

How Search and its Subtasks Scale in N Robots

Huadong Wang, Michael Lewis
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260
011-412-624-9426

huw16@pitt.edu ml@sis.pitt.edu

Prasanna Velagapudi, Paul Scerri, Katia Sycara
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
011-412-268-8825

pkv@cs.cmu.edu pscerri@cs.cmu.edu
katia+@cs.cmu.edu

ABSTRACT

The present study investigates the effect of the number of controlled robots on performance of an urban search and rescue (USAR) task using a realistic simulation. Participants controlled either 4, 8, or 12 robots. In the *fulltask* control condition participants both dictated the robots' paths and controlled their cameras to search for victims. In the *exploration* condition, participants directed the team of robots in order to explore as wide an area as possible. In the *perceptual search* condition, participants searched for victims by controlling cameras mounted on robots following predetermined paths selected to match characteristics of paths generated under the other two conditions. By decomposing the *search and rescue* task into *exploration* and *perceptual search* subtasks the experiment allows the determination of their scaling characteristics in order to provide a basis for tentative task allocations among humans and automation for controlling larger robot teams. In the *fulltask* control condition task performance increased in going from four to eight controlled robots but deteriorated in moving from eight to twelve. Workload increased monotonically with number of robots. Performance per robot decreased with increases in team size. Results are consistent with earlier studies suggesting a limit of between 8-12 robots for direct human control.

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

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 demands of decision making and control of a MrS are likely to exceed the cognitive capabilities of a human operator. Controlling robots that must act cooperatively, in particular, will likely be difficult because it is these activities [1] that theoretically impose the greatest decision making load. Because some functions of a MrS such as identifying victims among rubble depend on human input, evaluating the operator's span of control as the number of controlled entities scale is critical for designing feasible human-automation control systems.

Current estimates of human span of control limitations are severe. Miller [2], for example, showed that under expected target densities, a controller who is required to authorize weapon release for a target identified by a UAV could control no more than 13 UAVs even in the absence of other tasks. A similar breakpoint of 12 was found by [3] for retargeting Tomahawk missiles. Smaller numbers (3-9) [4] have been found for ground robots which typically require more frequent attention.

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. Because coordination demands can vary greatly across tasks [5], frequently dominate the demands on operator attention [6], and grow exponentially in the number of robots [1], a task involving independent operation of robots provides a clearer picture of the way performance scales with the number of robots. In the simplest case an operator controls multiple independent robots, interacting with each as needed. The *foraging* task [7] used for this experiment, in which each robot searches its own region, is of this category although minimal coordination may be required to avoid overlaps or gaps in coverage. Multirobot *foraging* with waypoint control was chosen for this experiment because search tasks have been widely cited as a likely field application and subject of study for multirobot systems and waypoint control represents the lowest level of automation compatible with independent control of multiple robots.

Control performance at such tasks can be characterized by the average demand of each robot on human attention [4].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'09, March 11–13, 2009, La Jolla, California, USA.

Copyright 2009 ACM 978-1-60558-404-1/09/03...\$5.00.

Table 1. Recent studies of the effects of robot team size on performance

Study	Task	World	Robots	Interaction
<i>Trouvain & Wolf (2002)</i> : User study of the impact of robot group size	Navigation	2D simulated office world	2, 4, 8 UGVs (homogeneous)	Waypoint
<i>Trouvain et al. (2003)</i> : user Study of map based and camera based user interface	Exploration	3D simulated outdoor world	1, 2, 4 UGVs (homogeneous)	Supervisory + waypoint control
<i>Olsen & Wood (2004)</i> : Fan-out independent study	Foraging	2D simulated office like world	18 UGVs (homogeneous)	Waypoint
<i>Humphrey et al. (2006)</i> : User study of robot team & halo interface	Foraging	3D simulated outdoor world (USARSim)	6, 9 UGVs (heterogeneous)	Teleoperation and scripted behaviors

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. Established methods of estimating MrS control difficulty, the neglect tolerance model, NT, [4] and the Fan-out measure [8], are predicated on such independence of robots and tasks. In the NT model, 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 can also provide an estimate of the number of robots that might be controlled. Fan-out, a term borrowed from the number of gates a logic gate can drive, refers to this maximum number of robots that can be advantageously controlled under particular conditions. Although Fan-out and the NT parameters should in theory be calculable from interaction with a single robot, data [8, 6] suggest that as the number of robots is increased operators relax the performance threshold to accommodate the control of additional robots. Following [8] we will estimate the number of controlled robots by inspecting for a plateau without further improvement, counting the number of robots actually being operated [8, 9] as well as by estimating the maximum number of controlled robots using parameters from the NT model [4].

