Title: Multiple robot-single human interaction: Effects on perceived workload

Author: Julie A. Adams

Affiliation: Department of Electrical Engineering and Computer Science
Vanderbilt University
VU Box 35 351824 Sta B
2301 Vanderbilt Place
Nashville, TN 37235-1824

Email: adamsj@ieee.org
Phone: 615 322 – 8481
Fax: 615 343 – 6702

Multiple robot-single human interaction: Effects on perceived workload

Abstract

This paper presents results from a user evaluation of a real multiple robot system in which the human's perceived workload and performance were measured. Participants completed tasks with one, two, and four real heterogeneous mobile ground-based robots for indoor material transportation tasks. Twelve participants completed four trials of each task over two days. Generally speaking, little difference was found between the one and two-robot tasks; however perceived workload significantly increased while performance decreased during the four-robot task. A correlation analysis found that perceived workload increased as the number of commands (in total and by command type) and completion times increased. The highest number of accidents and user errors occurred during the four-robot task. In order to increase the number of robots a single human can supervise, the human must be provided with capabilities that reduce the number of required commands while also optimizing the human's interactions with the robots.

Index terms – human-robotic interaction, perceived workload, NASA Task Load Index, multiple mobile robots

1 Introduction

There is an increased interest in human-robotic interaction that permits humans to supervise a robotic system. Such robotic systems may be employed for search and rescue (i.e. Murphy 2004), exploring mines (i.e. Thrun et al. 2003), and military applications (i.e. Rybski et al. 2001). Much of the research has focused on a single human interacting with a single robot (i.e. Fong et al. 2002) or multiple humans interacting with a single robot (i.e. Murphy 2004). There also has been a focus on a single or small number of humans interacting with a small number of robots (i.e. Yanco et al. 2004); however systems are still limited in their ability to permit a single human to simultaneously supervise more than four robots. This paper presents quantitative results from an evaluation in which a single human supervisor interacted with up to four robots. While this work and evaluation were initially completed in the mid-1990s, this work addresses the factors that effect perceived workload when interacting with multiple real robots.

This paper focuses on the factors that directly affect a human supervisor's perceived workload and the associated consequences when supervising multiple real mobile ground-based robots. One objective of this paper is to demonstrate which objective measures effect perceived workload. Previous research related to teleoperated[1] manipulator robots focused on quantitatively measuring perceived workload (Draper and Blair 1996, Kaber et al. 2000a, Kaber et al. 2000b). Draper and Blair (1996) focused on workload related to telepresence when teleoperating a single manipulator. Kaber et al. (2000b) measured perceived workload for a single teleoperated manipulator with communication latencies of up to four seconds. Kaber et al. (2000a) also focused on a single teleoperated manipulator and measured workload as participants completed tasks with four levels of autonomy. This evaluation found that workload decreased as the level of autonomy increased. However, it is well known that increased autonomy can lead directly to the introduction of new issues such as reduced situational awareness, complacency, and

[1] Teleoperation requires the human supervisor to provide each command the robot is to execute. In our evaluation, the operator provided straight forward and backward commands as well as individual rotational commands via mouse input.

skill degradation (Parasuraman et al. 2000, Wickens et al. 2004). While these results are important, they do not directly correspond to a human supervisor's perceived workload and performance when working with a multiple mobile robot system.

Schipani (2003) reported an evaluation that incorporated a single ground-based robot during field exercises and determined that when the human was required to interact with the robot, the human's workload significantly increased. This particular evaluation varied the terrain, task, robot speed, and the operator's ability to directly view the robot. These factors affect workload but the evaluation did not focus on the users' underlying system interactions.

Chadwick (2005) conducted an evaluation that included tasks requiring one or two simulated ground and aerial robots. The perceived workload results were not significant but indicated that conditions employing two robots resulted in higher workload values. Recently Parasuraman et al. (2005) evaluated a delegation-based interface for six simulated ground robots with autonomous capabilities. This evaluation found no conclusive results related to perceived workload. Both evaluations were simulation-based and were more general with regard to the operator interactions that affected perceived workload.

Ruff et al. (2002) evaluated the effects of levels of autonomy and decision-aid fidelity when a human supervisor interacts with one, two, and four simulated unmanned aerial robots. Their results found that as the number of robots increased, the human's perceived workload "generally increased;" however no actual statistical results related to perceived workload were provided.

This paper reports the results of a quantitative user evaluation in which participants completed tasks with one, two, and four real heterogeneous mobile ground-based robots in an indoor environment. The presented results relay the objective factors the directly effect the participants' perceived workload. As well, this paper presents results that indicate how objective measures are affected as the number of robots increased.

1.1 *System Description*

The Multiple Agent Supervisory Control (MASC) system was built to combine autonomous system capabilities with a human supervisor's ability to control the robots for indoor material transport tasks. The system included four heterogeneous mobile ground-based robots and the MASC human-robotic interface (Adams 1995, Adams et. al 1996). Each robot was composed of a TRC Labmate mobile base and two robots had heterogeneous sensing capabilities. The SensorBot and VisionBot were employed for environmental sensing. The SensorBot possessed sixteen ultrasonic sonar and infrared sensor pairs, a structured-light source and camera, and a stereo camera pair. The ultrasonic sonar and infrared sensor pairs were employed to detect wall and corner-like features (Kamberova et al. 1999, Mandelbaum and Mintz 1995). The remaining SensorBot sensing capabilities were presented as displays of their raw informational values. The VisionBot's sensors included a stereo camera pair and a camera mounted on a turntable. The stereo pair was employed for autonomous visually guided obstacle avoidance (Košecká and Bajcsy 1997). The obstacle avoidance process provided output in the form of a free-space map and a process state diagram. The two manipulatory robots (the PumaBot and the ZebraBot) were employed to transport objects from one location to another (Wang and Kumar 2001, Yamamoto and Yun 1996). These two robots relied upon the observation robots to guide them. The PumaBot robot was equipped with a Puma 260 manipulator, while the ZebraBot robot was equipped with a Zebra-ZERO manipulator. Robotic system details can be found in Adams et al. (1996).

The MASC interface (Adams 1995) allowed a human to interact with and observe multiple robots while the robots executed tasks at a remote location. The human supervisor provided commands or corrective actions to each robot via the interface.

[Insert Figure 1 about here]

The MASC interface provided the human supervisor with a primary window (Main Interface Window in Figure 1). The main window provided a three-dimensional graphical representation of the remote environment based upon a pre-configured world model. The display's viewing angle could be rotated, zoomed, and translated. The right hand side of the interface provided up to eight smaller windows that presented sensor information from the individual robots. These smaller windows, in Figure 1, represent: two virtual camera views from inside the main interface window, camera feedback from the sensorBot, camera feedback from the visionBot, the free space map created by the obstacle avoidance process, and a diagram indicating the state of the obstacle avoidance process. All robot camera images were provided in real-time at thirty frames per second. The autonomous obstacle avoidance process created a free space map of the area within the stereo cameras' field of view. The black area in the free space map indicated the open space through which the robot could navigate. The process also provided an indication of its current state via the obstacle avoidance process state diagram. Additional information available via the small display windows included the ultrasonic process state diagram and the path planning state diagram.

[Insert Figure 2 about here]

The main interface window provided the human with an indication of the robots' position in the world based upon the individual robots' odometry. The odometry readings were not corrected via a localization process[2] during the presented evaluations since it was a minimal factor in the evaluation tasks. The human supervisor provided teleoperation commands by first selecting the appropriate robot via the robot mode buttons and then entering the required commands directly on the main window via the mouse. If a move forward command was issued, the virtual robot would move from its initial position to the final goal location. As the robot executed the command, a phantom robot (Figure 2.a) indicated the actual robot's position along the commanded path.

