Evaluation of Human-Robot Interaction Awareness in Search and Rescue

Jean Scholtz and Jeff Young
NIST
Gaithersburg, MD, USA
{jean.scholtz; jeff.young}@nist.gov

Jill L. Drury
The MITRE Corporation
Bedford, MA, USA
jldrury@mitre.org

Holly A.Yanco
University of Massachusetts Lowell
Lowell, MA
holly@cs.uml.edu

Abstract—**In this paper we report on the analysis of critical incidents during a robot urban search and rescue competition where critical incidents are defined as a situation where the robot could potentially cause damage to itself, the victim, or the environment. We look at the features present in the human-robot interface that contributed to success in different tasks needed in search and rescue and present guidelines for human-robot interaction design.**

Keywords-human-robot interaction; urban search and rescue; human-robot awareness

I. INTRODUCTION

The use of robots in urban search and rescue (USAR) is a challenging area for researchers in robotics and human-robot interaction (HRI). Robots used in search and rescue need mobility and robustness. The environments in which they will be used are harsh with many unknowns. These robots must be able to serve as members of the USAR teams, sending back information to rescue workers about victims, the extent of damages, and structural integrity [1]. Operators of USAR robots will be working long shifts in stressful conditions. Fortunately, most USAR teams are infrequently called to service. This, however, means that human-robot interaction must support infrequent use. The user interactions in USAR robots need to be designed with these requirements in mind.

Robotics research is making progress in producing autonomous robots. A key to autonomy is perception capabilities. Robots must be able to recognize objects and to make decisions based on what an object is. For example, an off-road driving vehicle can recognize trees and plan a route to navigate around those trees. Current autonomous off-road driving performance is quite reasonable [2]. The objects that must be perceived are static and relatively few in nature. This is not true in the USAR domain. After fires or explosions, objects are difficult even for humans to recognize. Planning paths for navigation is not just locating trees or rocks but picking a path through or over a rubble-strewn area.

Completely autonomous robots for USAR are definitely not feasible in the near future. Operators must work as teammates with the USAR robots, with all parties contributing according to their skills and capabilities.

It is difficult to study actual USAR events. Casper and Murphy [3] documented efforts to use robots during 9/11 rescue efforts. Burke et al. [1] have conduced field studies during search and rescue training. Few robotics and HRI researchers are able to participate in such events. Moreover, given the nature of these events, data collection is difficult, if not impossible.

The National Institute of Standards and Technology has developed a physical test arena [4,5] that researchers can use to test the capabilities of their USAR robots. A number of international USAR competitions have used the NIST arena. We used these competitions to study a number of human-robot interfaces to determine what information helps the operator successfully navigate the course and locate victims. Although we have no control over the user interfaces, these competitions allow us to see a wide variety of designs and to determine how effective different features are in supporting USAR work. The competition simulates the stressful environment of a real disaster site by limiting the time periods that robots can be in the arena. Since it is a competition, additional pressure is added by the desire to do well. However, the safety issues that would be present in a real disaster are not present in the competition setting.

II. MEASURING EFFECTIVENESS

Olsen and Goodrich [6] offer six metrics to use in evaluating human-robot interaction: task effectiveness, neglect tolerance, robot attention demand, free time, fan out, and interaction effort. A brief description of each metric is provided in table 1.

In a study of a 2002 USAR competition, Yanco et al. [7] computed arena coverage, interaction effort, and amount of

time operators spent giving directions to the robots. These metrics are useful in helping us measure progress in human-robot interaction. However, it is difficult to extract information for designing more effective human-robot interactions from performance metrics.

TABLE I. METRICS FOR HRI, FROM [6]

Metric	Definition
Task effectiveness	How well a human-robot team accomplishes a task.
Neglect tolerance	How the robot's current task effectiveness declines over time when the operator is not attending to the robot.
Robot attention demand	The fraction of total task time a user must attend to a given robot.
Free time	The fraction of the task time the user does not need to pay attention to the robot.
Fan out	An estimate of the number of robots that a user can effectively operate at once.
Interaction effort	The time to interact plus the cognitive demands of interaction.

