Assessing Cooperation in Human Control of Heterogeneous Robots

Jijun Wang
University of Pittsburgh
School of Information Sciences
Pittsburgh, PA, USA
001-412-624-9426
Jiw1+@pitt.edu

Huadong Wang
University of Pittsburgh
School of Information Sciences
Pittsburgh, PA, USA
001-412-624-9426
Jwhd82@gmail.com

Michael Lewis
University of Pittsburgh
School of Information Sciences
Pittsburgh, PA, USA
001-412-624-9426
ml@sis.pitt.edu

ABSTRACT
Human control of multiple robots has been characterized by the average demand of single robots on human attention. While this matches situations in which independent robots are controlled sequentially it does not capture aspects of demand associated with coordinating dependent actions among robots. This paper presents an extension of Crandall's neglect tolerance model intended to accommodate both coordination demands (CD) and heterogeneity among robots. The reported experiment attempts to manipulate coordination demand by varying the proximity needed to perform a joint task in two conditions and by automating coordination within subteams in a third. Team performance and the process measure CD were assessed for each condition. Automating cooperation reduced CD and improved performance. We discuss the utility of process measures such as CD to analyze and improve control performance.

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
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 [10] 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 follow this approach to develop measures characterizing the demand imposed by tasks requiring cooperation among robots.

HRI'08, March 13-15, 2008, Amsterdam, Netherlands.

Applications for multirobot systems (MRS) such as interplanetary construction or cooperating uninhabited aerial vehicles will require close coordination and control between human operator(s) and teams of robots in uncertain environments. Human supervision will be needed because humans must supply the perhaps changing, goals that direct MRS activity. Robot autonomy will be needed because the aggregate decision making demands of a MRS are likely to exceed the cognitive capabilities of a human operator. Autonomous cooperation among robots, in particular, will likely be needed because it is these activities [1] that theoretically impose the greatest decision making load.

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 [2]. 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.

Because of the need to share attention between robots in MRS, teloperation can only be used for one robot out of a team [7] or as a selectable mode [8]. Some variant of waypoint control has been used in most of the MRS studies we have reviewed [2,7,8,11] with differences arising primarily in behavior upon reaching a waypoint. A more fully autonomous mode has typically been included involving things such as search of a designated area [7], travel to a distant waypoint [11], or executing prescribed behaviors [6]. In studies in which robots did not cooperate and had varying levels of individual autonomy [2,7,11] (team size 2-4) performance and workload were both higher at lower autonomy levels and lower at higher ones. So although increasing autonomy in these experiments reduced the cognitive load on the operator, the automation could not perform the replaced tasks as well.

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 MRS control difficulty, neglect tolerance and fan-out [2] 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 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 on cognitive resources. 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. MEASURING COORDINATION DEMAND

If robots must cooperate to perform a task such as searching a building without redundant coverage or act together to push a block, this independence no longer holds. Where coordination demands are weak, as in the search task, the round robin strategy implicit in the additive models may still match observable performance, although the operator must now consciously deconflict search patterns to avoid redundancy. For tasks such as box pushing, coordination demands are simply too strong, forcing the operator to either control the robots simultaneously or alternate rapidly to keep them synchronized in their joint activity. In this case the decline in efficiency of a robot's actions is determined by the actions of other robots rather than decay in its own performance. Under these conditions the sequential patterns of interaction presumed by the NT and fan-out measures no longer match the task the operator must perform. The measurable aspects of the interaction, IT and NT, however remain the same. Confronting a similar problem for measuring the impact of operators' attention allocation (AAE) Crandall and Cummings [3] turned to wait times; the time a robot remained below the performance threshold before intervention.

