Implementing Human-Robot Interaction Evaluation Using Standard Test Methods for Response Robots

Adam Norton[1], Brian Flynn[1], and Holly Yanco[1]

ABSTRACT

The standard test methods developed through the ASTM E54.09 Committee on Homeland Security Applications; Response Robots have served the robotics community for over a decade and a half. The test methods specified through this committee provide a common framework, scalable structure, and malleable test apparatuses that can be used to exercise many different robotic capabilities. In recent years, the committee has emphasized how the test methods can be used for training purposes to evaluate operator proficiency. The human-robot interaction (HRI) characteristics of a robot system, including its control interface, are particularly influential to operator performance. The operator's knowledge of the robot system and how to use it effectively will also affect performance. To this end, two potential standards efforts are proposed that can aid in characterizing and demonstrating effective HRI. The first is an expansion of an existing standard practice for recording robot configuration to include HRI characteristics and methods by which to measure the effectiveness of these characteristics. The second is a new standard practice that is aimed at demonstrating effective HRI by introducing variable test apparatus settings to elicit decision-making capabilities from the operator and system. The goal of this practice is to highlight HRI techniques to the standards community, such as utilizing assistive autonomous functionality for obstacle avoidance and error mitigation. By incorporating HRI evaluation in these ways, HRI can be seen as an integral component of a response robot system.

---

[1] New England Robotics Validation and Experimentation (NERVE) Center, University of Massachusetts Lowell, Lowell, MA, 01852, USA

Keywords

Robotics, human-robot interaction, response robots, standard test methods, disaster response, explosive ordnance disposal, interfaces, evaluation

Introduction

Response robots used by first responders for tasks such as urban search and rescue (USAR) and explosive ordnance disposal (EOD) are largely teleoperated. To effectively operate a response robot, the operator must have an understanding of the robot's capabilities, limitations, available options, and the environment the robot is operating in. Lack of situation awareness (SA) and high workload can increase the risk to the robot and potentially lead to terminal failures [1]. This is particularly important when operating in an environment with unknown characteristics. Performance of response robots can increase by utilizing effective techniques for human-robot interaction (HRI), meaning the relationship between the human and the robot through an interface. HRI characteristics include the manners in which information is presented to the operator, how the operator inputs commands to the robot, the operator's experience with using the robot, and the operator's SA of the robot, the environment, and the relationship between the two. Proper HRI aims to reduce operator workload, increase SA, reduce training time, increase accuracy of mental models, and increase fidelity of scene understanding.

The ASTM E54.09 [2] Standard Test Methods for Response Robots provide a set of common test apparatuses, procedures, and metrics that are used to holistically evaluate the capabilities of response robots and to measure operator proficiency. As part of this suite of test methods, a standard practice exists for recording a robot's configuration, meaning the properties of its physical components. While HRI is an integral component of response robot operation, its characteristics are not explicitly recorded using these standards. The test methods are typically run using pre-determined fixed parameters that are known to be physically achievable by the robot being tested (e.g., the minimum confined space to be traversed through). While this is useful for initial operator training and to increase system familiarity, the variable nature of real world environments is not able to be experienced. Such variability can be used to demonstrate effective HRI of a robot system, which could assist operator training for decision-making in unknown environments.

To these ends, two efforts are proposed that can aid in implementing HRI evaluation into the E54.09 standards community: characterization of HRI and introducing variability to further demonstrate effective HRI. The first expands the standard practice of recording robot configuration to include HRI characteristics and introduces a method to associate performance with those characteristics (see “Characterizing HRI”). The second is a new standard practice that uses variable test parameters and is aimed at highlighting effective HRI techniques, either by advanced interfaces, control methods, and/or competent operators (see “Demonstrating Effective HRI”). The goals for these efforts are to provide more contextual understanding to how robot test method performance was achieved, to enhance the standards for more effective operator training, and to highlight HRI as an important and integral set of characteristics for a robot system.

Disclaimer: Some commercial equipment is identified throughout this article when citing examples of different response robot characteristics and behaviors. In no case does such identification imply recommendation or endorsement by the authors.

Background and Related Work

HRI FOR RESPONSE ROBOTS

To qualify HRI for response robots, we must define the relationship between the two agents according to the role of the human and the proximity of the human to the robot. For the purposes of this article, the human is an operator (using the definitions provided by [3]; the supervisor role is also applicable, except that there is very little autonomy for response robots), meaning they command the robot to perform actions, and the robot is remote to the operator, requiring them to use an interface to interact with it. HRI refers to the relationships between the robot, the operator, and the interface, each of which can have many different characteristics that, if employed effectively, can increase performance. See Table 1 for a few examples of common HRI techniques used with response robots.

Input devices	Data displays	Operators	Autonomous functionality
Keyboard and mouse	Camera views	Number: single, multiple	"Fly the gripper" mode
Game controllers	Point clouds	Variation: fixed, changing	Waypoint navigation
Switch boards	Robot avatar (2D or 3D)	Experience levels: low to high	Error prevention (e.g., obstacle avoidance, alerts and warnings)
Touch screens	Sensor values (bar graph, numerical)	Experience with OCU analogs: computers, video games, etc.	Articulator management

Table 1. Examples of common HRI techniques used with response robots.

The interface is typically referred to as an operator control unit (OCU) that consists of one or more input devices, such as keyboards and gamepads, and one or more output devices, such as a video display screen. Information is presented on the display screen using a variety of modalities such as video feeds, sensor values, and/or visualizations of sensor data. The information presented corresponds to sensors on the robot, such as cameras for video or lidar sensors for distance data, visualized two or three-dimensionally. The operator interprets this data to build a mental model of the remote environment where the robot is, and gains a better understanding of the environment by maneuvering cameras to explore the space. The operator uses this data to understand the state of the robot, including details such as battery life and wireless communications connection strength. The interpretation of this data is referred to as situation awareness (SA) [4]. Input devices are used by the operator to command the robot, such as pushing on a joystick to drive the robot around or maneuver joints and limbs on the robot's manipulator. It is important for the operator to understand that there this is typically a delay – however small – between input commands on the input device and output device information displays [5], such as driving the robot forward and observing changes in a camera feed (i.e., moving "into" the video).

Many response robots are highly teleoperated, meaning the operator directly controls the movement of joints and motors on the robot using an input device. Some robots may employ some autonomous functionality, allowing the robot to perform some actions or tasks on its own, to varying levels. Typically, higher autonomy in a response robot means that less input is required by the operator to perform a task, which could refer to inputting less information continuously (e.g., a robot prevents itself from being driven into obstacles while the operator drives it around) or less information overall (e.g., the operator places a waypoint on a map and the robot navigates to the point on its own). While it is not common to see higher levels of autonomy in today's deployed response robots – autonomy is more prominent in the research community – some autonomous functionality is becoming common. For example, older response robots required that each individual joint of the robot's manipulator be controlled by the operator (e.g., Andros F6B) [6], whereas more advanced response robots can translate two degree of freedom (DOF) joystick movements from the operator's input device into coordinate trajectories of an end effector. In this case, the robot autonomously calculates the inverse kinematics equations required for determining individual joint positions and sends the necessary commands to move the end effector as desired, sometimes referred to as "fly the gripper" mode (e.g., Telerob Telemax [7]).

The operators of response robots consist of members of several communities that support emergency response operations, including the urban search and rescue (USAR), public safety bomb squads (PSBS), and explosive ordnance disposal (EOD) teams. The amount of training and experience these operators have with response robots will vary greatly, from new recruits to experts that have a better understanding of how to best control the robot and interpret its sensor data to utilize it effectively. Operator workload is always a concern; high workload can result from lack of training, overly complex robot control schemes, and/or long mission times. Response robot operations are sometimes hours or days long, with operators swapping in and out to continue the mission.

The relationships between each of these aspects influence one another and define the HRI. For example, an intuitively designed interface that employs a higher level of autonomous functionality could reduce the workload of an operator, but the system may also require so little interaction from the operator that they lose SA and have difficulty regaining it (referred to as the "out of the loop" problem [8]).