Some variant of automation through waypoint control was used in each of the reported human-in-the-loop simulation studies investigating performance as a function of the size of robot team [10,9,8,11] (see Table 1) with differences arising primarily in behavior upon reaching a waypoint. As Olsen and Wood [8] demonstrate, increased automation can drastically affect Fan-out extending it from two to nine in their study. In this review we differentiate between naïve waypoint control in which the robot moves in a single direction stopping upon encountering an obstacle (Fan-out=2 [8]) and standard waypoint control in which the robot employs obstacle avoidance and path planning to reach its waypoint. We also distinguish among three similar types of tasks: *navigation*, *exploration*, and *foraging* or *search*. In *navigation* [10, 8] the operator has a complete map or aerial view and only needs to direct robots to their goals. In *exploration* [9] the operator begins without knowledge of the terrain which must be built up by a process of exploration usually in a form resembling a laser rangefinder generated map. *Foraging* tasks [8, 11] are a variant of

exploration in which the robot seeks targets in the region being explored. Typically performance is reported as taking some action on these targets such as defusing a bomb or marking a victim on a map. These three tasks place different demands on the operator with *navigation* requiring simple path planning, *exploration* requiring monitoring and reactive path planning, and *foraging* adding the requirement of searching for targets.

Trouvain and Wolf [10] compared the navigation performance of two, four, and eight robots controlled by a single operator in a 2D simulated world. The operator’s task was to navigate robots using naïve waypoint control to reach *inspection points* that appeared on the map at varying locations and intervals. Upon reaching the *inspection point* the operator was required to execute an *inspect* command. The experiment found that system performance (ordering *inspections* after robots reached the *inspection point*) was higher for four robots than for two with only marginal further improvement in going from four to eight robots. Workload ratings showed the opposite pattern with a marginal increase between two and four and a significant one between four and eight. In a later experiment [9] using a substantially more realistic simulation the authors studied exploration for teams of 1, 2, and 4 robots. The operators performed a supervisory control task involving substantially greater automation than waypoint control in which the robots autonomously navigated toward their goal points with the operator allowed to override by directing them through new waypoints. Comparing individual robot performance they found significantly better performance in the one robot condition but no further decline in going from two to four robots. Participants in the multirobot conditions, however, performed less well than hands-off automation alone making interpretation difficult. The authors present a graph showing an increase in the mental demand scale of the NASA-TLX with increase in the number of robots.

Olsen and Wood [8] conducted a set of experiments in *navigation* (light worlds) and *foraging* (dark worlds) with automation ranging from naïve to standard waypoint control. In these experiments participants were supplied with up to 18 robots to navigate a maze to the location of targets. The task continued until all targets were found. Activity time, AT, was defined as the average time a robot was moving between interventions. Fan-out was determined empirically as the average number of active robots with

Interaction Effort, a stand in for Interaction Time, defined as $IE=AT/FO$. Because task completion times were not reported, aggregate team performance measures are not available as in the other studies. In going from a less cluttered to more cluttered environment they found Fan-out for waypoint control at the *foraging* task varied from 5 to 9 with mean activity times going from 14.2 to 24.18. The finding is easily understood. As the maze grows more complex, operators must intervene more frequently to retask robots. AT, largely determined by the time between these interventions, is reduced, leaving less time for operators to control additional robots, and hence the reduction in Fan-out. These results, however, illustrate how slippery the notion of Fan-out is, in that it is shown to depend not only on level of automation as mentioned earlier but on task characteristics as well.

