University of Pennsylvania Technical Report MS-CIS-96-11, March 1996

Human Factors Experimental Analysis of the MASC System [*]

Julie A. Adams

General Robotics and Active Sensory Perception Laboratory
Department of Computer and Information Science
The University of Pennsylvania
3401 Walnut Street, Room 301C Philadelphia PA 19104-6228

Abstract

The purpose of this document is to report on the Human Factors analysis of the MASC system supported by Sandia National Laboratories grant number AN-2317. This document provides brief reviews of the MASC interface, the mediation hierarchy and our multiagent test-bed. We will then provide a detailed description of the analysis development, experiments and experimental analysis. Finally, a discussion will be provided which describes the implication of these results for the various components of our system.

1 Introduction

The Multiple Agent Supervisory Control (MASC) system has been built to combine autonomous system aspects with the human's ability to control a system (teleoperation) via a human-machine interface. We have defined the mediation hierarchy which expands upon Sheridan's ([11]) definition of supervisory control. The purpose of this hierarchy is to permit the human operator to interact with the various system levels in order to allow the system to complete feasible tasks. This hierarchy is the underlying concept of the MASC interface development. The combination of the autonomous and telerobotic systems should create a more comprehensive semi-autonomous system which will successfully complete the execution of task assignments. We defined a hierarchy of supervisory intervention into the various system levels [2], which are shown in Figure 1. This hierarchy allows the supervisor to aide the agents when requested. It also permits the supervisor to govern an agent and repeal its processing decisions.

---

[*] 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.

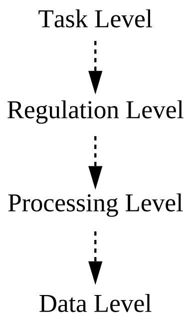

Figure 1: Hierarchical levels of human interaction.

MASC is the human-machine interface, shown in Figure 2, which was developed for the University of Pennsylvania's General Robotics and Active Perception Laboratory's multiagents project. The objective was to create a semi-autonomous system which successfully completes its assigned tasks.

The supervisor interacts with the agents solely via MASC. The supervisor may create autonomous commands via path planning methods or teleoperation commands via mouse interactions. The supervisor may also instruct any of the agents to pause, continue, halt or stop their current actions.

The multiagents system is composed of four robotic agents and MASC. Each mobile agent is composed of a TRC Labmate mobile base. Each agent is heterogeneous in its abilities and sensing capabilities. The two observation agents, shown in Figure 3, are employed for sensing the environment. The sensorBot agent possesses ultrasound and infrared sensors, a structured-light source and camera, and a stereo camera pair. The ultrasound and infrared sensors are employed by the ultrasound process which detects wall-like and corner-like objects. The other modalities are employed in their raw informational modes. The visionBot agent's sensing modalities include a stereo camera pair and a camera on a turn table. The stereo pair of this agent are employed for visually guided obstacle avoidance. The two manipulatory agents, as displayed in Figure 4, are employed to carry objects from one location to another. They must rely upon the observation agents to guide them throughout the environment. The pumaBot is equipped with a Puma 260 manipulator while the zebraBot is equipped with a Zebra-ZERO manipulator.

MASC permits the human to "supervise" [11] the actions of the agents during execution. Through MASC, the human supervises the system while observing sensory data and images. The supervisor is permitted to assist the agents when requested and may assume control of an agent when necessary. Each agent is composed of multiple control and processing levels. In order for the successful semi-autonomous execution of feasible task, MASC must permit the supervisor to interact with these levels which is based upon the mediation hierarchy theory.

The mediation hierarchy has been defined in [2]. This theory permits a human operator to interact with all levels of a robotic system. This mediation hierarchy defines four levels of interaction between the human supervisor and the robotic system. The higher levels of interaction are basic forms which are necessary for the simplest task executions and which are employed in many systems. The lower levels will create a higher level of rapport between the supervisor and the agents. This interaction at the lower system levels should permit the system to request assistance when it is

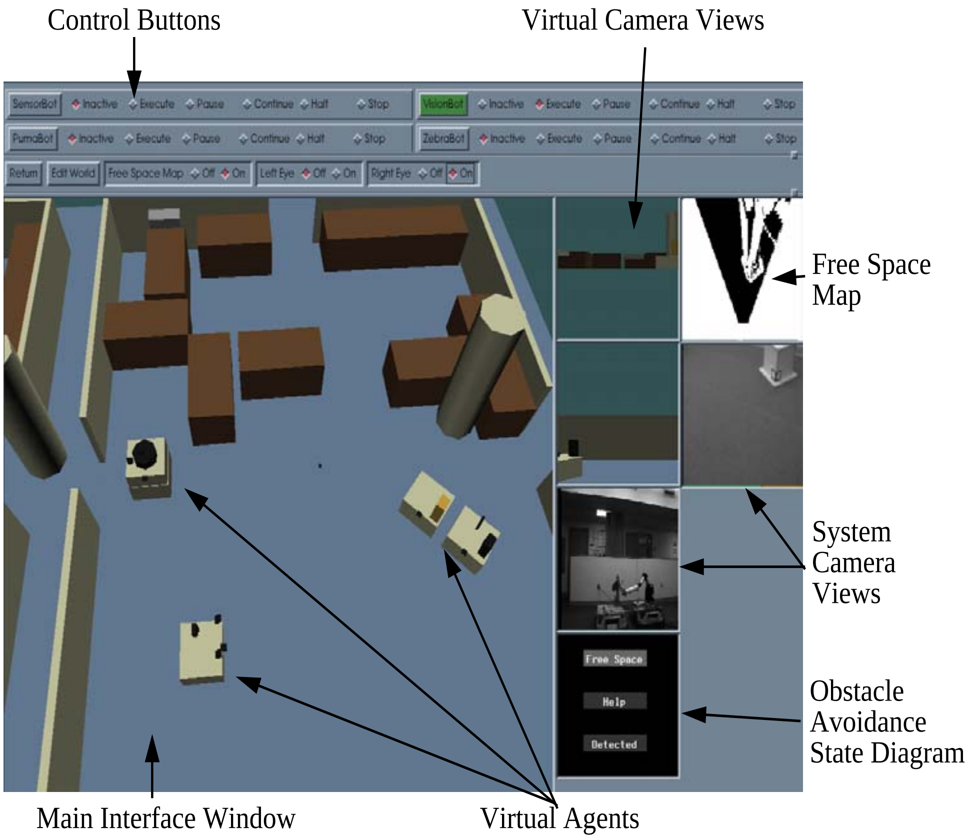

Figure 2: The MASC system interface.

unable to determine a proper action or make a decision based upon the given data. The lower levels of interaction permit the supervisor to interact with the lower levels of the robotic system to facilitate understanding the actions and decisions of these processes as well as permit the human to override process decisions and to supply alternatives to problem solutions. It is the goal of this hierarchy to supply the supervisor with the means to interact with all levels of a robotic system while permitting the system to work autonomously until such interaction is necessary.

The *task level* permits the supervisor to specify the actions an agent or a group of agents execute to complete an assigned task. The tasks may be defined as: exploration of the environment; follow an assigned path to a goal; observe the task execution assigned to another agent; march in formation; and cooperate with another agent to carry items, such as pallets, and the navigation necessary to transport the item from one location to another. Currently, the human decomposes the task into a set of actions which are assigned to the proper agent for execution. Another option is to integrate a task planner to relieve the human supervisor of this duty.

The *regulation level* permits the supervisor to control the multiple agents via three interaction forms. This level of interaction includes the most basic interactions required in such a semi-autonomous system. The *control interaction*

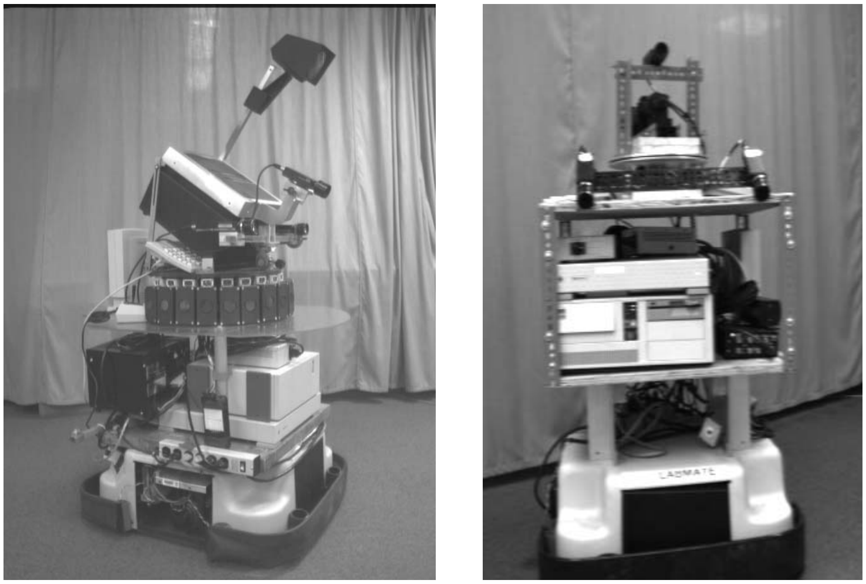

Figure 3: The Observation Agents: (a) SensorBot, and (b) VisionBot.

Figure 4: The Manipulatory Agents: ZebraBot (left) and PumaBot (right).

permits the human supervisor to directly control the agents actions. The supervisor may teleoperate an agent to avoid unstable agent states and assist the agents to avoid command executions which may endanger the agent as well as other agents. The *request interaction* permits the human supervisor to request sensory data and process information from the agents. This interaction allows the supervisor to request only the relevant information for the current situation. As the task execution proceeds, the supervisor is permitted to instruct the system to no longer display unnecessary data and may request relevant information. The *specification interaction* allows the supervisor to specify necessary information prior to processing which may be required by a process.

The *processing level* permits the supervisor to aid a process when it is unable to arrive at a decision and to rectify incorrect decisions deduced by a process either upon a process' request for assistance or as determined by the supervisor. This interaction level will protect the agents from entering unstable states.

The *data level* allows the supervisor to ensure accurate data is passed up through the system for interpretations and processing. It also allows the supervisor to reconfigure the system during a hardware failure.

This hierarchy has been the basis of the MASC system development. The purpose of this section has been to provide the reader with a brief background of the multiagents and MASC systems as well as the mediation hierarchy theory. For more detailed information on the MASC system see [4] and for the Multiagents system see [3]. The purpose of this grant was to permit us to conduct human factors experiments in which we intended to analyze this mediation hierarchy theory. The remainder of this document describes our experimental design, the experimental results, a discussion of the results and a description of future work.

2 Human Factors Experimental Design

The motivation for the mediation hierarchy's development was to create a semi-autonomous multiple robot system which can complete feasible tasks. Therefore, proof of the mediation hierarchy theory entails executing various tasks until the agents require supervisory assistance then demonstrating the supervisor's ability to assist and correct the problem through the MASC system interface followed by the agent's ability to continue with the task execution to completion. We designed a human factors experiment in an attempt to prove our hypothesis. During the experimental development, we found the multiagents system as a whole was not sufficiently sophisticated to fully test the mediation hierarchy theory. Also, it was determined that difficult experiments would require a vast amount of training for a novice user. Thus we developed the experiment to employ only a portion of the MASC system capabilities. The experiment was designed to provide data encompassing the subject's perceived workload, the MASC system's usability and preliminary feedback on the mediation hierarchy. This section's purpose is to provide the experimental design methodology. This section and Sections Three and Four follow the American Psychological Association's presentation standards, [1].

2.1 Purpose

We designed the experiment to follow human factors testing standards, [5, 7, 9, 12]. We employed a consultant to assist with the experimental design. During the design we determined if we wished the subjects to execute difficult tasks, the time and monetary requirements would be beyond our means. This difficulty level was associated with the overall multiagents system design. The multiagents system is fairly complicated and would require extensive training concerning the mechanisms and processes involved. Also, the overall multiagents system is not sophisticated enough to execute difficult tasks. Therefore, we concentrated the experiments upon the subjects workload levels and system usability issues. The experimental design permits some preliminary results to be drawn concerning the mediation hierarchy's role in the MASC interface.

The research question for this study was defined to be:

> Is a novice user with proper training able to effectively interact with the system levels (either when the system requests assistance or the user deems it necessary) such that feasible tasks can be successfully completed in a reasonable time frame with minimal human interaction?

Some questions we wished to answer through this study included:

1. Did the subjects workload levels increase as the number of agents increased?
2. Does workload level decrease over time and increased experience?
3. Did the time to complete assigned tasks increase as the number of agents increased?
4. Are there operator tasks which we should automate?
5. Did the human not detect problem situations which could have been averted?
6. Did the human create unnecessary interactions with the system?
7. What usability issues were detected?
8. Where the subjects able to interact effectively on the task and regulation levels of the mediation hierarchy?

2.2 Tasks

The subjects were required to carry out three tasks: the single agent task, the two agent task and the four agent task. These tasks were executed twice, sequentially during each session. Subjects were permitted to employ the MASC system's initialization and exploration modes. This was due to the immense training required to operate the system in the other system modes.

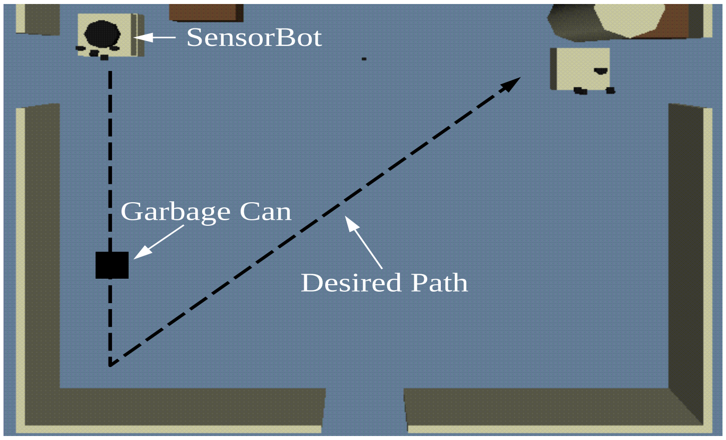

Figure 5: The single agent task

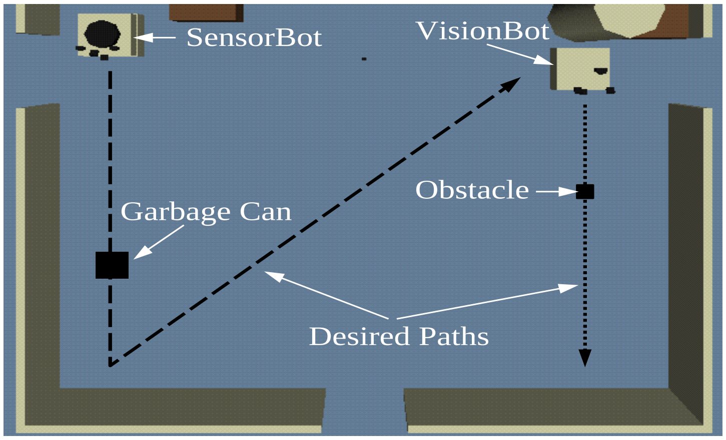

Figure 6: The two agent task

The single agent task required the subject to drive the SensorBot parallel to the southwest wall into the corner. Once the agent obtained the position in the corner, the subject was required to turn the agent and drive it diagonally across the room (the desired path is the dashed line in Figure 5). A tall garbage can was placed approximately two thirds of the distance between the agent's initial position and the corner (the garbage can is shown as the solid rectangle in Figure 5). The subjects were required to drive the agent around the obstacle into the corner. This requirement stems from the fact that the SensorBot is not equipped with an on-board obstacle avoidance procedure. Figure 5 displays the initial set up in the interface model and the actual environment was set up identically.