While there is no hardened science behind HRI design, usability studies have been conducted using quantitative and qualitative metrics to compare one HRI technique to another. For example, studies exploring the effectiveness of the use of exocentric cameras as opposed to egocentric cameras have shown that using exocentric cameras can improve robot driving performance and operator SA compared to using of egocentric cameras [9]. More generic design heuristics, such as those specified by Nielsen [10] (e.g., keep system status visible, match the system's representation of the world to the real world, prevent errors, allow for flexibility, etc.), can be followed to improve a robot interface, as can guidelines for HRI generated from performance studies (such as [9] [11], [12], and [13]). However, the manner in which these suggested design principles were derived varies, and additional considerations for robot morphology, degrees of freedom, and capabilities must also be considered. There are a series of workshops dedicated to designing standard HRI evaluation methods [14], but they are largely concerned with a particular application of HRI wherein the human is co-located with robot, such as in the social robots domain.

STANDARD TEST METHODS FOR RESPONSE ROBOTS

The test methods specified through ASTM E54.09 [2] use apparatuses that are designed to abstractly represent real-world environments in a measurable, repeatable, and easily replicable manner. The test methods are comprised of different suites for evaluating robot capabilities of different types, including mobility, dexterity, maneuvering, sensors, endurance, durability, and safety. Each test method is modeled after one or more robot competencies that are considered relevant for the response community, such as traversing over uneven terrain, manipulating objects inside of cars, or providing the operator with SA enough to maneuver through a confined space without colliding into the environment. The test methods are designed for ground, aerial, and aquatic robots (the efforts described in this article could apply to all three domains, even though only ground robots are discussed as examples).

When performing in a test method, the operator does not have line-of-sight of the robot and must rely only on the robot's interface to simulate the robot operating downrange. The operator successfully performs a task with the robot a number of times (i.e., repetitions) in order to reach a level of statistical significance for confidence and reliability in robot performance success. A certain number of faults are also allowed in order to achieve statistical significance. The baseline fault criteria for a test method is that the human must not physically intervene with the robot's performance, meaning that if the robot gets stuck while in the test method apparatus then the operator cannot enter the apparatus and try to fix it. All work must be done remotely, relying only on the robot's sensors and the HRI techniques implemented within the system. Some tests also have additional criteria for faults, such as falling off of the rails during the Maneuvering: Align Edges test [15]. The apparatus for each test method is typically configured using settings that are known to be achievable by the robot, determined by the dimensions of the robots or by having the operator practice ahead of time. Relevant characteristics are recorded and reported alongside a test metric, such as environmental conditions (e.g., lighting, temperature, wet/dry), the size of the apparatus (multiple sizes are available to account for small and large robots), and the configuration of the robot system at the time of the test (e.g., locomotion methods, manipulator characteristics, sensors on board). Elements of the ASTM E54.09 test methods have been used as part of competitions [16] and first responder training exercises [17]. When used for training, repeated usage of the test methods is intended to increase operator proficiency at robot control and its capabilities.

Characterizing HRI

This section details the first proposed effort of this article: to expand an existing ASTM E54.09 standard practice to allow for the characterization of HRI.

Many factors will determine if a response robot is fit for a given mission. A response robot's capabilities in terms of maneuvering, mobility, and manipulation are driven by a combination of robot hardware (e.g., wheeled or tracked locomotion methods, manipulator degrees of freedom), OCU hardware (e.g., input devices, display screens), software (e.g., interface design, control schemes, autonomous functionality), and the proficiency of the operator at using the robot system. All of these factors will impact if a response robot can perform as required. Characterization of a response robot system considers each of these factors and relates exhibited performance evaluation data to the particular characteristics of the robot according to each of these categories. Existing taxonomies [18] and metrics [19] for human-robot interaction have been developed and can be leveraged for characterization.

Efforts within the ASTM E54.09 subcommittee focus on the analysis of a response robot's hardware characteristics and relates them to performance data captured by that robot. Less importance is placed on the HRI characteristics of a response robot system, such as the information presentation techniques used and input device layouts. However, due to the highly teleoperated nature of response robots, the HRI characteristics of a system are integral to achieving performance, good or bad. The operator must have SA of the robot system, its environment, and the relationship between the two in order to properly exhibit the robot's capabilities during testing, all of which is driven by the robot system's HRI. To enable this type of evaluation, the HRI methods available for a given response robot system must be characterized (see "Configuration") and those which are engaged by the operator during test method performance must be captured (see "Utilization").

CONFIGURATION

Along with the performance data captured using the standard test methods specified through E54.09, there is also a standard practice for recording the system configuration of a response robot. Using this practice, ASTM E3132 [20], a variety of characteristics pertaining to the robot's hardware (e.g., locomotion method, manipulator degrees of freedom, sensor types) and operation infrastructure (e.g., shipping containers, setup time) are measured and recorded. The resulting filled out report form is then associated with all performance data captured using E54.09 standard test methods that were conducted while the robot was in that configuration. See Figure 1 for the report form used in the ASTM E3132 practice. This practice provides context for a set of performance metrics and communicates to the end user the manner in which the robot must be configured if they want to replicate the results of the test(s). It also provides a means by which to compare the performance data of robots with similar characteristics, or to point out the differences between two different robots' performance data.

By utilizing the ASTM E3132 practice, the consumer of the robot's performance data has a standard method by which to understand what features are available on the robot. The report form does not say that certain features of the robot were utilized during the recording of associated test method performance data, but that the features were there, which can make significant impacts. For example, if a robot with an optional manipulator is planning to be used for a mission that requires both manipulation of objects and stair climbing, then the robot should have its manipulator attached during all tests. Doing so may change its center of gravity, weight, and power consumption while climbing stairs, which is an important byproduct of robot configuration to capture during testing. For operator training purposes, an operator that is adept at controlling a robot with one set of characteristics may become more familiar with a different robot that has similar characteristics. Some procurements of response robots will also be dependent on certain characteristics, such as a certain number of degrees of freedom or the use of articulators in the front or back of the robot.

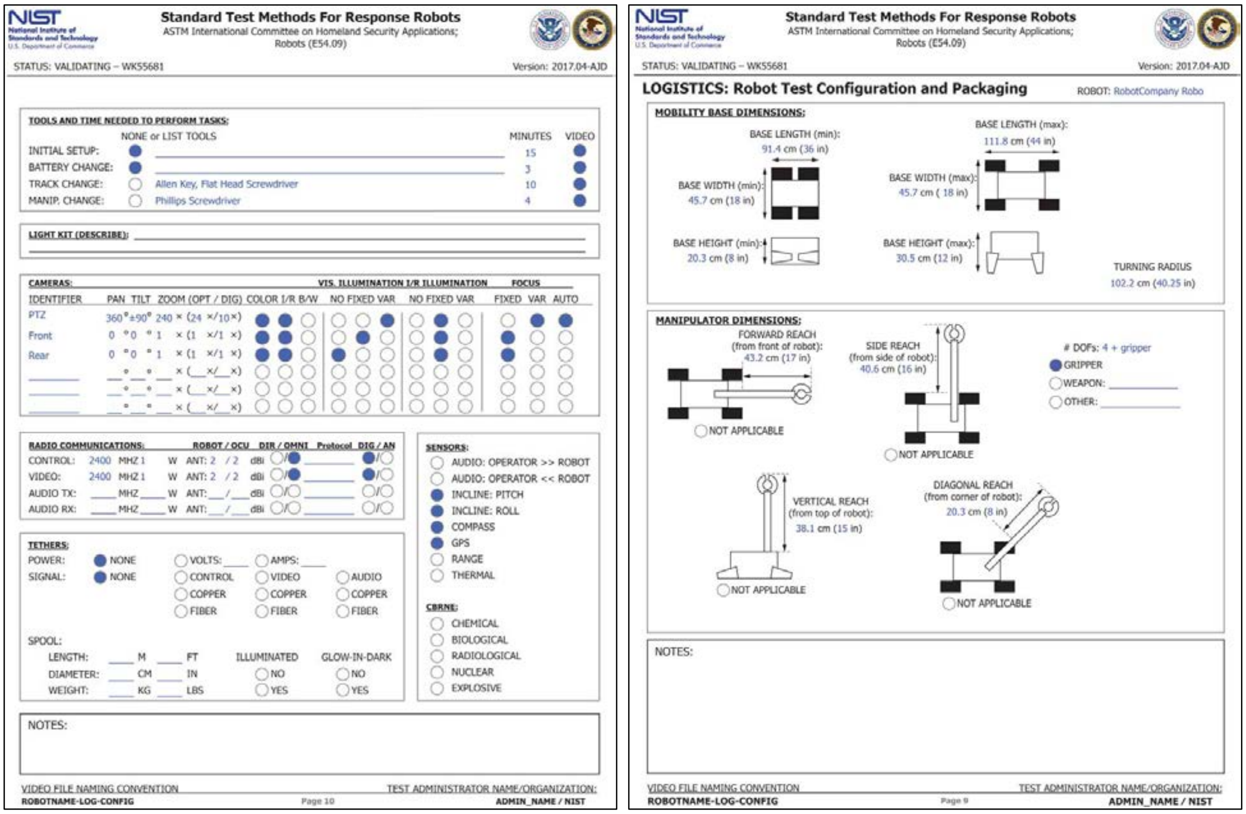


Figure 1. Report forms from ASTM E3132 with example data filled out. The left page shows information regarding set up, cameras, radio, tether, and other sensors. The right page shows dimensional measurements for the robot's body and manipulator.