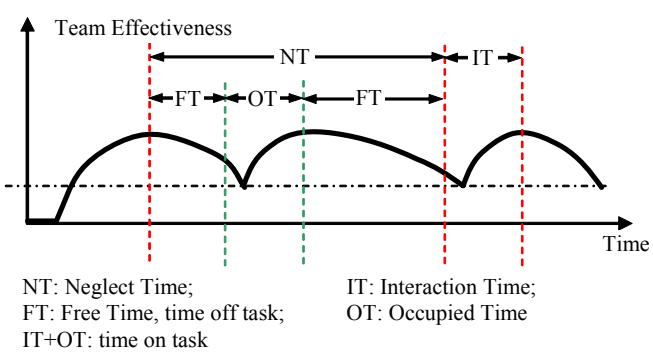

Figure 1. Extended neglect tolerance model for cooperative robot control

Although wait times can be observed they must then be differentiated as having arisen from poor SA (low AAE) or a queue of needy robots (too many or too difficult to control) to contribute to an understanding of the MRS. To separate coordination demand (CD) from the demands of interacting with independent robots we have extended Crandall's [2] 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 dependent tasks we introduce 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 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. A related measure, team attention demand (TAD), adds IT's to both numerator and denominator to provide a measure of the proportion of time devoted to the cooperative task; either performing the task or coordinating robots.

Unlike the basic neglect tolerance model that can be presumed to approximately mirror the time course of actual interactions; (the operator selects robot-1 gets it turned around and moving in the right direction and then moves robot-2 away from a wall so it can continue on its path, etc.) our model of cooperative robot control is not directly observable. While we can observe the IT for robot-1 the classification of the subsequent interaction with robot-2 as OT is conjecture. We cannot distinguish whether the interaction is being performed to keep robot-1's performance above threshold (OT) or for the benefit of robot-2's own performance (IT). What we can do is design tasks that maximize or vary CD and observe the effect on performance and the temporal measures.

One approach is to design experiments that allow comparisons between "equivalent" conditions with and without coordination demands. Wang and Lewis [12], for example, compared 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 [12] examined coordination demand indirectly by comparing performance between conditions in which it was filled either manually or through automation, our more recent experiments have attempted to manipulate coordination difficulty directly. In the first experiment reported in [13] robots performed a box pushing task in which CD was varied by control mode and robot heterogeneity. In the control condition participants controlled two robots through teleoperation requiring continuous coordination. In the medium difficulty condition they controlled heterogeneous robots via issuing waypoints making it difficult to synchronize actions due to differences in kinematics. In the low effort condition they controlled homogeneous robots easing the task of synchronizing distances and times. Because robots of the same type were used in the first two conditions and the heterogeneous robots had equivalent capabilities differences in performance could be attributed to variation in coordination demand between the conditions. Results were as expected with teleoperation leading to CD=1 and the heterogeneous condition producing higher CD than the homogeneous one.

2.1 Measuring weak cooperation

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 the box-pushing experiment, 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 case the Neglect Tolerance model must be further extended to measure coordination between different robot types.

We describe this form of heterogeneous MRS as a MN system with M robots that belong to N robot types, and for robot type i, there are m_i robots, that is $M = \sum_{i=1}^{N} m_i$. Thus we can denote a robot in this system as R_{ij} where $i = [1,N]$, $j = [1,m_i]$. Assume the operator serially controlled the robots for time T, and each robot R_{ij} was interacted with for l_{ij} times, then we can represent each interaction as IT_{ijk} where $i = [1,N]$, $j = [1,m_i]$, $k = [1,l_{ij}]$, and the following free time as FT_{ijk} where $i = [1,N]$, $j = [1,m_i]$, $k = [1,l_{ij}]$. The total control time T_i for type i robot should then be $T_i = \sum_{j,k}\left(IT_{ijk} + FT_{ijk}\right)$. Because robots belonging to the same robot type have identical capabilities, and substitution may cause uneven demand, we are only interested in measuring the average coordination demand CD_i, i=[1,N].

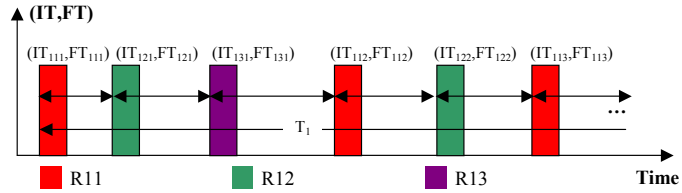

Figure 2. Distribution of (IT,FT)