In the most recent study [11], researchers used the UGV simulator, USARSim (also used in the current study), to compare robot control behaviors for six and nine independently controlled heterogeneous robots. The participants controlled the robots via teleoperation, through limited waypoint control by which a robot could be directed to orient or move to another robot's location, or through an *explore* command that caused the robot to move ahead, pausing at intervals to provide 360° camera views. This is a somewhat higher degree of automation than waypoint control alone so we would expect slightly higher values for Fan-out and neglect times. Although operators controlling six robots defused at least one bomb on the same number of trials the larger nine robot teams defused two or more bombs on 56.3% more trials. Times also favored nine robot teams with time per diffusion and time to second diffusion significantly lower for the nine robot teams. The results show that a higher number of robots led to higher workload measured using the NASA-TLX. The experimenters report improving SA with added robots, contradicting the common belief [4, 12, 10] that more robots lead to worse SA. However, because only the *attentional demand* scale of the 3D SART was significantly higher, more robots may have simply required greater effort to monitor without actually benefiting SA.

These studies suggest that for *foraging* tasks using waypoint control the Fan-out plateau lies somewhere between 4 and 9+ robots depending on the level of robot autonomy and environmental demands. Many of the tasks envisioned for robot teams, however, require larger numbers. To increase Fan-out we must increase robot autonomy, preferably in ways most compatible with operator capabilities. The *foraging* task can be decomposed into *exploration* and *perceptual search* subtasks corresponding to navigation of an unknown space and searching for targets by inspecting and controlling onboard cameras. In the reviewed studies, two [9,11] extend automation of the exploration subtask while one [8] leaves this task to the operator. Fan-out values, however, fall in the same range across these experiments. The present study investigates the scaling of performance with number of robots for operators performing either the full task or only one of the subtasks to identify limiting factors. Treating this as a resource sharing problem [12], we can determine whether:

- full task performance is being limited by performance on one

of the subtasks

$fulltask \approx subtask_1$ and $subtask_2 > fulltask$, in which case $subtask_1$ is a good candidate for increased automation

- full task performance shows some cost of concurrence for sharing the subtasks but subtasks scale at different rates
 $fulltask < subtask_1$, $fulltask < subtask_2$, and
 $subtask_1(N+) - subtask_1(N) > subtask_2(N+) - subtask_2(N)$, in which case $subtask_2$, which scales at a slower rate, is a good candidate for increased automation

If performance is equal across the conditions or the subtasks scale at the same rate, automation decisions would need to be based on other factors. The logic of our approach depends upon the equivalence among these three conditions.

In the *fulltask* condition operators used waypoint control to explore an office like environment. When victims were detected using the onboard cameras the robot was stopped and the operator marked the victim on the map and returned to exploration. Equating the *exploration* subtask was relatively straightforward. Operators were given the instruction to explore as large an area as possible with coverage judged by the extent of the laser rangefinder generated map. Because operators in the *exploration* condition did not need to pause to locate and mark victims the areas they explored should be strictly greater than in the *fulltask* condition. This discrepancy can be removed, however, by truncating these data by an interval corresponding to the pauses in the *fulltask* data.

Developing an equivalent *perceptual search* condition is more complicated. The operator's task resembles that of the payload operator for a UAV or a passenger in a train, in that she has no control over the platform's trajectory but can only pan and tilt the cameras to find targets. The targets the operator has an opportunity to acquire, however, depend on the trajectories taken by the robots. If an autonomous path planner is used, robots will explore continuously covering a wider area than when operated by a human (where pauses typically occur upon arrival at a waypoint). If human generated trajectories are taken from the *fulltask* condition, however, they will contain additional pauses at locations where victims were found and marked providing an undesired cue. Instead, we have chosen to use trajectories from the *exploration* condition since they should contain pauses associated with waypoint arrival but not those associated with identifying and marking victims. As a final adjustment, operators in the *perceptual search* condition must be able to pause their robots in order to identify and mark the victims they discover. While the equivalence of these trajectories cannot be directly established determining that for area covered $exploration > fulltask \approx perceptual search$, that for pause durations $exploration < fulltask \approx perceptual search$ supports equivalence.

