An Evaluation of Multi-Modal User Interface Elements for Tablet-Based Robot Teleoperation

Graeme Best[1,2] and **Peyman Moghadam**[1]
[1]Autonomous Systems, CSIRO
Pullenvale, QLD Australia
Peyman.Moghadam@csiro.au
[2]Australian Centre for Field Robotics, University of Sydney
Sydney, NSW Australia
g.best@acfr.usyd.edu.au

Abstract

For robot teleoperation systems, tablet and smart phone user interfaces provide portability and accessibility to allow anyone anywhere to connect to the system quickly and easily. This can be highly advantageous for disaster-relief robot systems where timing is critical. However, the small screen size and unconventional input methods mean that traditional teleoperation user interface elements, such as multiple visual windows, keyboards and mice, are no longer effective. In this paper we propose and investigate multi-modal user interface design principles as a solution for effective tablet-based teleoperation. We present our tablet user interface that features multiple complementary modalities of control and feedback mechanisms. It allows the operator to perceive the robot's environment and state by exciting their senses of vision, hearing and touch with overlaying visual displays, virtual 3-D audio and vibration feedback. The operator commands actions to the robot through the use of intuitive input methods utilising the touch screen and accelerometer. We analyse these design principles by performing a user study that focuses on finding what aspects of the multi-modal user interface affect the system performance during a navigation task. The study highlights advantages of multi-modal user interfaces for robot teleoperation and makes several findings to be considered when developing similar systems.

1 Introduction

Robot teleoperation systems connect humans and robots to allow a remote human operator to perform actions at a distance. These systems are particularly useful when the task is being performed in an environment that would be either too dangerous or impractical for a human. This can be seen in disaster-relief and rescue robotics systems, as highlighted by the use of robots during the recent Fukushima Daiichi nuclear disaster and the Christchurch earthquake, as well as the current research in the DARPA Robotics Challenge. Robot teleoperation systems also extend to other contexts such as military combat, underwater and outer-space [Ferre *et al.*, 2007; Kenny *et al.*, 2011; Fung, 2011]. The capabilities of fully autonomous systems are continuously improving, however there are still many unresolved complications for robots operating in unstructured, dynamic or unknown environments. In these situations semi-autonomous solutions that enable human operators to be a part of the system (human-in-the-loop) is usually more suitable since they can utilise the cognitive skills, perception and situational awareness of a human operator [Aracil *et al.*, 2007]. Also, in many situations, semi-autonomous systems are preferred due to humans having more trust in critical operator actions than the actions of an autonomous robot [Desai *et al.*, 2012; Dragan and Srinivasa, 2012], particularly when safety is a concern [Correa *et al.*, 2010].

For reliable and effective robot teleoperation systems, the human operator must have good situational awareness; that is to be able to accurately perceive the remote environment and the current state of the robot. Simultaneously, the human operator must also be able to decide the next actions to be made and transmit these actions back to the remote robot. Hence, the design of the user interface plays an important role in remote robot operations to improve awareness and control capabilities.

Conventional user interfaces for robot teleoperations usually are based on keyboard, mouse or joysticks. However, they often are not mobile and they require extensive training. Recent hand-held multi-touch technologies such as tablets and smart phones offer robot teleoperation systems increased portability and accessibility over conventional interfaces [Fung, 2011]. This allows anyone anywhere to connect to the robot system quickly and easily without requiring access to specialised equipment or being restricted to a particular location. This is particularly important in disaster situations since timing is crucial to the success of a mission. Portable user interfaces also allow the operator to move around, which is a clear advantage if the robot's environment

is in partial view [Randelli *et al.*, 2011].

These advantages for portable devices come at a price over more conventional user interfaces. The smaller screen size of portable multi-touch devices means that less information can be displayed visually at one time. An effective solution to this problem is to design user interfaces with multiple complementary feedback modalities to simultaneously excite multiple human senses; particularly the senses of vision, hearing and touch [Aracil *et al.*, 2007; Buss and Schmidt, 1999]. This allows the operator to perceive more information about the robot and remote environment in an efficient way. Typically, visual displays are used as the primary modality, while hearing and touch modalities are particularly useful for drawing the operator's attention to important events [Randelli *et al.*, 2011]. Similarly, a user interface with multi-modal input methods, such as touch screens, hand gestures and voice commands, allows the operator to send more complex commands or perform multiple actions simultaneously and therefore have better control of the robot.

In this paper we investigate these multi-modal user interface design principles for portable teleoperation systems. We first present our tablet based user interface system developed for this purpose; which features multiple complementary modalities of control and feedback mechanisms. Secondly, we analyse multi-modal design principles by performing a user study to evaluate the system performance during a remote navigation task.

In our user interface, the sensory data from the robot is presented to the operator through their senses of vision, hearing and touch. Three control methods are used: gesture based control using tilting mechanism, multi-touch based control and a semi-autonomous pointing method. The user interface utilises built-in hardware features; including the visual display, touch screen, accelerometer, vibrator, stereo sound and Wi-Fi. The application was inspired by common user interface designs, so the users would quickly become familiar with it and therefore eliminate the need for specialised training or experience.

The user study evaluates the design principles by having operators control the robot remotely under different conditions. The operators' task was to navigate the robot around an indoor environment as fast as they could. This task was performed several times by each operator and in each case a different combination of controls and complementary feedback mechanisms were presented to the operator. Performance was measured using task completion time and number of collisions. Each user completed a survey related to the tasks and the user interface.

A key restriction in disaster situations can be a limited communication bandwidth available to the teleoperation system. In remote operation systems, most of the data sent over the communication link is related to the video stream. Therefore part of our user study tested performance while the video stream was disabled to see how well the system performs in a reduced bandwidth configuration. This therefore tested the operators' ability to complete the navigation task while relying on the other feedback modalities.