The SensorBot was equipped with sixteen ultrasonic sonar and infrared sensor pairs. This information was displayed directly on the main window, as shown in Figure 2.b. The sonar and infrared cones were displayed using different colors for each sensor pair. Additional information provided directly on the main window included the walls and corners detected by the ultrasonic process, see Figure 3. The darker

[2] As mobile ground vehicles are driven, the recorded odometry information is not 100% accurate and this error is cumulative.

the detected wall or corner color, the higher the confidence in the presented information based upon the sensory feedback. The human had the ability to activate or deactivate all sensor displays as desired.

[Insert Figure 3 about here]

The human supervisor selected a particular robot to control using the command buttons provided for each robot via the top two rows of command buttons (as shown in Figure 1). The command buttons included the ability to pause or continue the robot's command execution, cancel all commands in the robot's command queue, or issue an emergency stop command. The third row of command buttons permitted the human supervisor to select the system mode commands (see Figure 1). The system mode commands included: replaying the past few minutes of recorded activities, initializing the system, using teleoperation to navigate the robots, and a limited autonomous robot navigation mode. This paper focuses on the system initialization and teleoperation system modes.

The culmination of the MASC system development was a user evaluation. At the time of this study, 1995, this evaluation was one of the first known multiple robot user evaluation employing real robots, not simulated robots (Endo et al. 2004). This paper focuses on the results related to the human supervisor's perceived workload and objective measures as the number of robots required for a task increased from one robot to four. Section 2 provides the evaluation apparatus and method while Section 3 details the evaluation results. Section 4 provides the results discussion and Section 5 provides conclusions.

2 Apparatus

The purpose of the evaluation was to understand how perceived operator workload was affected as the number of robots per task increased. Performance was quantified by analyzing the task completions (total number of completions and completions times), user errors, robot crashes, etc. as the number of robots increased from one to four. Three null hypotheses were defined. *Hypothesis one*: Perceived workload levels are not affected by the number of robots a human supervises. This hypothesis was evaluated via the NASA TLX perceived workload results (Section 3.1) and the correlation analysis (Section 3.7) of the perceived workload results with the objective measures (Section 2.1). *Hypothesis two:* Perceived workload levels are not affected by increased experience with the system (i.e. during the second session). This hypothesis was evaluated similarly to the first; however the analysis focused on the cross session analysis. *Hypothesis three:* Task performance is not affected by the number of robots. This hypothesis was evaluated via the number of task completions (Section 3.2), the task completion times (Section 3.3) and the number of errors (Section 3.6) analyses in addition to the associated correlation analysis (Section 3.7).

The single-robot task required that the SensorBot be commanded to drive parallel to a wall into a corner. Once the robot was in the corner, it was commanded to turn and move diagonally across the room. A tall garbage can, representing an obstacle, was placed between the robot's initial position and the corner. The participants were required to detect the obstacle and then command the robot to avoid the obstacle. This robot was not equipped with the autonomous obstacle avoidance process.

The two-robot task required the participant to teleoperate the SensorBot as described in the one-robot task while simultaneously commanding the VisionBot to move parallel to the opposite wall and into the corner. An obstacle was placed in the VisionBot's path and the autonomous obstacle avoidance process automatically avoided the obstacle.

The four-robot task required the VisionBot move along one wall and into the corner while autonomously avoiding an obstacle. The two manipulatory robots were in a side-by-side configuration ahead of the SensorBot on the opposite side of the room. The participants were instructed to control the manipulatory robots in the combined control[3] method for as much of the task execution as was feasible. All three robots were to move along the wall and into the corner. When this position was obtained, the robots were to turn in formation and move diagonally across the room to the goal position. The environment for these tasks was static and was intended to mimic a simplified indoor transportation task.

---

[3] Combined control permitted the supervisor to create a single command that was simultaneously executed by both manipulatory robots.

The task execution environment contained one obstacle for the one and four-robot tasks and two obstacles for the two-robot task. The SensorBot tasks required obstacle detection via the sensory displays (i.e. ultrasonic sonar and/or camera images) and commands to avoid the obstacle. Tasks employing autonomous obstacle avoidance permitted the observation of obstacle avoidance via the robot's deviation from the commanded path and via sensory displays as initialized by the participant (i.e. camera images, free space map, and obstacle avoidance process state diagram). The obstacles were not displayed on the main interface window with either obstacle avoidance technique. The obstacles were placed in the environment prior to the start of the task and were not visible via the cameras from the starting locations of the robots. Tasks employing autonomous obstacle avoidance used a small box as an obstacle that was visible within the stereo camera pair's field of view, while the other tasks that required the participant to manually perform obstacle avoidance. The manual obstacle avoidance tasks employed a large garbage can that was tall enough to be detected via the ultrasonic sonar and infrared pairs. The obstacles were located in similar locations for all tasks.

Errors occurred when the participant selected an incorrect item on the interface given the current system state. For example, the participant may attempt to command a robot that was not selected to receive commands. The participant received an error message via a pop-up window indicating the cause of the error. The participant had to dismiss the error window prior to resuming their tasks. All errors were presented in this manner. Accidents occurred when the robots ran into walls, obstacles, or each other. The robots' bumpers halted the robots' motion when they collided with an obstruction, thus resulting in the phantom robot's motion halting. This was the only indication that the robot had collided with an object.

The participants created four command types: locomotion, system mode, robot mode, and robot switch. Locomotion commands rotate a robot or move the robot forward and backwards via teleoperation. The system mode commands allow a transition from one system mode to another, such as from the system initialization to the exploration mode. The robot mode commands instruct a robot to conduct an emergency stop; pause a command execution; or resume a previous command execution. The robot switch commands allow the selection of a new robot to command and de-selection of the currently selected robot.

2.1 *Method*

A within-subjects design was employed in which participants completed three tasks: the single-robot task, the two-robot task, and the four-robot task. The independent variable was the number of robots per task. The dependent variables included task completion times, number of successful task completions, number of operator errors, number of sensors per task, number of commands issued per task (in total and by command type), and perceived workload. Each participant completed two sessions and each session entailed performing each task twice. The task presentation was counterbalanced across participants; however the task presentation was consistent for an individual participant across sessions.

Twelve University of Pennsylvania students completed the evaluation. Participants (3 females and 9 males) were novice mobile robot users and most had minimal graphical user interface experience. The participants' ages ranged between seventeen and thirty-three years and their educational backgrounds ranged from some high school to doctoral level education.

The MASC interface ran on a Silicon Graphics *Indigo*^{2} with 96 MB of memory. The interface computer was located such that the participants were unable to directly obtain visual or auditory feedback from the robots or the associated environment.

The underlying software automatically recorded participants' interactions with the interface along with the robots' positions and sensory readings. A second computer monitor, to the left of the interface monitor, duplicated all information and interactions from the primary monitor. The second monitor screen was videotaped for all tasks trials and the video was employed to validate errors and accidents. After each task trial the participants completed the NASA Task Load Index (TLX) workload demand rating scales (Hart and Staveland 1988, Hill et al. 1992, Nygren 1991). After two trials of each task, participants completed a general system and interface post-task questionnaire. Upon completion of the experiment (after Session two), participants completed the post-experimental questionnaire pertaining to all tasks, the system's abilities, usability issues, and the NASA TLX paired comparisons.

The automatic interaction recordings provided the following time stamped information: task duration, the number of created and executed commands by command type; the number of system mode changes; which robots, associated sensors and sensory displays were employed; and the errors participants committed. The robots' positions and the various sensor readings were also automatically recorded.

The experiment consisted of four phases: pre-experimental, training, and experimental sessions one and two. The first session lasted approximately two and a half hours and included the training and practice phases, execution of all tasks and trials, two breaks, and completion of data gathering tools. The second session occurred two days later and lasted approximately two hours. Participants were permitted a twenty-minute practice session on the second day. The session then continued as the first session and concluded with the post-experimental questionnaire.

The training session lasted thirty minutes and focused on the robots and the MASC interface. Each participant was shown the actual robots and surrounding environment. A detailed explanation and demonstration of the robots and the interface capabilities was also provided. After completing the training session, participants practiced with the MASC system for twenty-minutes.