The possible methods of executing this task include driving up close to the obstacle, turning away from the wall and then driving around the obstacle into the corner and across the room. Another method involved turning the agent at its initial starting point such that it was on an angle to avoid the obstacle. While executing the task, subjects required this agent's sensing modalities to detect the object. While most subjects relied upon the agent's real-time images, they could also employ the raw sonar readings as well as the ultrasound process. The raw sonar readings were useful for detecting the obstacle's position in relation to the agent. The ultrasound process was not as useful, as it requires a large amount of data before detecting objects. Instances of conservative driving did permit this process to provide the obstacle's approximate location information.

The two agent task required the subject to drive the SensorBot as described above while simultaneously driving the VisionBot parallel to the Southeast wall into the corner. The desired path for the SensorBot is the dashed line while the VisionBot's desired path is the dotted line in Figure 6. There was an obstacle placed in the VisionBot's path. The VisionBot's obstacle avoidance process was to be employed to automatically avoid the obstacle. This obstacle and the garbage can, in the SensorBot's path, are shown in Figure 6.

The possible methods to complete this task involve determining which agent to begin moving. As there exist more information available than can be displayed at one time the subjects were to choose the most relevant information for their current requirements. The options for solving the SensorBot's portion of the task are similar to the description for the singe agent task.

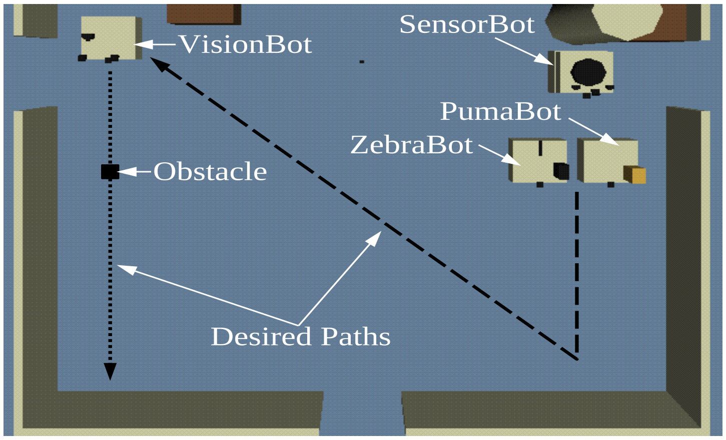

Figure 7: The four agent task

The four agent task required the subjects to simultaneously control all four agents. The VisionBot was positioned as in Figure 7. The subjects were instructed to drive the agent along the Southwest wall into the corner, (the dotted line in the Figure). There was an obstacle placed in front of the agent which the obstacle avoidance process was to avoid, the smaller rectangle in the Figure. The other three agents were positioned as in Figure 7. The two manipulatory agents were in a side-by-side configuration ahead of the SensorBot. The SensorBot's purpose was to observe the manipulatory agent's actions. Subjects were instructed to control the manipulatory agents in the combined control[1] method for as much of the task execution as was feasible. All three agents were to be driven along the Southeast wall into the corner, (the dashed line in the Figure). When this position was obtained, the agents were to turn and move diagonally across the room to the goal position marked in Figure 7.

The possible methods to complete this task are large. Some subjects attempted to turn the manipulatory agents in the combined control. This could not be completed successfully because of their positioning in relation to one another, their bumpers would hit and thus halt their progress. Thus, each agent must be turned individually. Subjects first positioned and turned either the PumaBot or the ZebraBot, this was left unspecified. The SensorBot's specifications for this task was to be maintained in a position such that the manipulatory agents could be observed and its final position was directly behind these agents. Thus, the SensorBot's position during the task was left unspecified. Subjects could have left the SensorBot at its initial position and rotated it to obtain the desired views or they could move the SensorBot along with the manipulatory agents.

2.3 Method

2.3.1 Subjects

The subject group was composed of thirteen computer literate members of the University of Pennsylvania community. Subjects were novice users with mobile robots and had various backgrounds in computer graphics. Most subjects had minimal training with a graphical user interface. The subject's ages ranged between seventeen and thirty-three years and their educational backgrounds ranged from some high school to doctoral level education. There were three female participants and ten male.

All subjects received identical training which included a system description. Training was based only upon those system portions which the subjects would employ for these experiments and lasted a total of thirty minutes. The subjects were paid a predetermined amount for the entire experiment. Payment was not contingent upon completion of the experiments or the amount of time required.

2.3.2 Apparatus

The MASC system version employed for these experiments was pared down from the complete system. It was determined that the entire system would require a training session significantly longer than thirty minutes. This

---

[1] Combined control permits the supervisor to create a single command to be executed by both manipulatory agents simultaneously.

imposed constraints upon acquiring subjects who would be willing to commit a vast amount of time to learning this system. If this was an industrial experiment, in which actually users were involved, it may have been feasible to use the entire system. The pared down version permitted the subjects to use all four agents and their sensing modalities. The locomotion command generation method was teleoperation and the autonomous locomotion methods were not employed.

The robotic agents employed are those described in Section One. The agent's cameras were calibrated prior to the experiments and were not modified during the experiments. All agent configurations were stable throughout the experimental period.

The agent's starting positions were marked on the laboratory floor to assure proper placement for each trail. The curtains surrounding the laboratory's eastern portion, in which the experiments would occur, were closed. Also, doorways into this section of the laboratory were blocked off. It was necessary to block the view of the area so that the agent's actions could not be observed during execution. Thus the subjects did not perceive pressure from others observing their experiments. The closed curtains also restricted the subject's ability to view the area between trails.

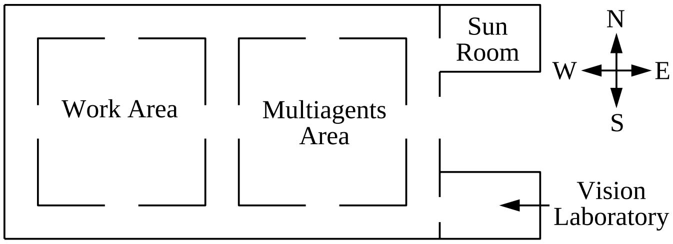

Figure 8: The generalized GRASP Laboratory's floor plan.

The MASC system was run on an Silicon Graphics *Indigo*^{2} with 96 megabits of memory in the Vision laboratory. This is the room to the east of the main laboratory, see Figure 8 for a laboratory diagram. Subjects were physically unable to view the multiagent work area during the experiments. Also, since this is not a high traffic area, subjects were not observed by spectators during the experiments.

Measurement	Method
Effectiveness	Automatic interaction recording Automatic robot status recording Video tapes of the sessions
Usability	Evaluation checklist Post-task Questionnaire Post-experimental questionnaire
Workload	NASA TLX Questionnaire Post-experimental questionnaire

Table 1: Data collection methods.

Data collection was performed employing the various methods displayed in Table 1. Automatic recording of the subjects interactions were built into the MASC system. Also, information pertaining to the agent's position and sensory readings were automatically recorded. Each session was video taped for later review. Prior to beginning the experiments, each subject completed a pre-experimental questionnaire pertaining to their previous experience with various systems which may be similar to the MASC system. After each task trail each subject completed a NASA TLX workload form ([6, 8]). After the completion of each task, subjects completed a post-task questionnaire which asked general questions pertaining to the system's ability to perform this particular task. Finally, upon completion of the entire experiment, subjects completed the detailed post-experimental questionnaire pertaining to the tasks, the system's abilities and usability issues.

The automatic interaction saves provided information such as:

- the task duration,
- how many commands were created,
- how many commands were executed,
- how many system mode changes occurred,
- what sensing modalities were employed,
- which agents were used at particular times, and
- how many and what type of errors did the subjects create.

Each of the above items was time stamped. The second form of automatically recorded data pertained to the positional update information of the agent's (x,y) and heading during task execution. Also, it included information pertaining to the various sensing modalities readings. This information was also time stamped.

A second monitor was placed to the left of the monitor on which the subjects worked. This second monitor was used to video tape the session. A SONY XC-999 Interline Transfer Hyper HAD CCD color camera with a 0.5 inch color sensor and an 8mm focal length lens was used to view the monitor. The information was transferred to an 8mm video recording device. All tasks and trials were recorded.

The NASA TLX workload form was employed to understand the subject's perceived workload level during a task. The questionnaire forms requested information which would help us determine system usability. Each questionnaire asked subjects to rate system aspects on a Likert scale and also provided a section for the subjects to state comments concerning the system and their experience.

I, the designer, completed Ravden and Johnson's, [10] usability evaluation check list. This was employed to raise usability issues and compare them with those detected by the subjects. It also pertained to various system aspects of which subjects had no knowledge.

2.3.3 Procedure

The experiment consisted of four phases: pre-experimental, training, experimental session one and experimental session two. The pre-experimental phase required ten minutes, the training phase lasted thirty minutes and each experimental phase took up to two hours. The pre-experimental, training and experimental session one phases occurred on the first day. The second experimental session phase occurred two days later.

The pre-experimental phase consisted of the subject reading and signing the required consent form to participate in the experiment. Then the subject completed the pre-experimental questionnaire.

The training phase consisted of: agent training and the MASC interface training. Each subject was taken into the experimental area where the agents would execute the tasks. The purpose of the multiagent's project was explained to the subjects. They were provided with information concerning the mobile bases, their non-holonomic features and general slippage information which occurs with these types of robots. Also, they were instructed on the computers employed to control the agents, and how the agents communicated with the interface. After this general introduction they were provided information pertaining to the individual agents. This information included describing the purpose of each agent (observation or manipulation), their sensing capabilities, and their abilities. Each sensing modality and associated processes were explained. Then the subject was taken into the Vision Laboratory where they would run the MASC interface. The interface was displayed on the screen and each step to running the interface was described in the following order:

1. How to chose the proper model file and have it appear on the screen.
2. What the various windows were that appeared on the screen and their purpose.
3. What the various system and agent command buttons were and how they could be used.
4. How to initialize the system by choosing the desired agents and processes.
5. How to display sensory information which is not provided automatically and its uses.
6. The automatic updates of an agent's position and heading.
7. The fact that differences may occur between the actual agent's position and the agent's virtual position.
8. How to switch between system modes.
9. How to determine which of the agents is currently within the human supervisor's control and all agents' execution status.
10. How to switch supervisory control between agents.
11. How to create individual locomotion commands for agents employing teleoperation.
12. How to create combined commands for the PumaBot and ZebraBot employing teleoperation.
13. How to issue emergency stop, continue or pause commands to the agents.

14. What the “phantom” agent is and its purpose.
15. How to shut the system down.

The first experimental session consisted of a practice session, followed by the three task executions. The practice session permitted the subjects to work with the interface, create commands for the agents and explore the various sensing modality displays. They were permitted to practice for twenty minutes. Subjects were not provided a script to follow during this portion of the session, but were permitted to “play” and familiarize themselves with the system.

After the practice session, subjects were provided instructions for their first task. The tasks were randomized among the subjects. The first five subjects completed the tasks in numerical order: single agent task, two agent task and four agent task. The remaining subjects were given task orderings which exhausted the various permutations for the three tasks. Originally, fifteen subjects volunteered to participate in the experiment. Two of the subjects did not participate. Based upon the assumption fifteen subjects would participate, the remaining ten subjects were separated into groups of two. Each group was assigned a different task set randomization. For instance, the four agent task followed by the single agent task followed by the two agent task.

Situations that would end the experiment, such as all involved agents crashing, were explained to the subjects. They were also instructed that if something happened to one agent to continue the experiment with the remaining agents. The task was described in a manner which stated what they were to accomplish, that they should complete the task as quickly and as efficiently as possible and they could achieve the goal in any manner. They were instructed that the environment may have changed since their practice session. Then the subjects were instructed as to which agents would be required for this task, which environmental model was to be used, which agent processes were required, and which sensing processes were available as options for this task. After receiving complete instructions, the subject then began the task. Upon task completion, the subject completed the NASA TLX questionnaire.

The subjects were then required to perform the same task a second time. Again they completed the NASA TLX questionnaire upon task completion. After the second task trial, the subject completed the post-task questionnaire. The second and third tasks were carried out in a similar manner.

The second experimental session, which occurred two days later, was similar to the first. The subjects were permitted a ten minute system re-orientation period. None of the subjects used the entire ten minutes, many only desired a couple of minutes to review the interface. The subjects then executed the two sequential trials of the three tasks in the same order as the first session. After each trial the subjects completed the NASA TLX questionnaire and after each task they completed the post-task questionnaire. At the completion of all tasks the subjects completed the post-experimental questionnaire.

Subjects were not given direction when they encountered problems or were unsure of what their next action should be. They were permitted to continue and use the system to determine what they should do to resolve the situation and obtain their goal.

This section has provided the detail design and methodology for our experiments.

3 Human Factors Experimental Results

This section presents the human factors experimental results in accordance with the standards established by the American Psychological Association [1]. The results are sectioned in accordance to the particular data sets. Then we present results in which various data were compared.

We calculated the various descriptive statistics for all data. We further statistically analyzed some of the data for significance. We created scatter plots of the data and then applied a linear least-squares fit. In most cases, a linear fit proved to be better than a quadratic or higher polynomial fit. This information was then employed to compute the various ANOVA statistics. Please note that it is infeasible to present all tables and graphs in this section, for full details please see [4]. Also, it should be noted that on some graphs the tasks are referred to numerically. Task one corresponds to the single agent task, task two to the two agent task and task three to the four agent task.

3.1 Pre-Experimental Questionnaire

The purpose of the pre-experimental questionnaire was to obtain the subject's background information. There were thirteen participants three female and ten male. The average age of the participants was twenty six years. Nine participants use a computer a majority of their day. Six participants possessed a fair amount of computer graphics knowledge and only two of these six subjects expressed reasonable expertise with three-dimension user interfaces. Only one participant reported knowledge of mobile robots above a beginner's level. The questionnaire and a graphical presentation of the responses are presented in [4].

3.2 Number of Commands

There exist many types of commands subjects may create when interacting with the system. We have broken these commands into four groups: locomotion, system mode, agent mode, and agent switch. Locomotion commands are move and rotation commands created via teleoperation. These are the commands which locomote the agent throughout the environment. The system mode commands are the commands the subject issues to move from one system mode to another, such as from the initialization mode to the exploration mode. The agent mode commands are the emergency stop, continue and pause commands for the individual agents. The agent switch commands are the instances when a subject chooses another agent to be under the human's control.

Table 2 presents the number of each command created by task for all sessions and trials. Locomotion commands composed 59.6% of all commands, system mode commands 11.8%, agent mode commands 18.8% and agent switches 9.8% of all commands. Figure 9 presents the average number of all commands created by task and session. As expected there is an increase in the number of commands from the single agent task to the four agent task. There is only a slight increase, 2.4, in the number of commands created for the single agent task compared to the two agent task in session one. The number of commands created for these two tasks is essentially identical for session two.