TABLE II. HRI AWARENESS, FROM [8]

HRI Awareness Type	Definition
Human-robot	The understanding that the humans have of the locations, identities, activities, status and surrounding of the robots.
Human-human	The understanding that humans have of the locations, identities and activities of their fellow human collaborators.
Robot-human	The robots' knowledge of the humans' commands and any human constraints.
Robot-robot	The knowledge that the robots have of the activities and plans of other robots.
Humans' overall mission	The humans' understanding of the overall goals of the joint human-robot activities and the progress towards the goal.

USAR competitions currently focus on the operator–robot interaction. The main task of the operator and robot is to locate as many victims as possible in the entire arena. Given this goal, there are a number of tasks that the operator has to do: navigate through the space, locate and identify victims, and deal with obstacles encountered. To accomplish these tasks, the operator-robot team needs a shared situational awareness. Drury et al. [8] developed a framework for HRI awareness. Table 2 shows the five areas of HRI awareness and their corresponding definitions. The framework is based on multiple robots and multiple humans working as a team.

We will use this framework to identify human-robot interaction features that contribute to maintaining sufficient awareness and to identify features that are lacking that could potentially contribute.

III. ROBOCUP2003 USAR COMPETITION

Thirteen teams competed in the USAR competition during Robocup 2003 in Padova, Italy. Twelve of these teams participated in our HRI study. Three arenas modeled on the NIST arena were constructed. The arenas were denoted as yellow, orange, and red and were of varying degrees of difficulty. Victims in the arenas are dummies, some of which have tape recorders so they can be identified using audio sensors. Other victims have heat for thermal identification. The yellow arena resembled an office environment that had suffered minor damage. Rubble consisted mainly of overturned furniture, papers, and venetian blinds. Victims could be located visually for the most part. The orange arena was multilevel and had more rubble, such as loose bricks, mesh, and wire. Victims were hidden so that only hands or feet might be visible. The red arena was multilevel with large holes in the upper level that robots had to avoid falling through. The floor was strewn with loose bricks, gravel, rubber tubing, and wire. Fig. 1 and 2 show the difficulty of the arena environment in the orange and red arenas.

Figure 1: The Orange Arena

Figure 2: The Red Arena

The teams had three chances to navigate through the arenas to locate victims. They were allocated 20 minutes for each of these runs. The six top scoring teams were allowed to move to the semifinals. These teams were given two runs each for the semifinals. The top four teams in the semifinals moved to the finals where again they were given two runs.

A. Data Collection

We focused our analysis on the top three teams. We selected only the runs they made during the semifinals and the finals for the analysis we present here. The competition was run for five consecutive days and some of the teams changed their robots and their human-robot interaction capabilities during the week. There were no changes for these teams between the semifinals and the finals so we are analyzing the same human-robot interaction capabilities in both sets of runs. As the teams were involved in a competition, we had to make our data collection as unobtrusive as possible. We were not able to collect think-aloud protocols [9] from the operators as they navigated the arena. We talked to operators after their runs, but time was limited as they had to vacate the area to get ready for the next team to set up. We were not allowed to put additional software on the teams' computers, so we used video equipment to capture the graphical user interface and any additional monitors or computer displays that were being used. Some teams developed maps that were kept on a different laptop. Often teams used a different display for the video being sent back from the robot. We also collected video of the robots moving in the arenas. This data, along with maps drawn by the competition judges of the paths the robots took, forms "ground truth" data. That is, we can tell exactly where and when (the video is time stamped) events occurred.

B. Team Descriptions

The three teams whose runs are discussed in this paper had extremely different user interfaces. All of the robots were teleoperated and used only one operator.

Team A used a virtual reality type of user interface. The operator used goggles to view the video being sent back from the robot. The goggles were used in conjunction with a head tracking device that allowed the operator to control one of three cameras mounted on the vehicle: low mounted front and back cameras with 1 degree of freedom and a higher mounted front facing camera with 2 degrees of freedom. This allowed the operator to view the wheels of the robot. The operator could select from a full display of information superimposed on the video display, a simpler view, or video only. Other information available included the camera selected, thermal sensor display, and an indicator of camera position relative to the robot body. The operator also had audio sensing available. There was the ability to capture still photos of the victims or the arena for later viewing. Figure 3a shows the full view of the user interface although the simpler view (Figure 3b) was the one used the majority of the time.

