A Human-Robot Interaction Testbed for Collaborative Robotics: Development Considerations and Test Method Concepts

Adam Norton and Holly Yanco

Abstract— **While standards for robot manipulator safety and test methods for performance are in development, the same considerations have not been applied to the collaborative elements between human and robot agents in manufacturing environments. To this end, we propose a testbed, specifying artifacts, apparatus, procedures, and metrics, for evaluating the performance of collaborative human-robot interaction. The testbed will be used to measure the capabilities of a collaborative robotic manipulator system (arm, end effector, and sensing) alongside a human, both for performance of task instruction from human to robot, and the resulting task execution performance. Considerations for development of the testbed and test method concepts are discussed.**

I. Introduction

Robots working alongside humans in manufacturing environments (i.e., sharing work cells) are becoming more prevalent. Evaluations of a collaborative robotics system (CRS) in this regard typically are focused on task performance throughput and safety [7]. Human-robot interaction (HRI) research has some standard procedures with respect to measurement techniques [16], but very little in terms of standardized experiment set-ups. A series of workshops are dedicated to this subject [5]. We propose a testbed of artifacts, apparatus, procedures, and metrics to be used for standardized and comparative evaluation of human-robot collaboration (HRC), specifically with robotic manipulators. In this paper, we discuss considerations for development and concepts for test methods.

II. Related Work

Test methods for industrial manipulator performance have been developed which evaluate elemental grasping [6] and functional assembly tasks [13]. Those test methods use simple artifacts to measure capabilities such as grasp strength and fastening, among others. There are also common benchmarks that are used throughout the robotic manipulation community such as the YCB Object and Model Set [4] and the Dex-Net 2.0 [11] performance dataset. The proposed testbed will leverage these existing efforts and will provide artifacts, apparatus, procedures, and metrics. This is similar to the test methods developed through ASTM E54.09 [1] and ASTM F45 [2], which abstract real world tasks that are used to measure robot capabilities. The standard test methods in those committees only specify measurement techniques; they do not specify performance thresholds.

Adam Norton is with the New England Robotics Validation and Experimentation (NERVE) Center, University of Massachusetts Lowell, Lowell, MA 01854 USA (adam_norton@uml.edu).

Holly Yanco is with the Computer Science Department, University of Massachusetts Lowell, Lowell, MA 01854 USA (holly_yanco@uml.edu).

III. Scope

The collaborative HRI testbed will focus initially on manufacturing and industrial environments where robotic manipulators are used to perform tasks alongside a human that are either shared between both agents or dependent on one another's participation. A CRS that can be used with the testbed will have a robotic manipulator with an end effector and sensors. It may also have knowledge representation of the human's actions and/or task progression. Performance of task instruction from human to robot and the execution of the resulting task will be considered. Both processes are tied together, and will be evaluated as such.

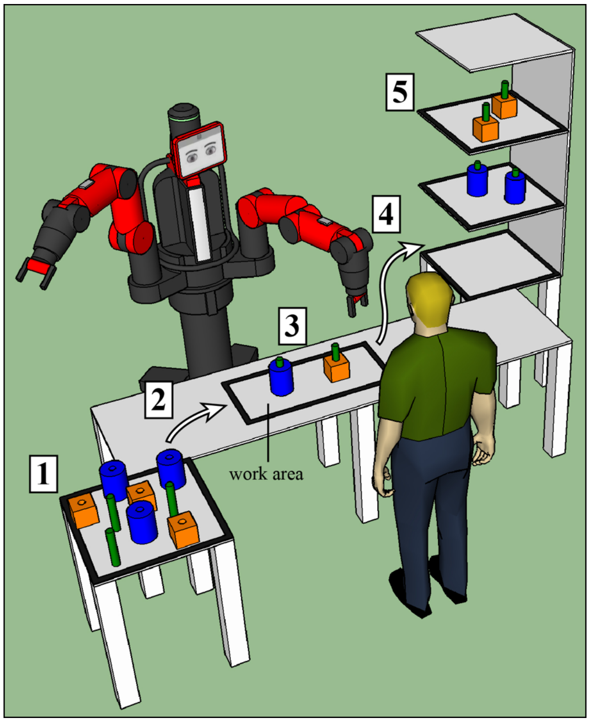
The testbed will comprise a series of test methods that will be designed for testing the performance of all HRC elements: the CRS (arm, end effector, sensing), the human, and the task. The test methods will be designed such that metrics can be derived for both task-based, holistic evaluation of the entire collaboration (e.g., task throughput, efficiency) and evaluation of the performance of individual elements in the context of the entire collaboration (e.g., the effect of the robustness of an object detection sensor on task execution, alleviated by human intervention).