Task	Locomotion Commands	System Mode Commands	Agent Mode Commands	Agent Switches	Total Commands
Single Agent	657	168	101	0	926
Two Agent	557	148	115	101	921
Four Agent	1121	144	522	280	2067
Total	2335	460	738	381	3914

Table 2: The break down of all commands created by task for all data.

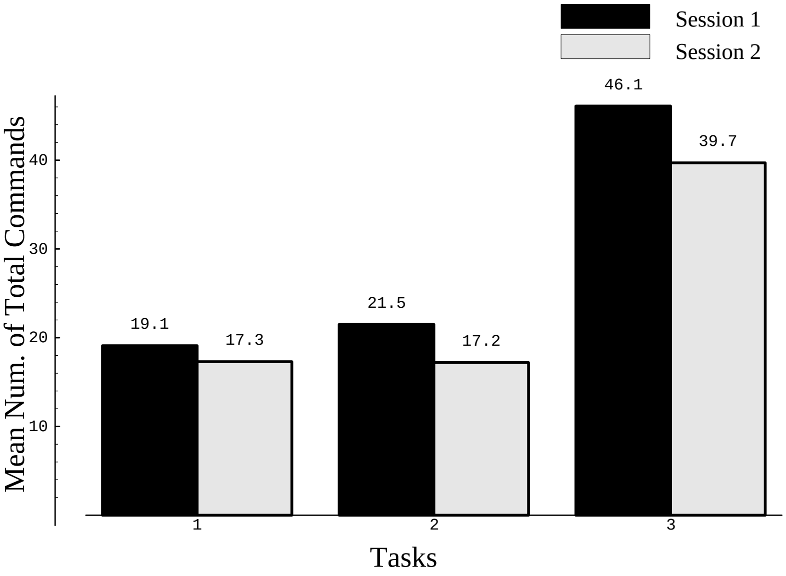

Figure 9: The means for all commands by task and session.

As expected, the average number of commands significantly increases from the single and two agent tasks to the four agent task. In both sessions, the mean number of commands doubled for the four agent task. We computed the linear least-squared fit and the ANOVA results by pairing the total number of commands for the single agent task with the total number of commands for the four agent task. As can be observed in Table 3 and Figure 10, even though there existed a large increase in the total number of commands, it was not statistically significant. The ANOVA analysis of the data model fit found ($f(1,50) = 0.26, P = 0.61$). Reviewing the information in Table 3 we find the relationship between this data for the number of commands is not significant as the x parameter's P value is high and the slope of the fitted line is small. It is interesting to note that the increase in the commands between the single and two agent tasks is either very small or non-existent. In fact, the average number of locomotion commands from the single agent to the two agent task falls in both sessions by approximately one command in session one and two and a half commands in session two. Again the ANOVA analysis shows this relationship is not significant. This is also true of the relationship between the two agent task and the four agent task.

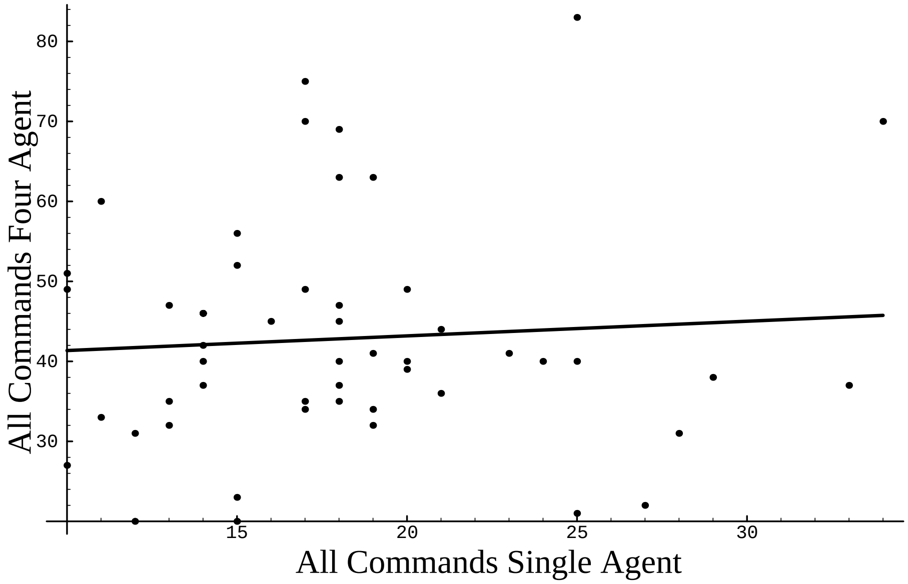

Figure 10: The total number of commands for single agent task plotted against four agent task for all data.

We also analyzed the total number of commands for the individual tasks between sessions. We found the decrease in the number of all commands between sessions one and two for the single agent task to be significant. The analysis results for the single agent task can be found in Table 4. These results are significant as both probability values are low and the slope is positive. The results showed the decrease in the number of total commands created in session two versus session one for the two agent and four agent tasks were insignificant. The two agent task produced $(f(1,24) = 0.26, P = 0.62)$ and the four agent task $(f(1,24) = 0.0003, P = 0.99)$.

Parameter	Slope Value	P Value
1	39.52	0
x	0.18	0.61

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	53.55	53.55	0.26	0.61
Error	50	10197.2	203.944		
Total	51	10250.8			

Table 3: Total number of commands between the single and four agent tasks for all data.

Parameter	Slope Value	P Value
1	11.28	0.001
x	0.31	0.05

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed *f*	P Value
Regression	1	91.81	91.81	4.12	0.0537
Error	24	535.303	22.3		
Total	25	627.115			

Table 4: Total number of commands for the single agent task between sessions.

3.3 Sensing Modalities

Each subject was required to chose which sensing modalities to employ during each task. We recorded which modalities they employed, the duration for which they were used, and how many of the available sensing displays were employed. Information regarding the agent's position and heading was not included in this data as the system provides this automatically. Each subject initialized their sensing displays prior to beginning the actual task and none of the subjects modified their displays during a task. Table 5 lists the various sensing modalities by the percentage of time the modality was employed by the subjects. Note that not all modalities may be employed in all tasks. The real-time images were the subjects predominant display choice. In fact, they were the subjects top five preferences. These were followed by the obstacle avoidance free space map, the raw ultrasound and infrared displays and the obstacle avoidance state diagram. Most subjects did not find the information provided by the ultrasound process useful and preferred not to use it. The path planning state diagram associated with the obstacle avoidance process

Sensor	Number Times Employed	Total Occurrences	Percentage
SensorBot Left Stereo Camera	134	151	88.7%
SensorBot Right Stereo Camera	130	151	86.1%
VisionBot Left Stereo Camera	78	100	78%
Light Stripe Camera	97	151	64.2%
VisionBot Right Stereo Camera	61	100	61%
Obstacle Avoidance Free Space Map	54	100	54%
Raw Ultrasound Display	79	151	52.3%
Obstacle Avoidance State Diagram	38	100	38%
Ultrasound Process	12	151	7.9%
Ultrasound Process State Diagram	7	151	4.6%
Path Planning State Diagram	4	100	4%

Table 5: The display methods and the percentage of time they were used.

was the least used modality.

There exist a number of reasons we have provided the human supervisor with the ability to turn on and off sensing displays. In particular, there exists limited real estate for such sensing displays. The system permits the supervisor to display up to six windows to the right of the main working window. If all sensing modalities are employed for the single agent tasks, they require only four of these windows but the two and four agent tasks encompass nine sensing displays. We collected information regarding how many of the available displays the subjects employed for each task. We found subjects employed all available displays, including those in the main working window, four out of the one-hundred fifty six trails, or 2.56% of the time. Table 6 presents the average number of displays employed by task for all trails and sessions. The total number of available sensing modalities and displays for the single agent task was six and for the two and four agent tasks there were eleven available sensing modalities and eight available displays. As displayed by the table, the average number of displays employed for the single agent task was only slightly below the total number of available displays while for the two and four agent tasks the difference was larger.

Task	Session One	Session Two
Single Agent	5.38	5.5
Two Agent	5.77	6.68
Four Agent	6.63	6.1

Table 6: The average number of displays employed by task and session for all trials

Figure 11 displays the number of displays used for the single agent task during both sessions. It is interesting to note that one subject attempted to execute the single agent task with no sensing modalities, this subject also ran into the garbage can placed in the experimental area. Also, all six displays were employed for two of the 26 trials. As the Figure shows, the preferred number of displays for this task was three in session one and four in session two.

The number of sensing displays available for the two agent task were eight. The subject's choices for all trials by session are displayed in Figure 12. The least number of displays employed was two and the largest number was seven. The most frequent number of displays employed was six. This increase from the single agent task would be expected as we have introduced another agent as well as other sensing modalities into the task.

The four agent task again increased the number of agents involved in the task but did not introduce any new sensing modalities or displays. Figure 13 presents the number of displays the subjects employed by session for all trials during the four agent task. Again, the lowest number of displays employed was two and in this task two trials employed all available displays. The subject's predominant choice was to employ six displays which is equivalent to the two agent task.

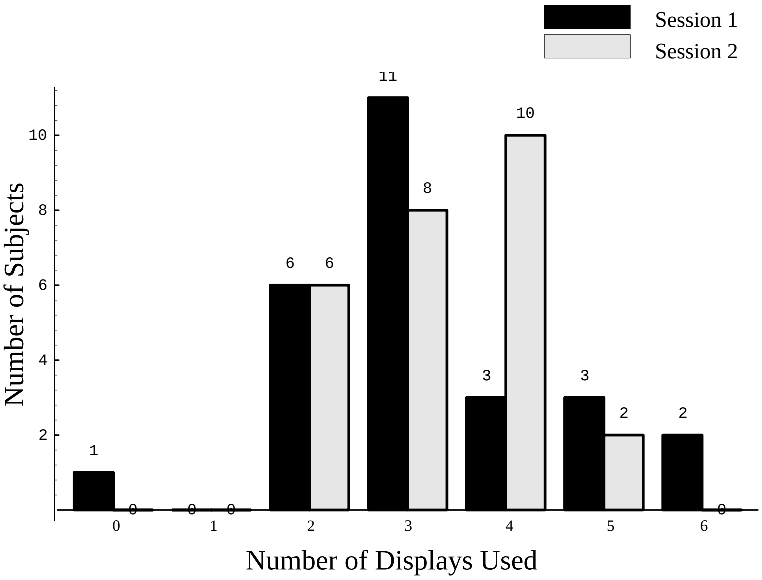

Figure 11: The number of displays employed for the single agent task by session.

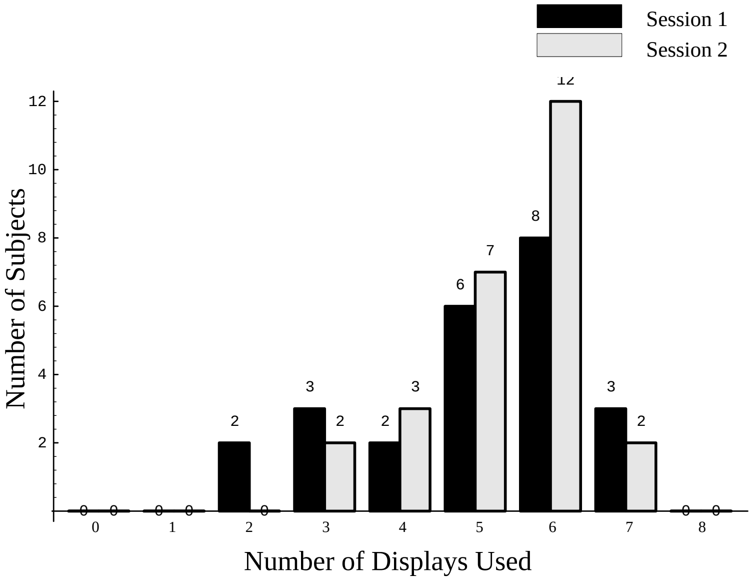

Figure 12: The number of displays employed for the two agent task by session.

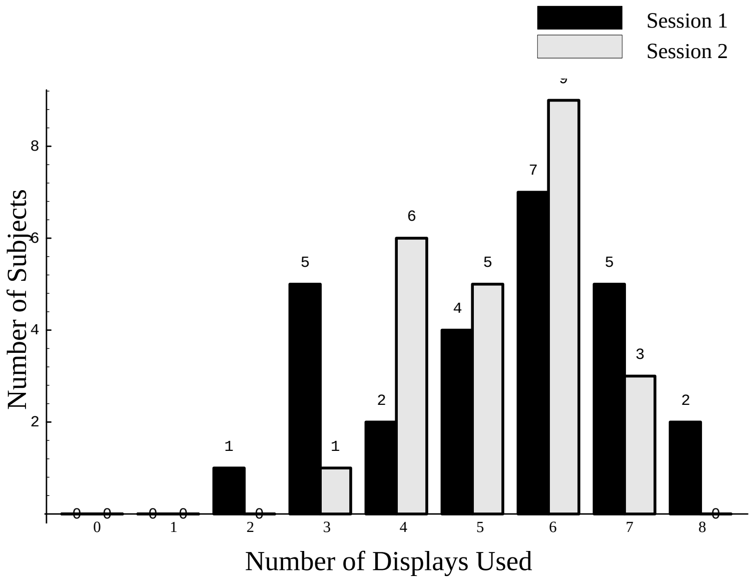

Figure 13: The number of displays employed for the four agent task by session.

3.4 Number of Errors

Every time a subject attempted to instruct the system to do something that was incorrect, they would receive an error message and this information was automatically recorded. These types of errors were encountered a total of thirty seven times, that rounds to approximately three errors per subject. Only one subject received no errors, two subjects received one error and four subjects received more than three errors.

Error	Occurrence
Attempting to work with un-chosen agent	20
Attempting to re-initialize an agent	7
No Camera Server	3
No SensorBot Robo Process	2
No ZebraBot Robo Process	3
No Path Following Process	1
No Ultrasound Process	1
Total number of errors	37

Table 7: Errors subjects received by frequency of occurrence.

Table 7 lists the errors which occurred by their frequency. The predominant error was attempting to instruct an agent, other than the agent currently under the supervisor's control, to stop, continue, pause or halt. This error accounted for over half, 54%, of all errors. The second most frequent error was attempting to initialize an agent and its processes after the subject had already completed the agent's initialization. The "No Camera Server", "No

Path Following Process" and "No Ultrasound Process" errors occurred when the subject attempted to display the processes sensory information when the process had not been requested. The errors relating to the SensorBot and ZebraBot occur when the subject's attempted to use those agents and had not initialized them for the task.

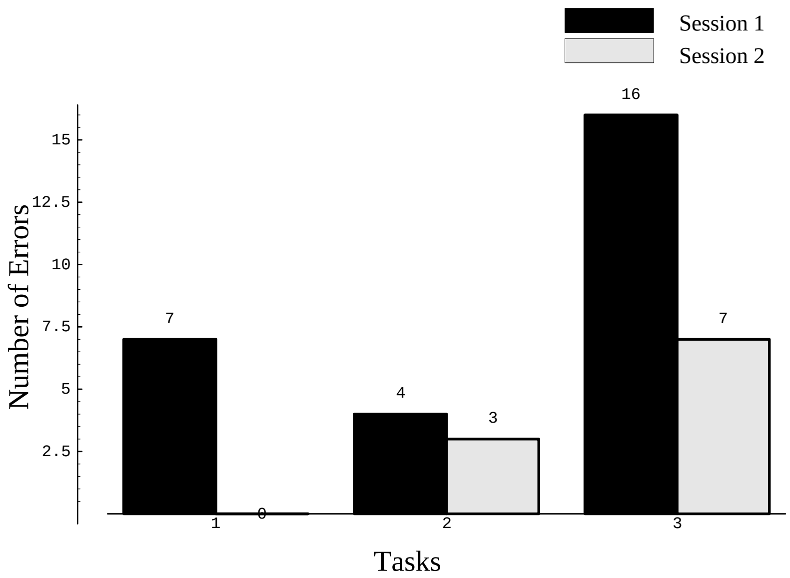

Figure 14: The number of errors by task and session for all trials.

