

in distributed networking. As a result, the pressing research issues are how to facilitate the larger human-robot system to reach the overarching goals of the rescue enterprise.

For more information about the R4 program and publications, please visit [www.crasar.org](http://www.crasar.org).

**ABOUT THE AUTHOR** *Robin Roberson Murphy received a B.M.E. in mechanical engineering, a M.S. and Ph.D in computer science from Georgia Tech, where she was a Rockwell International Doctoral Fellow. She is a professor in the Computer Science and Engineering Department at the University of South Florida with a joint appointment in Cognitive and Neural Sciences in the Department of Psychology. She is also director of the Center for Robot-Assisted Search and Rescue and the NSF Safety Security Rescue Research Center. She was awarded the NIUSR Eagle Award for her work using rescue robots at the World Trade Center disaster.*

© ACM 1072-5220/05/0300 \$5.00

**THE COMPETITION ENVIRONMENT.** The National Institute of Standards and Technology has developed a reference test arena for robots in USAR, urban search and rescue [1, 2]. Currently, there are three different arenas that differ by difficulty. We term them as the yellow, orange, and red arenas. Figure 1 shows an overhead view of the red arena used at the Robocup 2004 competition [3].

The arenas were of varying degrees of difficulty. The yellow portion represented a slightly damaged office building; the orange arena contained multiple stories, covered areas, more rubble, and negative obstacles (holes). The red portion was all rubble with multiple levels that had unstable access. Victims (simulated by mannequins) are more easily located in the yellow and orange arenas.

In the red arenas, some victims could not be located visually and teams had to use thermal, CO2 sensors, or sound. The image in Figure 2 is an example of a victim and rubble in the red arena.

Points were awarded for the number of victims found in a specified length of time weighted by the difficulty of the portion of the arena in which they were found, for the accuracy of the victim location map, and the description the robot operator constructs. Penalties were assessed for causing damage to any of the victims or to the arena while searching. For example, contact with any wall in the arena is a penalty as bumping into the structure could cause a further collapse in an actual Urban Search and Rescue (USAR) mission.

**History of HRI Data Collection.** In 2000, Robot Rescue was added to the AAI Robot Competition and Exhibition. In its second year, the event was co-sponsored by AAI and RoboCup, leading to two separate events starting in 2002. The competitions share their rules and disaster-scene simulation. The expanding competitions led to the possibility of studying human-robot interaction across a large number of robot systems in a longitudinal manner.

We began to collect data at the Robot Rescue competitions in 2002 and have continued our collection and analysis efforts for the past three years. Competitions provide a structured task with fixed time limits. The nature of competitions introduces performance-

## Using Competitions to Study Human-Robot Interaction in Urban Search and Rescue

**Jill L. Drury**  
[jldrury@mitre.org](mailto:jldrury@mitre.org)

**Holly A. Yanco**  
[holly@cs.uml.edu](mailto:holly@cs.uml.edu)

**Jean Scholtz**  
[jean.scholtz@nist.gov](mailto:jean.scholtz@nist.gov)



Figure 1: The red test arena used at Robocup 2004

study, we moved to hand-held cameras for taping robots in the arenas. In the last RoboCup competition, we used ultra-wideband tracking devices to automatically track the paths of the robots.

We have also refined our analysis over the years, largely due to the availability of better data. In our second year, with better tapes of the robot's progress in the arenas, we were able to identify critical incidents and their potential cause(s), allowing us to identify effective methods for acquiring situation awareness.

***What Can We Learn about HRI from These Studies?*** As there are no constraints on the hardware or software used in these competitions, we are able to see a number of different types of user interfaces and interactions, including a "direct manipulation" interaction and a virtual-reality interface.

Competitors perform the same tasks using different systems, so we can see how different user-interaction approaches can affect performance. In competitions, performance can be measured objectively, such as via the number of victims found and the number of penalties for bumping arena walls or victims. USAR competitions form a limitation on how well an interface can support users, in the sense that the people operating robotic systems in USAR competitions are the extremely technically-savvy developers of those systems.

Extensive data collection is possible during USAR competitions. We videotape the robot(s) in the arena and over the shoulders of robot operators. We capture operators' interactions with the interfaces via dynamic screen-capture software that allows us to play back what operators see and do with the interface displays. We use the maps of the robots' paths to assess coverage and penalties. Further, we conduct brief post-run interviews.

