Assessing Measures of Coordination Demand Based on Interaction Durations

Michael Lewis
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260 USA
ml@sis.pitt.edu

Jijun Wang
Quantum Leap Innovations, Inc.
3 Innovation Way, Suite 100
Newark, DE 19711 USA
jijunwang.cn@gmail.co

ABSTRACT
Controlling multiple robots substantially increases the complexity of the operator's task because attention must constantly be shifted among robots in order to maintain situation awareness (SA) and exert control. In the simplest case an operator controls multiple independent robots interacting with each as needed. Control performance at such tasks can be characterized by the average demand of each robot on human attention. In this paper we present several approaches to measuring, coordination demand, CD, the *added* difficulty posed by having to coordinate as well as operate multiple robots. Our initial experiment compares "equivalent" conditions with and without coordination. Two subsequent experiments attempt to manipulate and measure coordination demand directly using an extension of the Neglect Tolerance model.

Categories and Subject Descriptors
I.2.9 [**Artificial Intelligence**]: Robotics—*operator interfaces*

General Terms
Human Factors, Measurement, Experimentation

Keywords
Human-robot interaction, metrics, evaluation, multi-robot System

1. INTRODUCTION
Borrowing concepts and notation from computational complexity, control of robots by issuing waypoints, could be considered O(n) because demand increases linearly with the number of robots to be serviced. Another form of control such as designating a search region by drawing a box on a GUI (Graphical User Interface), being independent of the number of robots, would be O(1). From this perspective the most complex tasks faced in controlling teams are likely to be those that involve choosing and coordinating subgroups of robots. Simply choosing a subteam to perform a task (the iterated role assignment problem), for example, has been shown to be O(mn) [6]. The three experiments presented in this paper develop methods to assess the operator effort required to coordinate robots in tasks representative of expected application areas.

1.1 Coordination Demand
Despite the apparent analogy between command complexity and the workload imposed by a command task there is no guarantee that human operators will experience difficulty in the same way. The performance of human-robot teams is complex and multifaceted reflecting the capabilities of the robots, the operator(s), and the quality of their interactions. Recent efforts to define common metrics for human-robot interaction [11] have favored sets of metric classes to measure the effectiveness of the system's constituents and their interactions as well as the system's overall performance. In this paper we present new measures of the demand coordination places on operators of multirobot systems and three experiments evaluating our approach and the usefulness of these measures.

Controlling multiple robots substantially increases the complexity of the operator's task because attention must constantly be shifted among robots in order to maintain situation awareness (SA) and exert control. In the simplest case an operator controls multiple independent robots interacting with each as needed. A search task in which each robot searches its own region would be of this category although minimal coordination might be required to avoid overlaps and prevent gaps in coverage. Control performance at such tasks can be characterized by the average demand of each robot on human attention [5]. Under these conditions increasing robot autonomy should allow robots to be neglected for longer periods of time making it possible for a single operator to control more robots.

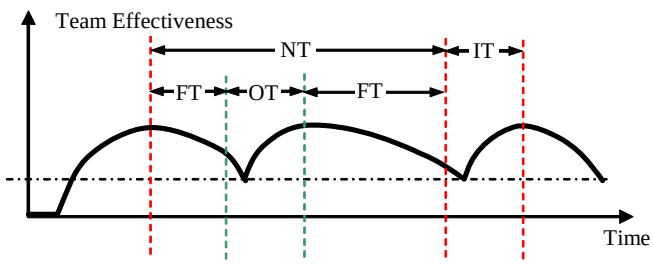
For more strongly cooperative tasks and larger teams, individual autonomy alone is unlikely to suffice. The round-robin control strategy used for controlling individual robots would force an operator to plan and predict actions needed for multiple joint activities and be highly susceptible to errors in prediction, synchronization or execution. Estimating the cost of this coordination, however, proves a difficult problem. Established methods of estimating multirobot system, MRS, control difficulty, neglect tolerance, and fan-out [5] are predicated on the independence of robots and tasks. In neglect tolerance, the period following the end of human intervention but preceding a decline in performance below a threshold is considered time during which the operator is free to perform other tasks. If the operator services other robots over this period, the measure provides an estimate of the number of robots that might be controlled. Fan-out, when measured

empirically, works from the opposite direction, adding robots and measuring performance until a plateau without further improvement is reached. Both approaches presume that operating an additional robot imposes an additive demand. These measures are particularly attractive because they are based on readily observable aspects of behavior: the time an operator is engaged controlling the robot, interaction time (IT), and the time an operator is not engaged in controlling the robot, neglect time (NT).