Situations that would end the experiment, such as all involved robots crashing, were explained to the participants. They were instructed that if a robot became inoperable to continue the trial with the remaining robots. The task descriptions stated what was to be accomplished and that the task should be accomplished as quickly and efficiently as possible. The participants were informed that the environment may have changed. After receiving the instructions, the participants began the task. Upon trial completion, the participant completed the NASA TLX questionnaire. The participants then performed the same task a second time. At the end of the second trial, they completed the NASA TLX and the post-task questionnaires. The second and third tasks were carried out in a similar manner.

3 Experimental Results

The experimental results focus on an analysis of the objective data and NASA TLX results. A total of 48 trials were completed for each of the individual tasks. The NASA TLX results provide insight into the perceived workload. The objective measures demonstrate the factors that do and do not influence perceived workload.

3.1 *Perceived Workload*

The NASA TLX tool was employed to assess perceived workload. The lowest recorded perceived workload value was 1.8, while the highest was 67.5. The perceived workload values for all tasks were lower during the second session. Three trials of the two-robot task had a lower perceived workload than the single-robot task. The four-robot task consistently had the highest perceived workload.

[Insert Table 1 about here]

Table 1 provides the mean (M) perceived workloads and the associated standard deviations (SD) by task, trial, and session. The within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found significant interaction effects across tasks ($F(2, 22) = 4.82$, $p = 0.03$) and sessions ($F(1, 11) = 8.1$, $p = 0.02$). A paired samples *t*-test found that the four-robot task was perceived as requiring significantly higher workload than either the single-robot ($t(47) = -3.5$, $p < 0.01$) or the two-robot tasks ($t(47) = -5.5$, $p < 0.01$).

[Insert Table 2 about here]

Table 2 presents the paired samples *t*-test across tasks within a session and demonstrates that the four-robot task resulted in significantly higher perceived workload than the other tasks during both sessions. This implies that the four-robot task required significantly more cognitive capabilities than the other two tasks. There was little difference when working with one or two robots; however increasing the number of robots to four increased the perceived workload. A paired samples *t*-test comparing the tasks across sessions found that only the two-robot task was perceived as requiring significantly lower workload during the second session, $t(47) = 5.6$, $p < 0.01$. There was no significant difference when comparing the single-robot and four-robot tasks across sessions; therefore as participants used the system for a longer period of time, the perceived workload did not significantly change.

3.2 *Task Completions*

It is very common when conducting evaluations with real robots that the robots will encounter failures. The system failures encountered during this evaluation included: communication failures, low batteries, and bumper activation. All trials with system problems were completed with the remaining robots.

[Insert Table 3 about here]

Table 3 provides the task completions by successful completions, trials that resulted in an accident and trials that encountered system problems. Additionally, each value is represented by session. One hundred twenty-five task trials were successfully completed, twelve trials were completed when a system failure was encountered, and seven trials ended as accidents. The participants successfully completed tasks with no system problems in 84.5% of the trials. The successful completion rate increased to 92.6% when the twelve system problem trials are included. Accidents occurred in 4.7% of the trials with the highest accident rate occurring with the four-robot task. A within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found a significant interaction effect across tasks $F(1, 11) = 5.09$, $p = 0.02$. A paired samples t-test comparing all data by task found that the number of completions was significantly higher for the single-robot task than for the four-robot task, $t(47) = 2.95$, $p < 0.01$. There were no significant differences when comparing the single-robot and the two-robot tasks or the two-robot and four-robot tasks. The above result is supported by a paired samples t-test comparing tasks within a session that found the single-robot task resulted in significantly more completions than the four-robot task during Sessions two, $t(23) = 2.4$, $p = 0.03$. A paired samples t-test found no significant differences by task across the sessions, thus indicating that task completion performance was relatively the same during both sessions.

3.3 *Task Completion Times*

The task completion times increased as the number of robots per task increased; however all completion times fell during the second session. Table 4 provides the mean completion times and associated standard deviations by task, trial, and session.

[Insert Table 4 about here]

A within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found significant interaction effects across tasks ($F(2, 22) = 27.24$, $p < 0.01$) and sessions ($F(1, 11) = 20.1$, $p < 0.01$). A repeated measures *t*-test comparing all tasks for all trials found that the single-robot task completion time was significantly faster than the completions for the two-robot ($t(47) = -3.52$, $p < 0.01$) and four-robot tasks ($t(47) = -8.0$, $p < 0.01$). Additionally, the two-robot task completion time was significantly faster than the four-robot task, $t(47) = -5.9$, $p < 0.01$).

[Insert Table 5 about here]

The repeated measures *t*-test comparing tasks within a session supported the finding that the four-robot task completions were significantly slower than the one and two-robot tasks, as shown in Table 5. The four-robot task was significantly slower than the other two tasks during both sessions. Additionally, it was found that the one-robot task was significantly faster than the two-robot task during Session two. A repeated measures *t*-test comparing the individual task completion times across sessions found that all tasks were performed significantly faster during the second session, as shown in Table 4. This result indicates that as the participants became more familiar with the system and tasks, their completion times significantly improved.

3.4 *Number of Commands*

[Insert Table 6 about here]

A breakdown of the number of commands by command type, as defined in Section 2.0, is presented in Table 6 by task. The total number of commands for the one- and two-robot tasks was very similar. The four-robot task required more than twice as many commands as the other two tasks. The majority of the commands issued (57.4%) were locomotion commands.

[Insert Table 7 about here]

Table 7 provides the mean and standard deviation for the total number of commands by task, trial, and session. The total number of commands was consistently higher for the four-robot task. There was very little variation in the mean number of total commands across trials and sessions. A within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found a significant interaction effect for tasks $F(2, 22) = 63.8$, $p < 0.01$. A paired samples *t*-test comparing tasks for all data reveals that the four-robot task requires significantly more commands than the single-robot ($t(47) = -10.1$, $p < 0.01$) and the two-robot tasks ($t(47) = -9.4$, $p < 0.01$).

[Insert Table 8 about here]

Table 8 provides the paired samples *t*-test results across tasks within a session. The results support the previous *t*-test results in that the four-robot task required significantly more commands than the single- and two-robot tasks during both sessions. Therefore, when the number of robots per task was four, the total number of commands required to complete the task was significantly higher than the total number of commands required for completing the other tasks. A paired samples *t*-test by task across sessions found no significant results; therefore the number of commands was relatively unchanged as participants became more experienced with the system and tasks. The results related to the total number of commands calls into question how the command types affected the overall results. Therefore, further analysis of each command type was conducted.

[Insert Table 9 about here]

An analysis of the locomotion commands found that on average the four-robot task required the most locomotion commands. Table 9 provides the mean locomotion commands broken down by task, trial, and session. A within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found a significant interaction effect for tasks, $F(2, 22) = 26$, $p < 0.01$. A paired samples *t*-test across all task data found that the four-robot task required significantly more locomotion commands than the single-robot ($t(47) = -5.9$, $p < 0.01$) and the two-robot tasks ($t(47) = -7.2$, $p < 0.01$).

[Insert Table 10 about here]

The paired samples *t*-test across tasks within a session supported this result. The results, in Table 10, demonstrate that the four-robot task required significantly more locomotion commands than the other tasks during both sessions. The number of locomotion commands per task did not significantly change across the sessions.

The system mode change commands were very consistent across tasks, trials, and sessions. Virtually every task required on average three commands. This is due to the system design. The individual results are not provided due to this fact and that no interesting statistical analysis results were found.

[Insert Table 11 about here]

Table 11 provides the robot mode command information by task, trial, and session. The four-robot task required the largest number of robot mode commands. The within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, and Trials) found significant interaction effects across tasks ($F(2, 22) = 17.68$, $p < 0.01$) and sessions ($F(1, 11) = 6.08$, $p < = 0.03$). A paired samples *t*-test revealed that the four-robot task required significantly more robot mode commands than the single-robot ($t(47) = -5.9$, $p < 0.01$) and the two-robot ($t(47) = -5.5$, $p < 0.01$) tasks.