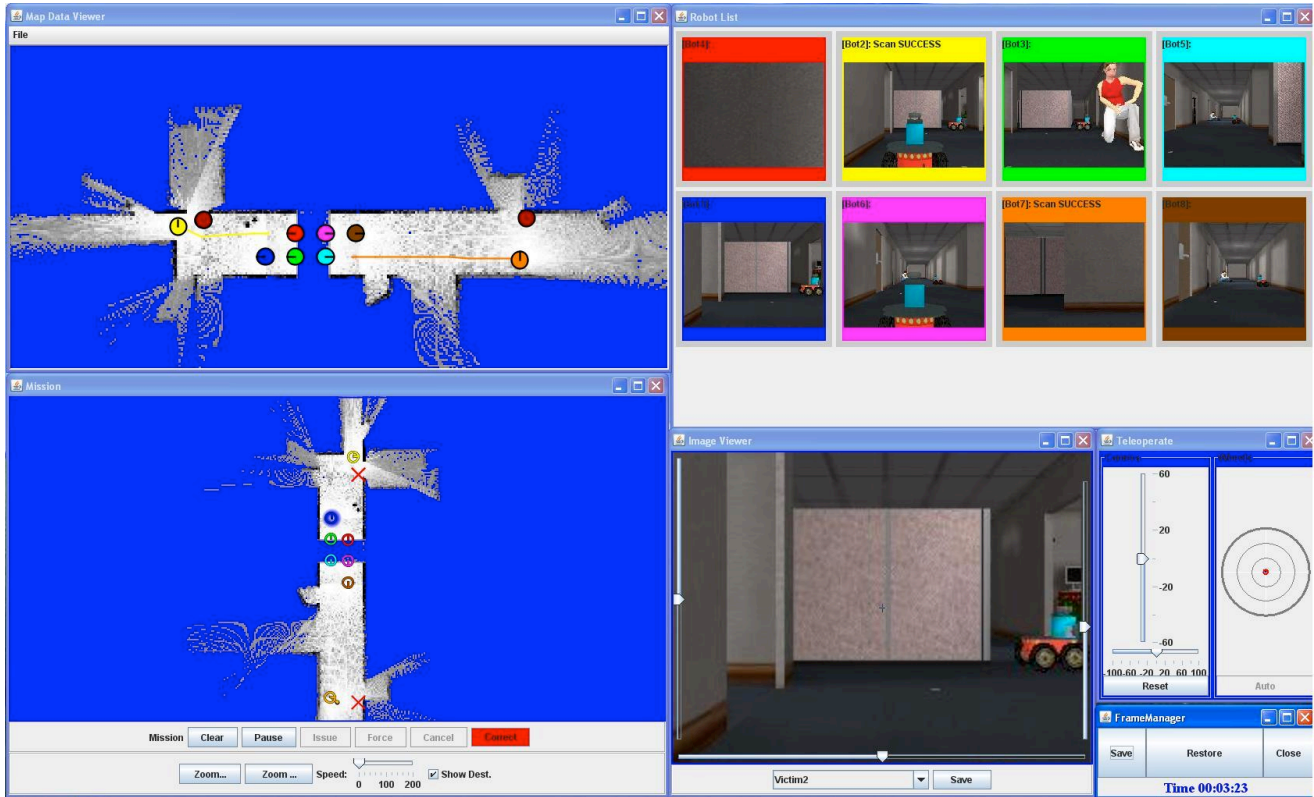


Figure 1. GUI for Multirobot control

2. METHODS

2.1 USARSim and MrCS

The reported experiment was performed using the USARSim robotic simulation with 4-12 simulated UGVs performing Urban Search and Rescue (USAR) *foraging* tasks. USARSim is a high-fidelity simulation of urban search and rescue (USAR) robots and environments developed as a research tool for the study of HRI and multi-robot coordination. USARSim supports HRI by accurately rendering user interface elements (particularly camera video), accurately representing robot automation and behavior, and accurately representing the remote environment that links the operator's awareness with the robot's behaviors. USARSim can be downloaded from www.sourceforge.net/projects/usarsim and serves as the basis for the Virtual Robots Competition of the RoboCup Rescue League. USARSim uses Epic Games'

UnrealEngine2 [13] to provide a high fidelity simulator at low cost. Validation studies showing agreement for a variety of feature extraction techniques between USARSim images and camera video are reported in Carpin et al. [14]. Other sensors including sonar and audio are also accurately modeled. Validation data showing close agreement in detection of walls and associated Hough transforms for a simulated Hokuyo laser range finder are described in [15]. The current UnrealEngine2 integrates MathEngine's Karma physics engine [16] to support high fidelity rigid body simulation. Several validation studies show close agreement in behavior between USARSim models and real robots being modeled [17,18,19,20,21].