2. COORDINATION DEMAND

To separate coordination demand (CD) from the demands of interacting with independent robots we have extended Crandall et al.'s (2005) neglect tolerance model by introducing the notion of occupied time (OT) as illustrated in Figure 1.

The neglect tolerance model describes an operator's interaction with multiple robots as a sequence of control episodes in which an operator interacts with a robot for period IT raising its performance above some upper threshold after which the robot is neglected for the period NT until its performance deteriorates below a lower threshold when the operator must again interact with it. To accommodate



NT: Neglect Time; IT: Interaction Time;
FT: Free Time, time off task; OT: Occupied Time
IT+OT: time on task

Figure 1. Extended neglect tolerance model

dependent tasks we introduce occupied time, OT, to describe the time spent controlling other robots in order to synchronize their actions with those of the target robot. The episode depicted in Figure 1 starts just after the first robot is serviced. The ensuing free time (FT) preceding the interaction with a second dependent robot, the OT for robot-1 (that would contribute to IT for robot-2), and the FT following interaction with robot-2 but preceding the next interaction with robot-1 together constitute the neglect time for robot-1. Coordination demand, CD, is then defined as:

$$CD = 1 - \frac{\sum FT}{NT} = \frac{\sum OT}{NT} \qquad (1)$$

Where, CD for a robot is the ratio between the time required to control cooperating robots and the time still available after controlling the target robot, i.e.; the portion of a robot's free time that must be devoted to controlling cooperating robots. Note that OT_n associated with $robot_n$ is less than or equal to NT_n because OT_n covers only that portion of NT_n needed for synchronization.

Most MRS research has investigated homogeneous robot teams where additional robots provide redundant (independent) capabilities. Differences in capabilities such as mobility or payload, however, may lead to more advantageous opportunities for cooperation among heterogeneous robots. These differences among robots in roles and other characteristics affecting IT, NT, and OT introduce additional complexity to assessing CD. Where tight cooperation is required as in box-pushing, task requirements dictate both the choice of robots and the interdependence of their actions. In the more general case requirements for cooperation can be relaxed allowing the operator to choose the subteams of robots to be operated in a cooperative manner as well as the next robot to be operated. This general case of heterogeneous robots cooperating as needed characterizes the types of field applications our research is intended to support. To accommodate this more general case, the Neglect Tolerance model must be further extended to measure coordination between different robot types and for particular patterns of activity.

The resulting expression [13] measures the way in which the team's capabilities or resources are combined to accomplish the task without reference to the operation or neglect of particular robots. So, for example, it would not distinguish between a situation in which one robot of type, X, was never operated while another was used frequently from a situation in which both robots of type, X, were used more evenly. The incorporation of action patterns further extends the generality of the approach to accommodate patterns of cooperation that occur in episodes such as dependencies between loading and transporting robots. When an empty transporter arrives, its brief IT would lead to extended OTs as the loaders do their work. When the transporter has been filled the dependency would be reversed.

3. SIMULATION ENVIRONMENT

The reported experiments were performed using the USARSim robotic simulation with 2-6 simulated robots performing Urban Search and Rescue (USAR), experiments 1 & 3, or box pushing (experiment 2) tasks. USARSim is a high-fidelity simulation of USAR robots and environments developed as a research tool for the study of Human Robot Interaction (HRI) and multi-robot coordination. Validation studies showing agreement for a variety of feature extraction techniques between USARSim images and camera video are reported in [3], showing close agreement in detection of walls and associated Hough transforms for a simulated Hokuyo laser range finder [2] and close agreement in behavior between USARSim models and the robots being modeled [4,8,9,12,15].

3.1 MrCS – The Multirobot Control System

A multirobot control system (MrCS) was developed to conduct these experiments. The system was designed to be scalable to allow of control different numbers of robots, reconfigurable to accommodate different human-robot interfaces, and reusable to facilitate testing different control algorithms.

The user interface of MrCS is shown in Figure 2. The interface is reconfigurable to allow the user to resize the components or change the layout. Shown in the figure is a configuration that used in the RoboCup 2006 competition in which a single operator controls six robots. On the upper and center portions of the left-hand side are the robot list and team map panels, which show the operator an overview of the team. The destination of each of robot is displayed on the map to help the user keep track of current plans. Using this display, the operator is also able to control regional priorities by drawing rectangles on the map. On the center and lower portions of the right-hand side are the camera view and mission control panels,

which allow the operator to maintain situation awareness of an individual robot and to edit its exploration plan. On the mission panel, the map and all nearby robots and their destinations are represented to provide partial team awareness so that the operator can switch between contexts while moving control from one robot to another. The lower portion of the left-hand side is a teleoperation panel that allows the operator to teleoperate a robot.