A similar type of characterization could occur for the robot's HRI, or more specifically, its interface. Elements such as the information presentation modalities that are available to the operator or a mapping of the input device to robot commands could be recorded. This type of characterization can further contextualize the performance data that it is associated with, which could be used for the same purposes as the existing hardware configuration practice: comparisons between robots with similar information presentation techniques, faster training for operators that have previous robot experience with robots of similar characteristics, etc. In order to develop a practice for characterizing the HRI techniques for a system, a common language and terminology must be established. The terms should provide categories of the differing characteristics for each distinct type of information presentation technique, input device features, and any correlations between the two (e.g., pushing the joystick forward drives the robot forward).

The following is a proposed list of terms with definitions that were developed based on observations of the features commonly found on response robot interfaces. The options for each category are meant to capture the higher-level groupings of different information presentation techniques, delineated by differences in the amount of information shown (e.g., single camera display vs. multiple), dimensionality of data presented (e.g., 1D raw sensor values, 2D or 3D visualization of sensor values and direction/orientation in relation to the robot), characteristics shown in the research to elicit different performance (e.g., exocentric vs. egocentric camera placement [9], robot avatar for pose information [21], sensor fusion [22] [23]), and differences in functionality (e.g., pressing buttons to move individual different parts of the robot vs. to move to a predetermined pose). For reference, examples of these characteristics can be seen in Figure 2.

- <u>Camera views</u>: video displays that correspond to one or more cameras on the robot, categorized according to the manner in which they are displayed and the type of view of the robot's body that is provided
  - Display methods: single, multiple, picture-in-picture
  - Exocentric: the robot's body and external edges are visible in the camera view
  - Egocentric: none or very little of the robot's body is visible in the camera view
- <u>Pose information</u>: pertaining to the current shape and orientation of the robot's body driven by the robot's onboard sensors, including manipulator positions, articulator angles (both typically derived from joint encoder measurements), and/or pitch and roll of the robot's base (typically derived from an inertial measurement unit)
  - Values: numbers or single dimensional graphic display (e.g., bar chart) pertaining

to a single feature of the robot

  - 2D avatar: 2D visualization of the robot from a side or top-down profile view
  - 3D avatar: 3D visualization of the robot from a perspective that is either fixed (typically an isometric view) or customizable (can be adjusted by the operator)

- Sensor fusion: combining information displays into a single view with a shared reference frame (e.g., overlaying distance visualization data with a camera view)
  - Specify what displays are fused together
- Input device features: elements of the input device that are used to command the robot
  - Joystick, directional pad, shoulder/rocker buttons, discrete buttons, switches, touch screen
- Input device functionality: correlating the features of the input device to the function they are used for when commanding the robot
  - Body movement: driving forward, in reverse, turning
  - Articulator movement: pitching articulators on base forward and backward
  - Arm movement: rotating or pitching joints in robot arm
  - Gripper movement: rotation or open/close of gripper
  - DOF selection: changing with specific degree of freedom is to be moved
  - Mast movement: raising, lowering, or rotating the mast on the robot
  - Pose selection: changing the robot's pose to a predefined configuration
  - Menu navigation: selecting and moving through software menu screens
  - If the correlation between input device features and their functions is fixed or can be customized by the operator
- Input device and display construction: whether the input device(s) and display(s) are

physically connected to one another

- Embedded: display screen(s) are mounted into the input device (i.e., both items have to move in tandem)
- Separate: display screen(s) may be connected to the input device(s) via wire or wirelessly (i.e., both items can move largely independent of one another)

For communicating these characteristics, a tabular format could be used that is similar to the tables already found in the ASTM E3132 report form; see Table 2 for an example. A diagrammatic method could also be used, similar to that in the robot configuration standard for detailing the components of a robot's manipulator; see Figure 3. In this diagram, a set of icons are used to represent the types of joints, links, and functionalities of the manipulator, which provides a condensed method to understand how the system operates. An experienced operator may be able to use this information to inform their own mental model of how to best perform certain manipulation tasks with the robot given its configuration. A similar method could be used to describe the HRI characteristics available for a robot, possibly overlaid on images of the OCU display screens and input devices, but this has not yet been developed.

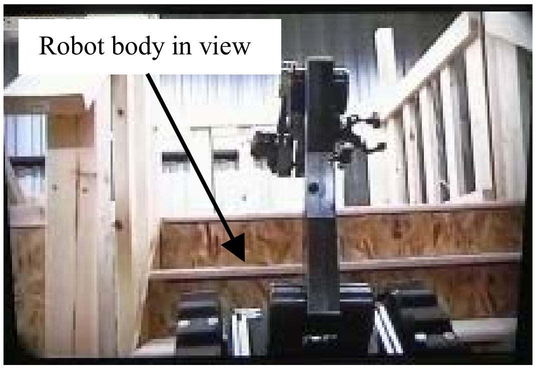
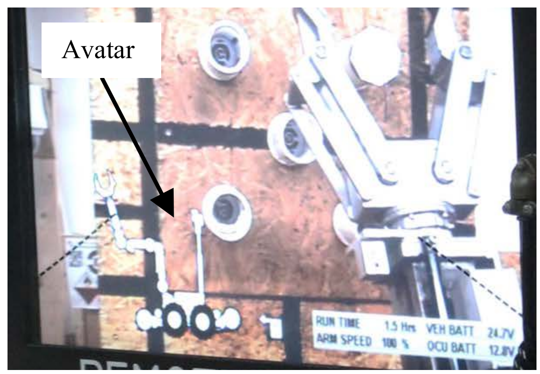
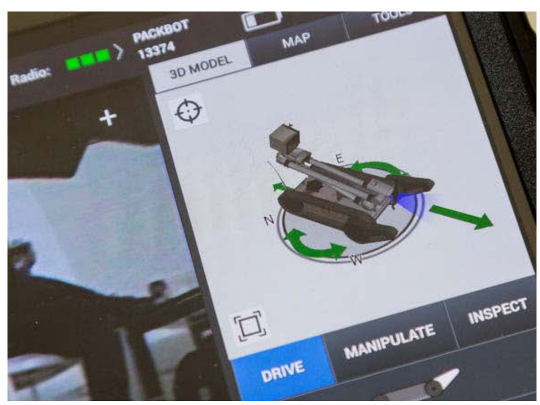
Camera views	Pose information
 Single, exocentric Robot: ICOR Caliber T5	 Values, numbers (top) and bars (bottom) Robot: Inuktun Delta Extreme
 Multiple, egocentric Robot: Andros FX	 2D avatar, side profile Robot: Andros F6B
 Picture-in-picture, exo and egocentric Robot: Telerob Telemax	 3D avatar, customizable perspective Robot: Endeavor Robotics Packbot

Figure 2. Examples of different HRI techniques that could be captured by a HRI characterization method.

