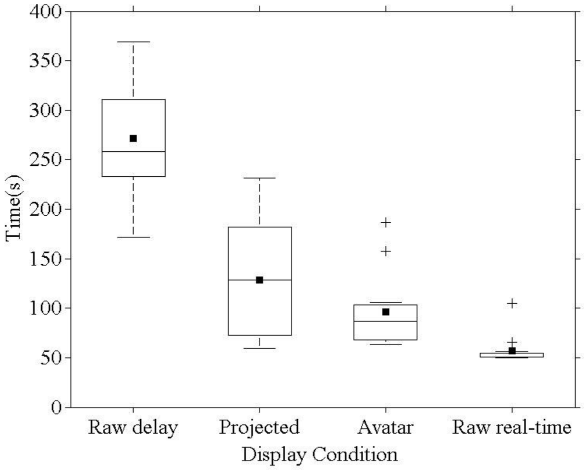

Figure 8 – Time taken to complete Course 2 (gates).

Course completion time was significantly different between the display conditions for both parts of the course (Course 1: F(3,11) = 51.2, p < .0001; Course 2: F(3,11) = 98.3, p < .0001), with both individual predictive display types yielding significantly lower times than the raw delay display for both courses, as hypothesized. Both predictive displays reduced the difference between median delayed video performance and median real-time video performance by at least half.

Post-hoc comparison between the predictive displays revealed a significant difference in times for Course 1 (t(11) = 3.27, p = .007), but none for Course 2 (t(11) = 1.92, p = .08). Performance in the real-time condition was so consistently strong that there seems to have been a ceiling effect in terms of course completion times near an optimal time based on the maximum speed of the rover and the ideal path. The resulting heteroskedasticity in the data was accounted for by the choice of an unstructured variance covariance structure.

Obstacle avoidance. Given the participant instruction to focus on course completion time (speed) and maintain a certain level of accuracy (do not collide with obstacles), course completion times should be the primary basis on which to make conclusions about predictive display effectiveness. It was expected that accuracy would vary somewhat between conditions, however

The analysis of accuracy here relies on the assumption that subjects were seeking to pass directly through the midpoint of each gate encountered in Course 2. This yields a useful 'best' path for comparison with actual driving, and allows conclusions about the effect of display on performance in terms of accuracy. It was expected that the gates were narrow enough that deliberately seeking to pass on one edge or the other of a gate in order to save a very small amount of time would not be attempted.

Ratio of actual obstacle distance to ideal distance. Gates in the second part of the course did not all have the same width (wider gates were placed on the sweeping turn), so evaluating accuracy involved calculating the ratio between the ideal and actual distances between the rover and the nearest obstacle for each participant and condition. The ideal case was the two obstacles of a gate being equidistant from the (perfectly centred) rover, and thus a ratio of 1.

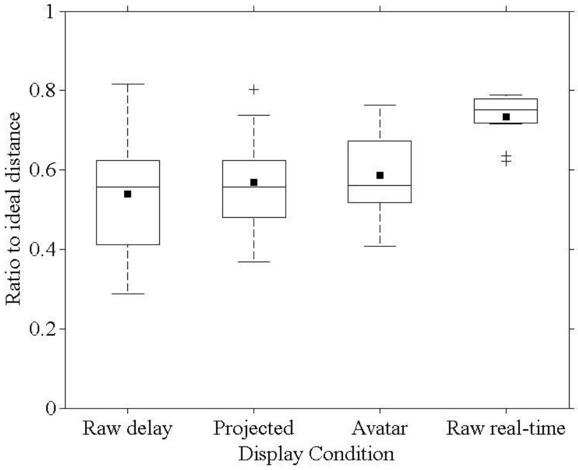

Figure 9 – Ratio of actual obstacle distance to ideal distance (1 = centred).

Display condition had a significant effect on this variable (F(3,11) = 14.3, p = .0004). An unstructured variance covariance structure was used to account for heteroskedasticity in the data. Follow-up contrasts did not reveal differences among the delayed conditions. However, as is visible in Figure 9, the real-time condition resulted in a better performance than the delayed conditions (p < .05).

Ratio of actual obstacle distance to distance in real-time condition. To compare the performance of test subjects with a more realistic benchmark than ideal performance, but more importantly to reduce some variation due to differing general task aptitude among study participants, the distances for each gate were expressed as ratios relative to each participant's performance with real-time video (Figure 10). Even this ratio was not significantly different between (delayed) display conditions, however (F(2,20) = 0.70, p = .51).

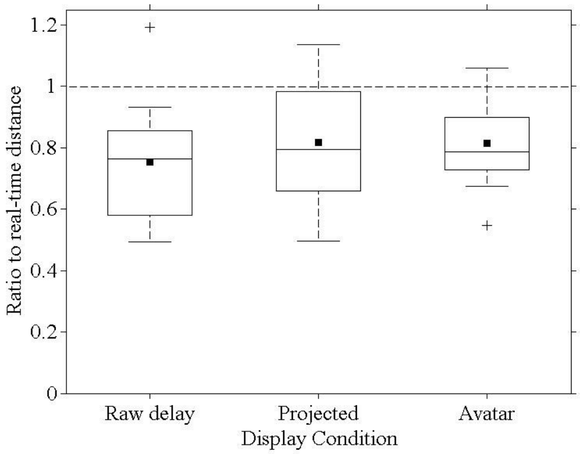

Figure 10 – Ratio of obstacle distance with delayed display (various) to obstacle distance with real-time display (< 1 indicates better in real-time).