2 Related Work

Traditionally, robot teleoperation interfaces have been based on conventional human-computer interfaces featuring multiple window displays and input devices such as a joystick, keyboard or mouse [Fong and Thorpe, 2001]. However, these conventional designs are less effective than tangible user interfaces when it comes to human-robot interaction [Randelli *et al.*, 2011]. Recent development in multi-touch sensing surfaces have evolved the way that humans interact with computing devices as well as robots [Kenny *et al.*, 2011; Micire *et al.*, 2011; Bristeau *et al.*, 2011]. Touch screens allow humans to interact with computers directly by fingers in more intuitive and natural way instead of using the traditional user I/O devices [Ketabdar *et al.*, 2012]. Despite tremendous progress in developing touch-based technologies, this form of human-computer interaction has not been extensively used for designing input user interfaces for interaction with a remote robot. Micire *et al.* [2011] designed a multi-touch table-top interface for a wheeled robot that can adapt to the user. An advantage of a multi-touch table-top is that it has large screen size that provides enough space for multiple side-by-side video panels to be displayed to the remote user simultaneously and interaction can happen outside the video screens to prevent occlusion by users' fingers.

However, the big size of the table-tops prevents them from being deployed to disaster and search and rescue missions which require portable and mobile user interfaces. Fung [2011] introduced a video-centric smart-phone user interface for robot teleoperation. Their system was tested by soldiers who gave positive feedback on the portability of the smart-phone interface, but commented that further development is required to make the controls as intuitive as conventional input devices. Kenny *et al.* [2011] presents a similar system and compares several control mechanisms. Whilst both of these smart-phone user interfaces mitigate the portability limitations of multi-touch table-tops, they suffer from the small size of the screen that restricts the number of interface elements that can be shown to the operator at one time. To address the limitations of small size screen one possible solution is to employ multi-modal complimentary forms of interaction for the human-robot interfaces, which can be applied to both the user controls and feedback mechanisms.

The main advantages for using multi-modal controls is to reduce the workload and increase the performance of the operator. Three multi-modal control methods are presented by [Randelli *et al.*, 2011] and their user study found a Wiimote based control method had better performance than more conventional keyboard and joystick controls. The user study presented in [Kenny *et al.*, 2011] compares four smart phone based control methods for teleoperating a mobile robot

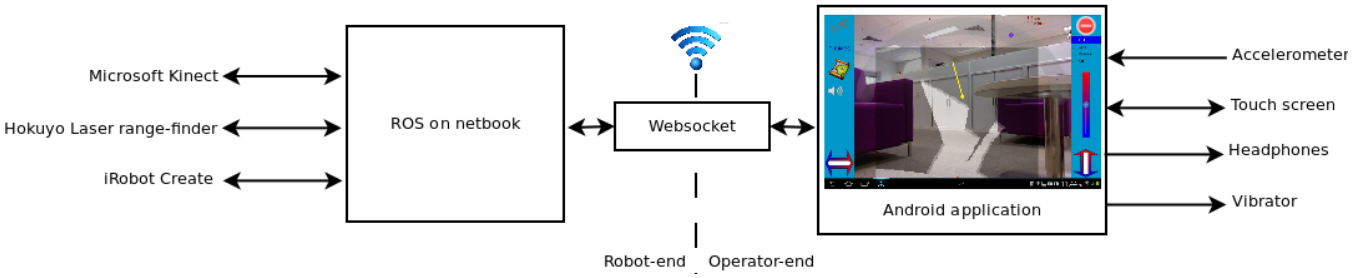

Figure 1: System architecture of the robot and user interface

around obstacles and found a touch button interface was less confusing and resulted in increased performance over accelerometer based controls. In this paper, we also propose three multi-modal interaction methods. However, our proposed user interface allows the operator to choose the most appropriate controls for the task at hand on-the-fly, with minimal effort from the operator. It is in accordance with the findings in [Leeper *et al.*, 2012] that they showed for a human-in-the-loop robotic grasping system the best control method is dependent on how cluttered the environment is with obstacles.

Additionally, robot teleoperation interfaces can be classified based on the types of sensory feedback information that are presented to the human operators. The main sensory information presented to the operator is a video stream and this is the most natural form of feedback for a human to understand. Complementary input information such as range information is vital in teleoperation interfaces since it can be difficult for the operator to judge distances to obstacles solely based on a video stream. Micire *et al.* [2009] compared different configurations of the complementary range information in multiple side-by-side panels on a table-top user interface. However, such representation is not possible in small screen user interfaces. Another way to present range data is to fuse it with the video to give a 3 dimensional representation of the robots environment [Ferland *et al.*, 2009; Ricks *et al.*, 2004]. We present an alternative to these approaches in our user interface, with the range information being displayed semi-transparently over part of the video display, while still allowing the operator to perceive both video and range information without switching their attention. The key advantage with our approach is that it does not require extensive 3D graphics processing and it is part of the main display therefore making it more suitable for tablets.

Sensory information can be presented to the operator through complementary non-visual modalities; particularly audio and haptics. Offering multiple modalities can be an effective way of avoiding overloading the operator with too much information to process at one time [Aracil *et al.*, 2007]. Prior work has shown that information overloading can cause the operator to suffer from mental and physical fatigue or a lack of perception and situational awareness and therefore become disoriented [Kaber *et al.*, 2000; Burke *et al.*, 2004]. The importance of non-visual modalities is significantly more important when using smaller screen user interfaces such as tablets and smart phones, since less visual information can be effectively displayed at one time. In our user study, we investigate the effectiveness of haptics and audio complementary feedback modalities.