<table>
<tr><td colspan="2">Camera view options:</td><td colspan="2">☑ Single</td><td colspan="2">☑ Multiple</td><td colspan="2">☑ Picture-in-Picture</td><td>☑ None</td></tr>
<tr><td colspan="2">Pose information:</td><td colspan="2">☑ Side profile</td><td colspan="2">☑ Isometric</td><td colspan="2">☑ Values</td><td>☑ None</td></tr>
<tr><td colspan="2">Sensor fusion:</td><td colspan="2">☑ Specify what displays are fused together:</td><td colspan="4"><open response></td><td>☑ None</td></tr>
<tr><td colspan="2">Input device features:</td><td>☑ Left shoulder</td><td>☑ Left joystick</td><td>☑ Right shoulder</td><td>☑ Right joystick</td><td>☑ Touch screen</td><td>☑ Buttons or Switches</td><td></td></tr>
<tr><td rowspan="9">Input device functionality:</td><td>Body movement:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Articulator movement:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Arm movement:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Gripper movement:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>DOF selection:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Mast movement:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Pose selection:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Menu navigation:</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑</td><td>☑ N/A</td></tr>
<tr><td>Layout:</td><td colspan="2">☑ Fixed</td><td colspan="5">☑ Customizable</td></tr>
<tr><td colspan="2">Input device and display construction:</td><td colspan="2">☑ Embedded</td><td colspan="5">☑ Separate</td></tr>
</table>

Table 2. Example tabular input format for recording HRI configuration characteristics.

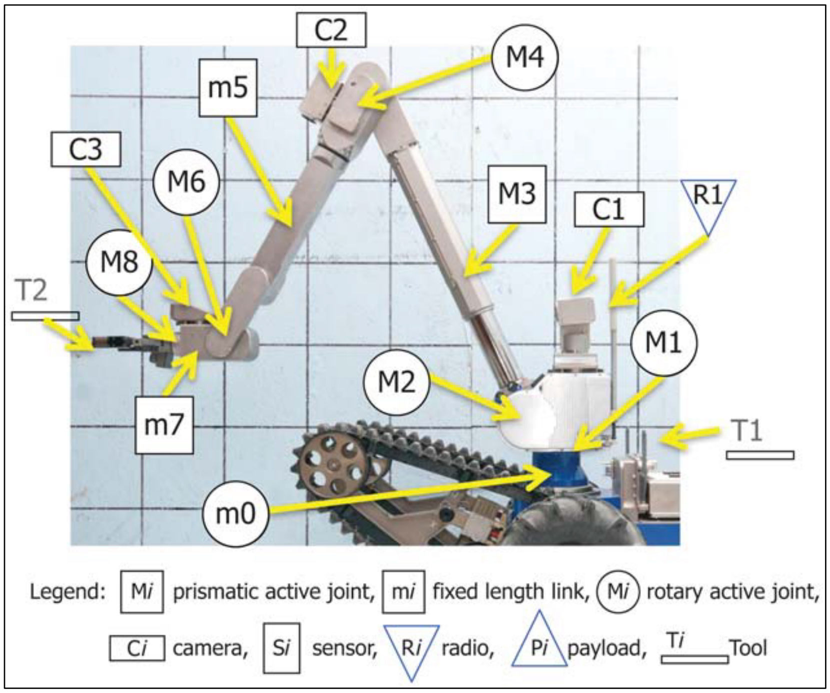


Figure 3. Manipulator degree of freedom and functionality identifier and labeling scheme from ASTM E3132.

UTILIZATION

When conducting evaluations with ASTM E54.09 test methods, a desired mission profile is typically defined, broken down into required capabilities, then matched to the test methods that can be used to measure those capabilities. Given that a response robot is unlikely to change its hardware configuration during the course of a mission downrange – save for when it becomes damaged – the hardware configuration of the robot, as recorded by the ASTM E3132 practice, is typically held static during testing. If the configuration changes (e.g., a robot can only climb stairs when its manipulator is not attached), then the new configuration is recorded and is associated only with the performance data captured while using that configuration. Response robot interfaces generally have many different modes and information presentation techniques available for the operator to choose to use based on what they think will optimize their HRI. So, if the HRI characteristics are to be recorded as described in the previous section, restrictions similar to those for robot hardware configuration changes will not be applicable. Instead, the HRI techniques that are utilized during test method performance can be recorded and associated with that exhibited performance.

The distinction between characterizing what is available and what is utilized is an important one to make, particularly when it comes to operator training. Response robot control is driven by operator decision-making to command the robot in a way that they judge to be appropriate based on the scenario. This is informed by the operator's knowledge of the how the robot system operates, availability of interface and control options, and their understanding of the robot's environment. All of these factors influence one another and can change over time with experience, but more effective and faster training can result from more effective HRI techniques compared to others. An operator may not know the best interface and control configurations to use for a particular scenario, so recording what settings are used when good (and bad) performance is exhibited can only assist in providing more context to replicate that performance. The same could be true when a robot operator is training on a new robot system that has similar HRI characteristics to one they are already familiar with (see Figure 4 for an example of similar HRI characteristics for two different robots performing the same task). The similarity in HRI characteristics may decrease training time for that operator. Correlating exhibited performance with utilized HRI techniques can also lead to the development of best practices for response robot HRI and guides for effective HRI design.

Recording what HRI techniques are utilized can be done for an entire test method performance, or more minutely into specific robot tasks (e.g., maneuvering the robot towards an object to be manipulated, positioning the robot's end effector to grasp an object, etc.) or activities (e.g., inspecting an area to determine the best driving path, changing the robot's pose to prepare for the next part of the task). See Figure 5 for an example of different HRI configurations used for different actions throughout a task. Delineating task performance according to the activities being performed will vary depending on the test method. For example, the ASTM E2991 Mobility: Traverse Sand Terrain standard test method [24] involves driving a figure-8 pattern through the test apparatus comprised of sand; the type of performance exhibited here is largely the same throughout (driving), but could be divided according to the type of driving (straight traversals or turning). In contrast, the ASTM E2804 Mobility: Obstacles: Stairs/Landings standard test method [25] involves ascending and descending a set of stairs; approaching the stairs, positioning the robot to climb stairs (such as angling articulators or changing the position of the manipulator), ascending the stairs, turning around, preparing to descend the stairs, and descending the stairs could all be different segments. More importantly, the distinction between each segment must be generally observable by the human eye, or parts of the apparatus where robot activity changes can be marked to delineate (e.g., a line on the ground separating when the robot should turn around while performing a figure-8).

The faults that are incurred during test method performance – such as when the robot falls off of the rails during the Maneuvering: Align Edges test [15] – can then be associated with each segment of the performed task, and the HRI techniques being used by the operator during that time. Additional faults can also be defined for happenings that do not cause a task repetition to fail, but are considered to be signs of diminished SA (e.g., colliding with a wall, pressing the incorrect button on the input device, etc.). This type of analysis has been performed for large scale robotics competitions to study effective HRI techniques, such as at the DARPA Robotics Challenge wherein a set of standard tasks were broken down into smaller segments (subtasks) and the HRI techniques used by each competitor were categorized according to characteristics of information presentation, input devices, and sensor fusion [23]. Each attempt at a task or subtask was recorded as a success or failure and any critical incidents (i.e., faults) were noted to calculate a percentage of success. The time taken to complete a task or subtask was also recorded. That study produced a set of guidelines for designing HRI for semi-autonomous humanoid robots, which were derived from the correlations between utilized HRI and exhibited performance. Implementing a similar evaluation method during testing may enable similar types of recommendations to be made for effective response robot HRI design.