NASA Task Load Index (TLX). NASA TLX data show a decrease in levels of mental and temporal demand, effort, frustration, and self-assessed lack of performance (physical demand was not assessed) moving from raw delayed to projected, to avatar, and finally to real-time display (Figure 11). These data provide more evidence of both predictive displays being effective.

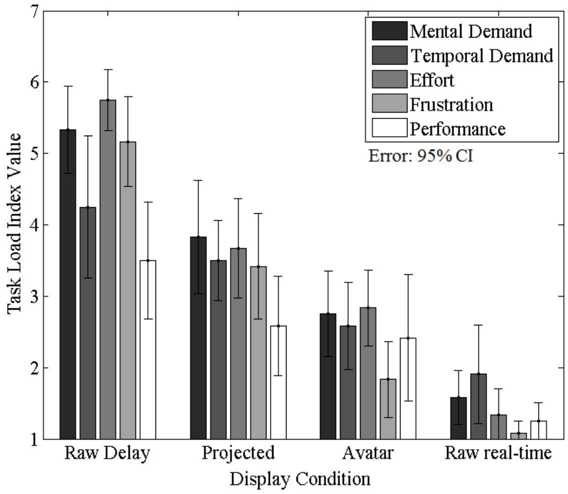

Figure 11 – NASA TLX responses (1: low, 7: high; "Performance" refers to performance being self-assessed as poor).

DISCUSSION

Since increased speed and accuracy and decreased levels of cognitive load are beneficial to tele-operators, predictive displays such as those under test in this study should continue to be explored for inclusion in delayed tele-operation systems.

The most limiting design factor behind both displays would appear to be the assumption of flat and level ground, which is not a given on Earth's moon (or nearly anywhere). Both display types could be improved through the use of 3-D sensors (such as the rangefinders and stereo cameras planned for the LELR), since topographical maps generated in real time could lead to more realistic avatar positioning and more realistic methods of view simulation through projection. Given only small departures from level for long distances however, and given no opportunity for errors to accumulate beyond the period of the transmission delay, even in their current states either of the current displays would likely be worth employing.

This experiment sought to separately test the usefulness of an augmented immersive display and of a simulated tethered (or avatar) display. The performance improvement caused by the avatar display in this study may be due to advantages inherent to a tethered display (such as being keenly aware of the vehicle width), while that caused by the projected display may be due simply to real-time video feedback. Future predictive displays could incorporate elements of both with the goal of mimicking augmented display types used in applications with negligible time delays to begin with (the vast majority of augmented displays). Such future displays might similarly raise performance beyond even that with raw real-time video, making this kind of design worth exploring.

REFERENCES

Bejczy, A. K., Kim, W. S., & Venema, S. C. (1990). The phantom robot: predictive displays for teleoperation with time delay. In *Proceedings of IEEE International Conference on Robotics and Automation* (pp. 546 – 551). Cincinnati, OH.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA_TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier Science.

Langley, C., Nimelman, M., Mukherji, R., L'Archeveque, R., & Milgram, P. (2010). Prototype development of an operator situational awareness and navigation air for manned planetary vehicles. In *Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space*. Sapporo.

Lim, J., Wu, W., Wang, J., Detre, J. A., Dinges, D. F., & Rao, H. (2010). Imaging brain fatigue from sustained mental workload: an ASL perfusion study of the time-on-task effect. *NeuroImage*, *49*(4), 3426–3435.

Lovi, D., Birkbeck, N., Herdocia, A. H., Rachmielowski, A., Jägersand, M., & Cobzaş, D. (2010). Predictive display for mobile manipulators in unknown environments using online vision-based monocular modeling and localization. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5792 –5798). Taipei.

McCoubrey, R., Langley, C., Chappell, L., Ratti, J., Ghafoor, N., Ower, C., … Barfoot, T. D. (2012). A Canadian lunar exploration light rover prototype. In *Proceedings of the International Symposium on Artificial Intelligence, Robotics, and Automation in Space*. Turin, Italy.

Nielsen, C. W., Goodrich, M. A., & Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, *23*(5), 927 –941.

Noyes, M. V., & Sheridan, T. B. (1984). A novel predictor for telemanipulation through a time delay. In *Proceedings of the 20th Annual Conference on Manual Control*. Moffett Field, CA.

Sheridan, T. B. (1993). Space teleoperation through time delay: review and prognosis. *IEEE Transactions on Robotics and Automation*, *9*(5), 592 – 606.

Wright, J. R. (2007). Challenges in robotic and robot-assistive activities on the Moon and Mars. In *Proceedings of the International Conference of Robotics and Automation*. Rome, Italy.