The number of errors for each task by session for all trials is presented in Figure 14. As is presented, there were twenty seven errors in session one and only ten in session two. In both sessions errors occurred most frequently during the four agent task. The overall number of errors between the single and two agent tasks was equal. When we computed the ANOVA's for the number of errors during the single agent task compared to the number of errors during the two and four agent tasks, there was no statistical significance. The comparison with the two agent task resulted in $(f(1, 50) = 0.044, P = 0.55)$ and for the four agent task $(f(1, 50) = 0.38, P = 0.54)$. This was also true when we compared the two agent task to the four agent tasks, $(f(1, 50) = 0.003, P = 0.96)$.

3.5 Task Completions

This section presents the results concerning the number of task completions and accidents. The number of overall completions is broken into two groups, those who experienced system failures and those that did not. System failure examples include instances when an agent would not move because it believed its bumper was activated or when an agent lost communications with the rest of the system. Subjects had been routinely instructed that when a system failure occurred they were to continue to execute the task with the remaining agents. All such instances resulted in what we quantify as a "completion under the circumstances".

There were one hundred thirty four successful completions, twelve system failure completions and a total of eight accidents. These results are presented in Table 8 by task and session. This data shows that the subjects successfully

	Completions Session		Incomplete Session		Accident Session		System Problem Session	
Task	One	Two	One	Two	One	Two	One	Two
Single Agent	25	25	0	0	0	0	1	1
Two Agent	20	25	2	0	2	1	2	0
Four Agent	20	19	0	0	2	3	4	4
Total	65	69	2	0	4	4	7	5

Table 8: Break down of task completions results by task and session.

completed the assigned tasks with no system problems 85.9% of the time. The successful completion rate increases to 93.6% if we factor in the twelve trials in which system failures occurred and the subjects continued on with the task. Accidents occurred in only 5.1% of the trials. All the single agent task trials were successful while two accidents occurred during the two agent task. The two occurrences of "Incomplete" for the two agent task were related to system problems and the subject's time constraints and thus did not complete these trials. The highest number of accidents occurred during the four agent task. This is not surprising as there are many more variables in this task.

The system failures which occurred were related to the following agent hardware: serial line communications, batteries, bumpers and power cables. The serial line communications between the two manipulatory agents and the rest of the multiagents system was the single most frequent failure occurring a total of six times. Three of the instances involved only the PumaBot while the remaining three instances involved both the PumaBot and the ZebraBots loosing communications. There was one instance when the SensorBot's front bumper was activated for an unknown reason and a similar instance with the VisionBot. There were two instances when the SensorBot's battery was dead and the agent was not able to function properly. Finally, there were two instances when the SensorBot's wheels became stuck on its power cable and it was unable to move. In all instances, the subjects were able to continue on with the remaining agents to complete the task.

3.6 Task Completion Times

The amount of time each subject required to complete a task was automatically recorded in seconds. Figure 15 presents the average completion times by task and session. As expected, the figure shows the time to complete a task increased as the number of agents involved increased. Also, there was a decrease in the means from session one to session two for all tasks. The single agent task completion times dropped by 96.2 seconds, the two agent task times by 68.5 seconds and the four agent task times dropped by 193.3 seconds between session.

We analyzed the completion time differences between tasks. We found the variances were significant between the single and the two agent tasks, $(f(1,50) = 6.04, P = 0.018)$ but the relationship between the two and four agent tasks was not as significant, $(f(1,50) = 3.7, P = 0.059)$. The relationship was insignificant between the single and four agent tasks, $(f(1,50) = 25.34, P = 0)$ as can be observed by reviewing table 9. The variance between the single versus four agent tasks is insignificant due to the large probability value of the constant.

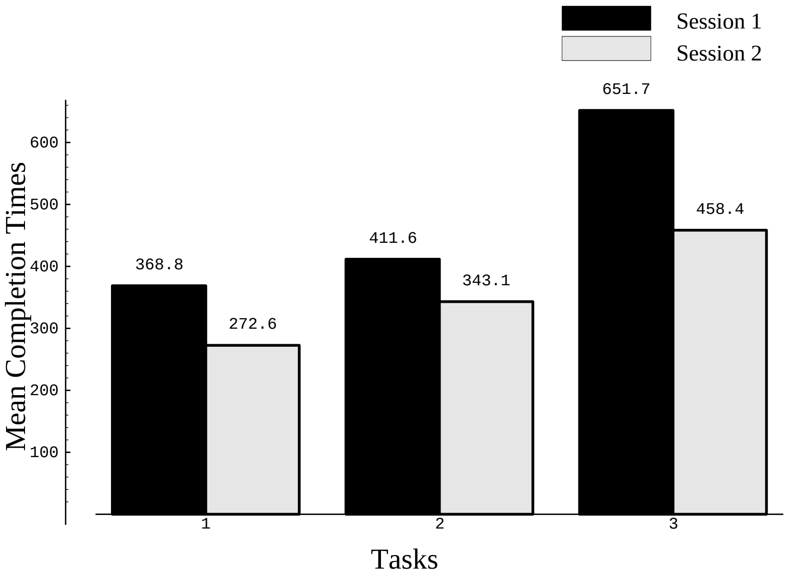

Figure 15: The mean completion times by task and session in seconds.

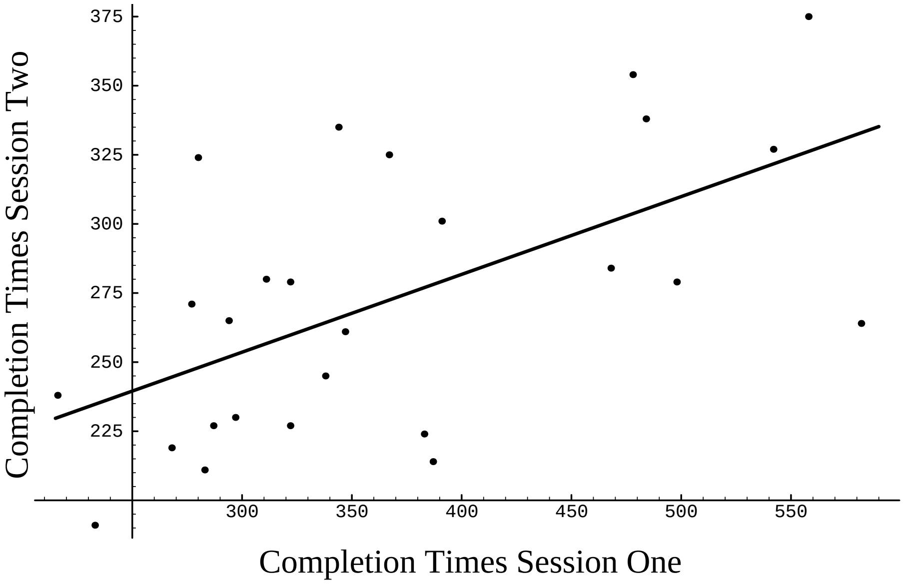

Figure 16: The completion times for the single agent task between both sessions for all trials.

Parameter	Slope Value	P Value
1	114.795	0.21
x	1.37	0

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	843668	843668	25.34	0
Error	50	$1.66483 * 10^6$	33296.6		
Total	51	$2.5085 * 10^6$			

Table 9: Task completion times between the single and four agent tasks for all sessions and trials.

We further analyzed the task completion time data between sessions for each task. The results showed that a significant result existed between sessions for the single agent task, ($f(1, 24) = 12.4, P = 0.002$). These results are presented in Figure 16 which shows the slope of the x parameter is positive. The constant probability value was 0.00001 and the the linear parameter probability was 0.002. The results were insignificant for a similar comparison between sessions for either the two agent task, ($f(1, 24) = 0.3, P = 0.59$) or the four agent task, ($f = 0.17, P = 0.68$).

3.7 Perceived Workload Measures

We recorded the subject's perceived workload employing the NASA TLX method. At the completion of each trial the subject completed a NASA TLX questionnaire. This information was combined with the subject's responses to the weighting measure pairs from the post-experimental questionnaire. The perceived workload values were calculated as the method prescribes. The resulting values range from zero to one hundred. The lowest perceived workload value was 1.83 and the highest was 67.49. Figure 17 displays the mean perceived workload values by task and session. As can be observed, the means increased between the single and two agent tasks as well as the two and four agent tasks for session one. During session two there was only a slight increase in the value between the single and two agent tasks. Again there existed an increase for the four agent task during session two. There was a decrease in the mean values for all tasks between sessions one and two. Figure 18 displays this result in graphical form.

We further analyzed the data for significance. We computed the ANOVAs between tasks and found the comparisons to be significant. The comparison between the single agent and the two agent task found the constant probability to be 0.0009 and the linear parameter to be 0 with a positive slope value. The comparison of the two and four agent tasks found the constant parameter $P = 0.0006$ and the x parameter $P = 0$ with a positive slope value. Table 10 presents the results for the comparison between the single and four agent tasks. Again the probability value was very close to zero with a positive slope value.

We also analyzed the data between sessions for a single task. The results showed that the differences in perceived workload for each task between sessions was not significant. The single agent task resulted in an a constant probability

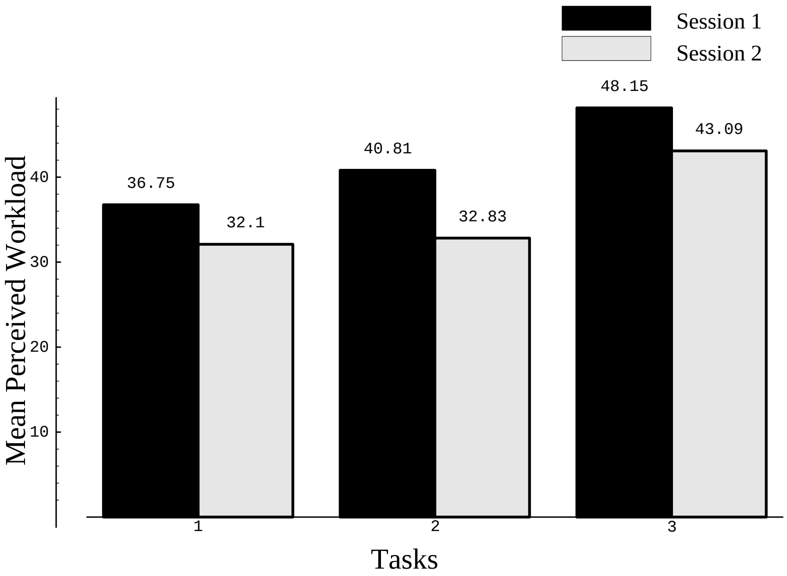

Figure 17: The perceived workload means by task and session for all trials.

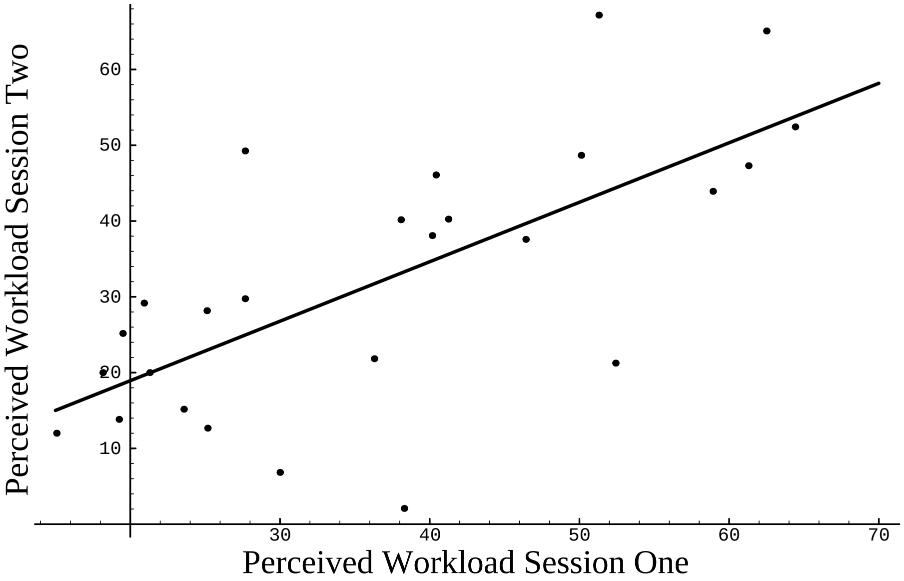

Figure 18: The perceived workload measures for the single agent task between both sessions for all trials.

Parameter	Slope Value	P Value
1	24.96	0
x	0.60	0.00001

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	4950.86	4950.86	23.28	0.00001
Error	50	10632.3	212.65		
Total	51	15583.2			

Table 10: Perceived workload measures between the single and four agent tasks for all data.

of 0.63, ($f(1,24) = 21.74, P = 0.00001$). The result for the two agent task computed the constant $P = 0.19$, (f(1,24) = 21.74, P = 0.00001). The four agent task results showed the constant $P = 0.58$, (f(1,24) = 87.65, P = 0). As the constant parameter probability values are insignificant, there exist no significant relationship.

3.8 Multiple Data Comparisons

The previous sections presented the results for the individual data collection items. This section will present the data analysis results between some of the data item groups. The purpose of this analysis was to determine the significance of these relationships.

3.8.1 Perceived Workload Measures

We were interested in determining what factors contributed to the subject's perceived workload measures. Thus we analyzed the workload data in comparison with the task completion times, the total number of commands and the number of errors subjects created.

Perceived Workload Measures Versus the Total Number of Commands We began by exploring the general relationship between the total number of commands created for all task compared with the perceived workload values for all tasks. This analysis determined there existed a significant relationship between the two data sets which is presented in Table 11. As this Table displays, the computed linear parameter probability value is 0.0004 and the constant value is 0. This combined with the positive slope value indicate a significant relationship. Based on this result we further analyzed the data for all tasks between sessions.