3 Teleoperation System

In this paper we seek to analyse the design principles of a multi-modal tablet user interface for robot teleoperation. This section details the system we developed for this purpose, and the following sections analyse the design principles through a user study consisting of a remote navigation task. The system developed, depicted in Figure 1, consists of a mobile robot connected via Wi-Fi to an application on an Android tablet device. The remote operator receives feedback from and sends commands to the remote robot through the Android tablet.

3.1 Mobile Robot

The robot, shown in Figure 2, is based on the design of a Turtlebot[1] with some modifications in both hardware and software. The robot platform is an iRobot Create that has differential drive wheels, bump sensors for detecting collisions and wheel-encoder odometry for dead-reckoning. The platform is connected to a netbook running ROS (Robot Operating System) via serial communication. The software makes use of the standard Turtlebot packages and the rosbridge package which creates an interface between ROS and a websocket. The Android tablet connects to the websocket via a Wi-Fi network. The robot is equipped with an RGB-D Microsoft Kinect camera to provide color video and depth map streams. The robot is equipped with a Hokuyo URG 04LX 2-D laser range-finder (180° field of view and 5 m range) used for collision avoidance and providing the operator with complementary range information.

3.2 Operator User Interface

The remote operator interacts with the robot through an Android application that we have developed for a Samsung

[1] see `http://turtlebot.com/`

Figure 2: The mobile robot used in the teleoperation system

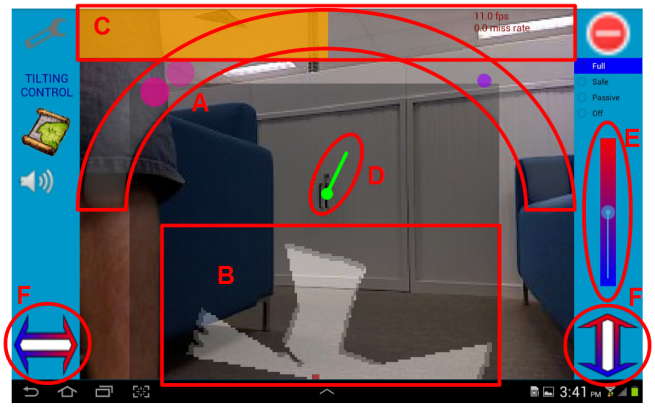

Figure 3: Screenshot of the tablet user-interface with annotation referred to in the text.

Galaxy tablet (see screenshot in Figure 3 and 4). The application makes use of the built in 10.1 inch multi-touch screen, stereo audio through headphones, vibration motors and accelerometer. Feedback from the remote environment is presented to the operator in several ways and there are various input methods available to send action commands to the robot. The main focus in the design was having an interface that did not require extensive training or assume any prior experience.

Feedback Mechanisms

The operator perceives the remote environment and the state of the robot through several complementary modalities in the tablet user interface. This includes visual, haptic and audio feedbacks.

The primary feedback given to the operator is a live color video stream (320x240 pixels/53x45° field of view throttled at 12 fps) from the Microsoft Kinect. This display fills most of the background of the screen. There are other partially transparent visual cues that overlay the video.

An occupancy grid map based on the laser scanner is displayed on the screen to give the operator a birds-eye-view of the environment in front of the robot. The map shows obstacles as black, and free space as white. The aim of this feedback is to give the user better depth perception and a wider field of view than from what they can see through the video feedback. This graphic is displayed transparently in front of the video (see Figure 3, label B) and the operator can enable/disable this view with a button in the side-panel on the touch screen on-the-fly.

In order to improve the situation awareness of the operator through complementary feedback mechanism we also present virtual 3-D audio warnings to the operator through stereo headphones. Audio feedback is particularly important for proximity warnings since the human auditory stimulus is faster than visual signals [Aracil *et al.*, 2007]. However, audio feedback can cause the operator to suffer from mental and physical fatigue and therefore become disoriented. Hence, we designed a virtual 3D audio warnings system that can place sound sources at a position within a virtual 3-D space using Head-Related-Transfer-Functions. The application plays short tones at a position based on the distance and angle of the closest obstacle detected by the laser range-finder. The distances are exaggerated so the operator can easily differentiate between close and far obstacles. Also, as the object moves closer, the frequency of the tone increases and the time between each tone decreases.

Visual obstacle warnings are also implemented to complement the audio warnings. Dots are drawn on the tablet screen at positions that indicate the direction of each detected obstacle, (see Figure 3, label A). The dots are arranged in a semi-circle centred around the centre of the screen. As the distance between the obstacle and the robot decreases, the size of the dot increases and the color changes from blue to red. This colormapping scheme is well established standards to draw the operator's attention to obstacles out of field of view of the camera, without distraction.

There are two bump sensors covering the front of the robot and when either of them are pressed, the tablet provides haptics sensation by vibrating to alert the operator to the collision. Also, a large orange bar at the top of the screen is illuminated (see Figure 3, label C). The operator can determine which bump sensor was triggered by looking at which part of the bar is illuminated (in this case the left bump sensor is triggered).

The operator is given immediate local feedback as to what driving command they are sending to the robot. This is necessary because for some input methods it may not be clear to the operator how their current actions are being processed into robot actions, which can lead to control-loop instability [Buss and Schmidt, 1999]. In this system, feedback is given in the form of a line extending from the centre of the screen (see Figure 3, label D) with a length proportional to the linear velocity and the angle indicates the direction. The colour of the line changes based on the current control method.