Given robots of the same type R_{ij}, j=[1,m$_i$], there are OT_i^* and NT_i^* that for each robot R_{ij} we can have $CD_{ij} = \frac{1}{l_{ij}}\sum_{k=1}^{l_{ij}} \frac{OT_{ijk}}{NT_{ijk}} = \frac{l_{ij}OT_i^*}{l_{ij}NT_i^*}$. Therefore, the CD$_i$ for type i robot is

$$CD_i = \frac{1}{m_i}\sum_{j=1}^{m_i} CD_{ij} = \frac{1}{m_i}\sum_{j=1}^{m_i}\left(\frac{1}{l_{ij}}\sum_{k=1}^{l_{ij}}\frac{OT_{ijk}}{NT_{ijk}}\right) = \frac{\sum_{jk} OT_{ijk}}{\sum_{jk} NT_{ijk}}$$

Assume all the other types robots are dependent with the current type robots, then the numerator $OT_i^*\sum_{j=1}^{m_j} l_{ij}$ is the total interaction time of all the other robot types, i.e. $OT_i^*\sum_{j=1}^{m_j} l_{ij} = \sum_{\substack{type=1\\ type\neq i}}^{N} IT$.

For the denominator, it is hard to directly measure NT_i^* because the system performance depends on multiple types of robots and an individual robot may cooperate with different team members over time. Because of this dependency, we cannot use individual robot's active time to approximate NT. On the other hand, the robots may be unevenly controlled. For example a robot might be controlled only once and then ignored because there is another robot of the same type that is available, so we cannot simply use the time interval between two interactions of an individual robot as NT. Considering all the robots belonging to a robot type, the population of individual robots' (IT,FT)s reveal the NT for a type of robot. Figure 2 shows an example of how robots' (IT,FT) might be distributed over task time. Because robots of the same capabilities might be used interchangeably to perform a cooperative task it is desirable to measure NT with respect to a type rather than a particular robot. In Figure 2 robots R11 and R12 have short NTs while R13 has an NT of indefinite length. (IT,FT)*, the distribution of (IT, FT) for the robot type, shown by the arrowed lines between interactions allows an estimate of NT for a robot type that is not affected by long individual NTs such as that of R13. When each robot is evenly controlled, the $(IT_i,NT_i)^*$ should be m$_i$ * (IT$_i$,FT$_i$) where (IT$_i$,FT$_i$) is the average (IT,FT) for type i robot, $\left(IT_i, FT_i\right) = \frac{T_i}{\sum_{j=1}^{m_i} l_{ij}}$. And when only one robot is controlled, the $(IT_i,NT_i)^*$ will be (IT$_i$,FT$_i$). In summary, the $(IT_i,NT_i)^*$ should be in the range $\left(IT_i, FT_i\right) \le \left(IT_i, FT_i\right)^* \le m_i\left(IT_i, FT_i\right)$. Here, we introduce weight

$$w_i = \frac{\sum_{j=1}^{m_i} l_{ij}}{\max_{j=1}^{m_i}(l_{ij})}$$ to measure how evenly the robots are controlled.

With the weight, we can approximate $(IT_i, NT_i)^*$ as:

$$(IT_i, NT_i)^* = w_i(IT_i, NT_i) = \frac{\sum_{j=1}^{m_i} l_{ij}}{\max_{j=1}^{m_i}(l_{ij})} \times \frac{T_i}{\sum_{j=1}^{m_i} l_{ij}} = \frac{T_i}{\max_{j=1}^{m_i}(l_{ij})}$$

Thus, the denominator in CD_i can be calculated as:

$$NT_i^* \sum_{j=1}^{m_i} l_{ij} = \left(\frac{T_i}{\max_{j=1}^{m_i}(l_{ij})} - IT_i^* \right) \sum_{j=1}^{m_i} l_{ij}$$

$$= \frac{\sum_{j=1}^{m_i} l_{ij}}{\max_{j=1}^{m_i}(l_{ij})} T_i - IT_i^* \sum_{j=1}^{m_i} l_{ij}$$

$$= \frac{\sum_{j=1}^{m_i} l_{ij}}{\max_{j=1}^{m_i}(l_{ij})} T_i - \sum_{type=i} IT$$

where $\sum_{type=i} IT$ is the total interaction time for all the type i robots.