Figure 19 presents the results for the comparison of the total number of commands versus perceived workload data for all tasks during session one. The relationship was found to be significant as the constant P value was 0 and

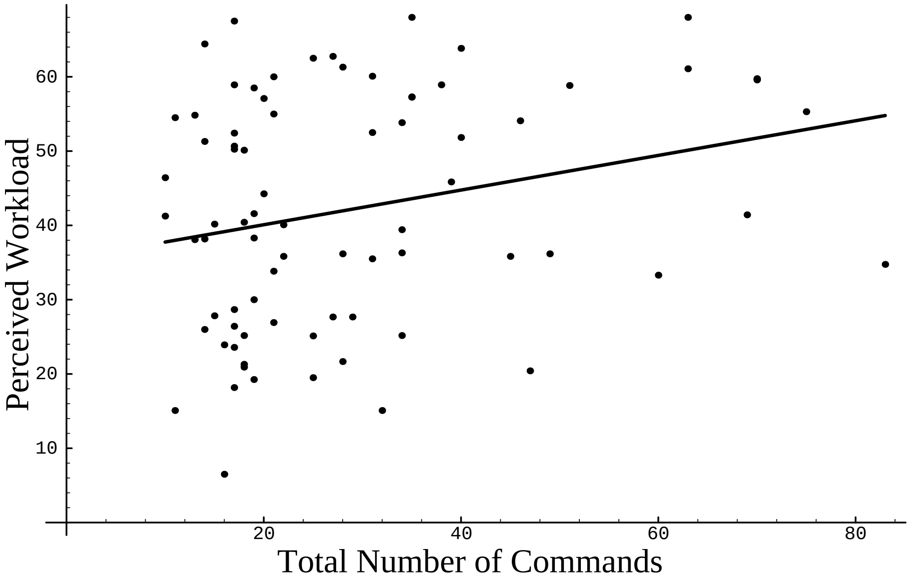

Figure 19: The perceived workload measures versus the total number of commands for all data during session one.

the x parameter's probability value was 0.026 with a positive slope value. Table 12 presents the results of the same comparison for session two. As the table displays, the computed linear coefficient's $P = 0.01$ while the constant P was 0 with a positive slope value. Since these relationships were found to be significant we further analyzed the data by the individual tasks.

The analysis of the individual tasks employed the data for all trials and sessions of the particular task. In general, we found none of these relationships to be significant. The analysis for the single agent task found $(f(1,24) = 0.07, P = 0.79)$. The analysis of the two agent task resulted in $(f(1,24) = 2.41, P = 0.127)$. Finally, the analysis of the four agent task found $(f(1,24) = 1.02, P = 0.318)$.

Parameter	Slope Value	P Value
1	30.45	0
x	0.32	0.0004

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	19381	19381	201.47	0
Error	154	14814.2	96.2		
Total	155	34195.2			

Table 11: Total number of commands versus perceived workload measures for all data.

Parameter	Slope Value	P Value
1	25.4	0
x	0.43	0.01

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	2069.42	2069.42	6.83	0.011
Error	76	23017.2	302.86		
Total	77	25086.6			

Table 12: Total number of commands versus perceived workload measures for session two.

Parameter	Slope Value	P Value
1	38.62	0
x	1.88	0.4

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	205.7	205.7	0.7	0.4
Error	154	45328.7	294.34		
Total	155	45534.4			

Table 13: Total number of errors versus perceived workload measures for all data.

Perceived Workload Measures Versus the Number of Errors The relationship analysis between the perceived workload measure and the number of errors the subjects created was necessary as one would believe there should be a significant relationship. Table 13 presents the results of this analysis. As the table displays, there was no significant relationship between the two data sets as the linear coefficient's probability was 0.4. As there existed no significant relationship we did not further pursue this relationship.

Perceived Workload Measures Versus the Task Completion Times Finally we explored the relationship between the perceived workload measures and the task completion times. We found that in general when comparing the data sets for all data there existed a relationship which was significant, as displayed in Table 14. Comparing the probability values and the slope value we found the P values low while the slope is rather small but still positive.

We continued to analyze the data by session as we found there was a significant relationship in the general analysis. The relationship for the data in the first session revealed a constant P value of 0 and a linear parameter P value of 0.052 with a small but positive slope value. The same analysis performed on the data from the second session found an insignificant relationship as the constant $P = 0.11$.

Parameter	Slope Value	P Value
1	26.36	0
x	0.03	0.00006

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	4565.08	4565.08	17.16	0.00006
Error	154	40969.3	266.035		
Total	155	45534.4			

Table 14: Task completion times versus perceived workload measures for all data.

The analysis of the individual tasks found the relationship for all data in all tasks was insignificant. The results for the single agent case found $(f(1, 50) = 2.73, P = 0.11)$, for the two agent task the constant $P = 0.12$, and for the four agent task, $(f(1, 50) = 1.45, P = 0.23)$. The results for the two agent task are presented in Figure 20.

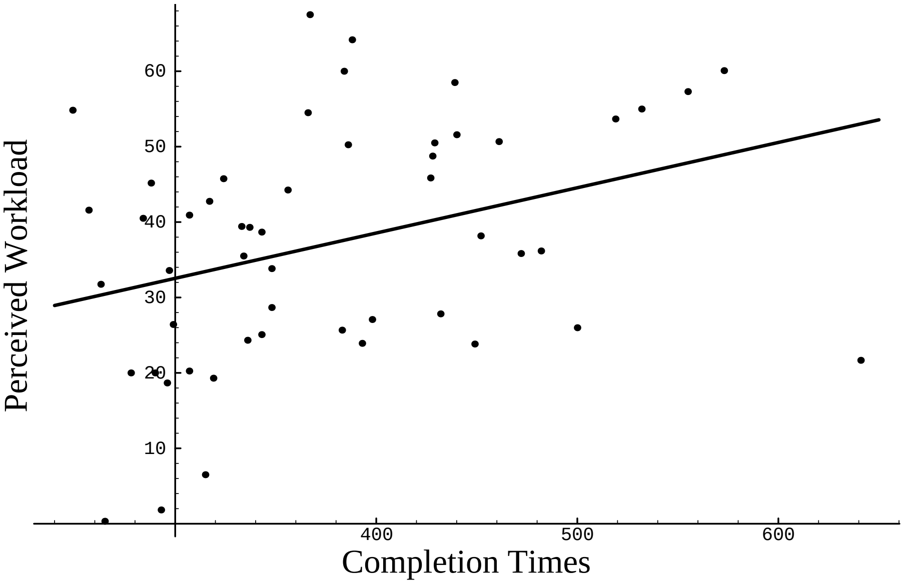

Figure 20: The perceived workload versus completion time for the two agent task for all data.

Based upon the combination of these results we can state in general the subject's perceived workload is effected by the total number of commands created for a task and by the task completion time. It does not appear to be adversely effected by the number of errors a subject committed.

3.8.2 Number of Commands

Next we explored the relationship between the total number of commands with the total completion time, and number of errors. As discussed in Section 3.8.1, we analyzed the relationship between the command data and the perceived workload.

Number of Commands Versus the Number of Errors We analyzed the relationship between the number of commands created versus the number of errors which occurred. The model results were significant with constant $P = 0$ and the linear term's $P = 0.026$, with a positive slope. This relationship is displayed in Figure 21.

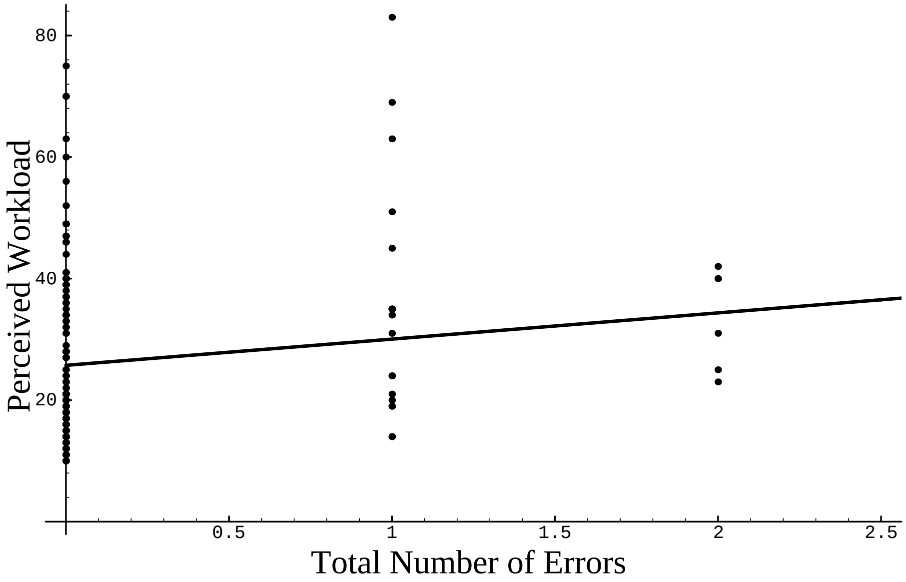

Figure 21: The number of errors versus the number of commands for all data.

Number of Commands Versus the Completion Times It was expected that we should see a significant relationship between the total number of commands created and the task completion times. As number of commands increases one would expect to see an increase in the completion times.

We analyzed the data for all tasks and session and found in general the linear term probability value was significant with a positive slope but that the constant probability value was not significant as shown in Table 15.

We then analyzed the data between sessions. The data from session one displayed a relationship which is insignificant with a constant $P = 0.91$. The constant probability value for the second session was 0.052 which is also considered to be insignificant.

Parameter	Slope Value	P Value
1	0.55	0.78
x	0.06	0

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed *f*	P Value
Regression	1	19381	19381	201.47	0
Error	154	14814.2	96.2		
Total	155	34195.2			

Table 15: Task completion times versus total number of commands for all data.

The data analysis between sessions for the individual tasks displayed significant results for the single and four agent tasks. The relationship for the single agent task found the constant P value to be 0.002 and the linear term's probability to be 0.00002 with a positive slope value. The results for the analysis of the two agent task were found to be insignificant as the constant probability value was 0.08. Figure 22 displays the relationship between sessions for the four agent task. The constant probability value was 0.00002 while the probability for the x term was 0 with a positive slope value.

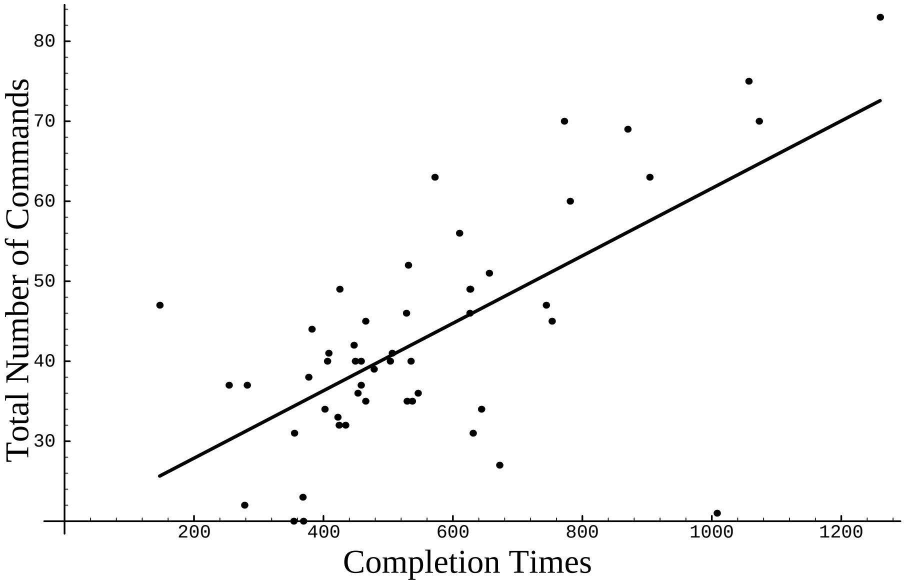

Figure 22: The total number of commands versus completion times for the four agent task between sessions.

Number of Commands Versus the Perceived Workload The analysis results of the number of commands versus perceived workload relationship were discussed in Section 3.8.1, therefore we will not repeat them in detail. It was found this relationship was significant in the general comparison as well as the between session and tasks.

3.8.3 Number of Errors

The relationship between the number of errors and perceived workload was found to be significant and was discussed in Section 3.8.1. A similar comparison between the number of errors and the total number of commands was found to be significant and was discussed in Section 3.8.2. The remaining relationship to report is the relationship between the number of errors and the task completion times.

Number of Errors Versus Completion Times The relationship between the number of errors and the completion times for a task was generally found to be significant. Table 16 presents the analysis results. As can be observed from the table, both probability values were significant and the slope was also significantly positive.

Parameter	Slope Value	P Value
1	400.7	0
x	70.8	0.0022

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	Computed f	P Value
Regression	1	291531	291531	9.69	0.0022
Error	154	$4.631 * 10^6$	30076.4		
Total	155	$4.923 * 10^6$			

Table 16: Total number of errors versus task completion times for all data.

As the results show, there was a significant relationship between the number of errors created and both the task completion time and the total number of commands created for a command. The relationship with the perceived workload was found to be insignificant.

3.8.4 Completion Times

The analysis results comparing task completion times have been previously discussed in this Section. The comparison with perceived workload was discussed in Section 3.8.1, with the total number of commands in Section 3.8.2 and with the number of errors created in Section 3.8.3.

The general relationship between completion times and perceived workload was found to be significant but the analysis results between sessions and tasks was mixed. The relationship analysis with the number of commands revealed no significant relationship in the general case as well as between sessions and was mixed between tasks. The relationship between this data and the number of errors created was also found to be significant.

3.9 Post-task Questionnaire

As subjects completed the two trials for a task during both experimental sessions, they then completed the post-task questionnaire. This questionnaire asked specific questions about the task's difficulty and the system's ability to perform this task. The subject's responded to questions by circling a number from 0 to 9 on a Likert scale.

The first question inquired as to the task difficulty. It was found that during the first session of the single agent task six subjects found the task fairly easy while during the second session eight subjects found the task "easy". No subjects found this task to be in the difficult to impossible range. The same question for the two agent task produce a majority of "easy" results during both sessions and none of the subjects found the task "difficult". The subjects' responses for the same question after the four agent task for session one found that most subjects did not find the task easy. This result flipped when asked the question again during the second session, with the majority finding it much easier.

When the subjects were asked to rate the task as "confusing" or "clear", roughly 73% of the subjects found all three tasks to be clear during both sessions. Only one subject found the four agent task somewhat "confusing" during the first session.

The subjects were asked to rate the task's stimulation factor. In general, the subject's replies fit into all spectrums of the scale for the single and two agent tasks during both sessions. The subjects found the four agent task to fall more towards "stimulating" than "dull". Many subjects commented that during the second session they were bored with executing the same tasks and would have preferred new tasks which were more challenging.

The subjects gave varying results when asked to rate the task on a scale from "frustrating" to "satisfying". At least one subject found each task somewhat frustrating during both sessions. The subjects who found the task somewhat satisfying during the first sessions tended to increase their satisfaction level during the second session.

The second question queried the subjects on their level of system control during a task. The scale for this question was between "rarely" and "always". During the single agent task of the first session most subjects felt they had considerable control over the system. This feeling increased during the second session for the same task. The results for the two agent task during the first session showed that most subjects did not feel they were always in control of the system and one subject appeared to only feel in control about 50% of the time. This again dramatically increased for the second session. While the subject's responses remained on the scale's high end for the four agent task during the first session only two subjects felt as though they "always" had control of the system. These responses increased for the second session. These results were likely due to the subjects' increased familiarity and practice with the system.

When the subjects were asked about their ability to understand and interpret the data readings during a task most replied they “always” understood during the single agent task for both sessions. During both sessions of the two agent task this number slightly decreased but all subjects replied in the scale’s upper portion. This would be expected as the two agent task introduces more information. The responses to the same question after completion of the four agent task proved similar to those of the two agent task. This particular result is not surprising as the types and number of sensing displays available between the two and four agent task is stagnant.

Subjects who received errors felt they were “always” or almost “always” capable of correcting their errors during all tasks for all sessions. Only one subject in both sessions of the four agent task did not feel this was true. This is a result which shows the interface provides the user with the information and capabilities to correct errors.

When asked if they felt as though they were “always” or “rarely” capable of completing a task, only one subject during the second session of the four agent task responded in the scale’s bottom half. All other subjects felt they were able to complete the tasks.