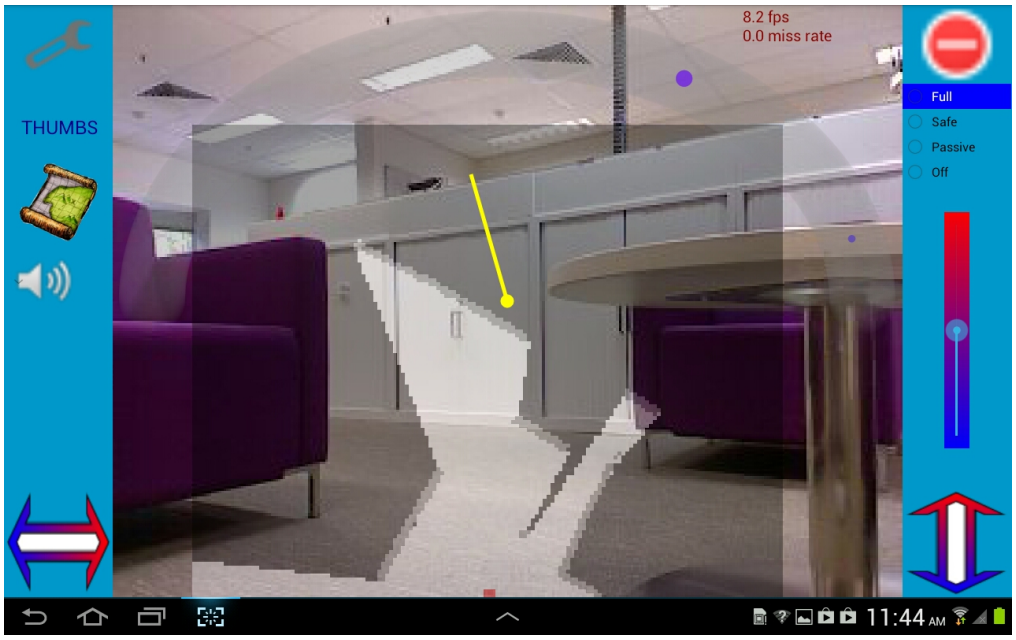

Figure 4: Screenshot of the tablet user-interface.

Control methods

We designed three multi-modal input methods for sending navigation commands to the robot in the proposed user interface. Two of these are direct control methods and one is a semi-autonomous method. The user interface can be configured to allow the operator to switch between the control methods on-the-fly.

The tilting method requires the operator to tilt the tablet based on the direction they want the robot to move. Tilting the device away from or towards the operator tells the robot to move forwards or backwards with a velocity proportional to the angle of the tilt. Similarly, tilting the device to the right or left tells the robot to turn clockwise or anticlockwise. The operator can adjust the sensitivity on-the-fly with a slider in the side panel of the screen (see Figure 3, label E) and the reference axes that the angles are measured against in the preferences menu. Tilting can occur on both axes simultaneously to combine linear and angular velocity. There is a threshold for the minimum angle before the robot starts moving to make it easier to remain stationary.

The touch method is based on controllers commonly seen in remote-control cars and video games. There are touch buttons placed on the bottom corners of the screen (see Figure 3, label F) that the user presses with their thumbs. The linear velocity of the robot is proportional to the up/down position where the operator is currently holding their thumb on the right button. Similarly, the angular velocity is proportional to the left/right position on the left button. When a user takes their thumbs off either of the buttons, the corresponding velocity component is set to zero.

The pointing method makes use of the semi-autonomous capabilities of the robot. It allows the operator to tell the robot to move to a position by clicking on a spot in the video on the touch screen. The robot calculates the required rotation angle based on the pixel coordinates and the distance based on the depth map from the Kinect. The robot then autonomously rotates on the spot by the calculated angle and then moves forward by the calculated distance based on the wheel-encoder odometry. The robot will stop if an obstacle is detected by the laser range-finder, a collision is detected by the bump sensors or the operator cancels the command by clicking the video again.

The user interface can be configured to allow the operator to switch between some or all of the control methods on-the-fly. In the situation where all three methods are used, switching between controls follows the state diagram in Figure 5. The tilting method has lowest priority and is in control only when the other methods aren't being used. If either of the thumbs are pressed then the touch method is in control and similarly if the video is pressed then the pointing method is in control. While the robot is performing a semi-autonomous pointing command, the operator can cancel the command by pressing the thumbs and then control immediately returns to the operator. When the operator release both thumbs from the screen or the robot finishes performing a pointing command, control returns to the tilting method. This configuration improves safety since the operator can quickly and easily cancel a semi-autonomous pointing command if they believe the robot's actions are unsafe. It also allows the operator to switch between controls if they believe another control method is more suitable for the current state of the robot in the environment.

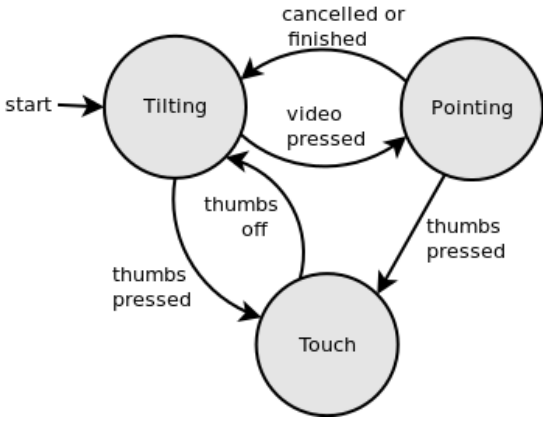

Figure 5: State diagram for switching controls on-the-fly.