[Insert Table 12 about here]

Table 12 provides the paired samples *t*-test across tasks within a session. The results support this result demonstrating that the four-robot task required significantly more robot mode commands than the other tasks during each session. There were no significant changes in the number of robot mode commands by task across sessions.

[Insert Table 13 about here]

Table 13 provides the results by task, trial, and session for the robot switch commands. This command category only applies to the two and four-robot tasks as multiple robots must exist in order to execute this command. The within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found significant interaction effects across tasks ($F(2, 22) = 27.3$, $p < 0.01$) and sessions ($F(1, 11) = 20.1$, $p < 0.01$). A paired samples *t*-test comparing the number of robot switches by task revealed that the four-robot task required significantly more switches between robots than the two-robot task, $t(46) = -13.1$, $p < 0.01$. The paired samples *t*-test across tasks within a session supported this result demonstrating that the four-robot task required significantly more robot switches than the two-robot task during Session one ($t(23) = -7.4$, $p < 0.01$) and Session two ($t(23) = -12.3$, $p < 0.01$). A paired samples *t*-test across sessions found that during the second session participants issued significantly fewer robot switch commands during the two-robot task, $t(22) = 3.8$, $p < 0.01$. A similar analysis for the four-robot task found no significant difference.

It would be expected that the total number of commands increase as the number of robots per task increased. It was found that locomotion, robot mode, and robot switch commands tend to significantly increase during the four-robot task as compared to the single- and two-robot tasks. There is some evidence that the number of commands decreased during the second session when the participants gained more system experience. However, there is no significant drop in the number of commands for the single and two-robot tasks across sessions.

3.5 *Sensor Display Usage*

Each participant chose the sensors on the robots and the associated sensor displays (recall that some sensors provide multiple displays), as defined in Section 1.1, to employ for each task based upon the available sensors for a specific robot. Automatic recording tracked the sensors and the associated displays employed, the duration of use, and how many sensor displays were employed per task. Each participant initialized the sensors and the associated sensor displays prior to beginning the task. None of the participants modified their displays during task execution.

[Insert Table 14 about here]

Table 14 presents the number of times a particular sensor display was employed by the participants (Usage), the number of times a display was available for use during a task (Occurrences), and the percentage of time a particular display was employed when available for use (Percentage). The predominant display choice focused on the camera images. The obstacle avoidance free space map, the raw ultrasound and infrared display, as well as the obstacle avoidance state diagram followed. Most participants did not find the information provided by the ultrasonic sonar process useful and did not use it. It was found that participants employed all available displays, including those within the main working window in less than 3% of the trials.

[Insert Table 15 about here]

Table 15 provides the mean number of sensor displays employed for all tasks. The participants in all tasks used roughly half of the total available sensor displays. The average number of displays employed for the two and four-robot tasks were virtually the same. The within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found a significant interaction effect for tasks, $F(2, 22) = 19.7$, $p < 0.01$. A paired samples *t*-test comparing the results across tasks found that the single-robot task employed significantly fewer sensor displays than both the two-robot ($t(47) = -7.99$, $p < 0.01$) and four-robot tasks ($t(47) = -8.2$, $p < 0.01$).

[Insert Table 16 about here]

The paired samples *t*-test across tasks within a session supported this result demonstrating that the single-robot task required significantly fewer sensor displays than the other tasks during both sessions. These results are provided in Table 16. The paired samples *t*-test analysis within a task across sessions found no significant differences. These results imply that there is no significant difference in the number of sensor displays employed for the two- and four-robot tasks; however both tasks required significantly more sensor displays than the single-robot task. Additionally, the participants did not significantly change the number of sensors and the associated displays that they employed as they gained system and task experience.

3.6 *Errors*

An error message was provided each time an incorrect action was requested and such errors were automatically recorded. A total of thirty-three errors were recorded. Only two participants had no errors, two participants had one error, and three participants committed more than three errors. The predominant error was attempting to control a robot, other than the robot currently selected to control. This error occurred twenty times and accounted for 54% of all errors. The second most frequent error, seven occurrences, was attempting to initialize a robot and its processes after the participant had already completed the robots' initialization. Another common error, five occurrences, was attempting to display sensor information when the associated sensor process was inactive. A similar error occurred when attempting to command robots that were inactive, five occurrences.

[Insert Table 17 about here]

Table 17 provides the mean number of errors by task, trial, and session. A within-subjects three-way repeated measures GLM procedure (Tasks, Sessions, Trials) found a significant interaction effect across tasks, $F(2, 22) = 8.3$, $p < 0.01$. The repeated measures *t*-test comparing all data by task found that the four-robot task resulted in significantly more errors than either the single-robot ($t(47) = -3.4$, $p < 0.01$) or the two-robot task ($t(47) = -2.7$, $p < 0.01$).

[Insert Table 18 about here]

The previous result is further supported by a paired samples *t*-test across tasks within sessions, as shown in Table 18. The four-robot task resulted in significantly more errors than either the single- or two-robot tasks during Session one and the single-robot task during Session two. The total number of errors was greater as the number of robots executing the task increased to four. While the mean number of errors fell during the second session, there were no significant differences by tasks across the sessions. Therefore, the participants did not significantly reduce their errors as they became more experienced with the system and tasks.

3.7 *Correlation Analysis*

A Pearson's correlation analysis was conducted to determine the significance of the relationships between the data. The primary intention of this analysis is to demonstrate which objective measures effect perceived workload. Additionally, the objective measures are analyzed. The correlation analysis included perceived workload, task completion times, number of mode changes, number of state changes, sensors per task, total number of commands, and number of locomotion, robot mode, and robot switch commands. Due to the small number of errors and non-completions, these variables are excluded from this analysis.

A majority of significant correlations over all the data were positive, thus indicating that both variables in question were affected, either both variables increase or decrease. Table 19 provides the correlation results for all data.

[Insert Table 19 about here]

The analysis for perceived workload found positive correlations with completions times ($r(144) = 0.29$, $p < 0.01$); the total number of commands ($r(144) = 0.25$; $p < 0.01$); the number of locomotion commands ($r(144) = 0.2$, $p = 0.02$), the robot mode commands ($r(144) = 0.25$, $p < 0.01$), and the robot switch commands ($r(144) = 0.26$, $p = 0.01$). Perceived workload also increased as the number of sensor displays employed increased ($r(144) = 0.26$, $p < 0.01$). Generally speaking, perceived workload increased as the completion times, number of commands, and the number of sensor displays increased.

The analysis for completion time found a number of positive correlations in addition to the correlation with the perceived workload. Completion time increased as the total number of commands ($r(144) = 0.5$, $p < 0.01$), the number of locomotion commands ($r(144) = .64$, $p < 0.01$), the robot mode commands ($r(144) = 0.72$, $p < 0.01$), and the robot switch commands ($r(144) = 0.54$, $p < 0.01$) increased. Completion time also increased as the number of sensor displays employed per task ($r(144) = 0.29$, $p < 0.01$) increased.

As mentioned, the analysis for total number of commands found positive correlations with the perceived workload and task completion times. Positive correlations also existed for the number of locomotion commands ($r(144) = 0.59$, $p < 0.01$) and number of robot mode commands ($r(144) = 0.54$, $p < 0.01$). As the total number of commands across all tasks increased so did the total number of locomotion and robot mode commands. During tasks that required multiple robots, there was a positive correlation between the total number of commands and the commands to switch between robots ($r(144) = 0.69$, $p < 0.01$). Therefore, the increased number of robots required more robot switch commands and this contributed to in an increase in the total number of commands.

The number of locomotion commands positively correlate with the number of robot mode commands ($r(144) = 0.5$, $p < 0.01$) and the number of robot switch commands ($r(144) = 0.7$, $p < 0.01$). Additionally, a positive correlation was found between the number of robot switch commands and the number of robot mode commands for the two and four-robot tasks, $r(144) = 0.54$, $p < 0.01$.