The development of the testbed will be guided by a characterization of CRS, the tasks they perform, and the conditions they perform in, to produce a taxonomy of common CRS elements. The taxonomy will drive the design of a set of malleable test artifacts, apparatus, procedures, and metrics that can be used to define various test methods, as previously done for obstacle avoidance test methods [14]. The test methods can be scaled and adjusted to fit varying configurations of CRS and task processes. The testbed will be fabricated using readily available materials such as aluminum tubing, wood, and 3D-printed models. The test artifacts will be designed such that they can be used to exercise challenges for task instruction and task execution.

IV. Development Considerations

The characterization taxonomy will drive the development of the testbed with four main outputs: physical elements of the testbed in the form of artifacts and apparatus, the procedures to carry out experiments, and the metrics to evaluate performance in the testbed. Throughout this section, an example HRC scenario is provided to illustrate each output. The scenario involves a collaborative assemble, pick, and place task, where the human assembles an object, places it into the work area for the CRS to detect, and the CRS organizes it based on the object's shape (see Figure 1). A scenario with reserved roles could also apply wherein the CRS assembles the object and the human organizes it.

Figure 1. The example HRC scenario in five stages, rendered in a notional test method apparatus. (1) The human grabs the artifacts from the start location in the apparatus. (2) The human places the green peg into the blue cylinder or orange block to create one of two possible unique assembled artifacts and places them in the work area. (3) The CRS detects the assembled artifact in the work area and grasps it. (4) The CRS lifts and transfers the combined artifact towards the possible end locations. (5) The CRS places the artifact on the shelf.



A. Characterization Taxonomy

The characterization taxonomy will be developed by conducting a literature review of relevant research papers and publications. The review will include papers that detail interaction design approaches for HRI/HRC (e.g., [9], [12]), surveys on the subject (e.g., [3], [10]), and existing taxonomies (e.g., HRI [15], robotic assembly [17]). From this literature review, we will extract the characteristics that are most relevant to CRS instruction and task execution. This includes characteristics for manipulator degrees of freedom (DOF), end effector type (e.g., gripper, tool), and sensor data types (e.g., 2D image, 3D point cloud). There will also be a breakdown of human roles and actions (e.g., adaptation to errors). For characterization of tasks to be performed, both generic (e.g., simultaneous or sequential actions) and specific elements (e.g., instruction type: kinesthetic teaching, learning by demonstration; end effector movements: pick and place, insert, slide) will be included. The qualities of individual components as well as the relationships and dependencies between them will also be covered.

The taxonomy will only feature the characteristics that can affect HRC performance as derived from the literature review. The variance in these characteristics between different CRS provides a possible area for developing a test method. The structure of the taxonomy may also be used as the basis for a recording technique to capture the configuration of a CRS prior to experimentation.

B. Artifacts

The artifacts are representative objects that the human and CRS interacts with, either to grasp and move around, use with a tool to change the artifact's state (e.g., sand or grind down), and/or are used in conjunction with another artifact (e.g., fastening two objects together). The configuration of the CRS being tested will influence how the artifacts are rendered with respect to color, shape, texture, dimensions, and mass. Depending on the use case, artifacts can be designed around a set of rules regarding how they are interacted with (e.g., must maintain orientation).

Using the example HRC scenario in Figure 1, there are three artifacts (stage 1): a blue cylinder with a hole, an orange block with a hole, and a green peg that fits the holes. By placing the peg in either of the holes to assemble the artifacts together (stage 2), the human can create two unique assembled artifacts. The two artifacts are held together by gravity. The characteristics of the assembled artifact influence performance parameters: it cannot be grasped by the peg only (stage 3) because the cylinder or block will fall out, it cannot be turned over during transfer (stage 4) because the peg will fall out, and the two unique assembled artifacts must be organized appropriately (stage 5). All of these parameters can be conveyed from the human to the CRS during the task instruction phase. Possible avenues for evaluation include the instruction of these parameters and adherence to these parameters during task execution.