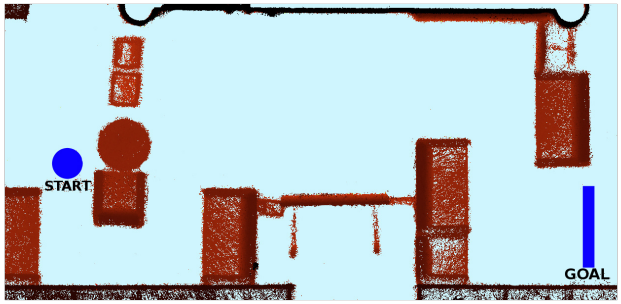

Figure 6: User study obstacle course - 15 m across.

4 Experimental Design

4.1 Hypotheses

In this study, we form fundamental hypotheses about the effects of an intuitive multi-modal tablet user interface for robot teleoperation. We form our hypothesis that a multi-modal user interface improves the task performance and the situation awareness of the participants during remote operations. Additionally, we seek to analyse how different aspects of the multi-modal system affect system performance.

Second, we hypothesised that in the presence of a low-bandwidth network that cannot support a live video stream, the human operators are able to complete navigation tasks within reasonable time compared to when the video stream is the centric information. We also hypothesized that participants who trained first without any video feedback would have better situation awareness through the complementary feedback mechanisms other than visual feedback.

4.2 Procedure

The user study proceeded as follows. Each participant was given an introduction to the study, the robot and the tablet user interface. Next, the robot was placed in a separate, remote location to the user. The only information that passed between them was through the tablet interface. Then, the users were trained on how to remotely control the robot with the multi-modal tablet user interface for about 5 minutes.

The main task was to navigate the robot around an indoor obstacle course, depicted in Figure 6, using the tablet user interface. The course contains two narrow sections with several corners that require finer control and one larger open area that allowed more rapid movement. Each participant performed this task 5 times, where on each occasion they had a different variation of the user interface. Four of the variations had all feedback mechanisms turned on with varying controls: A-tilting, B-touch, C-combined touch and pointing and D-combined tilting, touch and pointing. These combinations were chosen such that we could evaluate which single controls were most effective and how they compared to multi-modal combinations. The other variation (E) had all feedback mechanisms turned on except for the video stream and used combined tilting and touch control. Half of the users performed variation E first while the other half performed variation E last. The other 4 tasks were performed in a random order for each user.

During the experiments, the application logged various information about the operator's actions and robot sensor events, which were saved for later evaluation. At the end of experiments, participants were asked to complete a questionnaire that provided demographic information, experience with robots, touch screen devices, and video games, and finally qualitative feedback about the usability of the interface. For the driving controls, the users were asked what their preferred and least preferred control modality was for completing the task quickly, avoiding collisions, navigating open spaces and navigating confined spaces. For the feedback modalities, the users were asked to rate each modality from 'very useful' to 'not useful' for both completing the task quickly and avoiding collisions; and also asked how often they focussed on each modality from 'always' to 'never' while the camera was on and off.

4.3 Dependent Variables

Dependent variables employed in this study to assess the usability of several modalities of the robot user interface were completion time (highest priority) and number of obstacle collisions. The completion time is computed from the time the user touches the start button on the tablet screen to the time the robot reaches the goal position. We also recorded the time of reversing as an indicator of the time users spent to correct errors or mistakes. The number of obstacle collisions is measured based on the number of bump sensors being triggered, and the number of high risk proximity warnings as range-finder readings less than 10 cm.

4.4 Participants

Twelve male subjects with an average age of 25 years ($\sigma = 5.2$) participated in the user study. The participants include students and staff. Among all the participants, eleven people

Table 1: Mean of the performance measures for each variation (standard deviation in brackets)

Variation	Task Duration (seconds)	Number of bumps	High risk proximity warnings	Reversing time (seconds)
A - Tilting	91 (39)	2.0 (2.2)	95 (101)	11.2 (8.6)
B - Touch	74 (58)	2.1 (2.1)	37 (41)	2.1 (3.7)
C - Touch and Pointing	81 (24)	1.8 (1.3)	44 (48)	3.5 (5.1)
D - Tilting, Touch and Pointing	72 (19)	2.8 (2.5)	57 (50)	4.2 (4.3)
E - No video	113 (65)	2.8 (3.5)	122 (140)	5.4 (3.1)

Table 2: Survey results of the control variations, showing the number of votes for each variation.

Controls variation	Best for speed	Worst for speed	Best for avoiding collisions	Worst for avoiding collisions
A - Tilting	2	**6**	0	**10**
B - Touch	**5**	1	**9**	0
C - Touch and Pointing	3	2	3	1
D - Tilting, Touch and Pointing	2	3	0	1

used smartphones or tablets daily. Eight participants played video games monthly and two reported to use remote control vehicles at least once a month. All participants had experience with multi-touch screens.

5 Evaluation

We designed an experiment to determine if multi-modal user interface would improve situation awareness and enhance the user capabilities in remote teleoperation navigation tasks. We first analyse the performance metrics to compare the different control variations. Secondly, we describe our findings from the no-video task. Lastly, we analyse how the participants rated the various feedback mechanisms in the survey.

5.1 Control Methods

Multi-modal Controls

Table 1 illustrates the task completion time distributions across different variations of the user interface. We found that the participants had the shortest task duration when using multi-modal controls (variation D) ($\mu_D = 72$, $\sigma_D = 19$ seconds, $p = 0.43$). However, several factors in the user study prevents it to achieve statistical significance in task duration times. We noted a learning effect, such that participants improve their task performance on each subsequent run, independent of the order of the control methods (see Figure 7), and thus skewed the results.

An advantage of using the multi-modal controls was a higher performance consistency. Both variation C and D (multi-modal interfaces) had more consistent task duration (i.e., lower standard deviations, and fewer extreme values and outliers) compared to variations A and B (see Figure 8). We believe this difference is due to the ability of the participants to select the control methods that they thought were most suitable for the different parts of the course and were more confident with. Hence, the participants were never forced to use the least suitable control modality for any situation and therefore they were able to avoid or quickly recover from undesirable situations.