MrCS (Multi-robot Control System), a multirobot communications and control infrastructure with accompanying user interface developed for experiments in multirobot control and

RoboCup competition [22] was used in these experiments. MrCS provides facilities for starting and controlling robots in the simulation, displaying camera and laser output, and supporting inter-robot communication through Machinetta a distributed multiagent system. Figure 1 shows the elements of the MrCS. The operator selects the robot to be controlled from the colored thumbnails at the top of the screen. To view more of the selected scene shown in the large video window the operator uses pan/tilt sliders to control the camera. Robots are tasked by assigning waypoints on a heading-up map on the Mission Panel (bottom right) or through a teleoperation widget (bottom left). The current locations and paths of the robots are shown on the Map Data Viewer (middle left).

2.2 Experimental Conditions

A large USAR environment previously used in the 2006 RoboCup Rescue Virtual Robots competition [22] was selected for use in the experiment. The environment was a maze like hall with many rooms and obstacles, such as chairs, desks, cabinets, and bricks. Victims were evenly distributed within the environment. A second simpler environment was used for training. The experiment followed a between-groups repeated measures design with number (4, 8, 12) of robots defining the repeated measure. Participants in the *fulltask* condition performed the complete USAR task. In the subtask conditions they performed variants of the USAR task requiring only *exploration* or *perceptual search*. Participants in the *fulltask* condition followed instructions to use the robots to explore the environment and locate and mark on the map any victims they discovered. The *exploration* condition differed only in instructions. These operators were instructed to explore as much of the environment as possible without any requirement to locate or mark victims. From examination of area coverage, pausing, and other characteristics of trajectories in the

fulltask and *exploration* conditions a representative trajectory was selected from the *exploration* data for each size of team. In the *perceptual search* condition operators' retained control of the robots' cameras but robots followed the representative trajectory except when individually paused by the operator.

2.3 Participants

45 paid participants were recruited from the University of Pittsburgh community. None had prior experience with robot control although most were frequent computer users.

2.4 Procedure

After collecting demographic data the participant read standard instructions on how to control robots via MrCS. In the following 20 minute training session, participants in the *fulltask* and *exploration* conditions practiced control operations. Participants in the *fulltask* and *perceptual search* conditions were encouraged to find and mark at least one victim in the training environment under the guidance of the experimenter. Participants then began three testing sessions (15 minute each) in which they performed the search task using 4, 8, and finally 12 robots. After each task, the participants were asked to complete the NASA-TLX workload survey.

3. RESULTS

Data were analyzed using a repeated measures ANOVA comparing *fulltask* performance with that of the subtasks. Where measures were inappropriate for some subtask, comparisons are pairwise rather than tripartite. Number of robots had a significant effect on every dependent measure collected (smallest effect was N switches in focus, $F_{2,54} = 12.6$, $p < .0001$). The task/subtask conditions also had nearly universal effects. The N robots x task/subtask interaction that would indicate differential scaling in N robots was less robust but found for the most crucial performance measures. Overall, *fulltask* participants were successful in searching the environment at all team sizes, finding as many as 12 victims on a trial. The average number of victims found was 4.8 using 4 robots, 7.06 for 8 robots, but dropping back to 4.73 when using 12 robots. Participants in the *perceptual search* condition, however, were significantly more successful, $F_{1,28} = 27.4$, $p < .0001$, finding 6.93, 8.2, and 8.33 victims respectively.

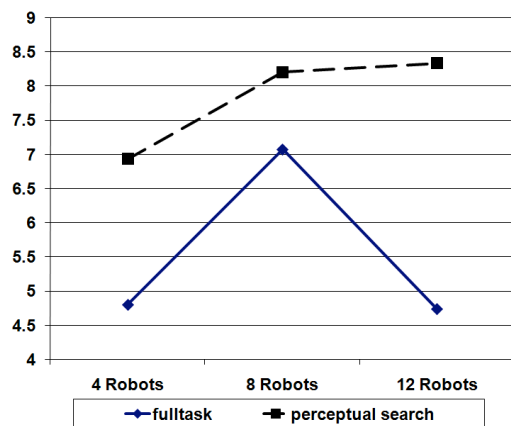


Figure 2. Victims found as a function of N robots

As shown in Figure 2 search performance in the *perceptual search* condition improved monotonically albeit shallowly while *fulltask*

performance peaked at 8 robots then declined resulting in a significant N robot x Task interaction, $F_{2,56} = 8.45$, $p = .001$.