Telerob Telemax

ICOR Caliber MK4

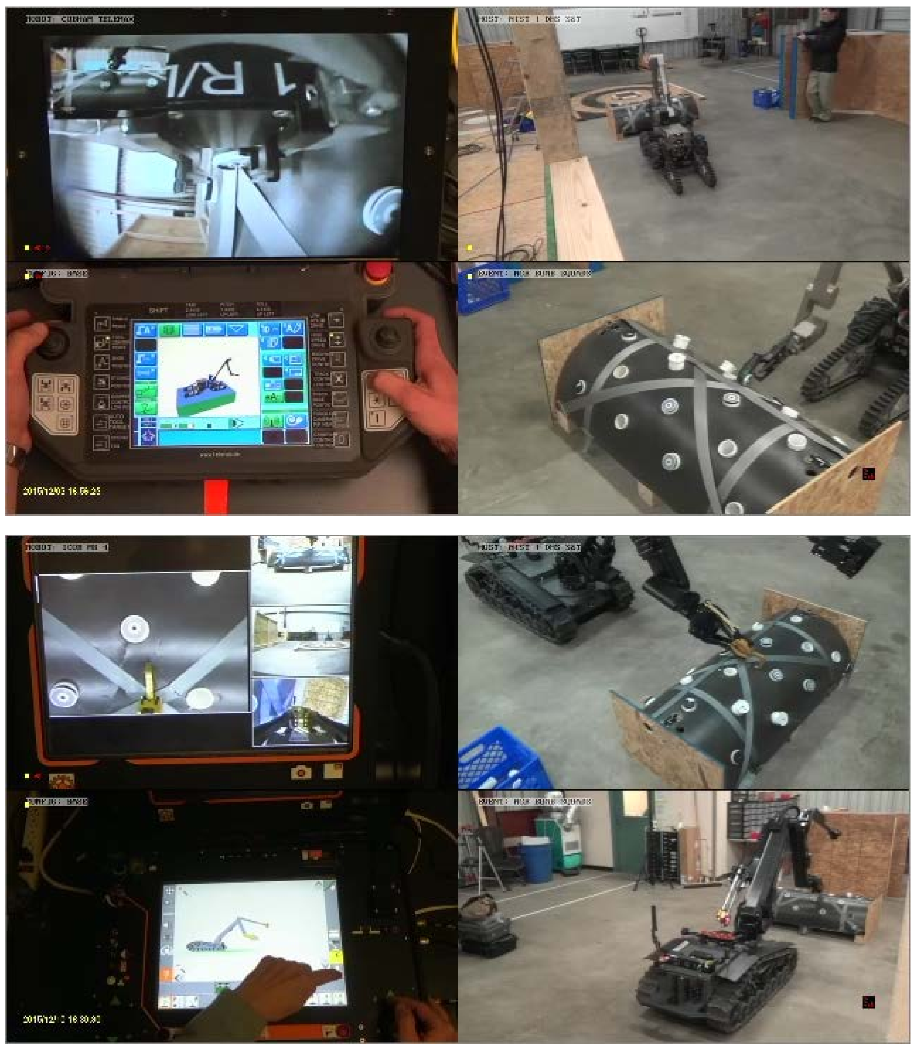
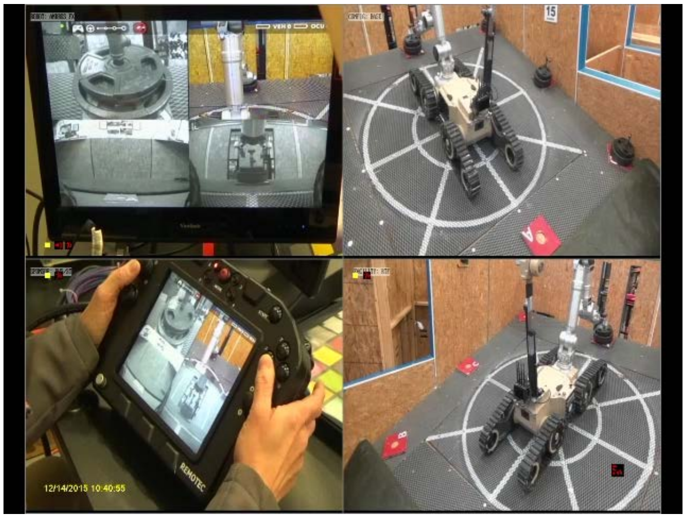


Figure 4. Examples of similar HRI characteristics (camera views, pose information, input device features) being utilized on different robots (Telerob Telemax, ICOR Caliber MK4) performing the same task (ASTM WK54278 Dexterity: Cut Strap [26]).

1. 

Actions:
Grasping and lifting object from start position in front of the robot

Camera views:
Multiple, exocentric (observing arm and gripper in relation to environment) and egocentric (out from body and gripper)

2. 

Actions:
Transferring object to target position behind the robot

Camera views:
Single, exocentric (observing arm and body in relation to environment)

3. 

Actions:
Placing object onto target position behind the robot

Camera views:
Multiple, exocentric (observing arm and gripper in relation to environment) and egocentric (out from body and gripper)

Figure 5. Examples of different HRI characteristics (camera views) being utilized on a robot (Andros FX) at different points throughout a test method (ASTM WK44323 Dexterity: Lift and Place [27]).

Demonstrating Effective HRI

This section details the second proposed effort of this article: the creation of a new standard practice that introduces variability into ASTM E54.09 test methods.

Operators of response robots must maintain SA of the robot, its environment, and the robot's relationship to the environment in order to perform effectively. Performing in one of the ASTM E54.09 standard test methods described previously can demonstrate an operator's ability to maintain SA, particularly those in the Maneuvering suite (e.g., [15]). While performing, an operator makes decisions on the best approach to perform a task with the robot, based on their SA of the robot's status and what actions are needed to perform the task, which includes whether or not the robot is physically capable of performing the task required. The HRI techniques employed by a robot system (see "Characterizing HRI") can assist an operator in gaining proper SA and maintaining it while performing a task. Effective techniques can range from camera placement on the robot and viewing those camera feeds on the interface at the appropriate times (e.g., exocentric camera view to aid in avoiding collisions in confined space) to semi-autonomous functionality such as obstacle avoidance (e.g., modifying input from the operator to drive the robot while preventing it from colliding with obstacles in the environment). The exocentric camera placement technique is very common on response robots, such as on the Endeavor Robotics 510 Packbot [28], while the semi-autonomous functionality technique is more common on robots in other domains, such as the Ava Robotics Ava [29] used for telepresence.

Some environmental factors may be known ahead of time and decisions on which robot to deploy can be made prior to the start of the mission. For example, if the mission takes place in a sandy area, the operator may already know if a robot will have issues with traversal over sand. The current manner in which the ASTM E54.09 test methods are used for operator training involves tuning the settings of the apparatus such that the task is achievable by the robot (e.g., adjusting the height of a target to be manipulated based on the robot's manipulator capabilities). This method is useful for initial training to familiarize the operator with the robot, however it may not sufficiently prepare operators for the potentially unknown elements of a real deployment that have to be dealt with in situ, such as the dimensions of confined spaces that need to be traversed or the dexterity required to interact with an object. The operator must acclimate and understand the requirements of performing the task, the capabilities of the robot, and any misalignments between the two.

A methodology for introducing variability into the ASTM E54.09 test methods is being developed whereby settings of the test method change during the test, varying between task settings that are achievable and those that are not achievable. The use of this methodology is intended to enable effective HRI to be demonstrated by tasking the operator with more explicit decision-making to determine if a task is achievable and, if it is, demonstrating their ability to perform the task. HRI characteristics that can assist the operator in maintaining proper SA to make more informed decisions when attempting to perform under changing conditions can benefit from information presentation techniques on the interface, semi-autonomous functionality and control methods, robot introspection methods, and well-trained, experienced operators. Effective HRI should mitigate errors from the operator/robot and improve task success.

The test methodology aims to:

1. Exercise an operator's ability to demonstrate SA by consistently using contextual information to determine whether the robot they are operating can complete tasks in unknown environments,
2. Exercise an operator's proficiency in performing tasks with their robot in scenarios where the given task is achievable,

3. Reward operators who demonstrate sufficient SA to promote further understanding of the capabilities of the robot, characteristics of the environment, and the relationship between the two, and
4. Reward the usage of robots that utilize HRI techniques which provide the operator with information that aids in furthering their SA through interface modalities and semi-autonomous functionality.

To accomplish this, a set of the ASTM E54.09 test methods that exercise basic system capabilities for maneuvering, mobility, and dexterity have been selected. The test methods chosen are those that require certain dimensional parameters of the task being performed to be tuned based on the robot's characteristics (e.g., a confined space test wherein the width of an aperture is adjusted based on the robot's size). This quality allows for variable conditions to be implemented, including those that are and are not achievable by the robot. Several new test methods are also being developed to expand upon the previously established suite.

VARIABILITY METHODOLOGY

The procedure for administering each test method is altered such that the tunable setting of the test method is varied within a defined threshold for each repetition. Similar to the methodology described by Jones et al. [30], ten unique conditions are generated that vary the tunable setting of each test method and are applied between repetitions in a randomized order. Each condition specifies a setting (S) for the test apparatus and is derived from a base characteristic (C) of the robot performing the test. Determining C for a given robot can either be done by measuring corresponding robot dimensions (e.g., the robot's turning radius), or can be determined in a pre-test (e.g., the widest gap that the robot is able to successfully cross), depending on the test method.

