Effects of Spatial Ability on Multi-robot Control Tasks

Shih-Yi Chien, Huadong Wang, Michael Lewis
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260 U.S.A.
shc56@pitt.edu, huw16@pitt.edu, ml@sis.pitt.edu

Working with large teams of robots is a very complex and demanding task for any operator and individual differences in spatial ability could significantly affect that performance. In the present study, we examine data from two earlier experiments to investigate the effects of ability for perspective-taking on performance at an urban search and rescue (USAR) task using a realistic simulation and alternate displays. We evaluated the participants' spatial ability using a standard measure of spatial orientation and examined the divergence of performance in accuracy and speed in locating victims, and perceived workload. Our findings show operators with higher spatial ability experienced less workload and marked victims more precisely. An interaction was found for the experimental image queue display for which participants with low spatial ability improved significantly in their accuracy in marking victims over the traditional streaming video display.

INTRODUCTION

Robots are increasingly being used in a wide variety of civilian and military applications. Although current systems often require many humans to control a single robot, future applications ranging from search and rescue to mining or farming are likely to depend on many robots controlled by a single operator. Demands on the operator are likely to be extremely high because of the need to switch attention between robots and develop an understanding of the environment from many different perspectives.

In widely studied urban search and rescue (USAR) foraging tasks, for example, operators control unmanned ground vehicles (UGVs) to search an environment and locate victims on a map using video from UGV cameras or images sampled from that video. By current practice, robots equipped with laser range finders use SLAM (simultaneous localization and mapping) to build a map based on laser scans and position the robot on the map relative to those scans. At the beginning the map is entirely unknown but as exploration continues features such as walls, objects, and open spaces appear on the map. The laser map, however, cannot provide sufficient resolution to perform complex perceptual tasks such as victim identification. The images or video can only provide the operator partial information about the environment because they are limited by the camera's field of view, the robot's orientation, and its trajectory through the environment. To coordinate multiple robots to search an environment in this way requires operators to continually shift their attention from robot to robot changing their perspectives to maintain situation awareness (SA) and locate victims as they appear within a camera's field of view. In order to locate the victims and clarify the relationship between a camera view and the robot's location on the map, operators have to expand their perceptive ability to maintain global and local mental models of the environment. Operator may encounter problems in integrating information because of the reference across different sources is not well presented, which has shown in several studies (Olmos, Wickens, & Chudy, 2000; Thomas & Wickens, 2001, Chen & Clark, 2008)

Performance on USAR foraging tasks could be affected by differences in orientation between the map and camera views which might require mental rotation. Several researches have shown the track-up map, ego-referenced with rotating viewpoints, is better for local navigation and the north-up map, world referenced with fixed viewpoint, is better for global awareness. (Aretz, 1991; Casner, 2005; Darken & Cevik, 1999; Lohrenz, Gendron, Edwards, Myrick, & Trenchard, 2004; Wang, 2004; Werner, 2002; Chen & Clark, 2008). Individuals with high spatial ability areable to adapt to using either type of map, track-up or north-up map, with less effort than individuals with low spatial ability (Darken & Cevik, 2002).

Werner et. al (1997) conclude that individual diversity in spatial cognition, which includes acqusition, organization, use, and revision of knowledge about the spatial environment, could affect the results directly. Individuals with higher spatial ability have been shown to perform significantly better at navigation tasks than those with lower spatial ability (Cassenti et. al, 2002). Another recent study, Chen (2010), found individuals with better spatial ability performed significantly better in atarget search task under the night conditions than those with a low sense of direction. Baldwin and Reagan (2009) note that "individuals with poor sense of direction relied more heavily on verbal rather than visuospatial working memory resources, and, conversely, individuals with good sense of direction exhibited more route-learning disruption from a tapping task, suggesting a greater reliance on visuospatial working memory resources."

Hegarty & Waller (2004), found most of the current spatial tests cannot examine either object rotation ability or perspective-taking ability purely and the variance in strategies could affect the performance directly. According to their findings, the revised version of the object perspective test (Kozhevnikov & Hegarty, 2001) is both reliable and a largely strategy free measure of perspective taking ability. After