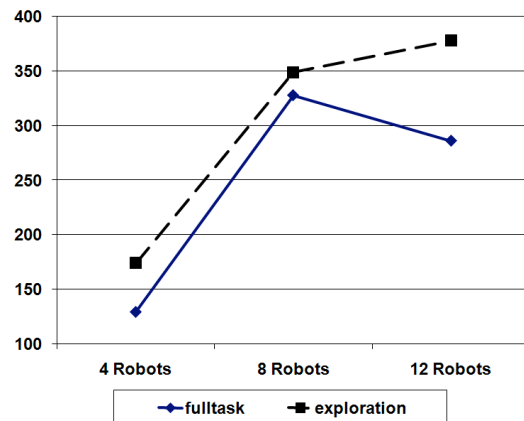


Figure 3. Area explored as a function of N robots

As figure 3 shows coverage was nearly identical at 4 and 8 robots but diverged at 12 robots, (between groups $F_{1,27} = 11.43$, $p = .002$, *fulltask* x N robots $F_{2,54} = 4.15$, $p = .021$) with *perceptual search* participants continuing to improve while those in the *fulltask* condition declined.

Of the process measures, switches in focus among robots (figure 4) most strongly differentiated the groups, with *perceptual search* participants switching far less than those involved in *exploration*. No difference ($F_{1,27} = 1.02$, $p = .32$) was found between the *fulltask* and *exploration* groups. The number of assigned missions (sets of waypoints) increased from 4 to 8 robots and declined slightly for 12. The *exploration* group who did not need to search for victims assigned significantly more missions, $F_{1,27} = 6.34$, $p = .018$. Average path length for missions increased monotonically in N robots with no difference between *fulltask* and *exploration* groups.

Robots that were completely neglected, $\chi^2_2 = 10.75$, $p = .005$, or operated only once, $\chi^2_2 = 31.3$, $p < .0001$, increased substantially more with N robots for *fulltask* than for *exploration* operators as determined by the Friedman test.

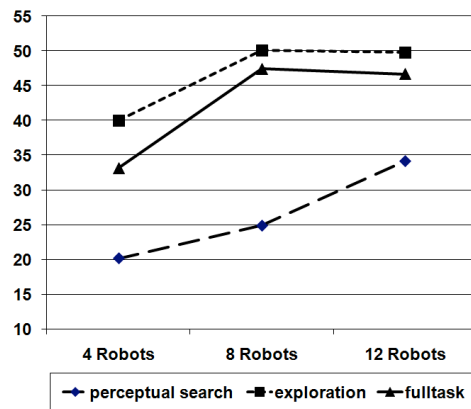


Figure 4. Switches in robot focus as a function of N robots

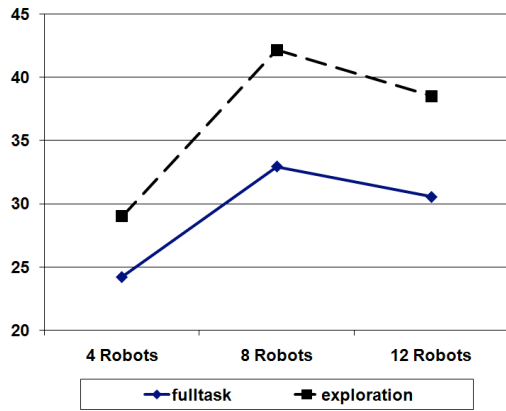


Figure 5. Assigned missions as a function of N robots

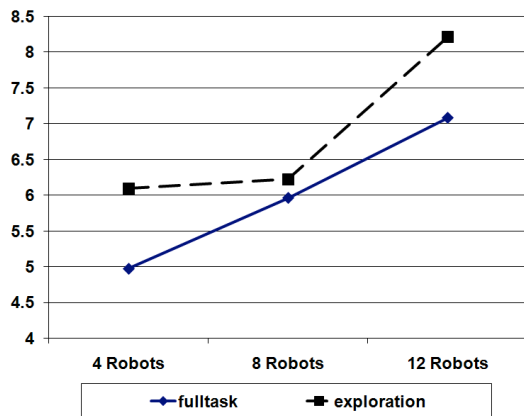


Figure 6. Average path length as a function of N robots

Workload increased monotonically in all groups but was substantially lower, $F_{1,27} = 21.17$, $p < .0001$, in the *perceptual search* condition.