3.2 Experiments

One approach to investigating coordination demand is to design experiments that allow comparisons between "equivalent" conditions

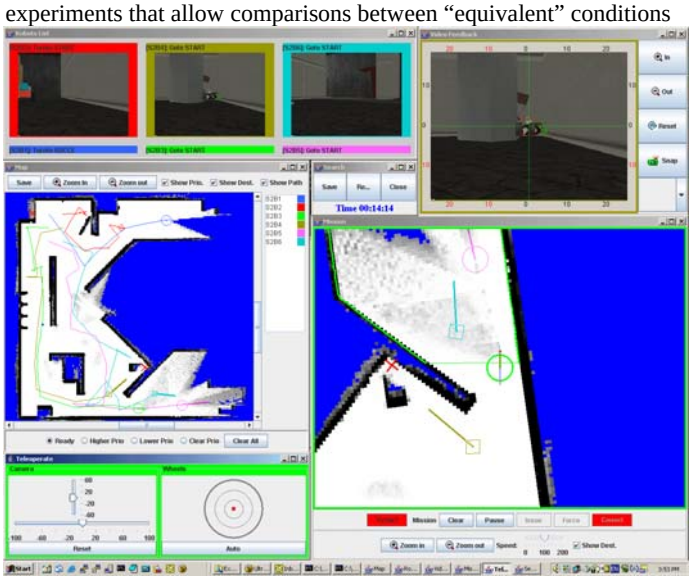


Figure 2. MrCS GUI

with and without coordination demands. The first experiment and one comparison within the third experiment follow this approach. The first experiment compares search performance between a team of autonomously coordinating robots, manually (waypoint) controlled robots, and mixed initiative teams with autonomously coordinated robots that accepted operator inputs. The impact of coordination demand was observable through the difference in performance between the manually controlled teams and the mixed initiative ones. The fully automated teams provided a control ensuring that the benefits in the mixed initiative condition were not due solely to the superior performance of the automation.

While experiment 1 examines coordination demand indirectly by comparing performance between conditions in which it was filled either manually or through automation, experiments 2 & 3 attempt to manipulate and measure coordination demand directly. In experiment 2 robots perform a box pushing task in which CD is varied by control mode and robot heterogeneity. By making the actions of each robot entirely dependent on the other, this choice of task eliminates the problem of distinguishing between interactions intended to control a target robot and those needed to coordinate with another. The third experiment attempts to manipulate coordination demand in a loosely coordinated task by varying the proximity needed to perform a joint task in two conditions and by automating coordination within subteams in the third. Because robots must cooperate in pairs and interaction for control needs to be distinguished from interaction for coordination for this task, CD is computed between robot types (equation 2) rather than directly between robots (equation 1) as done in experiment 2.

All three experiments used paid participants from the University of Pittsburgh and lasted approximately one and a half hours. All used repeated measures designs and followed a standard sequence starting with collection of demographic data. Standard instructions for the experiment were presented followed by a 10 minute training session during which the participant was allowed to practice using the MrCS. Participants then began their first trial followed by a second with a short break in between. Experiments 2 and 3 included a third trial with break. At the conclusion of the experiment participants completed a questionnaire.

4. EXPERIMENT *1*

Participants were asked to control 3 P2DX robots simulated in USARsim to search for victims in a damaged building. Each robot was equipped with a pan-tilt camera with 45 degrees Field of View (FOV) and a front laser scanner with 180 degree FOV and resolution of 1 degree. When a victim was identified, the participant marked its location on NIST Reference Test Arena, Yellow Arena [7]. Two similar testing arenas were built using the same elements with different layouts. In each arena, 14 victims were evenly distributed in the world. We added mirrors, blinds, curtains, semitransparent boards, and wire grid to add difficulty in situation perception. Bricks, pipes, a ramp, chairs, and other debris were put in the arena to challenge mobility and SA in robot control.

Presentation of mixed initiative and manual conditions were counterbalanced. Under mixed initiative, the robots analyzed their laser range data to find possible exploration paths. They cooperated with one another to choose execution paths that avoided duplicating efforts. While the robots autonomously explored the world, the operator was free to intervene with any individual robot by issuing new waypoints, teleoperating, or panning/tilting its camera. The robot returned back to auto mode once the operator's command was completed or stopped. While under manual control robots could not autonomously generate paths and there was no cooperation among robots. The operator controlled a robot by giving it a series of waypoints, directly teleoperating it, or panning/tilting its camera. As a control for the effects of autonomy on performance we conducted "full autonomy" testing as well. Because MrCS doesn't support victim recognition, based on our observation of the participants' victim identification behaviors, we defined detection to have occurred for victims that appeared on camera for at least 2 seconds and occupied at least 1/9 of the thumbnail view. Because of the high fidelity of the simulation, and the randomness of paths picked through the cooperation algorithms, robots explored different regions on every test. Additional variations in performance occurred due to mishaps such as a robot getting stuck in a corner or bumping into an obstacle causing its camera to point to the ceiling so no victims could be found. Sixteen trials were conducted in each area to collect data comparable to that obtained from human participants.