C. Apparatus

The environment that contains the CRS testbed will be comprised of apparatus that are representative of collaborative manufacturing settings. For our purposes, this largely refers to the geography of the testbed, such as the position of the human, robot, and task workspace in relation to each other. The task workspace includes the start, mid, and end states or locations of the artifacts that make up the steps to be taken to perform the task. The dimensions and configuration of the apparatus may restrict where certain interactions can take place. It can also obstruct approach angles for certain actions (e.g., picking an artifact off of a tabletop vs. out of a box).

The notional apparatus in Figure 1 shows a start location that contains artifacts (stage 1), a mid location for the artifacts to be interacted with by both agents (stage 3), and the end location for the artifacts (stage 5). The end location has multiple possible zones that can be used for measuring performance (i.e., if the CRS places the wrong assembled artifact on the wrong shelf). It also has vertical obstructions that restricts how the CRS is able to interact with it (e.g., approach trajectory must be parallel to the shelf). The obstructions could be detected by the CRS during operation, or the movements to place the assembled artifacts safely in each area can be programmed during the instruction phase.

D. Procedures

A set of commands for the human and robot to follow to perform the task in a measurable way is a procedure. We envision an underlying set of procedures that are agnostic to task specifics (e.g., grasping vs. welding), but capture the relevant qualities of the collaborative task scenarios. For instance, a set of testbed procedures may focus on handoff tasks where the human and robot are passing artifacts to one another, while another may involve the shared handling of a large artifact and placing it somewhere in the apparatus. Procedures will be defined for the instruction phase and the execution phase of task performance.

The procedure for performing a test in the example HRC scenario will first involve instructing the robot how to perform the task. This includes conveying the parameters for grasping the assembled artifacts as previously described, how to do so safely (i.e., when the human's hands are not in the way), and transferring to the appropriate end location. After instruction, the procedure will involve executing the task for a number of repetitions and recording performance. The manner in which the human agent performs can be scripted to exercise certain error states (e.g., the human intentionally places unassembled objects into the CRS work area) or left to perform "naturally," noting any errors incurred along the way. The procedure will vary depending on the test method.

E. Metrics

Measures related to task performance (e.g., speed, accuracy, time to recover from faults; individually for both agents and the task itself), task instruction (e.g., "amount" of instruction; number of steps, time), and correlations between the two will be implemented. The metric for "amount" of instruction could compare the number of manually instructed steps (e.g., using a teach pendant), those occurring more "naturally" between the human and the CRS (e.g., techniques like learning by demonstration), and those performed more autonomously by the CRS (e.g., scanning the area and to plan an optimal grasp, as is done in [8]). Metrics can also be developed to assess the "efficiency" of the collaboration. This could be expressed by comparing optimal task performance (i.e., both agents perform perfectly) with suboptimal task performance (i.e., one or both of the agents make mistakes that must be rectified). Additionally, measures such as ground truth data of the human and robot can be gathered through motion capture and/or wearable IMUs. The artifacts and apparatus designs will also contain inherent metrics related to their physical qualities, such as artifact size reflecting functional gripper width.

The design of the example HRC scenario contains some inherent metrics as they pertain to CRS task performance, such as proper grasp, transfer, and organization of the assembled artifacts. Using the proposed "collaboration efficiency" metric described previously, one test could involve the human making no mistakes when assembling and placing the artifacts in the CRS work area. The throughput or rate of performance of this test could be compared to another test where the human intentionally performs suboptimally. The CRS could correct for these mistakes, such as by adapting its behavior and/or alerting the human, to maintain a desired throughput measure.

V. Test Method Concepts

While development is still underway, concepts for test methods can be discussed using the considerations described previously to illustrate the testbed. A few test method concepts are described in this section, as well as the manner in which each test method can be applied to the example HRC scenario. The capabilities being evaluated in each test method are not mutually exclusive from one another. Rather, they use the example HRC scenario as the setting, where an experiment could take place to perform the assemble, pick, and place task. Depending on the desired evaluation, test methods like those described below could be used to elicit different types of performance.