The subjects were asked if they felt in control of the individual agents during the task. As would be expected, the single agent task resulted in the highest number of “always” replies. This task requires only one agent, therefore the subject does not have the opportunity to become distracted while working with another agent. This result was true during both sessions of this task. The responses for the two agent task showed that during the first session the number of replies stating “always” was nearly half that of the single agent task. The responses increased during the second session but still were not as high as the single agent task. In general, most subjects felt they maintained a reasonable amount of control over the agents during the two agent task. This was also true in the four agent task. The number of subjects that “always” felt in control of the individual agents was nearly half that of the single agent task and slightly lower than the two agent task.

When asked if the system’s power to complete this task was “inadequate” or “adequate” the general response was that it was mostly “adequate” for the single and two agent tasks. These responses increased between sessions one and two. A fair number felt the system maintained adequate power during the four agent task but this result was lower than that of the first two tasks.

The last question asked if the system was flexible enough to perform this task. The results were quite mixed for all tasks. While the responses remained in the scale’s upper portion, they were scattered for the single agent task during the first session. The second session resulted in a higher opinion of the system’s flexibility for this task. As the number of agents increased, the subject’s opinions of the system’s flexibility decreased but tended to increase slightly during the second session.

This questionnaire also provide a space for the subjects to provide comments about this particular task and the system. Their comments will be discussed in Section 3.11 along with their comments from the post-experimental questionnaire.

3.10 Post-Experimental Questionnaire

The post-experimental questionnaire was presented to the subjects at the completion of all tasks at the close of the second session. Its purpose was to obtain their general opinions of the system and to extract system usability issues. The questionnaire and a graphical presentation of the results is presented are [4]. Each question required the subject to circle a response between zero and nine on the Likert scale. This section will discuss the results for this questionnaire.

The first question asked for an overall reaction to the system based on six scales. The first nine value scale rated the system from “terrible” to “wonderful”. The majority of the responses, seven, rated the system a seven while the remaining six responses rated it eight and nine. Although the system is not “wonderful”, it still was rated in the scale’s upper half. The second scale rated the system from “frustrating” to “satisfying”. Two subjects found the system somewhat frustrating, two more subjects fell into the middle of the scale and the remaining subjects rated it in the scale’s higher third. In general, this tells us that the system tends not to frustrate the users. The third scale rated the system’s stimulation and dullness factors. All but one subject rated the system in the scale’s upper third, with seven responding with ratings of eight or nine. The responses on the scale’s lower portion may be due to the task repetition, as some subjects found this repetition boring. When asked if the system was “difficult” to “easy”, nine subjects responded in the scale’s upper third. One subject felt the system rated a three. The fifth scale rated the system’s overall power as “inadequate” or “adequate”. All but one response was in the scale’s upper third. The final scale rated the system’s “rigidity” and “flexibility”. Eleven subjects responded in the scale’s upper third for flexibility while two subjects felt the system rated a five. The overall reaction to the system was generally favorable.

The second set of questions related to the system’s screen or window layout. The first question asked if the current layout was “helpful”. Five subjects responded that it was while the remaining replies were scattered about the scale. This may be due to the user’s inability to reorganize the small windows to the left of the main window. Nine subjects found the one main working window “helpful” while the remaining four subjects responded in the scale’s upper portions. Subjects were queried as to the frustration level of turning on and off the data displays. Six subjects found this a “satisfying” feature while one found it “frustrating”. The remaining six responses were in the scale’s upper half. When asked if they found this ability “easy” or “difficult”, eight subjects felt it was “easy”, three subjects rated it somewhat easy and one subject rated it a five. When queried if this ability was “rigid” or “flexible”, eight rated it in the scale’s upper third as flexible while the remaining five rated it in the middle third.

Subjects were asked if they felt that the use of command buttons was “logical” or “illogical”. Seven responded that it was “logical” with a total of twelve responses in the scale’s upper third. The final response rated it a six. When asked if they found the command buttons “frustrating” or “satisfying”, eight subjects responded in the scale’s upper third while the remaining five fell into the scale’s middle third. Nine subjects found the use of the command buttons “easy” while the remaining responses were scattered in the scale’s upper half.

The subjects were questioned as to the “phantom” agent’s effectiveness during teleoperation. Nine subjects found its use “helpful” while the remaining four subjects found it moderately helpful. Nine subjects found its use “logical” while all responses remained in the scale’s upper third. No subjects found the “phantom” agent’s use “frustrating” but one subject rated it a five. Eleven subjects rated its use in the upper third of the “satisfying” scale and one subject had no opinion.

All subjects found the amount of information which could be displayed on the screen in the scale’s upper third from “inadequate” to “adequate”. Eight subjects rated it “adequate”. The subjects had mixed feelings as to the logic of the screen information arrangement. Four subjects stated it was “logical” while four more rated it in the scale’s upper third. Four subjects revealed mixed feelings and one subject felt it was much more “illogical” than “logical”.

The third question set was related to the terminology and messages throughout the system. Ten subjects rated the terminology usage as “consistent” while the remaining three rated it in the scale’s upper third. The same results were received when the subjects were ask about the consistency of the command button labels throughout the system. When asked their opinions of the clearness associated with the command button labels, eight subjects felt they were “clear”. Four other subjects rated it in the scale’s upper third and one subject responded with a five. In general, the results to the three questions show that the system’s messages and terminology is consistent, which is very important in the user’s understanding and ability to use the system.

The remaining questions in the questionnaire’s third section are related to the error messages subjects received. Consistently throughout these questions, four or five subjects had no opinion. When the subjects were asked about the appearance of error messages on the screen the nine subjects with opinions rated it in the scale’s upper third for consistency. The position at which the messages appear on the screen were rated as “consistent”. Subjects generally found the error message content to be “clear”, with eight responses between the scale values of eight and nine. One subject rated the content a six. The ten responses to the error messages helpfulness were in the scale’s upper third, with seven responses of “helpful” and two responses of the scale value eight. In general, the nine respondents found the error messages “easy” to read. Subjects felt error messages clarified the problem frequently with all responses falling in the the scale’s upper third, only two felt it was “always” helpful. Most subjects felt the phrasing of error messages leaned toward “pleasant”, while one subject rated the phraseology a five. The error message phrasing was rated somewhat “clear” by the respondents and the messages helpfulness was found to be more “helpful” than not. Only six subjects expressed an opinion to the final question in this section. The question asked if the instructions for correcting errors was “confusing” or “clear”. Only two subjects felt the instructions were “clear” and the remaining four fell towards the middle of the scale.

The questionnaire’s fourth section related to the subject’s ability to learn and remember the system’s functioning. Six subjects felt the system was easy to learn while the remaining seven felt it was moderately easy. Two subjects felt getting started using the system was slightly difficult while the remaining subjects rated it as fairly easy. Eleven subjects felt the time required to learn to operate this system was quite fast while the remaining two felt it was fairly fast. Eleven subjects believed remembering the command buttons uses and names was “easy” while the remaining two subjects found it a little more difficult. Only six subjects found remembering specific rules about entering commands “easy”. The remaining replies felt this was somewhat to fairly easy. Seven subjects felt tasks could “always” be performed in a straight forward manner while the remaining six subjects rated two a piece from six to eight on the scale. When the subjects were queried as to the number of steps required per task, we received mixed results. This is likely due to the fact that many subjects believed the system initialization should occur automatically. All subjects believed that the steps to complete a task followed a logical sequence of events from almost “always” to “always”.

The questionnaire's fifth section pertained to the system's overall capabilities. None of the subjects felt the general system speed was fast enough. In fact, seven subjects felt the system was particularly slow. Eight subjects felt the response time for most operations was fairly slow, while five subjects felt it was almost fast enough. The rate that information was displayed received mixed replies. Most subjects, eleven, felt it should occur faster. Five subjects felt system failures occurred fairly seldom while the remaining subjects felt the failures occurred fairly frequently, two responses, to somewhat seldom. Finally, subjects were asked if they felt novices could accomplish tasks after proper training. Seven responded they felt novices could accomplish this "easily", while the remaining six subjects ranged from somewhat easy to fairly easily.

Section six of this questionnaire asked the subjects to circle a member in a pair of the NASA TLX workload variables. This information is used to weight the information subjects provided on the NASA TLX questionnaires.

Finally, the subjects were provided a section in which they were permitted to write their own comments regarding what they liked or disliked about the system. These responses will be discussed in the next sub-section.

3.11 Subjects' Written Comments

In general, the subjects provided numerous comments. Some were as general as "This system is fun" to very detail oriented system related issues. All system oriented comments were constructive and would improve the system's usability. This section's purpose is to provide the reader with their comments. These comments were provided on both the post-task and post-experimental questionnaires. As many of them overlap between the two questionnaires, we are presenting them together in this section.

The subjects commented it was difficult to know the clearance a robot had in situations when it was close to a corner or object. The images did not provide enough sense of the agent's true position and that some form of localization was necessary.

The "blindness"[2] of the agents was noted as a negative. One subject suggested the addition of cameras on the agent's side. One subject states:

> Again, blindness was a problem. This time resulting in the failure of the task.

Another subject suggested providing the ability to pan, tilt and zoom the agent's cameras. In particular, this subject felt it would be useful when attempting to monitor the manipulatory agent's actions. Some subjects stated they would prefer a lower-quality image at a higher frame rate to the current images. This was particularly true when the system's communications were very slow.

One subject felt using the agent command buttons was difficult when switching between agents. This subject suggested that it may be better to permit the operator to choose the agent by clicking on the desired virtual agent.

---

[2] "Blindness" refers to the agent's inability to fully sense its environment.

Some subjects found the system mode control buttons confusing because they could not "visualize" where they were located in the control button tree. They were unsure which mode they were in and what they should do next. Some subjects disliked the requirement to specify that they wanted to create commands for both manipulatory agents or for a single agent. They found this time consuming and cumbersome.

Some subjects found it difficult to use the mouse to rotate the agent. In particular, if they placed the mouse on top of the agent, they did not have fine control of the rotation movement.

One subject stated that since the VisionBot was able to avoid obstacles with its on-board process, he basically ignored it after creating the necessary commands for its portion of the task. This is a problem as the agent may run into difficulties, if the human is not monitoring problems may arise. One subject commented on the fragility of the VisionBot's obstacle avoidance process. He stated:

> The VisionBot goes crazy when it sees the stain on the carpet in the corner and the user has no control over it.

This is a true statement. The process is very sensitive and is not fully integrated into the system, thus the human cannot completely control it in such situations. Another subject commented in relation to the two agent task:

> The only problem with this task was that the VisionBot always ends up turning in circles.

This again refers to the process' sensitivity.

One subject felt the raw sonar displays complicated the agent's navigation, thus he did not use them. Another subject only used the raw sonar displays "to watch the changing colors on the screen".

The initialization of the agents requires too much time and is very repetitive in the current system because the human must initialize the system every time they begin a task. Many suggested allowing the system to automatically initialize the agents either via a preference file choice or just those processes, such as the robo process, which are always required by the agents. This second suggestion is not feasible with the manner in which the multiagents system's communications have been established, one is unable to request processes as desired. This issue will be discussed further in the discussion section.

Double-clicking on buttons generally caused a system failure and is a system bug that appeared during the experiments. This problem was particularly difficult for those subjects who were accustomed to working with systems which require double clicking, such as a Macintosh system. In the particular case of requesting windows to the right of the interface, if the subject double clicked, two windows would appear. One window would display the image; the other would remain black and there was no mechanism to remove the useless window. Thus it would limit the user's ability to display information. There were also times when the subjects were unsure if they had clicked a button, as they could not observe that the system was taking a long time to process their request. Two suggestions where: change the button color when it is selected or display a message that the system received the command and is processing it.

One subject thought it would be preferable to be able to control the agent's speed. This would permit them to increase the agent's speed in situations where there was little concern of an accident and reduce the speed in populated environments.

Many comments were received on the small images to the right of the main window. Many subjects suggested that each image displayed have a title bar associated with it so one would know which agent's image they were observing. Another dislike was the inability to reorganize the windows such that image pairs could be arranged beside one another. In particular, subjects wanted to do this with the stereo camera image pairs. Subjects also thought that the ability to turn on and off the virtual camera images would be useful as it would provide two more display areas. One subject thought the smaller displays should be placed across the top of the main floor plan as it seemed to him that this would provide a more natural gaze from one to the other, thus eliminating the left to right gazing which this subject found distracting.

Some subjects did not feel the system provided enough sense of "presence" within the environment. Another subject stated they became impatient with the system as they gained familiarity with it because it was too slow.

The subjects liked many things about the interface. In particular, there were numerous favorable comments on the system's graphics and the "phantom" agent concept. While some subjects had difficulty using the mouse to create commands for the agents, others found it easy. Subjects found the interface to be "very user-friendly". One subject stated the system was "complete" and "user-friendly". Another felt:

> It is quite easy for someone to learn how to use this interface, even someone with little or no computer experience. It's also visually attractive and logical.

Still another subject stated the system was easy to learn and it was "very easy" to control all four agents during the four agent task.

3.12 Ravden and Johnson's Evaluation Check list

Ravden and Johnson's [10] evaluation check list was created to provide a practical tool with which to rate a system's usability. It is based upon a set of software ergonomics criteria which a well-designed user interface should encompass. As we wished to apply this check list to the entire system and use it as a reference point to the responses the subjects provided, we completed the check list ourselves. As has been previously stated, the interface version employed for the experiments was only a small portion of the entire interface.

The check list is composed of nine sections all of which relate to a separate usability criterion. These nine sections permit the respondent to choose from the following responses: Always, Most of the time, Some of the time, and

Never. At the end of each section the respondent gives an overall rating on this particular topic. Section ten explores system usability problems with responses of: No problems, Minor Problems and Major Problems. The last sections provide general questions with an open response. We completed this check list after the experimental sessions but prior to analyzing the data.

3.12.1 Visual Clarity

This section is related to the system's visual clarity: in particular, the clarity, organization, unambiguousness and ability to read information on the screen. As this is a general check list, we must regard the questions as they apply to our interface. All questions are related to the screen and window displays. We found from completing this list:

- The interface does not clearly identify windows with an informative title.
- Most of the time important information is highlighted on the screen.
- It is clear where the user should enter information and commands for the system as well as the required format.
- Information usually appears to be logically organized on the screen.
- All the various types of information are clearly separated on the screen.
- Large amounts of information are always properly separated from other information.
- Columns of information, such as the smaller windows, are always properly aligned.
- Colors are properly displayed and help to make the display clearer.
- Most display aspects would be easy to view if a low resolution screen was used and if a user is color blind. One subject was color blind and performed as well as the others.
- Information on the screen is always easy to see and read.
- The screen usually appears uncluttered.
- Pictorial displays are properly drawn but are not annotated.
- It is usually easy to find required information.

The overall system rating in terms of visual clarity was "moderately satisfactory". This rating was chosen because the images displayed from the agents are not labeled, thus a user may become confused. Also, if the system response time is slow, currently one may not know if they properly pressed a command buttons. A mechanism is required to indicate the system is busy.