The detailed data collection efforts allow us to characterize the operators' use of the interfaces. For example, in past competitions we have determined the percentage of time users spent in different activities, such as in manipulating the interface versus navigating the robot; also, we examined what parts of the interfaces were most heavily used, and for what purposes.

Figure 2: A victim in the red arena



related stress; competitors are likely to take their runs seriously when being scored.

Over the three years of data collection, we have refined our collection methods. For example, in the first year, we relied upon cameras mounted above the competition arena to tape the progress of robots. However, these overhead views did not provide the best angles and often robots would maneuver into a location not covered by any of the mounted cameras. In our next

Studies conducted at USAR competitions also help us learn about operators' situation awareness (SA) strategies. The most generally accepted definition of SA is Endsley's: the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future [4]. Having good SA is especially important in remote robot operations, meaning the robot operator cannot see the robot from their control terminal; this is the case for USAR competitions. We are able to observe what information operators extracted from the interface to attain or reacquire SA, such as the example in which an operator angled a camera downward to see the robot wheels' in relationship to obstacles and voids.

Even in cases where operators had good SA, they still occasionally damage the arena or their robots, or take actions that would harm victims (if they were real people rather than mannequins). Thus, USAR competitions afford us the opportunity to perform critical-incident analyses. For example, we observed more bumping incidents happening to the rear of robots for those systems that did not have rear-facing cameras. We also investigate whether operators are aware that critical incidents had occurred, for example via post-run interviews. Often the information operators received via the interface is not sufficient to alert them to the fact that a critical incident had occurred; we look at what information and/or presentation mechanism would have been necessary to provide the missing cues.

It is also worthwhile noting what we could *not* learn via studies of USAR competitions. Because we could not do anything that would jeopardize a team's competitive chances, we could not ask operators to "think aloud" [6] or interrupt them with questions during a competition run. Thus, competitions afford limited insight into operators' mental models or ways of thinking about the interfaces. Further, we could not use anything from the class of "explicit performance" techniques for SA measurement such as the Situation Awareness Global Assessment Technique [5] because these techniques involve short suspensions of the task during which operators answer questions. Further, we could not ask that the teams employ USAR personnel for competition runs. Thus, we could not normally see how well the interfaces could be used by their intended end users. In several cases, however, we were able to observe a domain expert use several of the robotic systems in the NIST test arena.



**ABOUT THE AUTHORS** *Jill Drury earned an Sc.D. in Computer Science at the Human-Computer Interaction Lab of the University of Massachusetts Lowell. Besides human-robot interaction, her research interests are awareness mechanisms for collaborative systems and decision-support for teams performing safety-critical missions. She is an Associate Department Head at The MITRE Corporation and an adjunct faculty member at the University of Massachusetts Lowell.*



*Dr. Holly Yanco is an Assistant Professor in the Computer Science Department at the University of Massachusetts Lowell. Her research interests include human-robot interaction, adjustable autonomy, assistive technology, and multiple robot teams. She graduated from MIT with her Ph.D. in Computer Science in 2000.*



*Dr. Jean Scholtz is a computer scientist at the National Institute of Standards and Technology. Her research interests are evaluation methodologies and metrics for interaction with intelligent systems. Her work with human-robot interaction includes on and off-road driving, urban search and rescue, explosive ordnance disposal, and assembly in space.*

- REFERENCES**
1. Jacoff, A., Messina, E., and Evan, J. A reference test course for autonomous mobile robots. *Proceedings of the SPIE-Aerosense Conference*, Orlando, FL, April (2001).
  2. Jacoff, A., Messina, E., and Evans, J. A standard test course for urban search and rescue robots. *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, August (2000).
  3. Robocup 2004 Search and Rescue Competition. <http://www.rescuesystem.org/robocuprescue/> accessed Sept. 9, 2004.
  4. Endsley, M. R. "Design and evaluation for situation awareness enhancement," Proc Human Factors Society 32nd Annual Meeting, Santa Monica, CA (1988).
  5. Endsley, M. R. (1987). SAGAT: a methodology for the measurement of situation awareness (NOR DOC 87-83). Hawthorne, CA: Northrup.
  6. Ericsson, K. A. and H. A. Simon (1980). "Verbal reports as data." *Psychological Review* 87: 215 - 251.