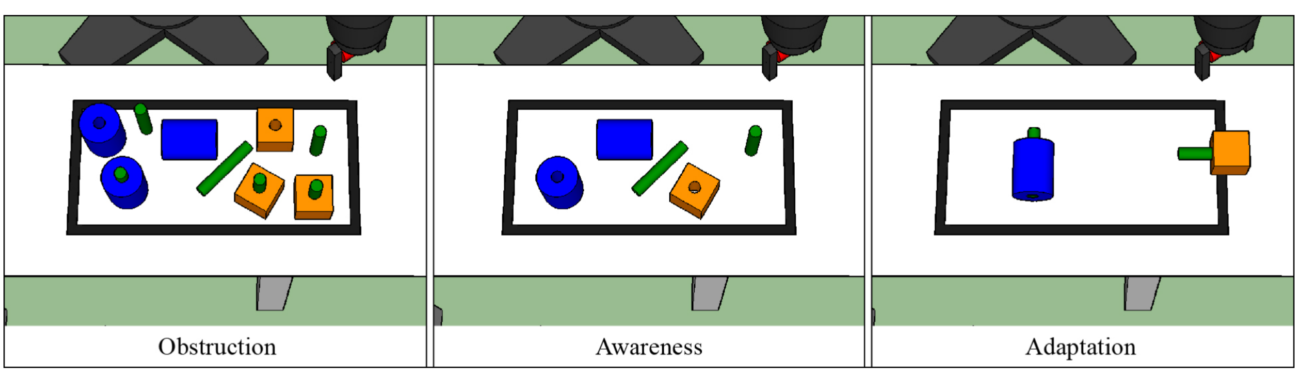
These test method concepts and more will continue to be developed as the characterization taxonomy is generated.

A. Obstruction

The layout of the apparatus and artifacts can require certain arm positions and end effector approach angles that could obstruct the robot's vision system, impacting task performance. Depending on how many DOF the robot arm has, an optimal position and approach could be used that is more beneficial to task performance, either in terms of productivity, user convenience, or both.

In the example HRC scenario, the human may assemble the artifacts at the start location, mid-air, or in the CRS work area. The work area could become obstructed if too many unassembled and assembled artifacts are present (see Figure 2). While the CRS is trying to detect the assembled artifacts in the work area, performance could be affected by increasing detection time or it could halt task progression if grasping an assembled artifact may not be achievable until the

Figure 2. Three test method concepts using the example HRC scenario depicted in Figure 1, showing the CRS workspace (stage 3). Left: Too many artifacts obstruct the work area. Center: The artifacts in the work area are not assembled. Right: The assembled artifacts are not oriented correctly or are not placed correctly inside of the work area boundary.



obstructions are moved. This type of condition could be scripted to occur in the procedure or it could happen naturally if the human performs faster than the CRS.

B. Awareness

For any given task, an artifact must be present with which the CRS end effector can interact. If the artifact is not present, the CRS could be aware and alert the human. Or, if an artifact is presented that the CRS is not able to handle (e.g., too big, too small, not what it was programmed for), the same reaction could occur.

The human is responsible for assembling the artifacts in the example HRC scenario. If he/she makes a mistake in assembling, such as placing unassembled artifacts in the work area (see Figure 2), the CRS could be aware of this error. The reaction of the CRS could be to pause task progression and/or alert the human to fix their mistake. If the CRS is not aware that an error has occurred, it could continue executing the task. The performance of the CRS could depend on how it was instructed to perform the task (e.g., the human did not program it to look for the green peg in the artifact) or a limitation of the CRS cognition.

C. Adaptation

If the human makes an error during task execution (e.g., hands an artifact to the robot in the wrong orientation), the CRS could adapt its behavior to correct the situation (e.g., grasp the artifact differently such that its orientation is corrected) and/or alert the human to the error so the human can adapt his or her behavior.

A parameter of the example HRC scenario is that the assembled artifacts need to be kept in an upright orientation. If the human does not adhere to this parameter (see Figure 2), such as by placing the assembled artifacts on their side, the CRS could detect this anomaly and adjust its grasp, or lift and transfer trajectories to reorient the assembled artifact. If the human does not place the assembled artifact in the right spot for the CRS to grasp (e.g., outside of the work area), the CRS could increase its detection area to look for it. Depending on other parameters of task performance, the CRS could continue to grasp the assembled artifact (if it were in reach) or alert the operator to move it into the work area.