[Insert Table 20 about here]

An analysis of each task found that within the single-robot task, a positive correlation between perceived workload and the number of sensor displays existed ($r(48) = 0.39$, $p < 0.01$). It was also found that completion time positively correlated with the number of total commands ($r(48) = 0.48$, $p < 0.01$), locomotion commands ($r(48) = 0.52$, $p < 0.01$), and the number of sensor displays employed ($r(48) = 0.33$, $p = 0.02$). The total number of commands increased as the number of locomotion commands increased during the single-robot task, $r(48) = 0.81$, $p < 0.01$.

[Insert Table 21 about here]

During the two-robot task perceived workload increased as the completion times increased, $r(48) = 0.32$, $p = 0.03$. The completion times increased as the number of locomotion ($r(48) = 0.51$, $p < 0.01$) and robot mode commands ($r(48) = 0.32$, $p = 0.03$) increased. The total number of commands issued during the two-robot task increased as the number of locomotion ($r(48) = 0.55$, $p < 0.01$) and robot switch commands ($r(48) = 0.4$, $p < 0.01$) increased. The number of locomotion commands increased with the number of robot switch commands ($r(48) = 0.34$, $p = 0.02$). Thus further supporting that as the number of robots increased, as did the number of total commands and locomotion commands.

[Insert Table 22 about here]

The analysis found no significant correlations between perceived workload and the objective variables during the four-robot task. However, the completion times positively correlated with the number of locomotion commands ($r(48) = 0.54$, $p < 0.01$), the robot mode commands ($r(48) = 0.78$, $p < 0.01$), and the robot switch commands ($r(48) = 0.31$, $p = 0.03$). The number of robot switches also increased with the number of locomotion commands ($r(48) = 0.55$, $p < 0.01$), the robot mode commands ($r(48) = 0.33$, $p = 0.03$) and the total number of commands ($r(48) = 0.34$, $p = 0.02$). Finally, positive correlations existed between the number of robot mode commands and the number of locomotion commands ($r(48) = 0.48$, $p < 0.01$) and the total number of commands ($r(48) = 0.36$, $p = 0.01$).

Similar results to the overall results were found when the data was analyzed by session across all tasks. These results indicate that when the number of robots increased from one to four robots per task, the effect was an increase to the values for the relationships found. Perceived workload increased with the completion times, number of commands, and the number of sensor displays employed. Therefore, in order to reduce perceived workload, a focus should be placed on reducing the number of required commands and sensor displays, which may result in reducing task completion times.

4 Discussion

It is important to understand the underlying factors that affect a human supervisor's ability to interact with multiple heterogeneous robots. Currently most deployed robot systems pair a single human with a robot or multiple humans with a single robot. Additionally, many human supervisors in these situations are trained robot operators. Examples include robots employed by bomb squads and urban search and rescue teams. However, as robots are deployed in greater numbers, the likelihood of a human supervisor being a highly trained robot operator will diminish. Emergency responders and the military do not have the capacity or the funds to maintain the level of training currently required to support such robotic systems.

Additionally, the department of defense has been calling for multiple robot systems controlled by a single human. In order to meet such objectives, it becomes very important to understand the demands placed upon the human supervisor as the number of robots under their supervision increases in order to ensure successful mission completion.

The evaluation in this paper included results from twelve novice users who completed one-, two-, and four-robot tasks in a simplified, static, indoor environment. The results provide two important insights. First, practice can improve the human's ability to supervise heterogeneous robots during tasks. Second, there is little difference between working with one or two robots; however as the number of heterogeneous robots required for a task increased to four, there were many consequences.

The four-robot task resulted in fewer successful task completions than the one and two-robot tasks. The number of accident and systems problems was greater for the four-robot task. The more robots, components, and processes that are involved, the higher the probability of a malfunction.

System problems actually make it very difficult to run a highly controlled user evaluation with real heterogeneous robots. In fact, this evaluation included fifteen participants but data for three participants had to be discarded because of problems with the robots that made them inoperable. This issue is not specific to running user evaluations; any type of experiment involving multiple robots can be very trying. Additionally, this evaluation was carried out in a static, indoor environment. These types of problems only increase as robots move into highly dynamic environments such as building collapses or military zones. One can hypothesize that this effect will be more dramatic in such environments.

The first null hypothesis indicated that participant workload levels would not be affected by the number of robots. The perceived workload analysis (Section 3.1) refutes this hypothesis. The analyses found that the four-robot task required significantly higher perceived workload than the one- and two-robot tasks across tasks and sessions. Therefore, the results refute the hypothesis that as the number of robots required for a task increases, perceived workload will not also increase.

The second null hypothesis stated that the participant workload levels are not affected by increased experience with the system. The reported perceived workload for all tasks fell during the second evaluation session. The two-robot task during the second session was the only task in which the perceived workload fell by a significant level. There was no significant difference for the one- and four-robot tasks across sessions. Therefore, while a drop in perceived workload was observed, this decrease was not statistically significant and the null hypothesis is not refuted.

The third null hypothesis indicated that task performance would not be affected by the number of robots. The results found that there was little difference in the task completion times for the one- and two-robot tasks while the four-robot task had significantly longer completion times. However, all completion times significantly decreased during the second session, thus demonstrating that with additional practice, the novice participants were able to improve their performance. The results also found that the total number of successful completions fell as the number of robots increased. The one-robot task had significantly more successful completions than the four-robot task. The participants also committed significantly more interaction errors during the four-robot task over the one- and two-robot tasks. There was little difference between the one- and two-robot tasks in either the number of successful completions or errors. Therefore, the results refute the null hypothesis that task performance would not be affected by the number of robots.

The total number of commands issued during the tasks was relatively unchanged for the one- and two-robot tasks. There also was no real decrease in the number of commands issued during the second session. However, the number of commands required when supervising four robots was almost twice as many as the one- and two-robot tasks, a significant increase. This interface required the participants to change the robot receiving commands before a command could be issued. This robot switch command almost quadrupled from the two-robot task to the four-robot task. Clearly an easier method of switching between robots for command purposes is required. At the same time, it is important to develop tools that will assist the human by guiding interactions and minimizing or optimizing the number of times the human switches between robots.

Generally speaking, when the participants completed the four-robot task, their perceived workload, task completion times, and number of generated commands rose while the number of completions fell. Perceived workload was found to increase with the task completion times, and the number of commands (specifically locomotion, robot mode, and robot switch commands). Therefore, it is vitally important to reduce the number of commands that the human supervisor must issue to a robotic system if the total number of heterogeneous robots supervised by a single human is to increase. This experiment was conducted in a static, indoor environment. As multiple robots are deployed in highly dynamic environments for much more complex tasks, these same factors will have an effect on the human supervisor’s performance. It is not clear from this evaluation the exact effects; however it is known that such environments include many more complications and challenges, placing more demands on the human.

One message from this evaluation is that in order to increase the number of robots that a single human can supervise, teleoperation cannot be the primary interaction technique. Automation is not the only solution to this problem, as it introduces a new set of issues. Interaction techniques that permit the human supervisor to optimize and manage the interaction are also necessary.

5 Conclusions

I have presented the Multiple Agent Supervisory Control (MASC) system. The test-bed and the general MASC human-robot interface were described. The system was designed for indoor material transportation tasks.

This evaluation was the first conducted with multiple real heterogeneous ground-based robots that looks specifically at the underlying factors that effect a human’s perceived workload while completing multiple robot tasks. This evaluation analysis included twelve novice users who completed the one-robot, two-robot, and four-robot tasks in a static, indoor environment. These participants employed teleoperation to command the robots and were able to complete the four-robot task. However, as the number of robots increased to four, there was a significant decrease in the number of tasks successfully completed. Tasks requiring four robots had a higher number of accidents and system problems. Additionally, the participants' perceived workload significantly increased during the four-robot task, which increased with the number of commands issued to the robots and the time to complete tasks..

As robots are deployed in highly dynamic environments, the demands placed upon the human supervisor will only increase. Teleoperation cannot be employed as the only commanding mechanism for teams of four (or more) robots. The human must be provided with tools that will support the decisions to be made and optimize the human's interaction with the robot team.