judging the requirements of the USAR task and comparing various spatial ability tests, we adopted the newer version of the object perspective test (Fig. 1) for measuring participants' spatial ability. In the Object Perspective Test seven objects are drawn on the top of a sheet and the bottom half page shows a circle marked with standing point and facing direction. Participants are asked to imagine being at the position of one object and facing another object, and then asked to indicate (draw) the direction to the target object. Participants are prevented from physically rotating their body or the booklet and have to identify the target object and then mark it on the circle (as the dotted line shown on the Figure 1), completing twelve questions within five minutes. The initial score on the spatial orientation test is an error score, a higher value represents less spatial ability. Scores are then linearly transformed (by subtracting the average error score from 180°) so that higher scores correspond to better performance. This transformed score in which higher scores correspond to higher spatial ability is used in the analyses we report.
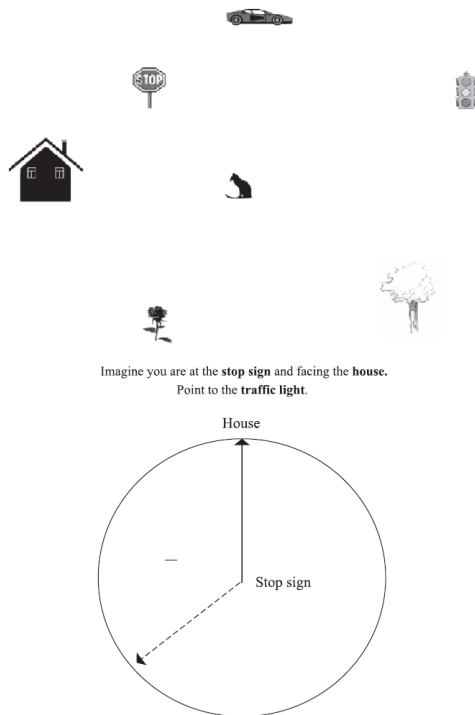

Imagine you are at the **stop sign** and facing the **house**.
Point to the **traffic light**.

House

Stop sign

Figure 1. Example of the Spatial Orientation Test (Hegarty & Waller, 2004)

METHODOLOGY

USARSim and MrCS

The reported experiments were conducted using the USARSim robotic simulation performing Urban Search and Rescue (USAR) foraging 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. USARSim supports HRI by accurately rendering user interface elements (particularly camera video), accurately representing robot automation and behavior, and accurately representing the remote environment. Validation studies showing close agreement in behavior between USARSim models and real robots being modeled are reported in (Carpin, Lewis, Wang, Balakirsky, & Scrapper, 2006; Lewis, Hughes, Wang, Koes, & Carpin, 2005; Pepper, Balakirsky, & Scrapper, 2007; Taylor, Balakirsky, Messina, & Quinn, 2007; Zaratti, Fratarcangeli, & Iocchi, 2006).

MrCS (Multi-robot Control System), a multirobot communications and control infrastructure with accompanying user interface developed for experiments in multirobot control and RoboCup competition (Balakirsky et al., 2007) 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. Figure 2 shows the MrCS user interface configured for 12 robots. Thumbnails of robot camera feeds are shown on the left, a video feed of interest in the top, middle. A GUI element in the top right allows teleoperation and camera pan and tilt. The bottom right shows the current map and allows operators to mark victims.
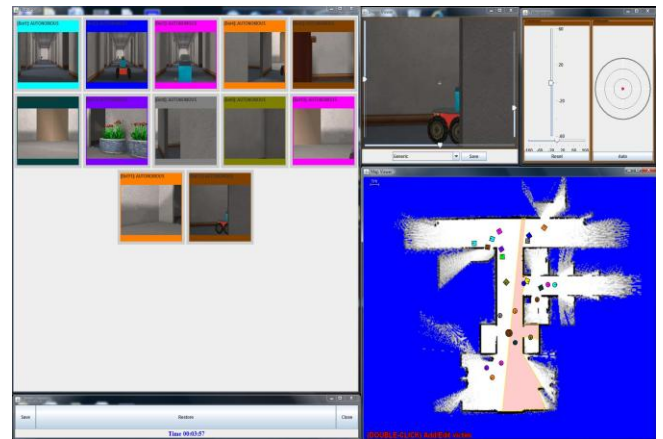

Figure 2. The MrCS user interface with 12 robots for Assigned Robots groups

Autonomous path planning was performed by a deterministic roadmap planner (Latombe, 1991) developed using the Carnegie Mellon Robot Navigation Toolkit (CARMEN) (Montemerlo, Roy, & Thrun, 2003) in the first experiment and by a Segment Voronoi Diagram (SVD) path planner providing smoother more "human like" paths in the second experiment. This distinction is not crucial; however, as Chien, Wang, & Lewis (2010) have shown that performance of the USAR task using the first planner was no worse than that of operators following human generated paths.

Both experiments used a large USAR environment previously used in the 2006 RoboCup Rescue Virtual Robots competition (Balakirsky et al., 2007). The environment, an office-like map, was a maze with a hall with many rooms and obstacles, such as chairs, desks, cabinets, and bricks. Victims were evenly distributed within the environment.