4.1 Results

All 14 participants found at least 5 of a possible 14 (36%) victims in each of the arenas. These data indicate that participants exploring less than 90% of the area consistently discovered 5 to 8 victims while those covering greater than 90% discovered between half (7) and all (14) of the victims. Within participant comparisons found wider regions were explored in mixed-initiative mode, $t(13) = 3.50$, $p < .004$, as well as a marginal advantage for mixed-initiative mode, $t(13) = 1.85$, $p = .088$, in number of victims found. Comparing with

"full autonomy", under mixed-initiative conditions two-tailed t-tests found no difference ($p = 0.58$) in the explored regions.

No difference was found between area explored in autonomous or mixed initiative searches, however, autonomously coordinating robots explored significantly, $t(44) = 4.27$, $p < .001$, more regions than under the manual control condition (see Figure 3). Participants found more victims under both mixed-initiative and manual control conditions than under full autonomy with $t(44) = 6.66$, $p < .001$, and $t(44) = 4.14$, $p < .001$, respectively (see Figure 8). The median number of victims found under full autonomy was five. Comparing the mixed-initiative with the manual control, most participants (79%) rated team autonomy as providing either significant or minor help.

Human Interactions

Participants intervened to control the robots by switching focus to an individual robot and then issuing commands. Measuring the distribution of attention among robots as the standard deviation of the total time spent with each robot, no difference ($p = .232$) was found between mixed initiative and manual control modes. However, we found that under mixed initiative, the same participant switched robots significantly more often than under manual mode ($p = .027$).

Across participants, the frequency of shifting control among robots explained a significant proportion of the variance in number of victims found for both mixed initiative, $R^2 = .54$, $F(1, 11) = 12.98$, $p = .004$, and manual, $R^2 = .37$, $F(1, 11) = 6.37$, $p < .03$, modes.

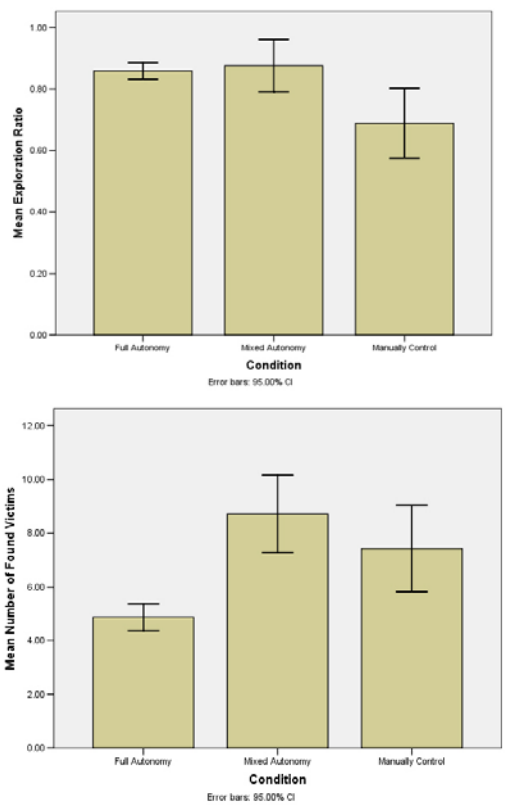


Figure 3. Victims Found and Regions Explored

In this experiment, cooperation was limited to deconfliction of plans so that robots did not re-explore the same regions or interfere with one another. The experiment found that even this limited degree of autonomous cooperation helped in the control of multiple robots. The results showed that cooperative autonomy among robots helped the operators explore more areas and find more victims. The fully autonomous control condition demonstrates that this improvement was not due solely to autonomous task performance as found in [10] but rather resulted from mixed initiative cooperation with the robotic team.

5. EXPERIMENT 2

Finding a metric for cooperation demand (CD) is difficult because there is no widely accepted standard. In this experiment, we investigated CD (as defined in Section II) by comparing performance across three conditions selected to differ substantially in their coordination demands. When an operator teleoperates the robots one by one to push the box forward, he must continuously interact with one of the robots because neglecting both would immediately stop the box. Because the task allows no free time (FT) we expect CD to be 1. However, when the user is able to issue waypoints to both robots, the operator may have FT before she must coordinate these robots again because the robots can be instructed to move simultaneously. In this case CD should be less than 1. Intermediate levels of CD should be found in comparing control of homogeneous robots with heterogeneous robots. Higher CD should be found in the heterogeneous group since the unbalanced pushes from the robots would require more frequent coordination. In the present experiment, we compared computed CDs between these three conditions.

