# Using Standard Test Methods for Response Robots To Evaluate Remote Human-Robot Interaction

Adam Norton
New England Robotics Validation and Experimentation (NERVE) Center
University of Massachusetts Lowell
Lowell, Massachusetts 01854
anorton@cs.uml.edu

Holly A. Yanco
Department of Computer Science
University of Massachusetts Lowell
Lowell, Massachusetts 01854
holly@cs.uml.edu

***Abstract*—Test beds for evaluating human-robot interaction (HRI) are generally developed to fit a particular experiment, lacking a common set of tasks. The standard test methods for response robots specified through ASTM E54.08.01 are used to evaluate robot mobility, manipulation, sensors, and operator proficiency. There are four test methods that focus on proper situation awareness (SA): Line Following, Center in Alleys, Align Edges, and Pan Tilt Zoom. These tests serve as candidates for standardized HRI experimental set ups as they simply and effectively capture many characteristics of the robot, interface, and operator. We discuss an example data set of test method performance and how they can be used to evaluate HRI.**

## I. Introduction

There currently exists no standard experimental set-ups for evaluating human-robot interaction (HRI). In order for such experiments to be standardized, they must be able to be applied to many different robots and interfaces of varying capabilities. The test metrics must be broadly applicable such that the performance of the robot, interface, and operator are captured.

The standard test methods specified through ASTM E54.08.01 Committee on Homeland Security Applications; Operational Equipment; Robots [1] are used to evaluate response robot capabilities and operator proficiency. They have been used to evaluate many teleoperated robots and to train end-users in urban search and rescue (USAR) and explosive ordinance disposal (EOD) domains. The test methods are performed without line-of-sight, meaning the operator must rely on the interface as they would in a real scenario. The settings for each test method are malleable such that they can be tuned to the characteristics of the robot. All aspects of the system (e.g., robot configuration, interface modalities, operator knowledge of system) have in impact on performance.

The test methods in the Maneuvering suite highlight the capabilities of the robot, interface, and operator at maintaining situation awareness (SA). Due to the test methods' malleability and holistic nature, they are good examples of potential standardized HRI experiments. To this end, we discuss a path forward for using them as such.

## II. Related Work

During the early development of the E54.08.01 standard test methods, physical and virtual implementations of test arenas were used [2] [3]. The arenas used versions of the test methods in an operational scenario, combining elements to form the challenges of each arena. These experiments have also been used to perform iterative HRI designs.

A test bed for evaluating HRI with EOD robots [4] distilled a set of tasks based on existing law enforcement training programs and real world incidents. The number of button presses and mode changes needed to perform each task were used to evaluate the HRI. From this it is suggested that the information required from the interface to perform each task be defined to determine if an operator can access it properly.

One of the four lessons learned from a longitudinal study of real world USAR events and training exercises highlighted that SA was a very prevalent issue with the HRI [5], citing that half of the operation time is spent gaining SA. These interactions could be evaluated using many of the common metrics for HRI [6], such as assessing the accuracy of mental models, had they occurred in a more controlled scenario.

## III. Standard Test Methods

There are four test methods in the Maneuvering suite:

- **LF**: Line Following (ASTM E2829)
- **CA**: Center in Alleys
- **AE**: Align Edges
- **PTZ**: Pan Tilt Zoom (ASTM WK33261)

An image of each test method can be seen in Figure 1. Each test method can be adjusted based on the characteristics of the robot and interface to allow for equally difficult challenges between platforms. Any teleoperated or semi-autonomous mobile robot with at least one camera can be used.

The performance metric used for the test methods is the number of tasks completed per minute, or rate of advance. For LF, CA, and AE, one task refers to traversing from one end of the test apparatus to the other and back, with some type of obstacle to be negotiated in between. The first half of a task is performed while traversing forward and the second half in reverse. For PTZ, one task refers to viewing near and far acuity targets in the apparatus. Each test has its own rules for fault conditions; if a fault occurs then that task is not counted and must be repeated. The time on task continues to increase, resulting in a decreased rate in advance.