For each condition, an increment (I), based on a percentage of the parameter C, is either added to or subtracted from C, producing the ten unique conditions, five where the value of S increases and five where the value decreases:

$$S = C\text{-}5I, C\text{-}4I, C\text{-}3I, C\text{-}2I, C\text{-}I, C+I, C+2I, C+3I, C+4I, C+5I$$

Of the ten conditions, five are achievable and five are not achievable. For some test methods, those where the value of S is increased will be achievable (e.g., when attempting to traverse through a confined space, increasing the size of the space will make the task less difficult), and for others the inverse is true (e.g., when attempting to climb up a step, increasing the size of the step will make the task more difficult). Each test method will have to be analyzed to determine which conditions are achievable and which are not. This also depends on the value used for I. Internal piloting has shown that 5% is a reasonable value; this means that the conditions with the highest deviation from C (i.e., most or least achievable, $C+5I$ or C-$5I$, depending on the test) is 25% more or less than C. For some tests, deviating further than 25% from C might produce a condition that is too obviously achievable or not achievable, requiring very little work from the operator to judge the scenario.

There are two stages to each repetition: first, the operator must declare if they believe the task to be achievable (judgment). If the operator believes that the task is achievable, they must then attempt it (execution). Each repetition and test (i.e., set of repetitions) can be evaluated across both axes. Performance metrics can be expressed as a percentage of correct judgment and successful completion of tasks. For example, if an operator judged correctly in eight of the repetitions and successfully performed the task on all repetitions that were achievable, their judgment score would be 80% and their execution score would be 100%.

Five filler conditions are also included (similar to [30]) which are randomly picked from the ten unique conditions, resulting in a total of fifteen repetitions per test. Filler conditions are added to discourage an operator from basing their judgments on the number of times they interpret the task as being achievable or not. Even if the operator is aware of the structure of the methodology (i.e., the number of unique conditions that exist), the addition of filler conditions should allow the results of the test to not be impacted. Due to the random picking of what *S* values are used for the filler conditions, performance in those conditions is not evaluated as part of an operator's score. The order of conditions used in each repetition is randomized to prevent the prediction or memorization of the test condition sequence.

APPLICATION TEST METHODS

The methodology is intended to be applicable to many different standard test methods available for response robots. Eight test methods are described here as examples, covering basic ground robot maneuvering, mobility, and dexterity capabilities for response robots; see Figure 6 for an image of each test method apparatus. In this section, each test is described in terms of its tunable settings (*S*), corresponding *C* variables, construction, and fault criteria.

Fit through vertical void

The robot traverses through a vertically confined space without colliding with the overhead boundary. The height between the ground and the overhead boundary varies, set to be taller or shorter than the height of the robot (*C*). This is a newly designed test method. A red panel is affixed to extruded aluminum bars that allow for its height to be adjusted and secured, forming the confined space between it and the ground. If the operator incorrectly judges the height of the void, the robot will collide with the red panel and incur a fault.

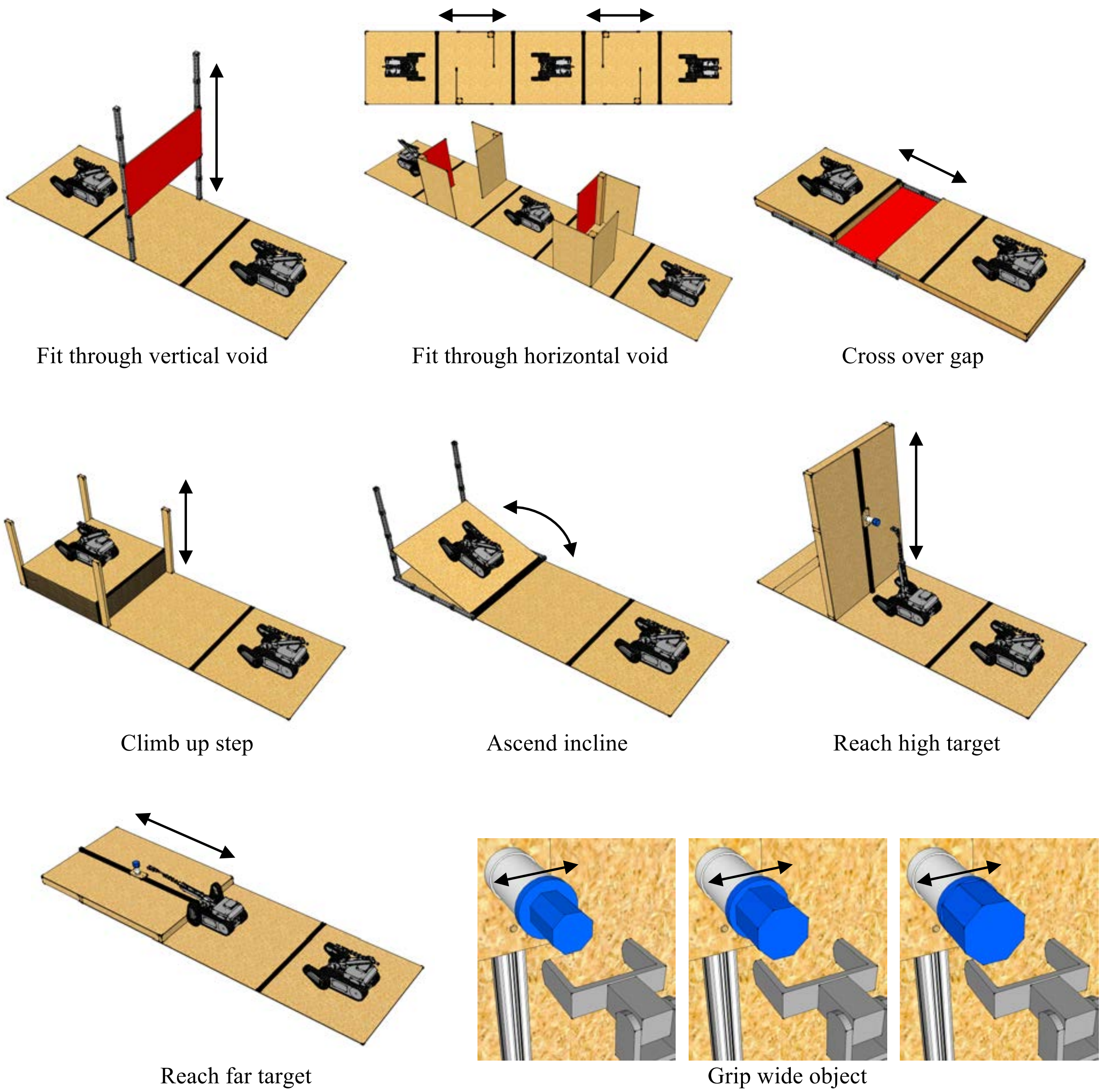


Figure 6. Renderings of eight example test methods with arrows noting which settings (*S*) of the apparatus can be varied during testing.

Fit through horizontal void

The robot traverses through a horizontally confined space without colliding with the boundaries on either side. The width between the wall boundaries varies, set to be wider or narrower than the width of the robot (*C*). This test method is currently under development within the ASTM E54.09 committee. Two pairs of red L-shaped walls define confined spaces that are traversed as left or right turns. If the operator incorrectly judges the width of the void, or struggles to maneuver the robot precisely, the robot will collide with the red walls and incur a fault.

Cross over gap

The robot traverses over a gap in the floor without falling in the gap. The distance between the two ground planes that form the gap varies, set to be larger or smaller than the largest gap that can be crossed by the robot (*C*). This is based on the ASTM E2801 test method [31]. The original ASTM test method design can be used to determine *C*. The start platform is fixed, and a second platform is moved along aluminum extruded bars that allow its position to be adjusted and secured, forming a gap between the platforms and revealing a red ground. If the operator incorrectly judges the length of the void, or struggles to maneuver the robot properly, the robot will fall into the void and touch the red floor, incurring a fault.

Climb up step

The robot traverses up a platform without becoming immobilized. The height of the platform varies, set to be taller or shorter than the tallest platform that can be climbed by the robot (*C*). This is based on the ASTM E2802 test method [32]. The original ASTM test method design can be used to determine *C*. A series of panels are stacked on top of one another in a frame, forming a platform. Panels can be added or removed to adjust the height of the platform. If the operator incorrectly judges the height of the platform, or struggles to maneuver the robot properly, the robot will fall backwards onto the ground. For some robots this can end their operation, in which case a fault would be incurred, while others can flip themselves back over and try again.