Figure 4. Box pushing task

Figure 4 shows our experiment setting simulated in USARSim [7]. The controlled robots were either two Pioneer P2AT robots or one Pioneer P2AT and one less capable three wheeled Pioneer P2DX robot. Each robot was equipped with a GPS, a laser scanner, and a RFID reader. On the box, we mounted two RFID tags to enable the robots to sense the box's position and orientation. When a robot pushes the box, both the box and robot's orientation and speed will change. Furthermore, because of irregularities in initial conditions and accuracy of the physical simulation the robot and box are unlikely to move precisely as the operator expected. In addition, delays in receiving sensor data and executing commands were modeled presenting participants with a problem very similar to coordinating physical robots.

We introduced a simple matching task as a secondary task to allow us to estimate the FT available to the operator. Participants were

asked to perform this secondary task as possible when they were not occupied controlling a robot. Every operator action and periodic timestamped samples of the box's moving speed were recorded for computing CD .

A within subject design was used to control for individual differences in operators' control skills and ability to use the interface. To avoid having abnormal control behavior, such as a robot bypassing the box bias the CD comparison, we added safeguards to the control system to stop the robot when it tilted the box.

5.1 Participants and Procedure

14 paid participants, 18-57 years old were recruited from the University of Pittsburgh community. None had prior experience with robot control although most were frequent computer users. Participants performed three testing sessions in counterbalanced order. In two of the sessions, the participants controlled two P2AT robots using teleoperation alone or a mixture of teleoperation and waypoint control. In the third session, the participants were asked to control heterogeneous robots (one P2AT and one P2DX) using a mixture of teleoperation and waypoint control. The participants were allowed eight minutes to push the box to the destination in each session. At the conclusion of the experiment participants completed a questionnaire.

5.2 Results

Figure 5 shows a time distribution of robot control commands recorded in the experiment. As we expected no free time was recorded for robots in the teleoperation condition and the longest free times were found in controlling homogeneous robots with waypoints. The box speed shown on Figure 5 is the moving speed along the hallway that reflects the interaction effectiveness (IE) of the control mode. The IE curves in this picture show the delay effect and the frequent bumping that occurred in controlling heterogeneous robots revealing the poorest cooperation performance.

None of the 14 participants were able to perform the secondary task while teleoperating the robots. Hence, we uniformly find TAD=1 and CD=1 for both robots under this condition. Within participants comparison found that under waypoint control the team attention demand in heterogeneous robots is significantly higher than the demand in controlling homogeneous robots, $t(13)=2.213$, $p=0.045$ (Figure 5). No significant differences were found between the homogeneous P2AT robots in terms of the individual cooperation demand ($P=0.2$). Since the robots are identical, we compared the average CD of the left and right robots with the CDs measured under heterogeneous condition. Two-tailed t-test shows that when a participant controlled a P2AT robot, lower CD was required in homogeneous condition than in the heterogeneous condition, $t(13)=-2.365$. $p=0.034$. The CD required in controlling the P2DX under heterogeneous demand (CD) condition is marginally higher than the CD required in controlling homogenous P2ATs, $t(13)=-1.868$, $p=0.084$ (Figure 5). Surprisingly, no significant difference was found in CDs between controlling P2AT and P2DX under heterogeneous condition ($p=0.79$). This can be explained by the three observed robot control strategies: 1) the participant always issued new waypoints to both robots when adjusting the box's movement, therefore similar CDs were found between the robots; 2) the participant tried to give short paths to the faster robot (P2DX) to balance the different speeds of the two robots, thus we found higher CD in P2AT; 3) the participant gave the same length paths to both robots and the slower robot needed more interactions because it trended to lag behind the faster robot, so lower CD for the P2AT was found for the participant. Among the 14 participants, 5 of them (36%) showed higher CD for the P2DX contrary to our expectations.