Once a victim entered into a camera's field of view and was potentially detectable, a series of actions need to be performed to develop sufficient situation awareness (SA) to perform the victim marking task. The operator first needs to identify the robot and regain SA of the robot by matching the

robot's color and numerical label on the map. Next the operator has to determine the orientation of the robot and match landmarks between camera and map views. In order to clarify the relationship between the robot and victim, the operator may choose to teleoperate the selected robot to help locate it on the map and determine its orientation through observing the direction of movement. The operator must then locate the victim on the map corresponding to the camera view.

EXPERIMENT I

The experiment followed a between groups design with 24 robots. Each task was performed by a team of 2 participants. In the *Assigned Robot* conditions participants were assigned 12 UVs each. In the *shared pool* conditions participants shared control of the 24 UVs and viewed the same screens (Figure 3). Robots generated their own waypoints using distributed path planning and participants were able to teleoperate the in-focus robot to extricate it when it became stuck.
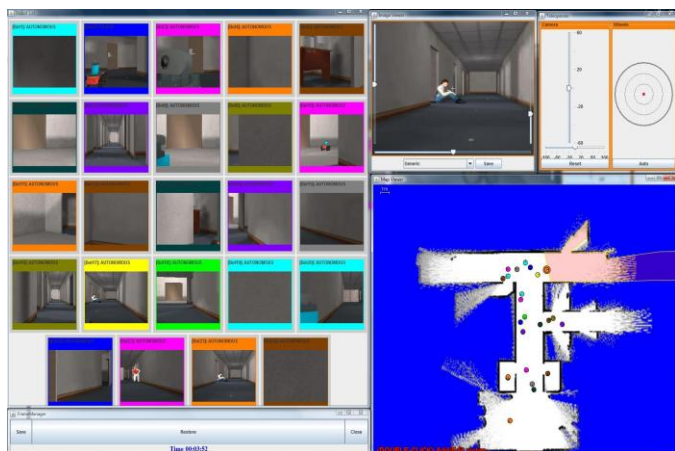

Figure 3. The MrCS user interface with 24 robots for Shared Pool groups

60 paid participants (30 teams), age 23 to 35 years, and balanced among conditions for gender were recruited from the University of Pittsburgh community. None had prior experience with robot control or were expert video game players although most were frequent computer users.

Participants completed a perspective taking test, then read standard instructions followed by a 30 minute training session. Participants then began the experimental session (25 minutes) in which they performed the search task controlling 24 robots in teams. After the task, the participants were asked to complete workload ratings on the NASA Task Load Index (NASA–TLX; Hart & Staveland, 1988).

Data were analyzed using a one-way ANOVA comparing the Assigned Robot condition with the Shared Pool condition. However, no significant difference was found for either victims found or region explored (Table 1). Accuracy in marking victims on the laser generated map (RMS errors) also showed no advantage for any of the conditions.

Differences were found on other measures related to the participants' monitoring and operational behavior. The ANOVA for "Selected to Mark", the time from when the operator selected the robot to successfully marking the victim, favored the assigned robots condition. The Assigned Robot condition also led to fewer mistakes as measured by deletion of redundant or imprecise markings.

A comparison of SOT results between the two conditions found no significant difference (F_1, 28 = 1.034, p = .318) indicating the samples were unbiased with respect to spatial ability. There was a significant negative correlation between scores on the SOT and the Delete measure (r = -.372, p = .047). The negative correlation between scores of SOT and RMS errors were marginally significant (r = -.318, p = .093). In addition, participants' SOT score was found to be an accurate predictor of their perceived workload, the score on SOT and workload correlated with each other (r = -.364, p = .005).

Table 1. One-way ANOVA results for Experiment I

Variables	Assigned Robots $\bar{x}$	Shared Pool $\bar{x}$	$F_{1,28}$	P
Victim Found	15.86	17.07	.951	.338
Region Explored	792.69	767.77	.750	.394
RMS Errors	.0200	.0220	1.803	.191
Display to Mark Time	31.35	40.56	4.866	.036
Delete	7.29	10.87	4.672	.040