### A. Line Following (LF)

The operator drives the robot over a figure-8 line on the ground, of which each end falls in one of the apparatus' end zones. The line must remain underneath the robot's body while traversing, meaning the operator must maintain a view of the

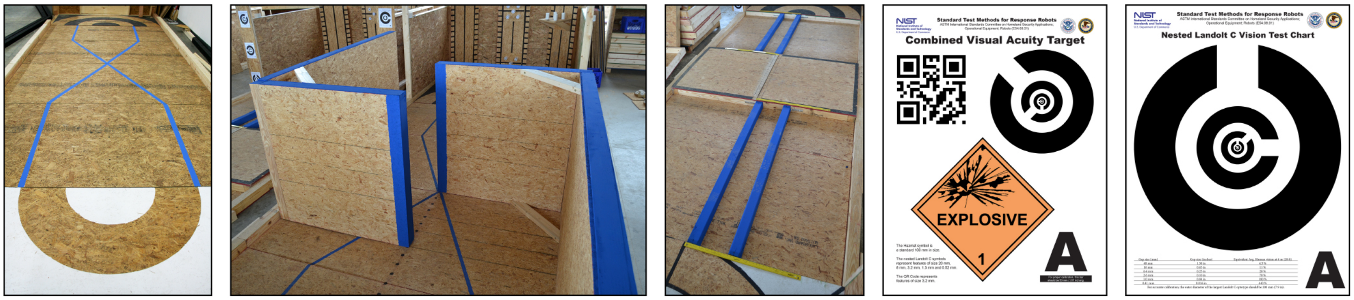


Fig. 1. Images of the standard test methods. Left to right: Line Following, Center in Alleys, Align Edges, and Pan Tilt Zoom near and field targets.

line with respect to the robot. The figure-8 shape forces the operator to match the direction of the line with the orientation of the robot. If the robot drives off of the line such that it is visible outside of the robot's body then a fault is incurred.

### B. Center in Alleys (CA)

The operator drives the robot between two walls that form a confined passageway sized to the turning diameter of the robot. The passageway is perpendicular to a straight path between the end zones, requiring the operator to turn the robot while traversing through. The walls are attached to the apparatus using vertical barrel bolts into the floor such that if they are bumped slightly by the robot then they will remain in place, but hard collisions will move them. If the walls are moved due to a robot collision then a fault is incurred.

### C. Align Edges (AE)

The operator drives the robot over two rails parallel to a straight path between the end zones. The distance between the outside edges of the rails is set to match that of the robot's wheels or tracks. Two sets of rails must be traversed; one on the right side of the lane and the other on the left. A platform in between each set allows the operator to orient the robot's approach to the next set. If the robot falls off of the rails, then a fault is incurred.

### D. Pan Tilt Zoom (PTZ)

While staying in a fixed location, the operator points the robot's camera(s) at targets with visual acuity artifacts. The near and far field targets are each labeled A-J; the operator alternates between looking at a near field target and then its corresponding far field target. The visual acuity artifacts used are Landolt C eye charts, which have concentric circles with cuts in them at varying orientations. Based on the robot's camera(s) resolution combined with the interface's display resolution, the operator is able to identify the orientation of the Landolt C eye charts to the level of acuity that is achievable. The robot is allowed to rotate in place if necessary.

## IV. Situation Awareness Factors

Each test method is defined by the level 1, 2, and 3 SA (as defined in [7]) it requires to be performed in Table I. The ability of an operator to acquire and maintain these SA elements is influenced by a variety of characteristics of the robot, its interface, and the operator.

### A. Robot Characteristics

For CA and AE, the dimensions of the footprint of the robot will affect how the walls and rails are positioned, respectively. Many robots in this domain use manipulators (generally on top of the base) and articulators (on the front and/or back of the base) that increase their overall size profile. If these components are able to be moved such that the footprint of the robot is made smaller (i.e., closer to the center of the robot's volume), it will aid in completing CA. The robot may also be tethered, most commonly on a motorized spool. These components introduce additional SA of the robot's status that the operator must maintain.

Teleoperated robots require at least one or two cameras that provide forward and rear facing views, but the number of cameras, where they are located on the robot, and their individual qualities (e.g., fixed or dynamic, field of view) can vary. Exocentric cameras located above the robot's body and provide an outside view of it have also been shown to increase spatial reasoning [8]. These cameras may be located on the robot's manipulator or a vertical pole referred to as a mast. The ability to pan, tilt, and zoom these cameras around to view the body of the robot (CA and AE) and the environment around it (PTZ) is also beneficial. Alternatively, for PTZ, one or more fixed cameras can used, but the robot will have to rotate in place in lieu of a rotational degree of freedom.