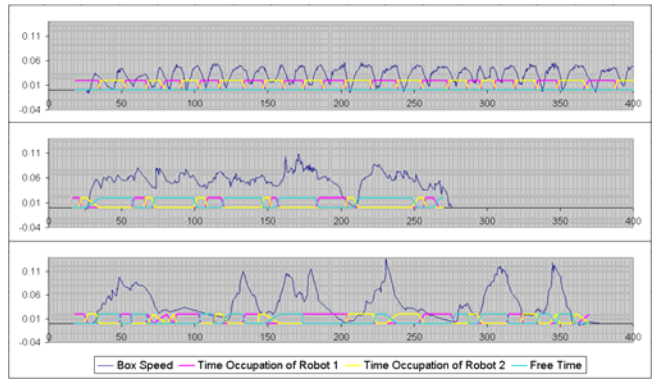


Figure 5 The time distribution curves for teleoperation (upper) and waypoint control (middle) for homogeneous robots, and waypoint control (bottom) for heterogeneous robots

6. EXPERIMENT 3

To test the usefulness of the extended CD measureme for a weakly cooperative MRS, we conducted an experiment assessing coordination demand using an Urban Search And Rescue (USAR) task requiring high human involvement and of a complexity suitable to exercise heterogeneous robot control. In the experiment, participants were asked to control *explorer* robots equipped with a laser range finder but no camera and *inspector* robots with only cameras. Finding and marking a victim required using the inspector's camera to find a victim to be marked on the map generated by the explorer. The capability of the robots and the cooperation autonomy level were used to adjust the coordination demand of the task..

6.1 Experimental design

Three simulated Pioneer P2AT robots and 3 Zergs [1], a small experimental robot were used. Each P2AT was equipped with a front laser scanner with 180 degree FOV and resolution of 1 degree. The Zerg was mounted with a pan-tilt camera with 45 degree FOV. The robots were capable of localization and able to communicate with other robots and control station. The P2AT served as an explorer to build the map while the Zerg could be used as an inspector to find victims using its camera. To accomplish the task the participant must coordinate these two types robot to ensure that when an inspector robot finds a victim, it is within a region mapped by an explorer robot so the position can be marked.

Three conditions were designed to vary the coordination demand on the operator. Under condition 1, the explorer had 20 meters detection range allowing inspector robots considerable latitude in their search. Under condition 2, scanner range was reduced to 5 meters requiring closer proximity to keep the inspector within mapped areas. Under condition 3, explorer and inspector robots were paired as subteams in which the explorer robot with a sensor range of 5 meters followed its inspector robot to map areas being searched. We hypothesized that CDs for explorer and inspector robots would be more even distributed under condition-2 (short range sensor) because explorers would need to move more frequently in response to inspectors' searches than in condition-1 in which CD should be more asymmetric with explorers exerting greater demand on inspectors.

We also hypothesized that lower CD would lead to higher team performance. Three equivalent damaged buildings were constructed from the same elements using different layouts. Each environment was a maze like building with obstacles, such as chairs, desks, cabinets, and bricks with 10 evenly distributed victims. A fourth environment was constructed for training. Figure 6 shows the simulated robots and environment. A within subjects design with counterbalanced presentation was used to compare the cooperative performance across the three conditions.

6.2 Results

Overall performance was measured by the number of victims found, the explored areas, and the participants' self-assessments. To examine cooperative behavior in finer detail, CDs were computed from logged data for each type robot under the three conditions. We compared the measured CDs between condition 1 (20 meters sensing range) and condition 2 (5 meters sensing range), as well as condition 2 and condition 3 (subteam). To further analyze the cooperation behaviors, we evaluated the total attention demand in robot control and control action pattern as well. Finally, we introduce control episodes showing how CDs can be used to identify and diagnose abnormal control behaviors.

6.2.1 *Overall performance*

Examination of data showed two participants failed to perform the task satisfactorily. One commented during debriefing that she thought she was supposed to mark inspector robots rather than victims. After removing these participants a paired t-test shows that in condition-1 (20 meters range scanner) participants explored more regions, $t(16) = 3.097$, $p = 0.007$, as well as found more victims, $t(16) = 3.364$, $p = 0.004$, than under condition-2 (short range scanner). In condition-3 (automated subteam) participants found marginally more victims, $t(16) = 1.944$, $p = 0.07$, than in condition-2 (controlled cooperation) but no difference was found for the extent of regions explored. In the posttest survey, 12 of the 19 (63%) participants reported they were able to control the robots although they had problems in handling some interface components, 6 of the 19 (32%) participants thought they used the interface very well, and only one participant reported it being hard to handle all the components on the user interface but still maintained

Figure 6 Scout and Explorer robots

she was able to control the robots. Most participants (74%) thought it was easier to coordinate inspectors with explorers with long range scanner. 12 of the 19 (63%) participants rated auto-cooperation between inspector and explorer (the subteam condition) as improving their performance, and 5 (26%) participants though auto-cooperation made no difference. Only 2 (11%) participants judged team autonomy to make things worse.