In variation D, the participants could switch between the three different controls and it is interesting to note that all participants (100%) at least used two different controls and five of the twelve participants (41%) used all three control modalities during the navigation task. Participants selected tilting and touch controls for about the same percentage of time ($\mu_{tilting} = 44\%$, $\sigma_{tilting} = 34\%$, $\mu_{touch} = 47\%$, $\sigma_{touch} = 41\%$, $\mu_{pointing} = 8\%$, $\sigma_{pointing} = 10\%$). Moreover, we found that all participants switched at least three times ($\mu_{switch} = 8.25$, $\sigma_{switch} = 4.99$ switches) and four of the twelve participants (33%) switched more than 10 times between the different controls. In variation C, we found that nine of the twelve participants (75%) switched controls at least three times ($\mu_{switch} = 7.6$, $\sigma_{switch} = 5.3$ switches) and four of the twelve participants (33%) switched more than 10 times between the two controls. These results indicate that no participants preferred single control for performing the entire navigation task.

Touch and Tilting Controls

We found that the mean task duration was faster for variation B (touch controls) than in variation A (tilting controls) ($\mu_A = 91$, $\sigma_A = 39$, $\mu_B = 74$, $\sigma_B = 58$ seconds, $p = 0.21$), although this was not statistically significant. Variation A had significantly higher number of high risk proximity warn-

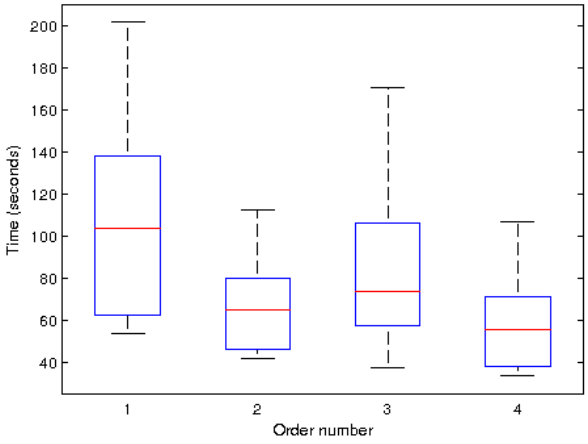

Figure 7: Task completion time based on the order each participant performed each variation (ignoring variation E).

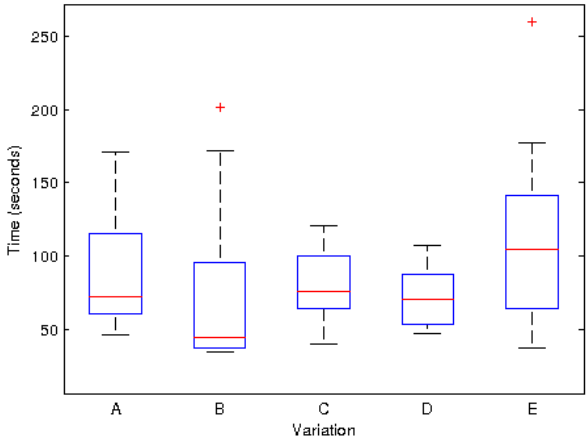

Figure 8: Task completion time for each variation.

ings issued compared to variation B ($\mu_A = 95$, $\sigma_A = 101$, $\mu_B = 37$, $\sigma_B = 41$ warnings, $p < 0.01$) which indicates that users lack full fine control of the robot. Also, participants using variation A on average spent significantly more time reversing ($\mu_A = 11.2$, $\sigma_A = 8.6$, $\mu_B = 2.1$, $\sigma_B = 3.7$ seconds, $p < 0.01$), which indicates they spent more time correcting errors or mistakes. The number of bumps triggered for each run were low ($\mu = 2.2$, $\sigma = 2.3$ bumps) and evenly distributed across all 5 variations.

Table 2 shows the survey results for how many participants selected each control variation as the best and the worst for completing the task quickly and for avoiding collisions. Variation B was considered to be the best by the most participants for both conditions, whereas A was the worst. However, the answers to these questions were well spread out, showing that different controls are better suited to different participants with different experience and prior knowledge. Therefore it is a good idea to offer multiple controls so the operator can choose their preferred modality.

Based on the above findings for the task duration, number of proximity warnings, time spent reversing and the participants' survey responses, we conclude that the touch controls method is more useful than the tilting controls for our navigation task.

Pointing Control

The main advantage observed for using the semi-autonomous pointing method was the reduced number of actions[2] required by the operator. Variation C had fewer actions than variation B ($\mu_C = 1.3$, $\sigma_C = 0.57$, $\mu_B = 1.8$, $\sigma_B = 0.73$ actions per second, $p < 0.2$) for the participants who utilised the pointing method (75% of the participants). This shows that having the ability to switch to the semi-autonomous pointing method reduces the workload for the operator.

We found the participants were significantly more likely to use the pointing control while there were fewer high risk proximity warnings ($\mu_{pointing} = 0.14$, $\sigma_{pointing} = 0.23$, $\mu_{other} = 0.74$, $\sigma_{other} = 0.78$ obstacle detections per second, $p < 0.01$). These results indicate that the participants preferred to use the pointing method in open sections of the obstacle course rather than closed spaces. Additionally, several participants commented in their survey that the pointing control was more useful in the open spaces. The course was mostly made up of tight spaces and we believe this is why the participants used the pointing method for on average only 11% of the task duration.