EXPERIMENT II

This experiment followed a two condition repeated measures design comparing the conventional MrCS displays (streaming video) with MrCS augmented by the experimental image queue display counterbalancing conditions. Automated path planning to improve search performance and autonomous exploration was used in both conditions. The operators performed a supervisory control task in which the robots navigated autonomously with the operator allowed to override by directing them through new waypoints. When necessary, participants were able to teleoperate the in-focus robot to extricate it when it became stuck.
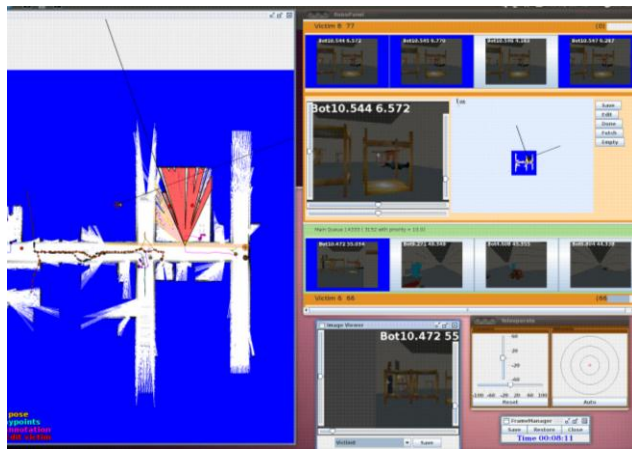
Figure 4. The MrCS user interface in the image queue condition

30 paid participants were recruited from the University of Pittsburgh community balanced among conditions for gender. None had prior experience with robot control although most were frequent computer users.

After completing a perspective taking test participants read standard instructions on how to control robots via MrCS. In the following training session, participants practiced control operations for both streaming video and image queue condition for 10 min each. After the training session, participants began the two 15 minute sessions in which they performed the search task controlling 12 robots using either the streaming video or image queue display. At the conclusion of each session, participants were asked to complete the NASA-TLX workload survey.

Data were analyzed using a repeated measures ANOVA comparing streaming video with the image queue condition. In addition to the performance measures reported in Experiment 1, victim markings were compared to ground truth. A mark made further than 2 meters away from any victim or multiple marks for one victim were counted as *false positives*. Victims that were missed, but present in the video feed, and not marked were counted as *false negatives*. The image queue condition was found to reduce both false positive and false negative errors as compared with the traditional synchronous display. However, victims found and the area explored were not significantly different (table 2).

Table 2. Repeated ANOVA results for Experiment II

Variables	Streaming Video $\bar{x}$	Image Queue $\bar{x}$	$F_{1,28}$	P
Victim Found	9.01	8.51	.733	.387
Region Explored	1058.08	1017.77	2.147	.154
False Positive	2.286	1.214	13.032	.001
False Negative	9.34	7.48	5.526	.026

An examination of the correlation between the SOT scores and performance measures, shows a significant negative correlation between scores on the spatial orientation test (SOT) and false positive errors in the streaming video conditions (r = -.408, p = .031). This correlation is only marginally significant in the image queue condition (r = -.345, p = .072).

CONCLUSION

In this paper we investigated how individual differences in spatial ability affect foraging performance in multirobot USAR control. One dominant finding was that participants with higher spatial ability experienced lower workload; while participants with low spatial ability experienced much higher workload. The task of marking victims on a SLAM generated map is a complex process initiated by detection of a victim in streaming video or sampled imagery. The operator first has to identify where the view was taken on the map and then transform the camera view into the map to determine the orientation and estimate the distance from the view to the victim. During this sequential process, operators may need to match the surrounding environment, such as identifying landmarks or encountered obstacles on the laser map, to get better situation awareness. Next, if other victim-marks are found near the planned marking location, the operator has to decide whether the victim had been previously marked. The results show that operators with higher perspective-taking ability were much more precise in marking victims than participants with lower perspective ability. As well as making significantly more incorrect marks, operators with lower spatial ability were forced to revise or delete their previous marks much more often.

The image queue display in Experiment 2 offers some promise for assisting low spatial ability operators. In this experiment the accuracy in marking victims improved for both low and high spatial ability participants when using the image queue display. Our results found operators with poor spatial ability did significantly worse in locating victims on the map when working with streaming video. Most of the errors were duplicated marks (multiple marks for single victim) or reversed rotation (confused left/right or front/back), the sorts of errors you would expect from persons with difficulty performing mental rotations. This problem is exacerbated for multirobot control where the operator must repeatedly process these transformations when switching between robots.

According to participants' feedback, although they felt streaming video to bemore intuitive and, image queue condition to be a relatively more complex display that feltunnatural, they, experienced higher pressure when moving from the practice session to controlling 24 robots in the streaming video condition. This was not reported in the image queue display since participantshad less interaction with robots.