VI. Conclusion

This paper presents a collaborative HRI testbed that is in its conceptual stage. More research and literature review is required to properly characterize CRS such that the taxonomy can be generated. The considerations presented here will be used during the design cycle to develop concepts for test methods, eventually to be prototyped and piloted with a series of CRS. While existing metrics commonly used for HRI and task performance can be applied in this testbed, more investigation is needed to develop new metrics that are specific to HRC. For example, those that consider task instruction, specifically how to characterize task instruction to measure its impact on the resulting task execution. Such a metric could aid in guiding developments for optimal HRI/HRC techniques.

References

[1] ASTM, "Subcommittee E54.09 on Response Robots," https://www.astm.org/COMMIT/SUBCOMMIT/E5409.htm, 2017.

[2] ASTM, "Committee F45 on Driverless Automatic Guided Industrial Vehicles," https://www.astm.org/COMMIT/SCOPES/F45.htm, 2017.

[3] A. Bauer, D. Wollherr, and M. Buss, "Human–robot collaboration: a survey," *International Journal of Humanoid Robotics*, vol. 5(1), 2008, pp. 47-66.

[4] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The YCB object and model set: Towards common benchmarks for manipulation research," *IEEE International Conference on Advanced Robotics (ICAR)*, July 2015, pp. 510-517.

[5] M. E. G. Dalgaard, M. Giuliani, T. Haidegger, A. Tapus, and G. S. Virk, "Workshop: Towards Standardized Experiments in HRI," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 2015.

[6] J. Falco, K. Van Wyk, S. Liu, and S. Carpin, "Grasping the performance: Facilitating replicable performance measures via benchmarking and standardized methodologies." *IEEE Robotics & Automation Magazine* 22, no. 4, December 2015, pp. 125-136.

[7] J. Fryman and B. Matthias, "Safety of industrial robots: From conventional to collaborative applications," *In Proceedings of 7th German Conference on Robotics; ROBOTIK*, May 2012, pp. 1-5.

[8] M. Gualtieri, A. ten Pas, K. Saenko, and R. Platt, "High precision grasp pose detection in dense clutter," *In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2016, pp. 598-605.

[9] W. Kim, J. Lee, L. Peternel, N. Tsagarakis, and A. Ajoudani, "Anticipatory Robot Assistance for the Prevention of Human Static Joint Overloading in Human-Robot Collaboration,". *IEEE Robotics and Automation Letters*, vol. 3(1), July 2017, pp. 68-75.

[10] P. A. Lasota, T. Fong, and J. A. Shah, "A Survey of Methods for Safe Human-Robot Interaction," *Foundations and Trends in Robotics*, vol. 5(4), May 2017, pp. 261-349.

[11] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. Ojea, and K. Goldberg, "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics," *In Proceedings of Robotics: Science and Systems*, July 2017.

[12] S. Nikolaidis, P. Lasota, G. Rossano, C. Martinez, T. Fuhlbrigge, and J. Shah, "Human-robot collaboration in manufacturing: Quantitative evaluation of predictable, convergent joint action," *In Proceedings of the 44th IEEE International Symposium on Robotics (ISR)*, October 2013, pp. 1-6.

[13] NIST, "Performance Metrics and Benchmarks to Advance the State of Robotic Assembly," https://www.nist.gov/programs-projects/performance-metrics-and-benchmarks-advance-state-robotic-assembly, 2017.

[14] A. Norton and H. Yanco. "Preliminary Development of a Test Method for Obstacle Detection and Avoidance in Industrial Environments," *Autonomous Industrial Vehicles: From the Laboratory to the Factory Floor (ASTM STP1594)*, April 2016, pp. 23-40.

[15] M. O. Shneier, E. R. Messina, C. I. Schlenoff, F. M. Proctor, T. R. Kramer, and J. A. Falco, "Measuring and Representing the Performance of Manufacturing Assembly Robots," *NIST Interagency/Internal Report (NISTIR)-8090*, 2015.

[16] 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, pp. 33-40.

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