Ascend incline

The robot traverses up an inclined ramp from the ground. The angle of the inclined ramp varies, set to be higher or lower than the steepest angle that the robot can traverse (C). This test is based on the ASTM E2803 test method [33]. The original ASTM test method design can be used to determine C. A floor panel is attached to extruded aluminum bars or a winch is used to easily change the angle of the platform. If the angle is too steep, then the robot will not be able to ascend the platform. Depending on the incline, some robots may fall backwards while trying to ascend the ramp. If this ends their operation, then a fault is incurred.

Reach high target

The robot reaches above to interact with a target. The height of the target from the ground varies, set to be higher or lower than the highest target that is reachable by the robot (C). This test is based on either ASTM WK54271 [34], WK54272 [35], or WK54274 [36] based on the task being performed with the target once it is reached (inspect, touch/aim, or extract/grasp, respectively). The original ASTM test method design can be used to determine C. There are also additional considerations for the orientation of the target to face forward, upward, or downward. A PVC pipe target is attached to a vertical extruded aluminum bar to allow the height to be easily changed. A fault is incurred if the PVC pipe is knocked off the apparatus.

Reach far target

The robot reaches forward to interact with a target. The distance of the target from the robot varies, set to be further or closer than the furthest target that is reachable by the robot (*C*). This test is based on the same ASTM test methods as "Reach high target," and has the same considerations for target orientation, apparatus construction (a horizontal bar instead of vertical, and faults). The target is placed on a raised platform so the robot's forward position towards the target is fixed.

Grip wide object

The robot grasps an object. The width of the object varies, set to be larger or smaller than the maximum gripping width of the robot's manipulator (*C*). This is a newly designed test method. A series of octagonal shapes are 3D printed and placed inside of a PVC pipe target, much like the "Reach high/far target" test methods. The 3D printed artifacts could also be used with those test methods for a combination test of both reach length and gripper width.

EXAMPLE PERFORMANCE DATA AND ANALYSIS

Internal exercising of the variability methodology was performed to pilot the concept and collect example performance data. In this example, a tracked response robot with a manipulator, front articulators, and four cameras providing both exocentric and egocentric views was operated via a gamepad controller to perform in the "fit through horizontal void" test method. See Table 3 for the example performance data. Performance in the ten non-filler conditions is used for evaluation. For evaluating the operator's judgment, all ten repetitions are used. The operator correctly judged seven of the ten, resulting in 7/10 = 70% accurate judgment. For evaluating the operator's execution, only the repetitions where they did not skip the execution phase (i.e., they judged, correctly or incorrectly, that the task was achievable) are counted. In this case, eight of the repetitions were judged to be achievable and were therefore attempted. Only five of those repetitions were successfully executed; three of the repetitions that were attempted were in unachievable conditions and therefore were not executed successfully and the other five attempted repetitions were in achievable conditions and were executed successfully, resulting in 5/8 = ~63% successful execution.

Repetition #	Condition	*S* (cm)	Judgment		Execution
			Expectation	Result	
1	*C+I*	23.5	attempt	attempt	success
2	*C+3I**	25.5	attempt	attempt	success
3	*C+4I*	26.5	attempt	attempt	success
4	*C-4I*	18.5	skip	attempt	fail
5	*C-3I*	19.5	skip	attempt	fail
6	*C+2I**	24.5	attempt	attempt	success
7	*C-I*	21.5	skip	skip	---
8	*C-2I*	20.5	skip	attempt	fail
9	*C+2I*	24.5	attempt	attempt	success
10	*C+4I**	26.5	attempt	attempt	success
11	*C+3I*	25.5	attempt	attempt	success
12	*C-5I**	17.5	skip	skip	---
13	*C+5I*	27.5	attempt	attempt	success
14	*C-5I*	17.5	skip	skip	---
15	*C-I**	21.5	skip	attempt	fail

Table 3. Example performance data from the "fit through horizontal void" test method using the variability methodology. For this data, *C* = 22.5 cm and *I* = 1 cm. An asterisk (*) notes a filler condition, during which the operator's performance is not evaluated as part of the test, so the table cells have been shaded.

If the same operator were to continuously perform this test with the same robot utilizing the same HRI techniques, judgment and execution performance may improve due to more experience with the system. This could increase the operator's SA with respect to the manner in which the robot is controlled (i.e., joystick movements translating to actual robot movements as interpreted through camera views) and the robot's dimensional relationship to the environment (i.e., spatial reasoning and understanding of the robot's width and confined spaces). Essentially, in this case, the operator must learn to mitigate their own potential errors because the robot is completely teleoperated and has no autonomous functionality. This is the case for many response robots deployed in the world today.

The errors that were encountered during this test could be attributed to the operator losing, or never properly gaining, SA, which can be further delineated at different levels. According to the levels defined in [4] as applied to this test method, each level would refer to:

- <u>Level 1 SA, perception of the elements in the environment</u>: the dimensions of the walls that defined the horizontal void, the dimensions of the robot, and the robot's status.
- <u>Level 2 SA, comprehension of the current situation</u>: the distance between the walls that define the horizontal void and the outer edges of the robot, and whether or not the robot can fit through the void.
- <u>Level 3 SA, projection of future status</u>: how to command the robot to avoid colliding with the walls that defined the horizontal void while traversing through (if the operator judges that the robot can fit through the void).

Using these levels, performance could be analyzed to further understand where things went wrong. Judging whether or not a condition is achievable is based on level 1 and 2 SA, which can refer to that of the environment or the robot. In the example scenario, the operator can only use spatial reasoning to perceive the dimensions of the horizontal void because only camera data is available. Their knowledge of the robot's dimensions alone may not aid with judgment, but it will when analyzed relative to the environment. The position of the cameras and the camera views utilized by the operator can assist with this aspect, applicable to level 1 SA regarding the operator's knowledge of how to use the robot system. On a more advanced robot system that employs distance sensors, a visualization of the distance between the robot and the walls could support the operator's judgment, or the intelligence to introspect could see the robot alerting the operator that it will not fit through the void, essentially making the judgment for them (similar to the technique described in [37]). Executing a task that is judged by the operator to be achievable is based on level 3 SA; in the example scenario, this is performed entirely by the operator and is informed by their level 1 and 2 SA. If the more advanced robot system previously mentioned were used, semi-autonomous functionality for obstacle avoidance could be utilized to assist the operator in not colliding with the walls.

Discussion and Future Work

In order to develop the HRI characterization and evaluation methods described previously, standard terminology will first need to be developed. There are terms that are used very commonly throughout the HRI literature, but many research developments are focused on more advanced and future-forward elements of HRI. The response robot domain is one that is currently active and has many robots deployed all over the world, so this type of development should be prominent. There are conferences dedicated to HRI research [38], but more advanced and future-forward robot systems are typically covered, such as social and assistive robots. More effort is needed to fully develop HRI for response robots for its affiliated characteristics that differentiate it from the other fields.

The HRI utilization evaluation method described previously (see "Utilization") is not yet very prescriptive. There are many nuances to recording the moment when a particular HRI technique is being used. For instance, if an interface is showing multiple camera views on screen at once, it should not be assumed that the operator is indeed using all of them to perform a task. With that said, if the performance of the task had gone differently – maybe the operator accidentally drops an object from the robot's arm – then the other camera views could be engaged.

The operator's characteristics, such as their experience with similar control systems (e.g., video games, construction equipment), could also be recorded possibly via a survey. As part of operator proficiency training when using the test methods, prior experience with the robot system being tested and the test method being performed should be considered, with considerations such as how the operator's performance metrics have changed over time. Development of the proposed methods should be sensitive to these factors.

The variability methodology (see "Demonstrating Effective HRI") needs to continue to be exercised. Varying task parameters with an expectation of increased or decreased difficulty is not necessarily a linear scale; for example, when applying the methodology to the "reach far target" test method, it is assumed that a target that is closer to the robot is less difficult to grasp. This may not necessarily be the case due to the placement of cameras on the robot or the construction of the robot's manipulator. However, if the robot is able to move around (perhaps within a certain set of boundaries), then the distance can be modified as needed by the operator. Clarification on this point may be needed, including consideration as to whether or not allowing this type of behavior is acceptable. The methodology also is currently only dealing with difficulty that is based on physical limitations of the robot system. The same type of difficulty could be introduced on the human's capabilities, the robot's sensor's capabilities to interpret the challenge, etc. These avenues can be further explored for possible methods to continue to vary test settings.