### B. Interface Characteristics

The interface used by the operator to control the robot (specifically the information it provides and the operator's knowledge of it) is the largest contributor to performance.

| Test | Level 1 SA | Level 2 SA | Level 3 SA |
|---|---|---|---|
| Line Following | Local environment (line) underneath the front and back of the robot | Alignment deviation of the line from underneath the robot | How to adjust the robot's position to maintain alignment while traversing |
| Center in Alleys | Local environment (walls) around the outside of the robot's volume | Distance from the walls to the robot | How to adjust the robot's positioning to avoid collisions while traversing |
| Align Edges | Local environment (rails) underneath the outside edges of the robot | Alignment deviation from the robot's position to the edge of the rails | How to adjust the robot's positioning to maintain alignment while traversing |
| Pan Tilt Zoom | Local (near field targets) and global (far field targets) environment around the robot | Proper positioning of the robot's position and its camera, and the pan, tilt, and zoom settings of its camera(s) | How to adjust the robot's camera and/or positioning to decipher the environment |

TABLE I. The specific level 1, 2, and 3 SA (as defined by [7]) that must be maintained to perform each test method.

| Robot | Exocentric Camera | Panning DOF | Manipulator | Articulators | Tether Option | Camera Presentation | Interface Pose Info | LF | CA | AE | PTZ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | n/a | Body rotate | n/a | Rear | No | Single, multiple | Side | * | * | * | 0.62, 0.82 |
| B | Mast, manipulator | Body rotate | 5 DOF | n/a | No | Single | Side | 1.3 | 2.5, 2.35, 2.45 | 0.8 | 1.1 |
| C | Manipulator | Body rotate | 4 DOF | Front | No | Single, multiple | Side | 1.4 | 2.0 | 0.3 | 0.83 |
| D | Manipulator | Body rotate | 5 DOF | Front | No | Single, multiple | Isometric | 2.1 | 3.3 | 0.7 | 1.7 |
| E | Mast, manipulator | Body rotate | 5 DOF | n/a | No | Single | Side | * | * | * | * |
| F | Mast, manipulator | Camera control | 8 DOF | Front | Yes | Single, multiple | Isometric | * | 0.13 | 0.1, 0.14 | * |
| G | Manipulator | Camera control | 7 DOF, telescoping limbs | Front, rear | No | Single, multiple | Isometric | * | * | * | * |
| H | Mast, manipulator | Camera control | 4 DOF | n/a | No | Single, multiple | Isometric | * | 0.5, 0.65 | 0.25 | 2.0, 1.82 |
| I | Mast, manipulator | Camera control | 6 DOF; telescoping limbs | Front, rear | Yes | Single, multiple | Side | * | 0.46, 0.22, 0.5, 0.27 | * | * |

TABLE II. ROBOT AND INTERFACE CHARACTERISTICS THAT ARE PERTINENT TO HRI AND EXAMPLES OF PERFORMANCE DATA IN EACH OF THE TEST METHODS. EACH PERFORMANCE METRIC IS A RATE OF ADVANCE, MEANING THE NUMBER OF TASKS COMPLETED PER MINUTE. * INDICATES THAT THE PERFORMANCE DATA FOR THAT ROBOT'S PERFORMANCE IN THE TEST METHOD WAS NOT AVAILABLE AT THE TIME OF PUBLICATION.

Input devices generally use at least one joystick of some kind in addition to a series of buttons or switches. The sensitivity of the joystick, complexity in changing control modes (e.g., navigating a series of menu screens vs. flipping a switch), and latency between command and feedback can affect the operator's ability to perform the test. If a system employed automatic direction reversal (ADR), which has the system maintain the orientation of its control when driving in reverse, it could improve performance [9] in LF, CA, and AE.

To perform CA, the robot's footprint must be reduced by adjusting the its manipulators and/or articulators. Some interfaces offer predefined poses that can be selected, which is beneficial for robots with many degrees of freedom. A visualization of the robot's pose is also common, either as a side profile or isometric representation, which can increase the operator's SA of the robot's status [10].