Acknowledgements

This research is funded in part by: ARO Grants DAAL03-89-C-0031; ARPA Grants N00014-92-J-1647; ARPA/NSF Grant IRI94-12913; NSF Grants STC SBR8920230 and CISE/CDA-88-22719; and Sandia National Laboratories AN-2317. The author thanks her Ph.D. advisor, Dr. Richard Paul for his support. The author also thanks the reviewers and Dr. Barrett Caldwell for their insightful feedback regarding the analysis and the paper.

References

Adams, J. A., 1995, Human Management of a Hierarchical System for the Control of Multiple Mobile Robots. Ph.D. Dissertation, University of Pennsylvania.

Adams, J. A., Bajcsy, R., Košecká, J., Kumar, V., Mandelbaum, R., Mintz, M., Paul, R., Wang, C. C., Yamamoto, Y., and Yun, X., 1996, Cooperative Material Handling by Human and Robotic Agents: Module Development and System Synthesis. *Expert Systems with Applications,* Pergamon, **11**(2), pp. 89-97.

Chadwick, R., 2005, The impacts of multiple robots and display views: An urban search and rescue simulation. In the *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, pp. 387-391.

Draper, J. V. and Blair, L. M., 1996, Workload, flow, and telepresence during teleoperation. In the *Proceedings of the 1996 IEEE Conference on Robotics and Automation*, pp. 1030-1035.

Endo, Y., MacKenzie, D. C., and Arkin, R. C., 2004, Usability Evaluation of High-Level User Assistance for Robot Mission Specification. *IEEE Transactions on Systems, Man and Cybernetics – Part C*, **34**(2), pp. 168-180.

Fong, T., Thorpe, C., and Baur, C., 2003, Multi-robot remote driving with collaborative control. *IEEE Transactions on Industrial Electronics*, **50**(4), pp. 699-704.

Hart, S. G. and L. E. Staveland, 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, P. A. Hancock & N. Meshkati (Eds.), North-Holland: Elsevier Science, pp. 139-183.

Hill, S., Iavecchia, H., Byers, J., Bittner, A., Zakland, A. and Christ, R., 1992, Comparison of four subjective workload rating scales. *Human Factors*, **34**(4), pp. 429-439.

Kaber, D. B, Onal, E., and Endsley, M. R., 2000a, Design of Automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing*, **10**(4), pp. 409-430.

Kaber, D.B., Riley, J.M., Zhou, R., and Draper, J., 2000b, Effects of visual interface design and control mode and latency on performance, Telepresence and workload in a teleoperation task. In the *Proceedings of the IEA 2000/HFES 2000 Congress*, **1**, pp. 503-506.

Kamberova, G., Mandelbaum, R., Mintz, M., and Bajcsy, R., 1999, Decision-Theoretic Approach to Robust Fusion of Location Data. *Journal of the Franklin Institute*, **336**(2), pp. 269-284.

Košecká, J. and Christensen, H., 997, Experiments in Behavior Composition. *Journal of Robotics and Autonomous Systems*, **19**, pp. 287-298.

Mandelbaum R. and Mintz, M., 1995, Feature-Based Localization using Fixed Ultrasonic Transducers (extended version). In the *Proceedings of the Autonomous Vehicles in Mine Countermeasures Symposium*.

Murphy, R. R., 2004, Human-Robot Interaction in Rescue Robots. *IEEE Transactions on Systems, Man and Cybernetics Part A*, **34**(2), pp. 138-153.

Nygren, T., 1991, Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, **33**(1), pp. 17-33.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., and Miller, C., 2005, A flexible delegation-type interface enhances system performance in human supervision of multiple robots: empirical students with RoboFlag. *IEEE Transactions on Systems, Man and Cybernetics Part A*, **35**(4), pp. 481-493.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D., 2000, A model for types and levels of human interaction with automation. *IEEE Transactions of Systems, Man and Cybernetics Part B*, **30**(3), pp. 286-297.

Rybski, P., Stoeter, S., Papanikolopoulos, N., Burt, I., Dahlin, R., Gini, M., Hougen, D., Krantz, D., and Nageotte, F., 2003, Sharing control: Presenting a framework for the operation and coordination of multiple miniature robots. *IEEE Robotics and Automation Magazine*, **9**(4), pp. 41-48.

Ruff, H. A., Narayanan, S., and Draper, M.H., 2002, Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence*, **11**(4), pp. 335-351.

Schipani, S. P., 2003, An evaluation of operator workload, during partially-autonomous vehicle operations. In the *Proceedings of the 2003 Performance Metrics for Intelligent Systems Workshop*.

Thrun, S., Hähnel, D., Ferguson, D., Montemerlo, M., Triebel, R., Burgard, W., Baker, C., Omohundro, Z., Thayer, S., and Whittaker, W., 2003, A system for volumetric robotic mapping of underground mines. In the *Proceedings of the IEEE International Conference on Robotics and Automation*, p.p. 4270-4275.

Wang C. C. and Kumar, V., 2001, The Performance of Repeatable Control Schemes for Redundant Robots. *Journal of Robotic Systems*, **18**(4), pp. 171-186.

Wickens, C. D., Lee, J. D., Liu, Y., and Gordon Becker, S. E., 2004, *An Introduction to Human Factors Engineering*, Second Edition (New Jersey: Prentice Hall.

Yamamoto Y. and Yun, X., 1996, Effect of the dynamic interaction on coordinated control of mobile manipulators. *IEEE Transactions on Robotics and Automation*, **12**(5), pp. 816 – 824.

Yanco, H. A., Drury, J. L., and Scholtz, J., 2004, Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition. *Journal of Human-Computer Interaction*, **19**(1 and 2), pp. 117-149.

List of Figure Captions

Figure 1 The MASC system interface.

Figure 2 a) The phantom robot displayed during a move command execution. b) The ultrasonic sonar and infrared displays for the sensorBot.

Figure 3 Detected walls and corners as provided by the ultrasonic sonar process.

Figure 1

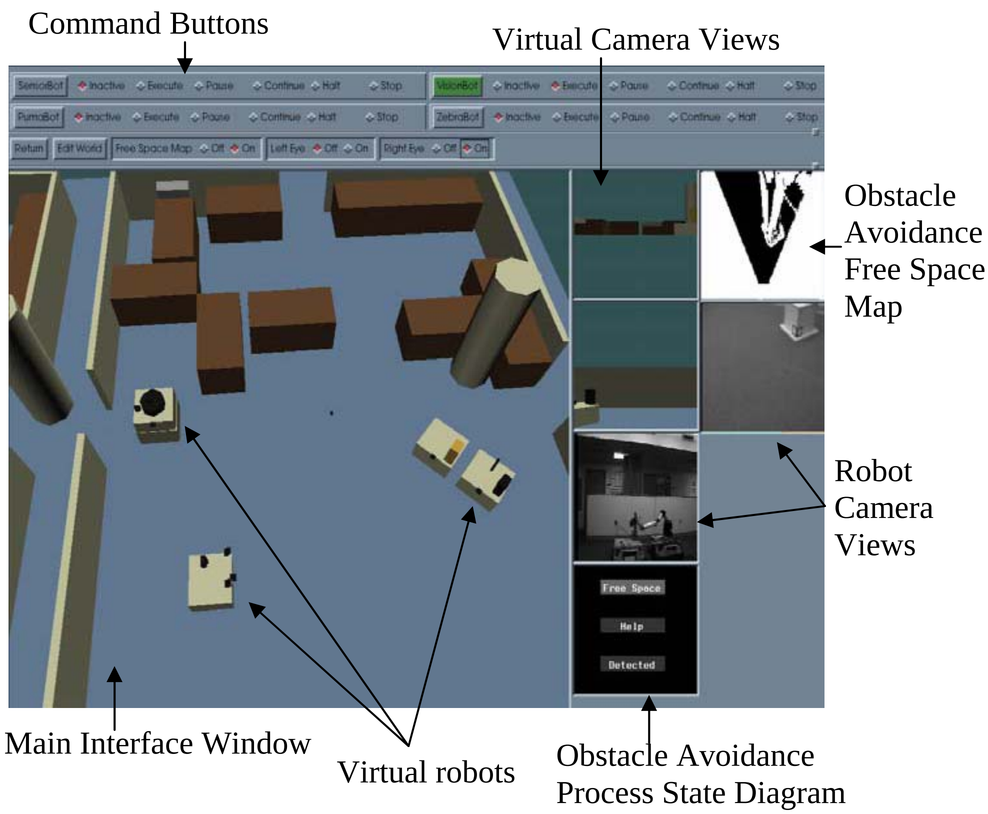
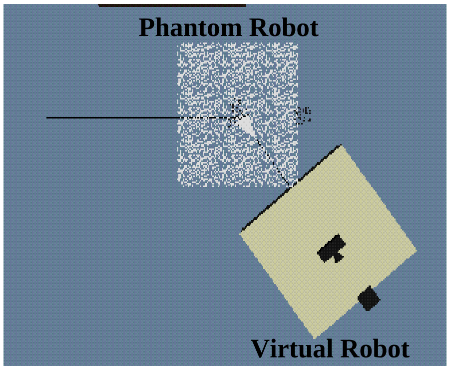
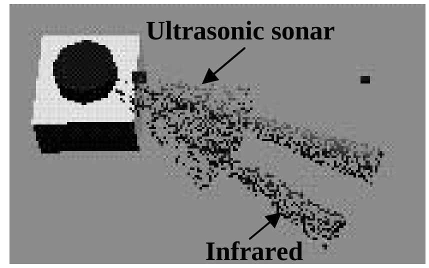