4. DISCUSSION

The purpose of this experiment was to examine the tasks which go into controlling foraging robots in order to identify subtasks which might benefit most from increased automation. Rather than looking for cognitive tasks that go into controlling robot teams [8,

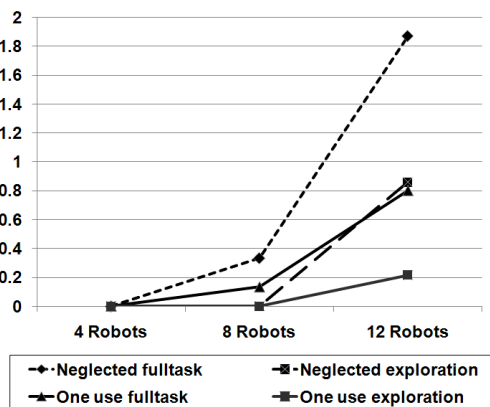


Figure 7. Robots neglected or used only once

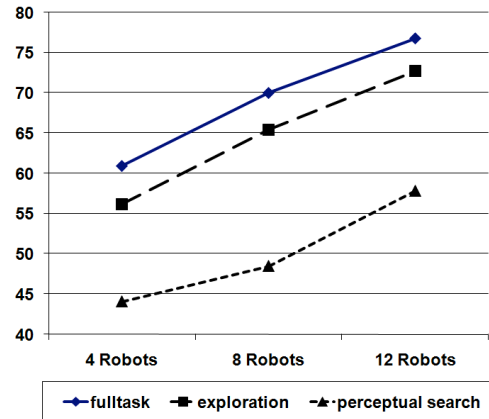


Figure 8. Workload as a function of N robots (NASA-tlx)

23] we have adopted a naïve model of the process by identifying two isolatable activities (each consisting of many cognitive tasks) that could be separated for an experiment.

Our data suggest that there is some cost of concurrence for performing the exploration and perceptual search tasks because both *victims found* (*perceptual search*) and *area explored* (*exploration*) were performed better alone than in the *full task* condition. In particular, for both measures performance at the subtasks was still increasing between 8 and 12 robots while both measures for *full task* performance decreased. Although there is no direct way to compare improvements in *perceptual search* and *exploration* examination of Figures 2 and 3 and comparison of η^2_p values (accounting for variance) for the N robots x subtask interaction shows that improvements in *perceptual search* ($\eta^2_p = .35$) contribute more to *victims found* than improvements in exploration ($\eta^2_p = .13$) do to *area explored*.

These results support automation of path planning and navigation over efforts to improve automation for target recognition and cueing providing the technical challenges are comparable. Because robots followed comparable trajectories in the *full task* and *perceptual search* conditions the larger number of victims found in the perceptual search condition represents a more thorough search and indicates that *full task* participants were missing victims they should have found. Since avoiding missed targets is crucial to many foraging tasks such as de-mining or search and rescue this advantage is more important than other performance gains such as widening the search area.

The finding of distinctly lower workload ratings for *perceptual search* is also encouraging in that it suggests that there remains reserve capacity to monitor additional robots. The top of the curve for *perceptual search* participants between 8 and 12 robots may have been artificially low due to a ceiling on the number of targets that could be observed given the limitations in trajectories. While suboptimal trajectories were needed to match the trajectories of other experimental conditions, a MrS using autonomous exploration could avoid the pauses and overlaps of our user generated trajectories.

The image of a MrS operator sitting in front of a bank of screens like a security guard monitoring surveillance cameras raises concerns over SA. Operators might be able to detect targets but be unable to relate these detections to the overall search. This was not the case in our data, however, because victims could only be scored by locating them on a map.