In summary, we can compute CD_i as:

$$CD_i = \frac{\sum_{type \neq i} IT}{\frac{\sum_{j=1}^{m_i} l_{ij}}{\max_{j=1}^{m_i}(l_{ij})} T_i - \sum_{type=i} IT} \quad (2)$$

3. EXPERIMENT

To test the usefulness of the CD measurement for a weakly cooperative MRS, we conducted an experiment assessing coordination demand using an Urban Search And Rescue (USAR) task requiring high human involvement [6] 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.

The experiment was conducted in simulation using USARSim, a high fidelity simulation of USAR robots and environments [5]. MrCS (Multi-robot Control System), a multirobot communications and control infrastructure with accompanying user interface developed for experiments in multirobot control and RoboCup competition [1] was used for the experiment. MrCS provides facilities for starting and controlling robots in the simulation, displaying camera and laser output, and supporting inter-robot communication through Machinetta [9] a distributed mutiagent system. The distributed control enables us to scale robot teams from small to large.

3.1 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 *explore*r 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 3 shows the simulated robots and environment.

A within subjects design with counterbalanced presentation was used to compare the cooperative performance across the three conditions. The same control interface shown in Figure 4 allowing participants to control robots through waypoints or teleoperation was used in all conditions.

3.2 Participants

19 paid participants, 19-33, years old were recruited from the University of Pittsburgh community. None had prior experience with robot control although most were frequent computer users. 6 of the participants (31.5%) reported playing computer games for more than one hour per week.

3.3 Procedure

After collecting demographic data the participant read standard instructions on how to control robots via MrCS. In the following

15~20 minute training session, the participant practiced each control operation and tried to find at least one victim in the training arena under the guidance of the experimenter. Participants then began three testing sessions in counterbalanced order with each session lasting 15 minutes. At the conclusion of the experiment participants completed a questionnaire.

Figure 3. The robots and environment

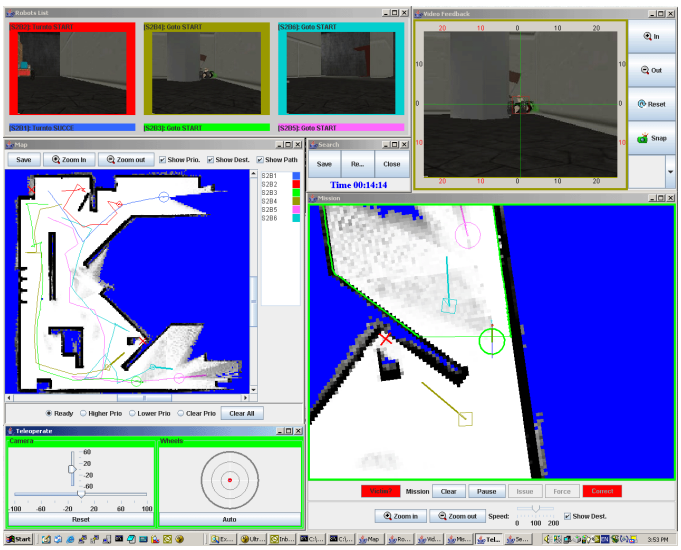

Figure 4. The GUI

4. 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.

4.1 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 she was able to control the robots. Most participants (74%) thought it was easier to coordinate *inspector*s 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.

4.2 Coordination effort

During the experiment, each control action, such as pressing a button on GUI or clicking a mouse button, and each robot state message from the simulation were logged in a file with timestamp to allow us calculate the time parameters, IT, FT and total task time. We then computed CDs for each type robot according to the equation in section 2.1. Figure 5 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 *inspector*s. After the participant controlled an *explorer*, he needed to control an *inspector* multiple times or multiple *inspector*s 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 6 shows the mean of measured CDs. We predicted that when the *explorer* has a longer detection range, operators would need to control the *inspector*s 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 *inspector*s. This may have occurred because under these two conditions the *inspector*s have exactly the same capabilities and the difference in *explorer* detection range was not large enough to impact *inspector*s' 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 *inspector*s 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.