5.2 No-video Variation

In the no-video variation (E) we investigated the impact of significantly reducing the communication bandwidth between the remote operator and the robot, which would be a realistic restriction in disaster situations. Eleven out of the twelve participants successfully completed variation E with reasonable task performance comparable to when the video is enabled ($\mu_E = 113$, $\sigma_E = 65$ seconds). One participant did not complete the navigation task and went off course at the beginning and did not recover. The results show that non-video feedback modalities alone provided enough information to be relied on by the operators to complete a navigation task and support our initial hypothesis. The advantage of the variation E is that the required bandwidth for the data being sent from the robot to the tablet was reduced from 300 KB/s to 100 KB/s and therefore more suitable for systems with limited bandwidth.

We tested if the no-video variation forms a good way of training operators for a multi-modal user interface by hav-

[2] Actions calculated as the number of large changes in velocity during the tilting and touch methods and points with the pointing method

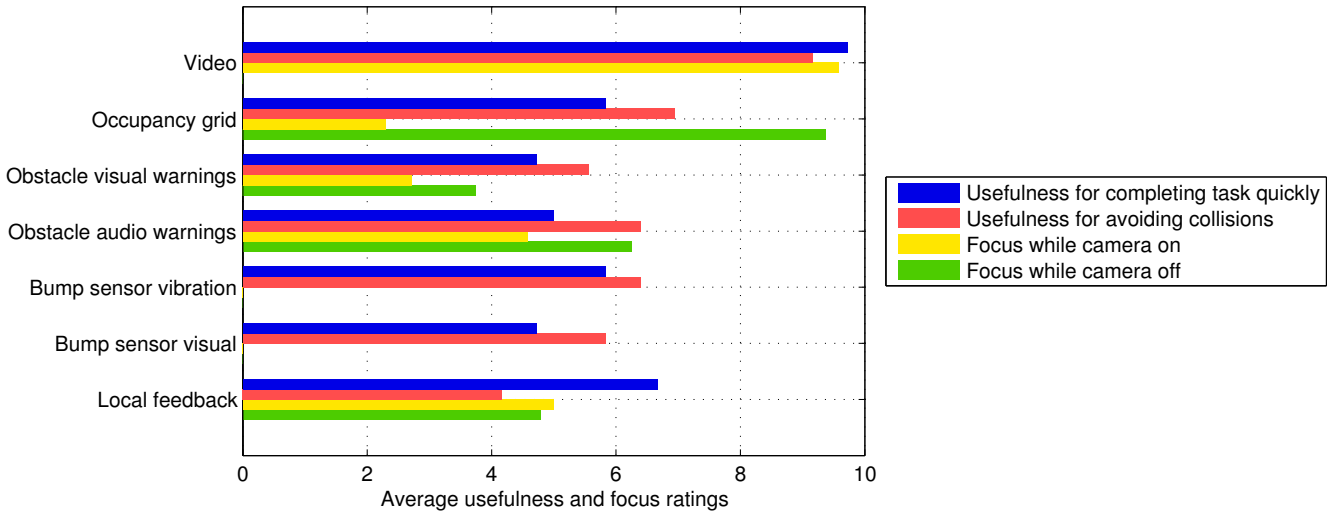

Figure 9: Mean usefulness and required focus of different parts of the user-interface, as rated by the participants. The vertical axis shows the aspects that the participants focussed on during the tasks. The scale of the focus aspects are 10 - "always focus on this component" to 0 - "never focus on this component". The scale of the usefulness aspects are 10 - "very useful" to 0 - "not useful and too distracting". The bars illustrate the mean of the participant responses for the relevant aspects.

ing half of the participants perform variation E first and the other half last. We found out of the participants who performed variation E first, the mean of their other 4 task durations were faster than the participants who performed it last ($\mu_{Efirst} = 73$, $\sigma_{Efirst} = 21$, $\mu_{Elast} = 89$, $\sigma_{Elast} = 30$ seconds, $p = 0.13$). Additionally, the amount of time spent reversing also decreased ($\mu_{Efirst} = 3.6$, $\sigma_{Efirst} = 3.7$, $\mu_{Elast} = 7.0$, $\sigma_{Elast} = 8.3$ seconds, $p < 0.05$). We believe the improvements occurred because the participants who performed variation E first would have a better understanding and awareness of the complementary feedback mechanisms other than the video and this would help them have better performance when completing the other tasks. Therefore we suggest this as an effective way of training operators for multi-modal user interfaces.

5.3 Feedback Mechanisms

We have compared the usefulness of our various feedback mechanisms based on the participants' responses in the survey. The results from the survey are summarised in Figure 9 and below we describe several observations from this chart.

We found that the video stream was rated the most useful and received the most focus, which is typical of video-centric user interfaces such as ours. In the no-video task, the participants' attention focussed primarily on the occupancy grid, which is not surprising since this gave the most information about the environment after the video stream.

For our complementary feedback mechanisms, the participants rated the non-visual modalities as more useful than the visual modalities. For the bump sensor alerts, the participants rated the vibration feedback as more useful than the visual feedback. Similarly for the proximity warnings participants rated the stereo audio feedback as more useful than the visual warnings. This highlights how the audio and haptic modalities provide a useful way for drawing the operators awareness to critical events.

6 Conclusion

In this paper we have evaluated design principles of a multi-modal tablet-based user interface for robot teleoperation. Our system featured multiple modalities of both controls and feedback. Evaluation was performed through a user study to analyse how various aspects of the system affected the performance. Overall, we have demonstrated several advantages for using multi-modal user interfaces for robot teleoperation.

For the controls, it was found that the multi-modal controls setup was preferred and had the most consistent performance. The semi-autonomous pointing control required fewer actions from the operator and was most useful in open spaces. Our system features a method to allow the operator to switch between control modalities on-the-fly with minimal effort, based on what is most suitable for the current situation.