Integration of individual differences in capabilities into the comparison of interfaces and methods for multirobot control systems is essential and an often ignored component. Given that production interfaces will be used by professionals, who require costly training, we should aim at designing

interfaces that minimize requirements for such training. Furthermore, we can also provide an objective and sound basis for recruitment of said professionals. For example, given an USAR system with multiple video streams it is highly advisable to select for individuals with good spatial ability. In future work we aim toinclud additional measures of cognitive ability and develop a catalogue of relevant tests for multirobot control systems. Ultimately, such tests may also be transformed into an online monitoring of performance and adaptation of the interface to the user's capabilities.

ACKNOWLEDGMENT

REFERENCES

Aretz, A. J. (1991). The Design of Electronic Map Displays. *Human Factors*, 33, pp. 85-101.

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

Baldwin, C. & Reagan, I. (2009). Individual differences in route-learning strategy and associated working memory resources. *Human Factors*, vol. 51, pp. 368-377.

Carpin, S., Lewis, M., Wang, J., Balakirsky, S., & Scrapper, C. (2006). 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*, pp. 1–12.

Casner, S. M. (2005). The Effect of GPS and Moving Map Displays on Navigational Awareness While Flying Under VFR. *International Journal of Applied Aviation Studies*, 5 (1), pp. 153-165.

Cassenti, D., Kelley, T. Swoboda, J. & Patton, D. (2009). The effects of communication style on robot navigation performance. *Human Factors & Ergonomics Soc Meeting*, San Antonio, TX.

Chen, J. Y. C. & Clark, B. R. (2008). Unmanned Aerial Vehicle (UAV)-Unmanned Ground Vehicle Teaming: UAV Guided Navigation. *Army Research Laboratory Report.*

Chen, J. Y. C. (2010). Effects of Operator Spatial Ability on UAV-Guided Ground Navigation. *IEEE.*

Chien, S. Y., Wang, H, & Lewis, M. (2010). Human vs. algorithmic path planning for search and rescue by robot teams, *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society* (HFES'10), 379-383.

Darken, R. P. & Cevik, H. (1999). Map Usage in Virtual Environments: Orientation Issues. *Proceedings of the IEEE Virtual Reality Conference*. pp. 133-140.

Hart, S., & Staveland, L. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.). *Human mental workload*, Amsterdam, The Netherlands: Elsevier, pp. 39-183.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32, pp. 175–191.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object-manipulation spatial ability and spatial orientation ability. *Memory and Cognition*, 29, pp. 745–756.

Latombe, J. C. (1991). *Robot motion planning*. New York: Springer.

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

Lohrenz, M. C., Gendron, M. L., Edwards, S. S., Myrick, S. A. & Trenchard, M. E. (2004). Demonstration of a Moving-Map System for Improved Precise Lane Navigation of Amphibious Vehicles and Landing Craft. *Sea Technology* pp. 39-43.

Montemerlo, M., Roy, N., & Thrun, S. (2003). Perspectives on standardization in mobile robot programming: the carnegie mellon navigation (carmen) toolkit. *In Proc. of IROS*, 3, pp. 2436- 2441.

Olmos, O & Wickens, C. D. (2000). Chudy, A. Tactical Displays for Combat Awareness: An Examination of Dimensionality and Frame of Reference Concepts and the Application of Cognitive Engineering. *The International Journal of Aviation Psychology 2000*, 10 (3), pp. 247-271.

Pepper, C., Balakirsky, S., & Scrapper, C. (2007). Robot Simulation Physics Validation, *Proceedings of PerMIS'07.*

Taylor, B., Balakirsky, S., Messina, E., & Quinn, R. (2007). Design and Validation of a Whegs Robot in USARSim. *Proceedings of PerMIS'07*.

Thomas, L. C. & Wickens, C. D. (2001). Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. *Proceedings of Human Factors and Ergonomics Society 45th Annual Meeting*, pp. 336-340.

Wang, W. (2004). Human Navigation Performance Using 6 DOF Dynamic Viewpoint Tethering in Virtual Environments. *Unpublished Ph.D. dissertation*, University of Toronto, Canada.

Werner, S. (2002). Cognitive Reference Systems and Their Role in Designing Spatial Information Displays. *Künstliche Intelligenz*. 16 (4), pp. 10-13.

Werner, S., Krieg-Brückner, B., Mallot, H., Schweizer, K. & Freksa, C. (1997). Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation. *Informatik aktuell*, pp. 41-50.

Zaratti, M., Fratarcangeli, M., & Iocchi, L. (2006). A 3D Simulator of Multiple Legged Robots based on USARSim. *Robocup 2006: Robot Soccer World Cup X. Springer, LNAI*. pp. 13–24.