6.2.2 *Coordination effort*

During the experiment we logged all the control operations with timestamps. From the log file CDs were computed for each type robot according to equation 2 in section 2. Figure 7 shows a typical (IT,FT) distribution under condition 1 (20 meters sensing range) in the experiment with a calculated CD for the explorer of 0.185, a CD for the inspector of 0.06. The low CDs reflect that in trying to control 6 robots the participant ignored some robots while attending to others. The CD for explorers is roughly twice the CD for inspectors. After the participant controlled an explorer, he needed to control an inspector multiple times or multiple inspectors since the explorer has a long detection range and large FOV. In contrast, after controlling an inspector, the participant needed less effort to coordinate explorers.

Figure 8 shows the mean of measured CDs. We predicted that when the explorer has a longer detection range, operators would need to control the inspectors more frequently to cover the mapped area. Therefore a longer detection range should lead to higher CD for explorers. This was confirmed by a two tailed t-test that found higher coordination demand, $t(18) = 2.476$, $p = 0.023$, when participants controlled explorers with large (20 meters) sensing range.

We did not find a corresponding difference, $t(18)=.149$, $p=0.884$, between long and short detection range conditions for the CD for inspectors. This may have occurred because under these two conditions the inspectors have exactly the same capabilities and the difference in explorer detection range was not large enough to impact inspectors' CD for explorers. Under the subteam condition, the automatic cooperation within a subteam decreased or eliminated the coordination requirement when a participant controlled an inspector. Within participant comparisons shows that the measured CD of inspectors under this condition is significantly lower than the CD under condition 2 (independent control with 5 meters detection range), $t(18) = 6.957$, $p < 0.001$. Because the explorer always tries to automatically follow an inspector, we do not report CD of explorers in this condition.

As auxiliary parameters, we evaluated the total attention demand, i.e. the occupation rate of total interaction time in the whole control period, and the action pattern, the ratio of control times between inspector and explorer, as well. Paired t-test shows that under long sensing conditions, participants interacted with robots more times than under short sensing

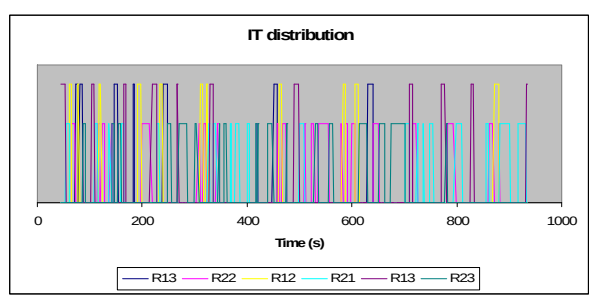


Figure 7 Typical (IT,FT) distribution (higher line indicates the interactions of explorers).

which implies that more robot interactions occurred. The mean action patterns under long and short range scanner conditions are 2.31 and 1.9 respectively. This means that with 20 and 5 meters scanning ranges, participants controlled inspectors 2.31 and 1.9 times respectively after an explorer interaction. Within participant comparisons shows that the ratio is significantly larger under long sensing condition than under short range scanner condition, $t(18) = 2.193$, $p = 0.042$.

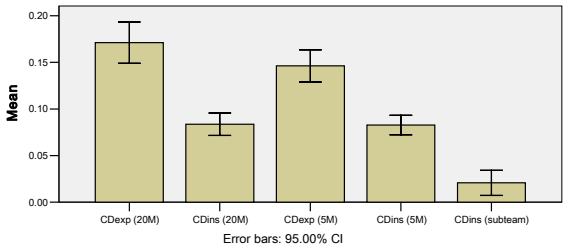


Figure 8 CDs for each robot type

6.2.3 *Analyzing Performance*

As an example of applying CDs to analyze coordination behavior, the performance over explorer CD and total attention demand under the 20 meters sensing range condition reveals three abnormal cases A, B, and C low on both CD and TAD. Associating these cases with recorded map snapshots, we observed that in case A, one robot was entangled by a desk and stuck for a long time, for case B, two robots were controlled in the first 5 minutes and afterwards ignored, and in case C, the participant ignored two inspectors throughout the entire trial.

7. CONCLUSIONS

We proposed an extended Neglect Tolerance model to allow us to evaluate coordination demand in applications where an operator must coordinate multiple robots to perform dependent tasks. Results from the first experiment that required tight coordination conformed closely to our hypotheses with the teleoperation condition producing CD=1 as predicted and heterogeneous teams exerting greater demand than homogenous ones. The CD measure proved useful in identifying abnormal control behavior revealing inefficient control by one participant through irregular time distributions and close CDs for P2ATs under homogeneous and heterogeneous conditions (0.23 and 0.22), a mistake with extended recovery time (41 sec) in another, and a shift to asatisficing strategy between homogeneous and heterogeneous conditions revealed by a drop in CD (0.17 to 0.11) in a third.