The presentation of camera views can vary greatly between interfaces. Most systems are able to display full screen views of a single camera if desired. If multiple cameras are used, different options for picture-in-picture are generally available, such as a smaller display overlaid in the corner of a larger display (referred to as the "rear view mirror"), or displaying many views at once. Cameras may also be displayed in panoramic to provide a very wide field of view, which may aid in performing CA and AE. Local distance sensors could reduce collisions [11] while performing CA, although not many have been exhibited on deployed response robots.

### C. Operator Characteristics

The operator of the robot must be able to acquire and maintain proper SA of the robot's surroundings and status. Continued use of the test methods is intended to increase an operator's understanding of the robot's capabilities and knowledge of how to control it. Given that poor exhibition of HRI with response robots has been observed during real world scenarios such as during the Fukushima Daiichi disaster response [12], the development of this work is pertinent.

The operator must have an accurate mental model of the robot, particularly if it has a manipulator and/or articulators, when performing CA. Some interfaces do not provide such information, so the operator must mentally update their mental model every time they move the robot. Operators may prefer a system that uses inverse kinematics to control a high degree of freedom manipulator, while others may prefer to control each joint individually. The operator must also have proper spatial awareness of the relationship between the environment features and the robot for all test methods.

Some of these characteristics can be aided or hindered by the interface being used. For instance, if an exocentric camera with a more angled view is used then the distance between the robot and local obstacles can be visually estimated. If one is not available, then the operator must interpret the distances using depth perception, which may be more difficult.

## V. EXAMPLE PERFORMANCE DATA

Nine commonly used response robots and their pertinent characteristics are detailed in Table II, along with available test method performance data culled together from a variety of public and private test events. Some test methods have multiple measures with the same robot due to being performed by many operators. The robots have been anonymized by granting each an alphabetic identifier, ordered approximately from smallest (A) to largest (I) to give the letters some meaning.

### A. Discussion

A comprehensive HRI study of these test methods has not yet been performed. Due to the settings of each test being tuned to the robot, all robot performance could theoretically be equivalent if they were all operated at the same pace. However, the difference in performance is due to the many varying characteristics of the robots, interfaces, and operators.

In general, smaller robots (A-D) tend to exhibit a higher rate of advance in CA, most likely due to faster traversing speeds. Operators may be more cautious with a larger robot. However, it also may be due to each robot's manipulator if it exceeds the base footprint even when positioned in the smallest form factor possible. It should be noted that some operators have been observed performing CA with the manipulator obstructing their camera view while operating robot I. This may be indicative of a lack of knowledge of how to use the system or is indeed the optimal way of performing the test with that particular system.

The differences between exocentric camera views from a robot's manipulator or mast are captured by the performance in AE, as both views can be used to approach the rails properly and maintain a constant direction of traversal. A manipulator camera generally requires more positioning (4-8 DOF) to swap between views for forward and reverse traversals than that of a mast (1-3 DOF), which may take more time. Some exocentric cameras are only placed high enough to provide a view of one side of the robot's body, which was observed for robots A, E, and H. This may provide enough SA for the operator, but could be insufficient for supporting their mental model.

For PTZ, some robot/operators rely solely on a high resolution camera that is able to pan and tilt without moving the body of the robot at all (robots F-I). Others rotate the robot's body and/or tilt the robot's body with their articulators to make up for missing degrees of freedom (robots A-E). This results in varying control schemes for different robots, and sometimes within the same robot.

In the test methods' current design, the number of faults is generally not reported. If they were, it would give more insight into the exhibited HRI, as many faults could be indicative of poor camera presentation on the interface, the operator's lack of skill at maneuvering the joystick properly, etc.

These test methods have not yet been exercised using robots with autonomous navigation capabilities. Increasing robot autonomy would allow for the evaluation of HRI with respect to sharing SA between the robot and the operator through the interface, including aspects like varying operator interaction frequency (potentially exhibiting "out-of-the-loop" performance [13]) and the communication of failures.