3.12.2 Consistency

This section explores the manner in which the system looks and works at all times for consistency. The results of this section were:

- Colors are always used consistently throughout the system.
- Abbreviations, acronyms, etc. are always used consistently.
- Pictorial and graphical information are always used consistently.
- The same types of information, (error messages) are always displayed in the same location and in the same layout.
- Information which appears in numerous places, such as the agent data display buttons, are always displayed in the same format.
- The input information format and method is always consistent throughout the system.
- The action required to move the cursor (mouse) around the screen is consistent.
- The method of selecting options throughout the system is consistent.
- There exist standard procedures for performing similar and related activities. An example is the method of choosing way points for the various planning methods.
- The manner in which the system responds to a user action is somewhat consistent throughout the system. There exist cases where a command button may be chosen that does not cause the command line menu to change.

We gave the system a “very satisfactory” rating in this category. The last item mentioned is a minor problem with consistency and does not appear frequently.

3.12.3 Compatibility

This section requires one to examine the system’s compatibility with the appearance and functionality of other systems a user may use regularly. As we do not know of another interface of this exact nature, we have based our responses on general computer systems. We found:

- Colors assigned to graphical objects were as close as possible to the real object’s colors. Most of the time colors associated with actions were the conventional associations. The active agent’s color in the agent control buttons is green but it does not change to red when the operator issues an emergency stop command.

- Graphical representations are usually easy to recognize but images from the agents may not be easy to recognize because they are not labeled.
- There exist established conventions for the information displays format.
- Presented information is always in units which the user would be familiar with such as millimeters.
- When the user inputs information using the mouse, the movement on the screen usually corresponds to the mouse's movement. There are instances when rotating agents with the mouse positioned directly above the agent that the agent will move much farther than intended.
- Most system information is presented in a form which fits the user's task view.
- Most of the time the graphical display is compatible with the user's view of what they are representing. Discrepancies can occur due to slippage of the agent's wheels.
- The sequence of events to complete a task usually follow what the user would anticipate. Working with the manipulatory agents is slightly different, as they may be controlled simultaneously.
- In general the system works as the user would expect.

The system rated "Moderately satisfactory" in terms of compatibility. It is possible the system does not fit the user's perception of the task as every user differs. It is also possible users will become very confused when first attempting to control the manipulatory agents in the simultaneous control. While the command generation method for these agents is the same as for a single agent, there are many more factors to consider: such as, if I turn the agents and they are too close together then they will hit each other thus ending the task.

3.12.4 Informative Feedback

This section examines a system's ability to provide the user with clear and informative information concerning their location within the system, what actions they have already completed and the action's success and failure and what actions should be taken next. Our system does not attempt to instruct the user as to which actions should be taken next but the other aspects are relative to the MASC system. The results showed:

- Instructions and messages are always displayed in a concise and positive manner.
- Displayed messages are always relevant.
- Instructions usually clearly indicate what is required.
- It is usually clear what actions a user can take at a particular instance. There are some system command button states which are not always clear to users.
- It is usually clear what the user needs to do in order to create a desired action.

- It is always clear what information should be entered when data is requested by the system.
- There do not exist any short cuts.
- It is not always clear what changes occur on the screen as a result of an action. This can occur when the system response rate is slow, thus an appropriate system response may not appear as a result of the input action.
- The system does not provide status messages while it is busy.
- It is not always clear when the system has completed the requested action. Some of the system command buttons do not change the system state and the user may not notice a change in the button.
- In general, the error messages state what the error was, where it occurred and why it occurred.
- It is generally clear to users what action is required to correct an error.
- The system does not clearly indicate which system mode it is in at all times.

As there are many interface aspects which could be improved to provide the user more informative feedback, the overall rating for this criterion is "Neutral". In particular, when the system is slow to respond, there should exist an indication that the system is busy and will return to the user shortly.

3.12.5 Explicitness

This section explores the manner in which the system is structured and works for user clarity. The results of this section are:

- It is not always clear what stage the system has reached in executing a task.
- It is usually clear what is required of the user to complete a task.
- Lists of options generally have clear meanings to the operator.
- It is not always clear to the user in which system state they are in.
- It is very clear what the various system modes are and their functions.
- It is not always clear how changes may affect other system aspects.
- Generally, the system organization and structure is clear.
- The system structure may not be immediately obvious to a user. After working with the system, it is presumed this understanding could be gained.
- The system is not necessarily well organized from a user's point of view. There exist system aspects, such as initialization, which could be automated. The differences in the robot controls is also not well organized from a user's perspective.

- It is generally clear what the system is doing during a task.

There are instances where the system could be more explicit, thus we give it an overall rating of "Neutral". When the system is busy it currently is not clear what the user should do to continue with the task. The user must wait for a response before continuing with the task.

3.12.6 Appropriate Functionality

Appropriate system functionality refers to the ability to meet the user's needs and requirements while executing tasks. The results from this section are:

- The use of the mouse as an input device is generally appropriate. Some users experience difficulties in properly rotating the agents employing the mouse.
- The information presentation manner regarding the task is generally appropriate. Particularly as we permit the user to choose the information which they choose to display.
- In general, the system provides the user with information and necessary options at any particular stage. The inability to add new processes once an agent has been initialized is one instance when this is not true.
- Users generally feel the system provides them with the ability to complete tasks.
- Aside from system delays, the feedback from the system is appropriate for tasks.

The system does not provide help and tutorial facilities which many users may find helpful. This combined with the system difficulties listed above lead us to chose an overall rating of "Moderately satisfactory" for the MASC system functionality.

3.12.7 Flexibility and Control

An interface should provide the necessary flexibility to meet the requirements of all possible users and their preferences while also permitting them to feel in control of the system. We found from this section:

- There is not always an easy manner for an instruction to be "undone".
- There exists no mechanism to redo an undone action.
- Users generally have control over the order in which information is requested and activities can be carried out after system initialization.
- It is easy for the user to return to the general system state from any other system state.

- The user can easily move between different system modes.
- The user is only able to effect the rate information is displayed with the ultrasound process. This would be a useful feature for the other sensing modality information.
- The system does not permit the user to store user preferences for later user.
- Users can tailor the sensing displays for their preferences by turning them on or off. It would be useful to permit them to rearrange the small windows layout to their preference.

The overall system rating based upon this aspect is “Moderately satisfactory”. There are options, such as the redo and the display rates, which would be useful additions to the system while other aspects do permit the user to customize the interface and provide them with a sense of control.

3.12.8 Error Prevention and Correction

This evaluation section examines the system design to determine if it minimizes the possibility of user error and provides the user with the ability to verify their inputs and correct potential error situations. The following lists this section’s results:

- The system validates most user input.
- The system clearly and promptly displays an error box on the interface when an error is detected.
- The system permits the user to verify most inputs prior to instructing the agent to execute the command, for instance the path planners.
- The system does not provide a cancel key to reverse an error situation, but the user is able to instruct an agent to stop if its locomotion commands will lead to an undesired state.
- The system ensures the user corrects all detected errors before the input is processed.
- The user can explore possible path planning options without instructing the agent to execute them. This is not true with the teleoperation control.
- The system catches trivial errors such as choosing an inactive agent but does not deal with instances of double clicking, thus it does not always protect against trivial errors.
- We have found that double clicking on items in the interface can cause a system failure.
- Aside from a few errors, the system is generally bug and error free.

In general, the system is quite efficient at detecting and preventing errors thus we have chosen an overall rating of “Moderately satisfactory” for this section.

3.12.9 User Guidance and Support

This section explores the system's availability of informative, easy-to-use and relevant guidance and support to assist the user. This support should be provided in both hard copy and on-line documents. We have already stated that there exists no on-line help facility. At this time there also exist no formal hard-copy documentation. The documentation that will be provided, in hard-copy, will consider the factors explored in this section. As we currently lack documentation, the overall rating for this system aspect is "Very unsatisfactory".

3.12.10 System Usability Problems

This section explores possible system usability issues. The results from this section are:

- Users encountered minor problems when learning how to use the system and while attempting to understand how to execute tasks.
- Some users experienced minor difficulties finding the information they desired and then determining how this information related to other system aspects.
- Colors and other information appear clearly on the screen and do not over populate the system.
- The system is fairly flexible.
- Users experienced situations in which they felt lost in the system during the experiments. We imagine this "feeling" would increase when using the complete system.
- The system does not require the user to retain significant information about the task in memory.
- System response times are slow enough that the user generally knows what is happening in the system. In situations where the system response time is too slow the user does not feel as they know what the system is doing.
- All textual information which appears on the screen remains until the user has read it and supplied a response.
- The user may experience very slow system response times.
- If the user is accustomed to a double-click oriented system, this system will generally fail when they double-click.
- Employing the mouse as an input device makes it easier to use in most instances.
- The user always knows where to input information.
- System initialization requires too much input time but other input aspects do not require a significant amount of time.
- In general, the user does not have to be extremely careful about causing error as there are mechanisms to detect them.

The final evaluation section asks open questions about the system. Each of the questions in this section have been previously answered in the other sub-sections of this Section, therefore, I will not reiterate.

This evaluation provided deeper insights into the entire system abilities beyond what we were capable of testing with the human factors experiments. Therefore, it has been a useful tool.

This section has presented the human factors experimental data analysis results. We have presented the results of the subject's responses to the various questionnaires, our responses to Ravden and Johnson's usability evaluation check list, as well as the statistical analysis of data related to task completion times, number of commands and errors created and perceived workload. We also presented information regarding the sensing modalities subjects employed as well as their task completion data.

4 Discussion

The purpose of this section is to discuss the results presented in Section Three. The first section discusses the results for the MASC interface, then the mediation hierarchy and finally the Multiagents system.

The results from the human factors experiments may only be generalized for a population of computer literate novice MASC system users. This is based upon the backgrounds of the subjects who participated. Also, the experimental results may only be generalized to the reduced MASC system version. We can only predict what the results would be when the subjects encountered the complete system. Thus these experiments provide significant analysis for the *task* and *regulation levels* of the mediation hierarchy, while permitting preliminary evaluation of the *processing* and *data levels*.

4.1 MASC System Discussion

4.1.1 General Discussion

There exist significant evidence that a novice MASC and multiagent systems user can successfully complete tasks. There were only eight accidents during one hundred fifty six trail runs. We can also state, as the number of agents involved in the task increased, the number of accidents increased. There does not exist enough evidence that the number of accidents would decrease over time and practice, but one would conjecture this would be true.

As was expected, the number of commands required by the users increases significantly for the four agent task, in fact, it practically doubled from the other two tasks. As the subjects moved from performing the single and two agent task to the four agent, there was an increase in the task completion time. Also, there was a decrease in both the number of commands and task completion time during the second session. This shows as the subjects became more familiar and practiced with the system they were able to improve their performance. Surprisingly, there did not exist a vast difference in the number of commands created between the single and two agent tasks. It was expected there would be more commands for the two agent task. During session two, the average number of commands created was essentially equal to that created for the single agent task.

It was interesting to note which sensing modalities subjects preferred. It was expected that a majority would gravitate toward a few sensing modalities. It is plausible to believe the subjects preferred the agent's images because they were more familiar with that display format and did not feel they could properly interpret the other data display types. For those subjects without a formal training in computer science, they commented on the difficulties of understanding the concept and interpretation of a state diagram. The subjects felt the ultrasound process took too long to provide information and therefore was not useful. As some subjects did not feel certain sensing modalities were useful this would account for the fact subjects only used all available displays four of the one hundred fifty six trials. What was also interesting was the subject's preference to establish all sensing displays prior to beginning their work with the agents. They also did not change the displays during the task executions.

It was expected the subjects would commit many more errors when first beginning to use the system. While the number of errors during the first experimental session was more than double that in the second session, the total number of errors was fairly low for the given number of trials. This also shows that as the subjects increased their experience level they committed fewer errors. One would like to conclude the number of errors would still decrease, but this experiment does not provide enough evidence to support this statement.

The subject's perceived workload was anticipated to increase as the number of agents required for a task increased. The analysis showed that the values did generally increase from the single agent to the four agent tasks and that this was a significant relationship. We also expected as the subjects became more familiar with the system these values would fall in the second session. The results showed this was true and in fact, the mean workload values between the single and two agent tasks were essentially equivalent during the second session. The analysis did not find the decrease in the perceived workload value between sessions significant.

We anticipated that some of these variables would be related. In particular, we expected the subject's perceived workload measure would be effected by the number of commands and the amount of time the task required as well as the number of errors a subject committed. The analysis showed the perceived workload was generally effected by the number of commands and the amount of time required to complete the task. It was interesting to note the errors the subject committed did not display a significant effect on the subject's perceived workload measure. Since the most frequent types of errors occur during the formal task execution, not the initialization period, one would expect a relationship.

We have already stated the number of commands required for the tasks were essentially equal for the single and two agent tasks and increased for the four agent task. It was anticipated there would exist a significant relationship between the number of commands created and the amount of time required to complete the task. This was generally not found to be the case. The analysis generally found this relationship was insignificant. This implies that some subjects were capable of executing many commands just as quickly as others who used fewer commands. As the agent's speed remained the same throughout the experiments, this is exceptable explanation. There also exists a significant relationship between an increase in the number of errors with increases in the number of commands and task completion time. These relationships can be explained by the fact that 62% of all errors occurred during the four agent task which required the greatest number of commands and completion times.

Based upon these results and the subject's responses to the post-task questionnaire, we can also deduce that indeed the four agent task was more difficult than the single and two agent tasks. As we encountered hardware difficulties with the VisionBot which limited its abilities, the anticipated result that the two agent task would be more difficult than the single agent task is not upheld.

It was shown that the subject's workload values increased between tasks, particularly between the easier tasks and the four agent task. The perceived workload measures are based upon a scale from zero to one hundred. It is interesting to note, the highest perceived workload mean value was 48.15 while the highest actual value was 67.49, and these values dropped during the second session. This shows the system does not significantly overload the user's workload capabilities.

It was found that only twice did users not detect a problem situation. The first time the subject was attempting the task without sensors and the second, the subject forgot the SensorBot was unable to avoid obstacles automatically. So, for these given instances the subjects were unable to detect problem situations. One situation which many subjects encountered was the lack of knowing exactly where the agent was located. This could be overcome with the addition of a localization process.

It was observed that subject's approaches to driving the agents were as varied as the humans who participated. Some subjects preferred to create many commands and then issue the emergency stop command if they did not like the agent's actions. Others preferred to drive very conservatively, creating one small motion, waiting for the agent to complete it and then creating another. In both extremes, the subjects were theoretically creating what would be classified as "unnecessary" commands. Of course, these particular subjects would likely disagree with this statement.

Subjects generally found the tasks they were asked to execute interesting during the first session but were somewhat bored during the second session. They would have preferred new tasks. They also found they generally felt in control of the system during the tasks. Both results were not anticipated and can be accounted for in two manners. Either the tasks were too easy or the system effectively provides the user with the capabilities required for the tasks. The second assumption was substantiated through their responses to questions concerning the system's abilities and flexibility to complete the required tasks.

While we would have preferred the subjects to give the system a "terrific" overall rating, this is not realistic. The results do show that the overall rating was generally high. The results also showed that the system tends not to overly frustrate users.

The subjects raised several usability issues concerning the screen layout and the display of information in the smaller windows. Many subjects desired the ability to rearrange the windows into their preferred customization once they had displayed all the information. They also would have preferred the ability to turn off the views from the virtual cameras so that those windows could be used to display other sensory data. The subjects agreed the main working window was preferable and useful. These results agreed with our results from the usability evaluation check list's "visual clarity" section.

While the subjects found the command buttons use logical, they sometimes lost track of their current position in the system state tree. Given this fact, subjects still responded the buttons were generally easy to use. This is a result we are happy to find.