As most target applications such as construction or search and rescue require weaker cooperation among heterogeneous platforms the second experiment extended NT methodology to such conditions. Results in this more complex domain were mixed. Our findings of increased CD for long sensor range may seem counter intuitive. Our data show, however, that this effect is not substantial and provide an argument for focused metrics of this sort which measure constituents of the human-robot system directly. Moreover, this experiment also shows how CD can be used to guide us to identify and analyze aberrant control behaviors.

We anticipated a correlation between performance (found victims) and process (measured CDs) measures. However, we did not find the expected relationship in this experiment. From observation of participants during the experiment we believe that high level strategies, such as choosing areas to be searched and path planning, had a significant impact on the overall performance. The participants had few problems in learning to jointly control *explorers* and *inspector*s but needed time to figure out effective strategies for performing the task. Because CD measures control behaviors not strategies these effects were not captured. These experiments have demonstrated the utility of measuring the process of human-robot interaction as well as outcomes to diagnosing operator performance and identifying aspects of the task, particularly for multiple robots, that might benefit from automation. While our approach to measuring CD was supported in the last two experiments the third experiment suggests the need for more sophisticated measures that can take into account strategies and patterns of actions as well as their durations.

8. ACKNOWLEDGEMENTS

This work was supported in part by the Air Force Office of Scientific Research under Grant FA9550-07-1-0039 and Office of Naval Research Grant N00014-03-1-0580.

9. REFERENCES

[1] Balakirsky, S., Carpin, S., Kleiner, A., Lewis, M., Visser, A., Wang, J., and Zipara, V. Toward hetereogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue, Journal of Field Robotics, 2007, 24(11-12): p. 943-967.

[2] Carpin, S., Wang, J., Lewis, M., Birk, A., and Jacoff, A. (2005). High fidelity tools for rescue robotics: Results and perspectives, Robocup 2005 Symposium.

[3] S. Carpin, T. Stoyanov, Y. Nevatia, M. Lewis, M.,J. Wang, (2006a). Quantitative assessments of USARSim accuracy". Proceedings of PerMIS 2006

[4] Carpin, S., Lewis, M., Wang, J., Balakirsky, S., Scrapper, C. (2006b). Bridging the gap between simulation and reality in urban search and rescue. Robocup 2006: Robot Soccer World Cup X, Springer, Lecture Notes in Artificial Intelligence

[5] J. W. Crandall, M. A. Goodrich, D. R. Olsen, and C. W. Nielsen. Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics, Part A, 35(4):438–449, 2005

[6] B. Gerkey and M. Mataric. A formal framework for the study of task allocation in multi-robot systems. International Journal of Robotics Research, 23(9):939–954, 2004.

[7] A. Jacoff, E. Messina and J. Evans, Experiences in deploying test arenas for autonomous mobile robots. In Proceedings of the 2001 Performance Metrics for Intelligent Systems (PerMIS) Workshop, Mexico City, Mexico, September 2001.

[8] M. Lewis, S. Hughes, J. Wang, M. Koes and S. Carpin, Validating USARsim for use in HRI research, Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society, Orlando, FL, 457-461, 2005.

[9] C. Pepper, S. Balakirsky and C. Scrapper. Robot Simulation Physics Validation, Proceedings of PerMIS'07, 2007.

[10] N. Schurr, J. Marecki, M. Tambe, P. Scerri, N. Kasinadhuni, and J. Lewis, The future of disaster response: Humans working with multiagent teams using DEFACTO, In AAAI Spring Symposium on AI Technologies for Homeland Security, 2005.

[11] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz and M. Goodrich, M. (2006). Common Metrics for Human-Robot Interaction, 2006 Human-Robot Interaction Conference, ACM, March, 2006.

[12] B. Taylor, S. Balakirsky, E. Messina and R. Quinn. Design and Validation of a Whegs Robot in USARSim, Proceedings of PerMIS'07.
[13] Wang, J. Human Control of Cooperating Robots, dissertation, University of Pittsburgh, http://etd.library.pitt.edu/ETD/available/etd -01072008-135804/unrestricted/Wang_EDC_2007-final2.pdf (accessed 7/22/2008), 2007.
[14] J. Wang and M. Lewis (2007) Human control of cooperating robot teams, 2007 Human-Robot Interaction Conference, ACM.
[15] M. Zaratti, M. Fratarcangeli and L. Iocchi. A 3D Simulator of Multiple Legged Robots based on USARSim. Robocup 2006: Robot Soccer World Cup X, Springer, LNAI, 2006.