## VI. Conclusion and Future Work

These test methods are examples of designs that can be used towards standardized HRI experiments, particularly for robots that operate remotely. The many reasons for performance discussed in the previous section highlight the type of information that can be extracted from using the test methods. The next step is to conduct more comprehensive testing and analyzation that investigates the exhibited HRI, such as:

- Interface settings (e.g., camera presentation, robot speed) used during each test method performance;
- Specific robot actions performed (e.g., change of robot pose, rotation of manipulator or mast camera views);
- Faults incurred on each test method caused by incorrect SA of the robot, its environment, and/or understanding of the interface;
- Varying operator experience levels and how performance changes over increased use of the test method;
- Additional sensors to provide SA (e.g., LIDAR); and
- Trends in the previous points to determine optimal HRI for maintaining each type of SA.

From the proposed testing, a set of common robot, operator, and interface characteristics that correlate with specific faults (e.g., backing into a CA wall due to not switching to the rear-facing camera) can be distilled from the performance data, categorized by the type of SA loss. These can be used to form guidelines for determining broader HRI-specific faults.

Evaluating SA in terms of spatial awareness is an important aspect of HRI with response robots that calls for a confined space apparatus. Test methods for HRI in other domains that use more autonomy or tasks in more structured environments may not require this type of SA and will need additional considerations. For instance, HRI of an autonomous space rover may depend more on proper alert techniques and treatment of historical information. The current test method designs do not necessarily highlight those types of capabilities.

Further experimental set-ups will need to be structured to capture these aspects, while still using the same design principles of the test methods discussed in this paper: malleable settings that enable fairness across many solutions, simple performance metrics that can capture many types of errors, and using baseline tasks to evaluate foundational HRI.

## Acknowledgments

This research has been supported in part by the National Institute of Standards and Technology (60NANB14D286). The authors would like to thank Adam Jacoff and Ann Virts from NIST for their development of the test methods and assistance in gathering performance data.

## References

[1] ASTM, "Subcommittee E54 on Operational Equipment," 2015, http://www.astm.org/COMMIT/SUBCOMMIT/E5408.htm.

[2] M. Lewis, K. Sycara, and I. Nourbakhsh, "Developing a Testbed For Studying Human-Robot Interaction in Urban Search And Rescue," in *Proceedings of the 10th International Conference On Human Computer Interaction*, 2003, pp. 22–27.

[3] S. Carpin, M. Lewis, J. Wang, S. Balakirsky, and C. Scrapper, "USAR-Sim: A Robot Simulator for Research and Education," in *2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 1400–1405.

[4] J. Scholtz, M. Theofanos, and B. Antonishek, "Development of a Test Bed for Evaluating Human-Robot Performance for Explosive Ordnance Disposal Robots," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. ACM, 2006, pp. 10–17.

[5] R. R. Murphy and J. L. Burke, "Up From the Rubble: Lessons Learned about HRI from Search and Rescue," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 3. SAGE Publications, 2005, pp. 437–441.

[6] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common Metrics for Human-Robot Interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*.

[7] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, 1995.

[8] B. Keyes, R. Casey, H. A. Yanco, B. A. Maxwell, and Y. Georgiev, "Camera Placement and Multi-camera Fusion for Remote Robot Operation," in *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*, 2006, pp. 22–24.

[9] B. Keyes, M. Micire, J. L. Drury, and H. A. Yanco, "Improving Human-Robot Interaction Through Interface Evolution," *Human-Robot Interaction*, 2010.

[10] J. L. Drury, H. A. Yanco, W. Howell, B. Minten, and J. Casper, "Changing Shape: Improving Situation Awareness for a Polymorphic Robot," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. ACM, 2006, pp. 72–79.

[11] H. A. Yanco, B. Keyes, J. L. Drury, C. W. Nielsen, D. A. Few, and D. J. Bruemmer, "Evolving Interface Design for Robot Search Tasks," *Journal of Field Robotics*, vol. 24, no. 8, p. 779, 2007.

[12] K. Nagatani, S. Kiribayashi, Y. Okada, K. Otake, K. Yoshida, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, M. Fukushima *et al.*, "Emergency Response to the Nuclear Accident at the Fukushima Daiichi Nuclear Power Plants Using Mobile Rescue Robots," *Journal of Field Robotics*, vol. 30, no. 1, pp. 44–63, 2013.

[13] M. R. Endsley, "Automation and Situation Awareness," *(Automation and Human Performance: Theory and Applications*.