The SA analysis discussed previously (see "Example Performance Data") could be used as part of a study to develop more detailed guidelines for effective response robot HRI. This could be done by combining the HRI characterization techniques with the variability methodology: record the robot's available HRI configuration, run a test using the variable methodology, record what HRI techniques are utilized for each phase of task performance, and correlate performance with the HRI to determine what aspects of HRI are working properly and why. If faults occur during certain points of task execution – as related to the different levels of SA – can it be discerned where errors occurred and why. The method by which these efforts are able to be combined and work simultaneously will continue to be explored.

The efforts described in this article are continuing to be developed such that they can be proposed as potential standards for consideration by the ASTM E54.09 committee. More specifically, the proposals are:

1. Expand the ASTM E3132 standard practice for recording robot configuration to include HRI characterization and method for recording which of those HRI techniques were utilized during test method performance, and
2. Develop standard practice for implementing variable test method settings to further evaluate operator proficiency in task judgment and execution.

Conclusions

This article reviews two proposals for potential standards: one for expanding the current standard practice for recording robot configuration to include HRI characteristics and associated performance with them (see "Characterizing HRI"), and another for introducing variability into test method settings in order to elicit decision-making from the operator and allow effective HRI to be demonstrated (see "Demonstrating Effective HRI"). The goal of these efforts is to highlight the importance of HRI in the world of response robots. Robot operators can benefit greatly from understanding how to best utilize the HRI options available to them for a given scenario, and more advanced HRI techniques such as the introduction of semi-autonomous functionality could make response robot deployments even more effective.

ACKNOWLEDGMENTS

This research has been supported in part by the National Institute of Standards and Technology (NIST) under award 60NANB14D286. The authors would like to thank Adam Jacoff, Jennifer Marshall, and Ann Virts from the National Institute of Standards and Technology (NIST) for their support and assistance through the ASTM E54.09 subcommittee.

References

[1] J. Carlson and R. R. Murphy, "How UGVs Physically Fail in the Field," *IEEE Transactions on Robotics,* vol. 21, no. 3, pp. 423-437, 2005.

[2] ASTM, "Subcommittee E54.09 on Response Robots," ASTM International, 2018. [Online]. Available: https://www.astm.org/COMMIT/SUBCOMMIT/E5409.htm. [Accessed August 2018].

[3] J. Scholtz, "Theory and Evaluation of Human Robot Interactions," in *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003.

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

[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 49th Annual Meeting*, 2005.

[6] Northrup Grumman, "Andros F6B: The Industry's Most Versatile Platform," Northrup Grumman, 2018. [Online]. Available: https://www.northropgrumman.com/Capabilities/Remotec/Products/Pages/F6B.aspx. [Accessed August 2018].

[7] Telerob, "telemax PRO," Telerob, 2018. [Online]. Available: https://www.telerob.com/en/products/telemax-family/telemax-pro. [Accessed August 2018].

[8] M. Endsley, "Automation and situation awareness," in *Automation and Human Performance: Theory and Applications*, 1996, pp. 163-181.

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

[10] J. Nielsen, "10 Usability Heuristics for User Interface Design," Nielsen Norman Group, 1995.

[11] C. W. Nielsen, M. Goodrich and R. W. Wicks, "Ecological interfaces for improving mobile robot teleoperation," *IEEE Transactions on Robotics,* vol. 23, no. 5, pp. 927-941, 2007.

[12] E. Clarkson and R. C. Arkin, "Applying Heuristic Evaluation to Human-Robot Interaction Systems," in *FLAIRS Conference*, 2007.

[13] M. A. Goodrich and D. R. Olsen, "Seven Principles of Efficient Human Robot Interaction," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2003.

[14] EMSHRI, "Third International Workshop on Evaluation Methods Standardization for Human-Robot Interaction," EMSHRI 2017, 2017. [Online]. Available: https://sites.google.com/view/emshri2017/home. [Accessed August 2018].

[15] ASTM, *Draft Standard WK53649, New Test Method for Evaluating Ground Response Robot Maneuvering: Align Edges, under development of Subcommittee E54.09 on Response Robots.*

[16] R. Sheh, S. Schwertfeger and A. Visser, "16 Years of RoboCup Rescue," *KI - Künstliche Intelligenz,* vol. 30, no. 3-4, p. 267–277, 2016.

[17] A. Jacoff, A. Virts and K. Saidi, "Counter-Improvised Explosive Device Training Using Standard Test Methods for Response Robots," National Institute of Standards and Technology, 2015.

[18] H. A. Yanco and J. L. Drury, "Classifying human-robot interaction: an updated taxonomy," *IEEE International Conference on Systems, Man & Cybernetics,* vol. 3, pp. 2841-2846, 2004.

[19] 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*, 2006.

[20] ASTM, *ASTM E3132 / E3132M-17, Standard Practice for Evaluating Response Robot Logistics: System Configuration, ASTM International, West Conshohocken, PA, 2017, www.astm.org.*

[21] 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*, 2006.

[22] H. A. Yanco and J. L. Drury, "Rescuing Interfaces: A Multi-Year Study of Human-Robot Interaction at the AAAI Robot Rescue Competition," *Autonomous Robots,* vol. 22, no. 4, pp. 333-352, 2007.

[23] A. Norton, W. Ober, L. Baraniecki, E. McCann, J. Scholtz, D. Shane, A. Skinner, R. Watson and H. Yanco, "Analysis of Human-Robot Interaction at the DARPA Robotics Challenge Finals," *International Journal of Robotics Research,* vol. 36, no. 5-7, pp. 483-513, 2017.

[24] ASTM, *ASTM E2991 / E2991M-17, Standard Test Method for Evaluating Response Robot Mobility: Traverse Gravel Terrain, ASTM International, West Conshohocken, PA, 2017, www.astm.org.*

[25] ASTM, *ASTM E2804-11, Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Stairs/Landings, ASTM International, West Conshohocken, PA, 2011, www.astm.org.*

[26] ASTM, *Draft Standard WK54278, New Test Method for Evaluating Ground Response Robot Capabilities: Manipulation: Cut Strap, under development of Subcommittee E54.09 on Response Robots.*

[27] ASTM, *Draft Standard WK44323, New Test Method for Evaluating Ground Response Robot Dexterity: Lift and Place, under development of Subcommittee E54.09 on Response Robots.*

[28] Endeavor Robotics, "Endeavor Robotics Products," 2018. [Online]. Available: http://endeavorrobotics.com/products#510-packbot. [Accessed August 2018].

[29] Ava Robotics, "Ava Robotics," 2018. [Online]. Available: https://www.avarobotics.com/. [Accessed August 2018].

[30] K. S. Jones, B. R. Johnson and E. A. Schmidlin, "Teleoperation Through Apertures: Passability Versus Driveability," *Journal of Cognitive Engineering and Decision Making,* vol. 5, no. 1, pp. 10-28, 2011.

[31] ASTM, *ASTM E2801-11, Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Gaps, ASTM International, West Conshohocken, PA, 2011, www.astm.org.*

[32] ASTM, *ASTM E2802-11, Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Hurdles, ASTM International, West Conshohocken, PA, 2011, www.astm.org.*

[33] ASTM, *ASTM E2803-11, Standard Test Method for Evaluating Emergency Response Robot Capabilities: Mobility: Confined Area Obstacles: Inclined Planes, ASTM International, West Conshohocken, PA, 2011, www.astm.org.*

[34] ASTM, *Draft Standard WK54271, New Test Method for Evaluating Ground Response Robot Dexterity: Inspect, under development of Subcommittee E54.09 on Response Robots.*

[35] ASTM, *Draft Standard WK54272, New Test Method for Evaluating Ground Response Robot Dexterity: Touch or Aim, under development of Subcommittee E54.09 on Response Robots.*

[36] ASTM, *Draft Standard WK54274, New Test Method for Evaluating Ground Response Robot Dexterity: Extract and Place, under development of Subcommittee E54.09 on Response Robots.*

[37] H. Kono, "Warning System and Storing Method of Perilous environment for a Teleoperated Robot Operator," *International Journal of Control Systems and Robotics,* vol. 2, pp. 16-22, 2017.

[38] Human-Robot Interaction, "Conference | Human-Robot Interaction," 2018. [Online]. Available: http://humanrobotinteraction.org/category/conference/. [Accessed August 2018].