Out of the feedback mechanisms, the users reported they focussed on the video the most, and the occupancy grid when the video was disabled. The participants found the vibration and audio complementary feedback modalities were more useful than the visual feedback for the bump sensors and proximity warnings, and therefore highlights their advantage at alerting the operator to critical events. When the video was disabled the required bandwidth was significantly reduced, however the users were able to rely on the other feedback mechanisms to complete the navigation tasks with reasonable

performance. We also found the participants who completed the no-video variation first had increased performance for the other variations and therefore we suggest this as a way of training operators for multi-modal user interfaces.

For future work, we suggest further evaluating the advantages of multi-modal design principles for portable robot teleoperation systems. In particular, focus on further requirements for realistic large-scale missions; such as how situational awareness, information overload and fatigue over the course of a long mission may be improved by using multiple control and feedback mechanisms. Systems designed for these realistic missions are likely to benefit from other additional modalities, particularly those that introduce more high-level semi-autonomous capabilities. Overall this will allow the operators to perform more complex tasks such as navigating more difficult environments and operating multiple robots simultaneously.

Acknowledgements

The authors would like to thank Paul Flick, Stefan Hrabar, Pavan Sikka and David Haddon from CSIRO Autonomous Systems for their support. Support for this work has also been provided by CSIRO and the Realtime Perception Project. We also thank Robert Fitch from ACFR for insightful comments.

References

[Aracil *et al.*, 2007] R. Aracil, M. Buss, S. Cobos, M. Ferre, S. Hirche, M. Kuschel, and A. Peer. *Advances in Telerobotics*, chapter The Human Role in Telerobotics. Springer, 2007.

[Bristeau *et al.*, 2011] P.J. Bristeau, F. Callou, D. Vissière, and N. Petit. The navigation and control technology inside the ar. drone micro uav. In *World Congress*, volume 18, pages 1477–1484, 2011.

[Burke *et al.*, 2004] J.L. Burke, R.R. Murphy, M.D. Coovert, and D.L. Riddle. Moonlight in miami: a field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human Computer Interaction*, 19(1):85–116, 2004.

[Buss and Schmidt, 1999] M. Buss and G. Schmidt. Control problems in multi-modal telepresence systems. In Paul M. Frank, editor, *Advances in Control*, pages 65–101. Springer, 1999.

[Correa *et al.*, 2010] A. Correa, M.R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis. Multimodal interaction with an autonomous forklift. In *Proc. ACM/IEEE HRI*, pages 243–250, 2010.

[Desai *et al.*, 2012] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco. Effects of changing reliability on trust of robot systems. In *Proc. ACM/IEEE HRI*, pages 73–80, 2012.

[Dragan and Srinivasa, 2012] A.D. Dragan and S.S. Srinivasa. Assistive teleoperation for manipulation tasks. In *Proc. ACM/IEEE HRI*, pages 123–124, 2012.

[Ferland *et al.*, 2009] F. Ferland, F. Pomerleau, C.T. Le Dinh, and F. Michaud. Egocentric and exocentric teleoperation interface using real-time, 3d video projection. In *Proc. ACM/IEEE HRI*, pages 37–44, 2009.

[Ferre *et al.*, 2007] M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer. *Advances in Telerobotics*, chapter Introduction to Advances in Telerobotics. Springer, 2007.

[Fong and Thorpe, 2001] T. Fong and C. Thorpe. Vehicle teleoperation interfaces. *Auton. Robots*, 11(1):9–18, 2001.

[Fung, 2011] N. Fung. Light weight portable operator control unit using an android-enabled mobile phone. *Unmanned Systems Technology XIII, SPIE*, 8045, 2011.

[Kaber *et al.*, 2000] D.B. Kaber, E. Onal, and M.R. Endsley. Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 10(4):409–430, 2000.

[Kenny *et al.*, 2011] C.W.L. Kenny, C.H. Foo, and Y.S. Lee. Interactive methods of tele-operating a single unmanned ground vehicle on a small screen interface. In *Proc. ACM/IEEE HRI*, pages 121 –122, 2011.

[Ketabdar *et al.*, 2012] H. Ketabdar, P. Moghadam, and M. Roshandel. Pingu: A new miniature wearable device for ubiquitous computing environments. In *Proc. Int. Conf. Complex, Intelligent and Software Intensive Systems (CISIS)*, pages 502 –506, 2012.

[Leeper *et al.*, 2012] A. Leeper, K. Hsiao, M.T. Ciocarlie, L. Takayama, and D. Gossow. Strategies for human-in-the-loop robotic grasping. In *Proc. ACM/IEEE HRI*, pages 1–8, 2012.

[Micire *et al.*, 2009] M. Micire, J.L. Drury, B. Keyes, and H.A. Yanco. Multi-touch interaction for robot control. In *Proc. Int. Conf. on Intelligent User Interfaces, ACM*, pages 425–428, 2009.

[Micire *et al.*, 2011] M. Micire, M. Desai, J.L. Drury, E. McCann, A. Norton, K.M. Tsui, and H.A. Yanco. Design and validation of two-handed multi-touch tabletop controllers for robot teleoperation. In *Proc. Int. Conf. on Intelligent User Interfaces, ACM*, pages 145–154, 2011.

[Randelli *et al.*, 2011] G. Randelli, M. Venanzi, and D. Nardi. Evaluating tangible paradigms for ground robot teleoperation. In *RO-MAN, IEEE*, pages 389–394, 2011.

[Ricks *et al.*, 2004] B. Ricks, C.W. Nielsen, and M.A. Goodrich. Ecological displays for robot interaction: a new perspective. In *Proc. IEEE/RSJ IROS*, volume 3, pages 2855 – 2860, 2004.