Team B used two robots. One robot, on a tether, was used only as a communications relay. The other robot returned two video feeds on two separate displays. One video feed was from a movable camera and was controlled, as was the robot, using a joystick. The camera was mounted relatively high on the robot

and allowed the operator to view the front portion of the robot as it moved. The second camera was fixed and pointed down from the top of the robot giving a view of the robot and several inches of space surrounding the robot. The user interface on the laptop was only used for starting up the robot.
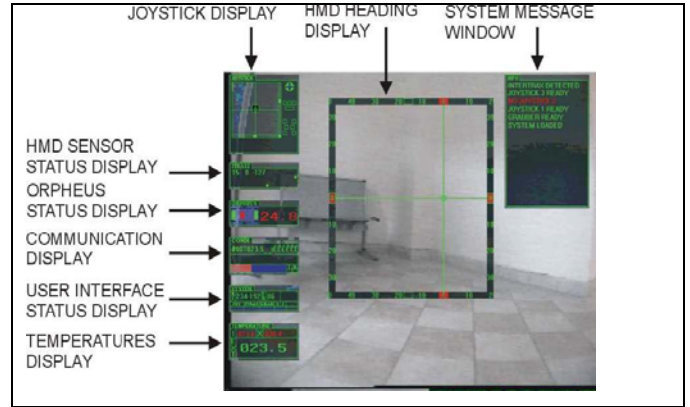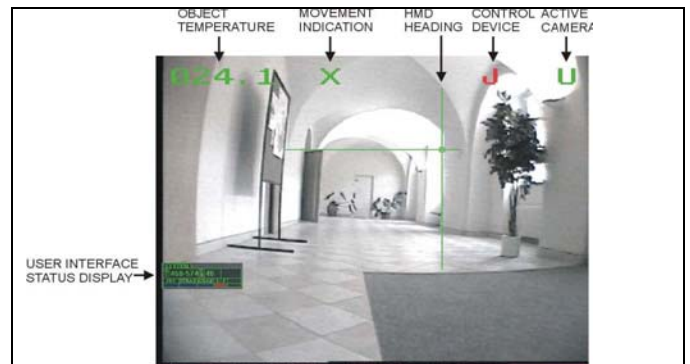

Figure 3a: Team A's Full User Interface

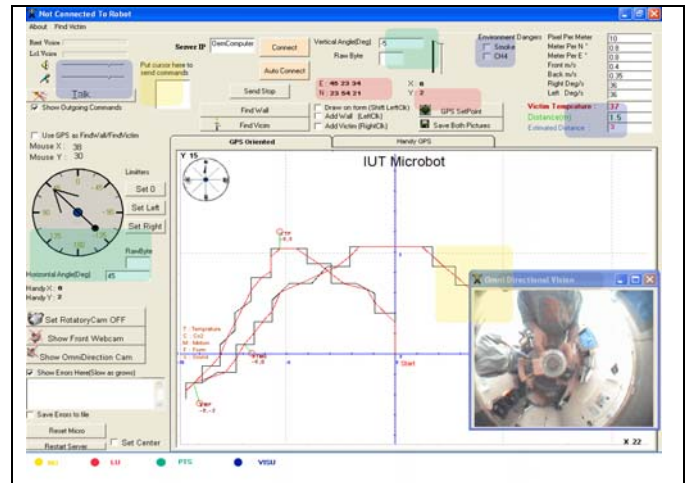Figure 3b: Team A's Normal User Interface

Figure 4: Team C's User Interface

Team C also used two robots. A small robot was teleoperated using a joystick. The larger robot was controlled and tracked on the laptop GUI. A second window on the GUI shows an omni directional camera view. A separate display was used for video being sent back from the robot. Other sensors available on the larger robot were laser range finder, full duplex audio, and sonar. The GUI had a map background that the operator could use to mark locations of victims found. When the robot is used outdoors the map can be automatically generated using the robot's GPS. Figure 4 shows Team C's GUI.