The subjects found the messages and terminology used throughout the system consistent. This upheld our belief that such information was consistent. They also agreed the position of error information was consistent. We anticipated this and it substantiated the evaluation check list results.

It was interesting the find the subjects believed the system was easy and quick to learn. When we were explaining the system during the training session, many subjects appeared to be completely confused. These same subjects stated later that once they began to work with the system, it was much easier than they had anticipated.

The issue of the system speed and information update were identified by users as a problem in the usability check list. The experimental results substantiated this finding.

In general, the users found many constructive usability issues. The usability evaluation check list, which we completed, also identified many of these same issues. It also brought notice to many good system aspects, such as the consistency of the various system aspects. This was not an issue we significantly considered when developing the interface.

4.2 Mediation Hierarchy Discussion

While the above results show many good and bad things with the MASC interface, we were primarily interested in substantiating the mediation hierarchy. As we have mentioned, more research would be required to fully substantiate the hierarchy.

The fact novice subjects were able to successfully instruct the agents to complete the tasks shows they were capable of communicating with the agents at the *Task level.* This result would be expected, as we have succeeded in communicating tasks to the agents, but this could not be assumed. This implies the system is straightforward enough for the user to act as the task planner.

Generally speaking, the concepts behind the three interaction types of the *Regulation level* were found to be upheld. The *command interaction* permitted the subjects to effectively teleoperate the agents employing the mouse while also permitting them to control the agent's actions with the agent command buttons. The subjects were also able to effectively switch control between the agents. Two issues which subjects raised were, they sometimes found the creation of rotation commands difficult and an easier method of switching between agents would be to click directly upon the virtual agent.

The results show the subjects found the concept behind the *request interaction* logical. While they did not fully exercise this capability, they responded that requesting sensory information was easy. We also anticipated we may find there was some information displays which should be automated. The subjects substantiated this fact in stating it would be preferred if the system remembered their preferences and displayed them upon initialization. The user could then customize the displays by turning on and off those which were and were not desired for the particular task.

While the subjects did not employ the path planning processes which are generally associated with the *specification interaction*, they were required to specify the agent processes required to interact with the agent. The analysis of their abilities to specify the initialization provides us with some preliminary results for this interaction type. This is true because the *specification interaction's* purpose is to permit the operator to specify necessary information prior to the commencement of processing. The processes required for a task must be initialized prior to commencing the task execution. The subjects were able to effectively specify the necessary information required to establish the agents. This specification is fairly simple, as they must only choose the processes, but this result displays the successful completion of the interaction. This result does not permit us to extrapolate to the subject's ability to complete such an interaction with the path planner specifications.

These experiments did not formally incorporate any of the *Processing level's* capabilities, but we can deduce some preliminary conclusions to this level's necessity. For instance, we did not provide the subjects with the training necessary to change the ultrasound process clustering variables. As a results of the fairly high default values for these variables, the subjects did not prefer to use this sensing modality because they found it provided very little or no information. Thus, if the subjects had obtained the knowledge to interact with the process to change these variables, it is conceivable they would have succeeded in doing such. Also, a few subjects suggested the ability to modify the image quality and frame rate of images received from processes. This type of interaction would reside on the *processing level*. Subjects also demonstrated a certain frustration level when the VisionBot's obstacle avoidance process detected a wall and proceeded to turn in circles in an attempt to avoid the obstacle. Many subjects desired the ability to instruct the agent to stop attempting to avoid the wall. This would be an instance of the human overriding the process's decision. Also, if the process was able to deduce it was in a situation where it could not find a way out, the process could request the human's assistance. Again, these are all preliminary conclusions as the experiments did not formally test this level's capabilities.

The experiments also were not capable of formally providing results to substantiate the mediation hierarchy's *data level*. We are able to establish some informal preliminary results based upon the subject's reactions and comments on the system. The most common complaint amongst the subjects was associated with the difficulty of determining exactly where the agent was in the real world verses the model. This was primarily due to wheel slippage on these mobile agents. Subjects stated they would like a mechanism (a localization procedure) to determine the proper location and then either reposition the virtual agent or reset the actual agent's odometry and heading. This particular aspect is one of the primary examples we use to substantiate this level's existence. As this issue arose frequently, we can deduce the *data level* is a necessary element.

4.3 Multiagents System Discussion

Throughout the experiments many issues were raised within the entire multiagents system. The purpose of this section will be to discuss these multiagent system issues.

In general, the overall system was found to be very fragile, both in a software as well as a hardware sense. In some instances, the software processes were not sophisticated enough or did not exist. The hardware problems range from difficulties with dead batteries to loss of communications.

The agent's ability to sense its environment as it exists is not sufficient for difficult tasks. In the tasks we required, the sensing capabilities were not tested to the envelope of their abilities. For instance, the obstacle avoidance process version which is integrated into the system is very fragile and unreliable. It does not always avoid obstacles but when it detects a wall it will go around in a circle attempting to avoid it. Also, if the focus or aperture of either of the two cameras is slightly different than the other, this process produces ambiguous data. It should be noted that a newer version of this process does exist but is not integrated into the MASC system.

Another major difficulty is the agent's narrow view provided by the cameras. Many times subjects and agents are unable to acquire a significant environmental view which would assist them in their task. We as humans have a 180^o field of view but when we explore what we can obtain from the agent's field of view; it is significantly less. There exist some people who have difficulties performing tasks with their own field of view, it is infeasible to expect a robotic agent to properly understand its environment with such a narrow field of view. Perhaps the new 180^o field of view camera will assist with this difficulty.

It is absolutely necessary to have a localization procedure. Many times throughout the experiments we observed subjects working with an agent when they were not positive of its location. In one instance, the virtual agent appeared to be almost into the GRASP laboratory's front office while the real agent was located just in front of the pillar in the eastern portion of the laboratory. This discrepancy was due to the fact the agent had become stuck on its power cable and then came free. While the agent was stuck, the odometry and heading readings were updated thus updating the virtual agent's positional information. It is interesting to note, the subject eventually did successfully complete the task relying upon the agent's real time images. We believe this subject was an exception and most subjects would not have been able to complete the task in a similar situation. This is an extreme example, and we observed that even the normal slippage of an agent is significant enough that localization would be very helpful to the human operator.

One complaint from the subjects was that they disliked the requirement of initializing all of an agent's processes at once. They would have preferred to start one process for an agent and then as was needed, add new processes. This currently is not permitted in the multiagents system because of the TCP/IP communication protocol employed. This protocol requires that a process know all other processes it will communicate with when it begins. This is further complicated when various processes must communicate with more than one process. All processes must be started before processing can begin. This is not realistic, the human may not initialize a process when they start a task and then may find it is necessary and is unable obtain the required information. Also, the current configuration does not permit the operator to shut down a process and restart it later as required. Originally, the human operator interface was written in such a manner as to start up and shut down processes as needed, but this had to be changed as more processes were integrated into the system because of the communication protocol. This is a very great restriction of the multiagent's system.

Another difficulty associated with the system's communications is the very unreliable communications with the manipulatory agents. Currently, the manipulatory agents are controlled by two personal computers and communicate with a Sun workstation via serial communications through the personal computers. This communications channel is very unreliable and frequently either one or both of the manipulatory agents would loose communications with the personal computers and hence the rest of the system. This leaves the human unable to work with these agents. As the experiments showed, this was a significant problem.

Currently the system runs the VisionBot and the SensorBot's communication through radio ethernet. It was found this medium was not sufficient to support communications for both agents to the MASC interface. In fact, communications became so slow we were required to connect the agents to their respective ethernet cables in order to obtain reasonable responses during the human factors experiments.

The system's agents are heterogeneous. This leads to difficulties as we attempt to execute more difficult tasks as some agents have no sensing capabilities. The human must rely upon information from the observation agents to monitor these other agents. This limits the types of tasks the system can perform.

At the time of the experiments, the VisionBot's batteries were dead. Thus for these experiments the agent was only able to move forward and backwards. If one attempted to turn the agent on a zero radius, the agent would fault. This limited the activities we could request the agent to perform and thus reduced the experimental tasks difficulty. New batteries were purchased but arrived too late to be used for these experiments.

There were a few instances when the agent's bumpers were activated for no apparent reason. In one instance, the problem was so consistent we disconnected the bumper. Also, the ability to create a software override of the bumper which permits the operator to move the agent away from the object and continue on with a task is necessary.

There are also problems associated with the consistency between system processes. In some instances, processes require information in millimeters, in other centimeters. There should be a standard established across all system code developed. This standard should not only apply to measurements but also other aspects such as communication protocols. There is also a need for developers to freeze code versions and place not only the executable but all other associated files into a directory where it will not be modified. During the MASC system development this was a recurring problem. We would integrate a process and then it would change and we would be left with the executable and no manner to rebuild it or repair its bugs. Also, in order for this type of a system to work effectively, all system developers must consider not only their current needs but also their own future needs, the needs of the entire system, the established standards and the human operator's requirements. As much of this system's code was not developed in this manner, it was extremely difficult to integrate it into the MASC system in a useful manner. For instance, if a process was developed and calculated information in a local coordinate frame, it was not useful for the human operator who observes the world in the global coordinate frame. The developer should be responsible for translating this information into the global coordinates so it is useful to other system processes. As the developer best understands the algorithm and the process, they are best equipped for such a translation. This section discussed many problematic issues raised by these experiments as well as the multiagent experiments.

Overall, the experimental results, combined with knowledge we have gained through the formal multiagents experiments (found in [4]), indicate the basic mediation hierarchy concept and purpose are upheld. We found the subjects were able to interact effectively with the upper levels of the hierarchy while demonstrating a need for the types of interactions the *processing* and *data level* provide. We also found the experiments upheld our basic research question as the novice subjects were capable of effectively interacting with the available system levels such that feasible tasks could be successfully completed in a reasonable time frame. These experiments also displayed the system is usable while there exist aspects which could be changed which would improve its overall usability. We were also pleased to find subjects did not encounter significantly high workload levels. As these experiments were considered a feasibility study, the feasibility of the concepts behind the development of the MASC system appear reasonable. This section has provided a discussion of the experimental results. We discussed the results in terms of the MASC interface, then in relation to the mediation hierarchy, and finally their relation to the entire multiagents system.

5 Future Work

There exist many extensions to this work and the multiagents system which can be included in this section. It is necessary to discuss the future work required for the entire multiagents system as it ultimately effects the MASC system.

5.1 MASC Interface Future Work

The human factors experiments raised many usability issues which could be improved upon in the MASC interface. Many of these issues improve the system's usability, but would not include further process integrations or develop the mediation hierarchy. These are very important issues within the MASC system development and should be considered during the development of later interface versions.

It appears that the inclusion of multiple display types for sensory information would be useful. While this currently exists for images, small image or overlay onto the model, other sensory data displays could be created. Also, the ability to control information in the displays, such as image resolution, would be helpful.

5.2 Mediation Hierarchy Future Work

The current MASC system does not fully implement the mediation hierarchy. In order to show a complete proof of our hypothesis this implementation is required.

We decided not to integrate a task planner for this work because this was previously shown to be feasible. In order to fully integrate the *task level* such a planning mechanism should be incorporated.

While the current integrations on the *regulation level* effectively permit the three interactions there exist revisions which would simplify and improve these interactions. For instance, the human factors experimental results provided ideas for improved *control interactions.*

We have integrated some interactions on the *process level* but there exist many more interactions which could be created. Almost all currently integrated processes contain further aspects which could be integrated at this level. Also, with the need to integrate new processes, there will arise the need for further development at this level.

The MASC system currently does not have processes available which would permit interactions on the *data level.* As the human factors experiments demonstrated, there exists a need for processes such as localization, this would provide opportunities to develop the interactions at this level.

5.3 Human Factors Analysis

There are aspects of the human factors experiments which also fall into the future work section. An enormous amount of data, beyond what was reported in this document, was collected during the experiments. This data could be analyzed for further results. Also the data which was reported could be further analyzed for other factors beyond those reported. We can foresee further human factors testing. For instance, we could conduct experiments with the entire MASC system as it currently exists. Also, upon resolving the above issues with the overall multiagents system and increasing its capabilities more testing could be conducted. The most vital human factors testing would occur upon completion of further mediation hierarchy integration. The purpose of these tests would be to fully substantiate our hypothesis which led to the development of the mediation hierarchy and its integration into the MASC interface.

This section has touched upon some of the future work issues beyond the scope of this thesis. There exist enough issues that work could continue for years.

6 Conclusions

We have briefly presented the Multiple Agent Supervisory Control (MASC) system which has been developed in conjunction with the University of Pennsylvania General Robotics and Active Sensory Perception Laboratory's multiagent project. We have also provided a brief overview of the mediation hierarchy. This hierarchy is the underlying basis for the MASC system development. The mediation hierarchy provides the supervisor with the ability to interact with the various multiagent system processing levels. We also presented a description of our test bed and the available processes as well as the general MASC human-machine interface.

This document provides a detailed description of the human factors experimental design and implementation, a complete review of the data analysis as well as an in depth discussion of the results. This information was provided in a format which adheres to the standards published by the American Psychological Association.

While developing the mediation hierarchy concept, we intended that it would improve the supervisor's abilities and create a more robust system. The experimental results substantiate the human operator's need to work within all system levels. They also partially substantiate our claim this hierarchy will permit the supervisor to interact with all system levels in order to correct problems and permit the system to successfully complete assigned tasks. It was found that the *task* and *regulation level* interactions and the need for the *processing* and *data levels* were substantiated.

References

[1] *Publication Manual of the American Psychological Association.* American Psycholocical Association, Washington, D.C., fourth edition, 1994.

[2] J. A. Adams and R. Paul. Human management of a hierarchical control system for multiple mobile robots. In *Proceedings of the 1994 IEEE International Conference on Systems, Man and Cybernetics*, pages 2780 – 2785. IEEE, Oct. 1994.

[3] J.A. Adams, R. Bajcsy, J. Kosecka, V. Kumar, R. Mandelbaum, M. Mintz, R. Paul, C. Wang, Y. Yamamoto, and X. Yun. Cooperative material handling by human and robotic agents: Module development and system synthesis. In *IEEE/RJS International Conference on Intelligent Robots and Systems.* IEEE/RJS, Aug. 1995.

[4] Julie A. Adams. *Human Management of a Hierarchical System for the Control of Multiple Mobile Robots.* PhD thesis, University of Pennsylvania, Dec. 1995.

[5] R.E. Eberts. *User Interface Design.* Prentice Hall, Englewood Cliffs, N.J., 1994.

[6] S. Hill, H. Iavecchia, J. Byers, A. Bittner, A. Zakland, and R. Christ. Comparison of four subjective workload rating scales. *Human Factors*, 34(4):429 – 439, Aug. 1992.

[7] D. Hix and H.R. Hartson. *Developing User Interfaces: Ensuring Usability Through Product & Process.* John Wiley & Sons, Inc., New York, 1993.

[8] T.E. Nygren. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33(1):17 – 33, Feb. 1991.

[9] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey. *Human-Computer Interaction.* Addison-Wesley, New York, 1992.

[10] S.J. Ravden and G.I. Johnson. *Evaluating Usability of Human-Computer Interfaces: A Practical Method.* Ellis Horwood Limited, Chichester, England, 1989.

[11] T. Sheridan. *Telerobotics, Automation, and Human Supervisory Control.* MIT Press, Cambridge, Mass., 1992.

[12] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley, New York, 1992.