Command Buttons
Virtual Camera Views
Obstacle Avoidance Free Space Map
Robot Camera Views
Main Interface Window
Virtual robots
Obstacle Avoidance Process State Diagram

Figure 2



(a)



(b)

Figure 3

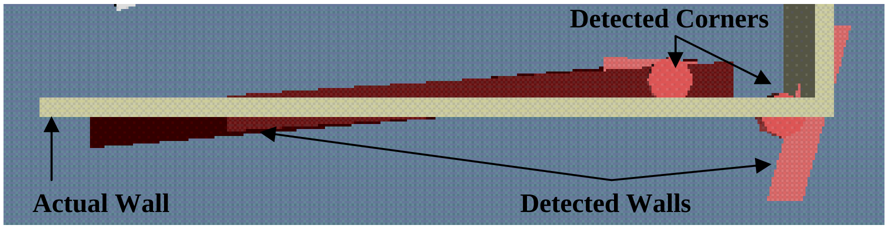


Table 1 The mean (M) perceived workload and standard deviations (SD) by task, trial, and session.

	Session One				**Session Two**			
	Trial One		**Trial Two**		**Trial one**		**Trial Two**	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	37.8	14.4	38.6	16.3	33.3	18.2	32.8	18.2
Two-Robot	39.9	12.9	40.2	16.6	30.6	15	32.1	16.5
Four-Robot	46.7	14.3	48	17.6	43.7	18.9	44.9	21.3

Table 2 The *t*-test comparisons across tasks within a session for perceived workload.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	-0.6	0.55	23	0.8	0.44
Single- vs. Four-Robot	23	-2.1	0.04	23	-2.7	0.01
Two- vs. Four-Robot	23	-2.9	<0.01	23	-4.8	<0.01

Table 3 The task completions by completion type, task, and session.

Task	Completions		Accident		System Problems	
	Session One	**Session Two**	**Session One**	**Session Two**	**Session One**	**Session Two**
Single-Robot	23	23	0	0	1	1
Two-Robot	20	23	2	1	2	0
Four-Robot	19	17	1	3	4	4
Total	62	63	3	4	7	5

Table 4 The task completion time means and standard deviations by task, trial, and session along with paired *t*-test results by task across sessions.

	Session One				Session Two				*t* test		
	Trial One		Trial Two		Trial One		Trial Two				
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	***df***	***t***	***p***
Single-Robot	398	111	338	100	288	56	257	43	23	5.3	<0.01
Two-Robot	439	108	384	91	339	78	332	48	23	3.0	<0.01
Four-Robot	670	205	564	277	459	114	458	111	23	3.0	<0.01

Table 5 The *t*-test results across tasks within a session for the task completion times.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	-1.6	0.14	23	-5.4	<0.01
Single- vs. Four-Robot	23	-5.3	<0.01	23	-7.2	<0.01
Two- vs. Four-Robot	23	-4.2	<0.01	23	-4.2	<0.01

Table 6 The number of all created commands by command type and task.

Task	**Locomotion**	**System Mode**	**Robot Mode**	**Robot Switches**	**Total**
Single-Robot	627	153	98	0	878
Two-Robot	542	145	112	96	895
Four-Robot	1014	144	280	491	1932
Total	2183	442	493	587	3805

Table 7 The mean total number of commands by task, trial, and session.

	Session One				**Session Two**			
	Trial One		**Trial Two**		**Trial One**		**Trial Two**	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	19.5	6.5	18.1	6.3	18.8	6.4	16.8	3.8
Two-Robot	21	10	16.3	4.4	19.8	8.7	17.5	4.5
Four-Robot	40.9	15.3	37.1	12.3	43	22.2	40	8.1

Table 8 The *t*-test comparison across tasks within a session for the total number of commands.

Tasks	**Session One**			**Session Two**		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	0.08	0.9	23	-0.5	0.6
Single- vs. Four-Robot	23	-6.6	<0.01	23	-7.6	<0.01
Two- vs. Four-Robot	23	-6.2	<0.01	23	-7.1	<0.01

Table 9 The mean number of locomotion commands by task, trial and session.

	Session One				Session Two			
	Trial One		Trial Two		Trial One		Trial Two	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	13.6	5.1	13.9	6.1	12.8	6.4	11.9	4.1
Two-Robot	12.4	8.5	11.7	5.5	9.9	3.5	11.2	3.7
Four-Robot	20.3	9.1	23.3	12.8	18.9	8.4	22	5.5

Table 10 The *t*-test comparison across tasks within a session for the number of locomotion commands.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	0.9	0.4	23	2.0	0.06
Single- vs. Four-Robot	23	-3.4	<0.01	23	-5.4	<0.01
Two- vs. Four-Robot	23	-4.3	<0.01	23	-6.2	<0.01

Table 11 The mean number of robot mode change commands by task, session, and trial.

	Session One				Session Two			
	Trial One		Trial Two		Trial One		Trial Two	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	2.5	2.8	1.8	2.0	2.3	2.1	1.5	1.6
Two-Robot	2.5	1.5	3.3	2.4	1.8	1.7	2.1	1.7
Four-Robot	7.3	5.7	6.5	6.2	4.8	1.8	5	2.1

Table 12 The *t*-test comparison across tasks within a session for the number of robot mode change commands.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	0.9	0.4	23	0.0	1.0
Single- vs. Four-Robot	23	-2.4	<0.01	23	-5.6	<0.01
Two- vs. Four-Robot	23	-7.0	<0.01	23	-6.4	<0.01

Table 13 The mean number of robot switch commands by task, trial, and session.

	Session One				Session Two			
	Trial One		Trial Two		Trial One		Trial Two	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Two-Robot	3.1	1.9	2.3	2.2	1.6	0.9	1.3	0.87
Four-Robot	10.3	4.5	11.1	4.9	10.4	4.1	10	3.4

Table 14 The usage and occurrences of the sensor displays including percentage used.

Sensor	**Usage**	**Occurrences**	**Percentage**
sensorBot Left Stereo Camera	127	144	88%
sensorBot Right Stereo Camera	124	144	86%
visionBot Left Stereo Camera	74	96	77%
visionBot Light Stripe Camera	91	144	63%
VisionBot Right Stereo Camera	59	96	61%
Obstacle Avoidance State Diagram	54	96	56%
Raw Ultrasonic Sonar Display	75	144	52%
Obstacle Avoidance Free Space Map	45	96	47%
Ultrasonic Process	12	144	8%
Ultrasonic Process State Diagram	7	144	5%
Path Planning State Diagram	4	96	4%

Table 15 The mean number of sensor displays employed by task, trial, and session.