TABLE III.   CLASSIFICATION SCHEME

Type of awareness	Overall mission Human-human Robot-human Human-robot Robot-robot
Type of task	Global navigation Local navigation Obstacle extraction Vehicle state Victim identification

IV. ANALYSIS

We identified "critical incidents" that we saw during the runs. We defined critical incidents as a situation where the robot could potentially cause damage to itself, the victim or the environment based on Leveson's definition of safety-critical situations [10]. Critical incidents can have positive as well as negative outcomes. A positive outcome for a critical incident could be managing to navigate safely through a very narrow space. A negative outcome could be moving a wall enough to cause a secondary collapse. In this study, we only coded critical incidents with negative outcomes. However, we did note a number of critical incidents with positive outcomes to help use understand how elements of the user interaction helped the operators' successes.

Table III shows the classification scheme we used in classifying critical incidents. Note that we employed a two part classification scheme: by HRI awareness type and task type. The HRI awareness types are defined in Table II. We define our task-related codes as follows:

- *Global navigation:* The operator's knowledge of the robot's position in the world. If this is inadequate it may be manifested by driving out of bounds or by covering areas already searched.

- *Local navigation:* The operator's understanding of the local environment and the ability to maneuver in constrained or difficult situations. A limited understanding may result in the robot's sliding, slipping or bumping, but without a significant delay in navigation.

- *Obstacle encounter:* The robot is hindered in moving towards a goal, e.g. by being stuck on something.

- *Vehicle state:* Robot is in a degraded state; not stable or upright or sensors impaired or broken. The operator may be able to still accomplish the task if this state is known.

- *Victim ID:* Operators have to locate victims and to identify whether a victim is conscious or not. It is possible to misidentify a victim based on inaccurate interpretation of sensor data.

A. Quantitative Results

We coded 12 runs; two semifinal runs for the three teams and two final runs for each team. Overall there were 52 critical incidents found by one or more coders. There were 15 incidents that were missed by one of the coders. Coder one found 6 incidents that coder two did not; coder two found 9 incidents that coder one did not. Overall the coders both found 71% of the incidents.

The two coders independently coded eight runs using the critical incident definitions and computed the agreement on those incidents using the Kappa coefficient. The Kappa coefficient computed for the 22 incidents found by both coders was 0.926. The agreement for coding the incidents that both coders found was extremely high but finding the same incidents initially was more problematic.

The coders associated only one type of HRI awareness with the critical incidents. In all cases, problems were due to a lack of human-robot HRI awareness (per our previous definition: "The understanding that the humans have of the locations, identities, activities, status and surrounding of the robots."). We will provide examples of these problems as we discuss critical incidents below. Human-human and robot-robot HRI awareness were not applicable because only one operator was employed by each team, and, while multiple robots were sometimes fielded, the robots did not communicate interactively with each other. Similarly, robot-human HRI awareness was not applicable because the robots were not autonomous; therefore, they were not responsible for interpreting humans' commands other than basic teleoperation and sensor operation commands. Finally, humans' overall mission understanding awareness remained high in all cases due to the straightforward nature of the task (locate and map victims), so no problems were traceable to this type of awareness.

In contrast, the coders associated three out of five of the task type classifications with the critical incidents. Table IV shows the breakdown of critical incidents by task type and team. To produce this table, the two coders discussed the critical incidents that they originally disagreed on and arrived at an agreement. Table IV contains all critical incidents found by all coders. Obstacle encounters were the most frequent type of critical incident, followed by local navigation and vehicle state.

TABLE IV. CRITICAL INCIDENTS WITH NEGATIVE OUTCOMES BY TEAM

Critical Incident	Overall	Team A	Team B	Team C
Local Navigation	15	4	1	10
Global Navigation	0			
Obstacle Encounter	26	6	9	11
Victim Identification	0			
Vehicle State	12	5	2	5

In contrast, the coders associated three out of five of the task type classifications with the critical incidents. Table IV shows the breakdown of critical incidents by task type and team. This table contains all critical incidents found by both coders with any differences in coding resolved. Obstacle encounters were the most frequent type of critical incident, followed by local navigation and vehicle state.