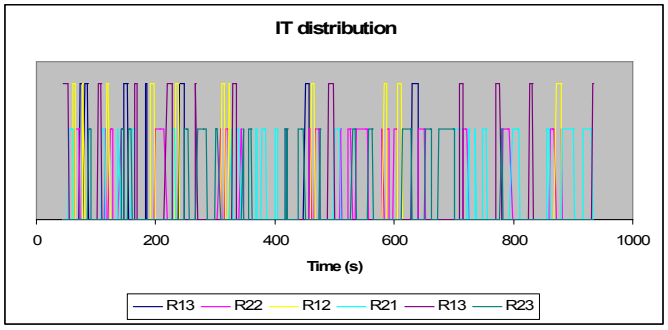

Figure 5 Typical (IT,FT) distribution (higher line indicates the interactions of explorers).

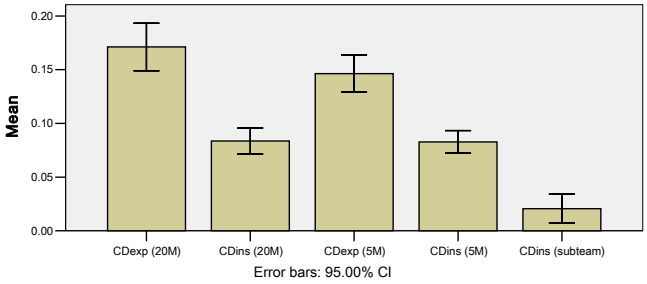

Figure 6 CDs for each robot type

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. Total attention demand measures the team task demand, i.e.; how hard the task is. As we expected paired t-test shows that under subteam condition, participants spent less time in robot control than under short sensing range condition, $t(18)=3.423$, $p=0.003$. However, under long sensing conditions, paired t-test shows that participants spent more time controlling robots than under the short sensing condition, $t(18) = 2.059$, $p = 0.054$. This is opposite to our hypothesis that searching for victims with shorter sensing range should be harder because the robot would need to be controlled more often. Noticing that total attention demand was based on the time spent controlling not the number of times a robot was controlled we examined the number of control episodes. Under long and short sensing range conditions two tailed t-tests found participants to control explorers more times with short sensing explorers, $t(18)=2.464$, $p=.024$, with no differences found in frequency of inspector control, $p=.97$. We believe that with longer sensing explorers participants tend to issue longer paths in order to build larger maps. Because the sensing range in condition 1 is five times longer than the range in condition 2, the increased control time under the long sensing condition my overwhelm the increased explorer control times. This is partially confirmed by a paired t-test that found longer average control time for explorers and inspectors under the long detection condition, $t(18)=3.139$, $p=.006$, $t(18)=2.244$, $p=.038$, respectively. On average participants spent 1.5s and 1.0s more time in explorer and inspector control in the long range condition. 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 *inspector*s 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, $t(18) = 2.193$, $p = 0.042$.

4.3 Analyzing Performance

As an example of applying CDs to analyze coordination behavior, Figure 7 shows the performance over *explorer* CD and total attention demand under the 20 meters sensing range condition. Three abnormal cases A, B, and C can be identified from the graph. 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 *inspector*s throughout the entire trial.

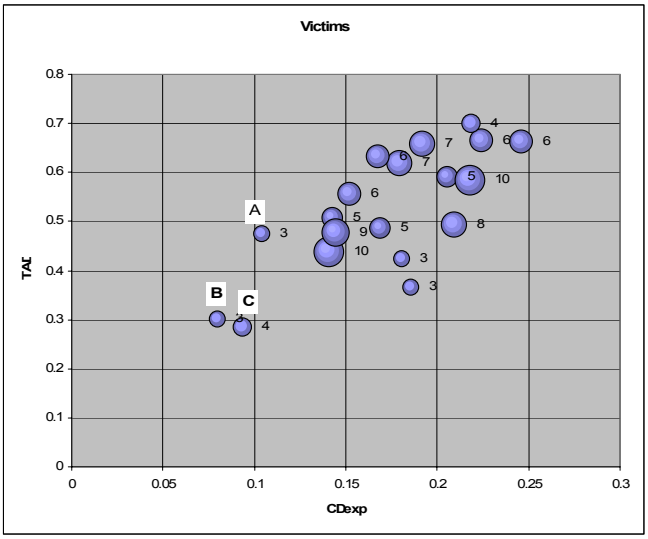

Figure 7 Found victims distribution over CDexp and TAD (total attention demand).