These results are encouraging for the future of large scale MrS in suggesting that systems could scale beyond the 12 robots we tested by using autonomous exploration. For largely independent tasks such as *foraging* this would allow very large teams by adding additional operators. While our results do not directly address more complex cooperative tasks such as assembly or construction our participants ability to maintain SA without direct control over exploration behaviors holds out hope that they may be able to successfully monitor and direct large teams of indirectly controlled cooperating robots as well.

5. ACKNOWLEDGMENTS

This research has been sponsored in part by AFOSR FA9550- 07-1-0039, AFOSR FA9620-01-1-0542, ONR Grant N00014-03-1-0580 and L3-Communications (4500257512)

6. REFERENCES

- [1] 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.
- [2] Miller, C. Modeling human workload limitations on multiple UAV control, *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, 526-527, 2004.
- [3] Cummings, M. and Guerlain, S. An interactive decision support tool for real-time in-flight replanning of autonomous vehicles, *AIAA Unmanned Unlimited Systems, Technologies, and Operations*, 2004.
- [4] 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.
- [5] J. Wang, M. and Lewis. Assessing coordination overhead in control of robot teams, *Proceedings of 2007 IEEE International Conference on Systems, Man, and Cybernetics*, 2645-2649, 2007.
- [6] J. Wang, H. Wang, and M. Lewis, M. Assessing Cooperation in Human Control of Heterogeneous Robots, *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction (HRI'08)*, ACM, 2008.
- [7] Y. Cao, A. Fukunaga, and A. Kahng. Cooperative mobile robotics: Antecedents and directions, *Autonomous Robots*, 4, 1-23, 1997.
- [8] D.R Olsen and S.B. Wood, Fan-out: measuring human control of multiple robots, in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, ACM Press: Vienna, Austria. p. 231-238.
- [9] B. Trouvain, C. Schlick, and M. Mevert, Comparison of a map- vs. camera-based user interface in a multi-robot navigation task, in *Proceedings of the 2003 International Conference on Robotics and Automation*. 2003. p. 3224-3231.
- [10] B. Trouvain and H. Wolf. 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.
- [11] C.M. Humphrey, C. Henk, G. Sewell, B. Williams, J. A. Adams. Evaluating a scaleable Multiple Robot Interface based on the USARSim Platform. 2006, Human-Machine Teaming Laboratory.
- [12] C. Wickens and J. Hollands. *Engineering Psychology and Human Performance* (3rd ed), NJ: Prentice-Hall Inc., 2000.
- [13] (UE 2) UnrealEngine2, <http://udn.epicgames.com/Two/rsrc/Two/KarmaReference/KarmaUserGuide.pdf>, accessed February 5, 2008.
- [14] S. Carpin, T. Stoyanov, Y. Nevatia, M. Lewis and J. Wang. Quantitative assessments of USARSim accuracy". *Proceedings of PerMIS 2006*
- [15] S. Carpin, J. Wang, M. Lewis, A. Birk and A. Jacoff. High fidelity tools for rescue robotics: Results and perspectives, *Robocup 2005 Symposium*, 2005.
- [16] Mathengine, *MathEngine Karma User Guide*, <http://udn.epicgames.com/Two/KarmaReference/KarmaUserGuide.pdf>, accessed May 3, 2005.
- [17] S. Carpin, M. Lewis, J. Wang, S. Balakirsky, C. Scrapper. (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*, 2006.
- [18] 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, 2005, 457-461
- [19] C. Pepper, S. Balakirsky, and C. Scrapper. Robot Simulation Physics Validation, *Proceedings of PerMIS'07*, 2007.
- [20] B. Taylor, S. Balakirsky, E. Messina and R. Quinn. Design and Validation of a Whegs Robot in USARSim, *Proceedings of PerMIS'07*, 2007.
- [21] 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.
- [22] S. Balakirsky, S. Carpin, A. Kleiner, M. Lewis, A. Visser, J. Wang and V. Zipara. Toward heterogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue, *Journal of Field Robotics*, 2007
- [23] J. W. Crandall and M. L. Cummings. Developing Performance Metrics for the Supervisory Control of Multiple Robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2007.