Below we discuss each type of critical incident.

1) Global Navigation

In this study, Team C had a global view of the arena in their omni directional camera. Team C also provided a map that the operator used to mark the location of victims. However, we didn't find any critical incidents in these runs that involved global navigation. We did note several incidents of this type in earlier runs, not analyzed in this paper, but there were few of these incidents. The arenas in this particular competition were smaller than the standard NIST test arena, so global navigation was not a major issue. This will not be true in general in USAR environments and we will need to devise some experiments to study information needs for global navigation.

2) Local Navigation

All the teams had critical incidents involving local navigation, though Team A had fewer incidents than Team C. Team B had only one incident involving local navigation during these runs.

Team A was very successful in large part because the operator had the ability to construct a frame of reference by using the 2 degree of freedom camera to view the robot's front wheels in relation to obstacles in the arena. We saw a number of instances where this strategy allowed the robot to go through extremely tight spaces; Team A maintained excellent HRI awareness of the robot's location and surroundings.

Team B's overhead camera was also used by the operator to view the space directly beside the robot and obtain HRI awareness, though this view was fixed and less flexible than Team A's. However, using a strategy on the part of the operator rather than an automatic behavior on the part of the robot places cognitive demands on the operator. One idea to mitigate cognitive demands is to integrate the camera output with sonar data so that when an obstacle is sensed, the obstacle is automatically displayed in the camera view.

Rear cameras also helped operators maintain HRI awareness of the robot's surroundings. Team A backed up a number of times when the space was too tight to turn around.

However, the operator had to manually switch to the rear camera even when backing up. Team C had a 360 degree view but this clearly did not help in local navigation, although we did see one instance in an earlier run where this was useful in global navigation.

Trying to navigate steep and slippery slopes is also an issue. Indicators of traction would be useful. Operators could also benefit from a referent to provide awareness of the slope or steepness of a ramp or incline. Using only the video feed places a large cognitive load on the operator because the operator must use subtle visual cues from the environment to estimate slope. Having sensors and referents to gauge the difficulty of a slope could be beneficial.

3) Obstacle Encounter

Obstacle Encounter incidents were fewer for Team A than for the other teams. Team A had both front and rear cameras as well as the front facing camera that could be manipulated. This gave the operator excellent awareness of obstacles at virtually any angle to the robot, including to the rear. The ability to point his movable camera at various angles while navigating through the environment also gave Team A's operator an advantage; he was able to maintain awareness of obstacles while on the move.

There were many instances of robots getting stuck or entangled with obstacles, while the operators lacked sufficient HRI awareness to understand the cause of the entanglement. Operators infer that something is wrong if the video sent back from the robot doesn't change even though they are commanding the robot to move. Sound is one means that operators use to determine that something is amiss; they hear motors revving, for example. If the environment is extremely noisy, as could certainly be the case in search and rescue, sound becomes useless. In other runs during this competition the operator in Team A mentioned that he used the audio to provide information about movement.

We also saw incidents where obstacles were stuck in the robot mechanism. While this did not prevent mobility in some instances, it could cause robots (and/or the obstacles stuck to the robots) to hit walls or victims. To the extent that operators did not understand the size or nature of the stuck obstacles, they lacked HRI awareness of the robot's status. A means of self-inspection seems necessary to successfully extract robots from these obstacles. Information such as the amount of tread on the ground or the number of wheels on the ground might be helpful.

4) Vehicle State

Vehicle state is closely related to obstacle encounters. We saw incidents where robots were on their side, did "wheelies" or had parts wedged under platforms. While information such as battery life and sensor status is displayed by some teams, sensors on different parts of the robots and pitch and roll indicators would be useful to provide HRI awareness of the robot's status and positions.

The number of vehicle state incidents was the same for Team A and C, with Team B's count being lower. This is counter-intuitive because Team B had by far the least

information presented in their user interface. However, their vehicle had impressive mobility. This is an area where we plan to conduct future studies to determine not only the type of information being presented but the presentation of that information. Team B's user interface presented a top down view of the robot and it may be possible that the operator was able to gain awareness of the robot's angle or instability using that view.