5. CONCLUSION

We proposed an extended Neglect Tolerance model to allow us to evaluate cooperation effort in applications where an operator must coordinate multiple robots to perform dependent tasks. Our previous study [13] validated CD measurement for an extended model under tight cooperation, such as box pushing. However, most target applications such as construction or search and rescue are likely to require weaker cooperation among heterogeneous platforms. The present study further extends NT methodology to such weak cooperation conditions. On first blush our findings on CD for sensor ranges may seem counter intuitive because inspectors would be expected to exert greater CD on explorers with short sensor range. 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 identify and analyze aberrant control behaviors.

We anticipated a correlation between found victims and the measured CDs. 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, have significant impact on the overall performance. The participants had few problems in learning to jointly control *explorers* and *inspector*s but they needed time to figure out effective strategies for performing the task. Because CD measures control behaviors not strategies these effects were not captured.

On the other hand, because the NT methodology is domain and task independent our CD measurement could be used to characterize any dependent system. For use in performance analysis, however, it must be associated with additional domain and task dependent information. As shown in our examples, combined with generated maps and traces CD provides an excellent diagnostic tool for examining performance in detail.

In the present experiment, we examined the action pattern under long and short sensing range conditions. The results reveal that it can be used as an evaluation parameter, and more important, it may guide us in the design of multiple robot systems. For instance, the observation that one *explorer* control action was followed on average by 2 *inspector* control actions may imply that the MRS should be constructed by n *explorer* and 2n *inspectors.*

In summary, the proposed methodology enables us evaluate weak or tight cooperation behaviors in control of heterogeneous robot teams. The time parameter based measurement makes this methodology domain independent and practical in real applications. The lack of consideration of domain, other system characteristics and information available to the operator, however, makes this metric too impoverished to use in isolation for evaluating system performance. A more complete metric for evaluating coordination demand in multirobot systems would require additional dimensions beyond time. Considering human, robot, task and world as the four elements in HRI, possible metrics might include mental demand, situation awareness, robot capability, autonomy level, overall task performance, task complexity, and world complexity.

6. ACKNOWLEDGMENTS

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

7. 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, in press.

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

[3] Crandall, J. and Cummings. M. Developing performance metrics for the supervisory control of multiple robots, *Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction* (HRI'07), 33-40, 2007.

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

[5] Lewis, M., J. Wang, and S. Hughes, USARSim : Simulation for the Study of Human-Robot Interaction. Journal of Cognitive Engineering and Decision Making, 2007. 1(1): p. 98-120.

[6] Murphy, R.R. and J.L. Burke, Up from the Rubble: Lessons Learned about HRI from Search and Rescue, in Proceedings of the 49th Annual Meetings of the Human Factors and Ergonomics Society. 2005: Orlando.

[7] Nielsen, C., Goodrich, M., and Crandall, J. Experiments in human-robot teams. In *Proceedings of the 2002 NRL Workshop on Multi-Robot Systems*, October 2003

[8] Parasuraman, R., Galster, S., Squire, P., Furukawa, H. and Miller, C.. A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with roboflag. *IEEE Systems, Man and Cybernetics-Part A*, 35(4):481–493, July 2005

[9] Scerri, P., et al., Coordinating large groups of wide area search munitions, in Recent Developments in Cooperative Control and Optimization, D. Grundel, R. Murphey, and P. Pandalos, Editors. 2004, Singapore: World Scientific. p. 451-480.

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

[11] Trouvain, B. and Wolf, H.. Evaluation of multi-robot control and monitoring performance. In *Proceedings of the 2002 IEEE Int. Workshop on Robot and Human Interactive Communication*, pages 111–116, September 2002.

[12] Wang, J. and Lewis, M. Human control of cooperating robot teams, *2007 Human-Robot Interaction Conference*, ACM, 2007a.

[13] Wang, J. and Lewis, M. Assessing coordination overhead in control of robot teams, *Proceedings of the 2007 International Conference on Systems, Man, and Cybernetics*, 2007b, in press.