	Session One				Session Two			
	Trial One		Trial Two		Trial one		Trial Two	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	3.7	1.4	3.1	1.1	3.3	0.97	3.3	0.97
Two-Robot	5	1.7	5	1.4	5.4	1.1	5.2	1.1
Four-Robot	5.5	1.98	5.3	1.6	5.3	1.3	5.3	0.99

Table 16 The *t*-test comparison across tasks within a session for the number of sensor displays.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	-4.3	<0.01	23	-7.7	<0.01
Single- vs. Four-Robot	23	-4.7	<0.01	23	-7.9	<0.01
Two- vs. Four-Robot	23	-1.2	0.23	23	0.0	1.0

Table 17 The mean number of errors committed by task, session and trial.

	Session One				Session Two			
	Trial One		Trial Two		Trial one		Trial Two	
Task	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Single-Robot	0.25	0.6	0	0	0	0	0	0
Two-Robot	0.33	0.5	0	0	0.1	0.3	0.17	0.4
Four-Robot	0.75	1.0	0.58	0.8	0.17	0.4	0.4	0.8

Table 18 The *t*-test comparisons across tasks within a session for the number of errors committed.

Tasks	Session One			Session Two		
	df	***t***	***p***	***df***	***t***	***p***
Single- vs. Two-Robot	23	-0.3	0.75	23	-1.8	0.08
Single- vs. Four-Robot	23	-2.6	0.02	23	-2.3	0.03
Two- vs. Four-Robot	23	-2.9	<0.01	23	-1.2	0.3

Table 19 Pearson correlations, means, and standard deviations for all data (N = 144). Note that the numbers along to the top of the table correspond to the variable numbers along the leftmost column. The gray cells represent significant results.

	1	2	3	4	5	6	7
1. Completion Time	$r = 1$						
2. Total Number of Commands	$r = 0.49$ $p < 0.01$	$r = 1$					
3. Number of Locomotion Commands	$r = 0.64$ $p < 0.01$	$r = 0.59$ $p < 0.01$	$r = 1$				
4. Number of Robot Mode Commands	$r = 0.72$ $p < 0.01$	$r = 0.5$ $p < 0.01$	$r = 0.5$ $p < 0.01$	$r = 1$			
5. Number of Robot Switch Commands	$r = 0.54$ $p < 0.01$	$r = 0.69$ $p < 0.01$	$r = 0.7$ $p < 0.01$	$r = 0.54$ $p < 0.01$	$r = 1$		
6. Number of Sensor Displays	$r = 0.29$ $p < 0.01$	$r = 0.15$ $p = 0.07$	$r = 0.87$ $p = 0.3$	$r = 0.10$ $p = 0.21$	$r = 0.06$ $p = 0.54$	$r = 1$	
7. Perceived Workload	$r = 0.29$ $p < 0.01$	$r = 0.25$ $p < 0.01$	$r = 0.2$ $p = 0.02$	$r = 0.25$ $p < 0.01$	$r = 0.26$ $p = 0.01$	$r = 0.26$ $p < 0.01$	$r = 1$
Mean	410.53	25.73	15.16	3.45	6.24	4.6	39.06
Standard Deviation	167.7	14.45	8.13	3.5	5.27	1.58	17.15

Table 20 Pearson correlations, means, and standard deviations for all data during Single-robot task (N = 144). Note that the numbers along to the top of the table correspond to the variable numbers in the leftmost column. The gray cells represent significant results.

	1	2	3	4	5	6
1. Completion Time	$r = 1$					
2. Total Number of Commands	$r = 0.48$ $p < 0.01$	$r = 1$				
3. Number of Locomotion Commands	$r = 0.52$ $p < 0.01$	$r = 0.81$ $p < 0.01$	$r = 1$			
4. Number of Robot Mode Commands	$r = 0.14$ $p = 0.33$	$r = 0.26$ $p = 0.08$	$r = -0.44$ $p = 0.8$	$r = 1$		
5. Number of Sensor Displays	$r = 0.33$ $p = 0.02$	$r = 0.03$ $p = 0.8$	$r = -0.07$ $p = 0.9$	$r = -0.60$ $p = 0.69$	$r = 1$	
6. Perceived Workload	$r = 0.24$ $p = 0.1$	$r = -0.03$ $p = 0.8$	$r = -0.07$ $p = 0.65$	$r = 0.07$ $p = 0.62$	$r = 0.39$ $p < 0.01$	$r = 1$
Mean	320.21	18.29	13.06	2.04	3.31	35.6
Standard Deviation	96.17	5.77	5.4	2.15	1.1	16.52

Table 21 Pearson correlations, means, and standard deviations for the two-robot task data (N = 144). Note that the numbers along to the top of the table correspond to the variable numbers in the leftmost column. The gray cells represent significant results.

	1	2	3	4	5	6	7
1. Completion Time	$r = 1$						
2. Total Number of Commands	$r = 0.19$ $p = 0.18$	$r = 1$					
3. Number of Locomotion Commands	$r = 0.51$ $p < 0.01$	$r = 0.55$ $p < 0.01$	$r = 1$				
4. Number of Robot Mode Commands	$r = 0.32$ $p = 0.03$	$r = 0.12$ $p = 0.41$	$r = 0.22$ $p = 0.14$	$r = 1$			
5. Number of Robot Switch Commands	$r = 0.19$ $p = 0.2$	$r = 0.39$ $p < 0.01$	$r = 0.34$ $p = 0.02$	$r = 0.21$ $p = 0.15$	$r = 1$		
6. Number of Sensor Displays	$r = -0.03$ $p = 0.83$	$r = -0.03$ $p = 0.85$	$r = 0.0$ $p = 0.99$	$r = -0.08$ $p = 0.58$	$r = 0.06$ $p = 0.71$	$r = 1$	
7. Perceived Workload	$r = 0.32$ $p = 0.03$	$r = 0.08$ $p = 0.59$	$r = 0.13$ $p = 0.36$	$r = 0.18$ $p = 0.21$	$r = 0.16$ $p = 0.27$	$r = 0.21$ $p = 0.14$	$r = 1$
Mean	373.56	18.65	11.29	2.38	2.04	5.15	35.73
Standard Deviation	91.92	7.34	5.6	1.87	1.7	1.3	15.47

Table 22 Pearson correlations, means, and standard deviations for the Four-robot task data (*N* = 144). Note that the numbers along to the top of the table correspond to the variable numbers in the leftmost column. The gray cells represent significant results.

	1	2	3	4	5	6	7
1. Completion Time	$r = 1$						
2. Total Number of Commands	$r = 0.13$ $p = 0.37$	$r = 1$					
3. Number of Locomotion Commands	$r = 0.54$ $p < 0.01$	$r = 0.22$ $p = 0.14$	$r = 1$				
4. Number of Robot Mode Commands	$r = 0.78$ $p < 0.01$	$r = 0.36$ $p = 0.01$	$r = 0.48$ $p < 0.01$	$r = 1$			
5. Number of Robot Switch Commands	$r = 0.31$ $p = 0.03$	$r = 0.34$ $p = 0.02$	$r = 0.55$ $p < 0.01$	$r = 0.33$ $p = 0.03$	$r = 1$		
6. Number of Sensor Displays	$r = 0.04$ $p = 0.81$	$r = -0.27$ $p = 0.06$	$r = -0.12$ $p = 0.42$	$r = -0.12$ $p = 0.28$	$r = 0.01$ $p = 0.93$	$r = 1$	
7. Perceived Workload	$r = 0.11$ $p = 0.48$	$r = 0.13$ $p = 0.39$	$r = 0.11$ $p = 0.45$	$r = 0.16$ $p = 0.28$	$r = -0.02$ $p = 0.88$	$r = 0.09$ $p = 0.51$	$r = 1$
Mean	537.81	40.25	21.13	5.9	10.45	5.33	45.84
Standard Deviation	203.99	15.01	9.2	4.45	4.14	1.46	17.68