Team A's interface provided another mechanism for obtaining awareness of the robot's status: their interface showed camera position relative to the robot body. Although this particular competition environment did not allow us to determine how useful this was, we have seen incidents in other competitions where navigation was unsuccessful because the operator did not realize where the camera was pointing [7].

5)   Victim Identification

Victims in the NIST arena could be located using vision, thermal, sound, and motion. In several instances teams used sound and thermal signatures to identify possible victims. In other runs in the competition we saw an incident where a robot with audio was able to detect a victim using sound. We did not see any misidentification of victims in these runs, although the competition rules are expanding to include identifying the state of the victim. This will necessitate a close inspection by the robot to determine if the victim is conscious.

V.   Conclusions

We have developed definitions of critical incidents and a coding scheme and used these to compare the performance of three teams in the USAR competition. Based on this assessment we examined the user interaction and identified potential information displays that, if implemented, may reduce the number of critical incidents. Based on this analysis we have generated five guidelines for information display for USAR robots.

Information displays for USAR should include:

- a frame of reference to determine position of robot relative to environment (and provide awareness of the robot's surroundings)

- indicators of robot health/state, including which camera is being used, the position(s) of camera(s), traction information, and pitch/roll indicators (to provide better awareness of the robot's status)

- information from multiple sensors presented in an integrated fashion (to avoid relying on the operator devising strategies to overcome information fragmentation and facilitate better awareness of the robot's location and surroundings)

- the ability to self inspect the robot body for damage or entangled obstacles (to provide enhanced awareness of the robot's status)

- automatic presentation of contextually-appropriate information, such as automatically switching to a rear camera view if the robot is backing up

Many other competitions were co-located with the USAR competition at Robocup and we saw many instances of wireless interference and degraded video. This is not unlike conditions during actual search and rescue activities. Therefore, heavy reliance on video will impair the operator's ability to teleoperator for periods of time. We recommend that feedback from other sensors be used to supplement video.

VI.   Future

The USAR competitions have allowed us to assess problems with current human-robot interaction and to develop some hypotheses about information that appears useful. The next step is to determine experimentally what information and what presentation of that information is helpful in providing awareness for operators of USAR robotics. We are working with several USAR teams to develop these experiments and test them in the NIST arena.

Acknowledgments

References

[1]   J. Burke, R. Murphy, and M. Coovert, "Moonlight in Miami: an ethnographic study of human-robot interaction in the context of an urban search and rescue disaster response training exercise," Journal of Human-Computer Interaction, in press.

[2]   C. Shoemaker, Plenary, PERMIS 2003, September 2003.

[3]   J. Casper and R. Murphy, "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center," IEEE Transactions on Systems, Man, and Cybernetics, Part B, June 2003, vol. 33, pp. 367-385.

[4]   A. Jacoff, E. Messina, and J. Evans, "A reference test course for autonomous mobile robots," Proceedings of the SPIE-AeroSense Conference, Orlando, FL, April 2001.

[5]   A. Jacoff, E. Messina, and J. Evans, "A standard test course for urban search and rescue robots," Proceedings of the Performance Metrics for Intelligent Systems Workshop, August 2000.

[6]   D. R. Olsen, Jr., and M. A. Goodrich, "Metrics for evaluating human-robot interactions," Proceedings of PERMIS 2003, September 2003.

[7]   H. A. Yanco, J. L. Drury, and J. Scholtz, "Beyond usability evaluation: analysis of human-robot interaction at a major robotics competition," Journal of Human-Computer Interaction, in press.

[8]   J. L. Drury, J. Scholtz, and H. A. Yanco. "Awareness in human-robot interactions," Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Washington, DC, October 2003.

[9]   Ericsson, K. A. &Simon, H.A. *Procotol Analysis: Verbal Reports as Data.* Cambridge, MA: The MIT Press. 1993.

[10] Leveson, N.G. "Software safety: why, what and how." ACM Computing Surveys, vol. 18,(2), 125-162. June, 1986.