An Introduction to the MASCAR (Multi-Agent Synchronized Collaborative Assembly Replication) Testbed

Megan Zimmerman*
U.S. National Institute of Standards and Technology
Gaithersburg, MD, USA

Jeremy Marvel
U.S. National Institute of Standards and Technology
Gaithersburg, MD, USA

Shelly Bagchi
U.S. National Institute of Standards and Technology
Gaithersburg, MD, USA

Snehesh Shrestha
University of Maryland, College Park, MD, USA
U.S. National Institute of Standards and Technology, Gaithersburg, USA

Abstract

A purpose-built testbed for human-robot interaction (HRI) metrology is introduced and discussed. This testbed integrates multiple sensor systems and precision manufacturing to produce high-quality HRI datasets of human volunteers working with robots to complete collaborative tasks in a shared environment. Sensors include audio, video, motion capture, robot information, and user entries, and may also incorporate task-specific object tracking. Data collected will be replicable in identical testbeds, and will enable more robust findings in future HRI studies.

CCS Concepts

• **Computer systems organization** → **Robotics**; • **Human-centered computing** → **User studies**; • **Information systems** → Data replication tools.

Keywords

testbed, replicability, reproducibility, human-robot interaction

ACM Reference Format:
Megan Zimmerman, Jeremy Marvel, Shelly Bagchi, and Snehesh Shrestha. 2026. An Introduction to the MASCAR (Multi-Agent Synchronized Collaborative Assembly Replication) Testbed. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI Companion '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3776734.3794511

1 Introduction

Repeatability and reproducibility continue to present fundamental challenges within the human-robot interaction (HRI) research community. Overwhelmingly, publications surrounding datasets, designs of experiments, and even systems descriptions lack the detail and metadata sufficient for the replication and extension of HRI research to one's own environment and/or research domain. While existing documentation is important and meaningful, it is not well-suited for generalizability and reproducibility, which is holding back future advances.

To address this limitation, the National Institute of Standards and Technology (NIST) is working to develop a reporting methodology that establishes the vital information necessary to enable research replication, data scalability, and result generalization. This report reflects an application of this methodology to a new HRI testbed designed and built by NIST to serve as a data collection, reporting apparatus, and validation tool. In particular, details are provided that thoroughly document design decisions, product[1] selection criteria, impetuses and results of iterative changes, and the overall design settled on as a result of this process.

The Multi-Agent Synchronized Collaborative Assembly Replication (MASCAR) testbed provides a traceable and repeatable means of collecting human-robot interaction data for the generation of high-quality datasets of human-robot interactions. "Traceability" in this context reflects the expectation that all components of the testbed, their capabilities (and limitations), and the data they produce (including measurement uncertainties) are all thoroughly documented. "Repeatability" specifically refers to providing the metadata (including formatting and installation details) in the datasets that captures the nuances of our specific installation (configuration, settings, environmental conditions, etc.) such that researchers can exactly replicate, model, or modify NIST-generated datasets.

2 Motivation

Within the field of HRI there exists a fundamental issue regarding research replicability and reproducibility [2, 4]. This replicability crisis is not unique to HRI, and is inherent within many fields of research. Although an individual laboratory cannot single-handedly address the replicability crisis, it can serve as an example to encourage others within the field to do the same. To accomplish this, it is necessary to hold oneself accountable and adhere to practices that address replicability and reproducibility. Similarly, while a single laboratory cannot force others to repeat its work as a step towards proper verification and validation, it should present its work in a way that enables and encourages replication. To this end, it is our goal to add full transparency to our own ongoing research, and present the design and structure of our own testbeds and study plans.

Given our goals of putting into practice the recommendation for full documentation and disclosure of testing conditions and enabling research replication and reproduction, we present the

*Corresponding author contact: megan.zimmerman@nist.gov

HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3794511

[1]Certain commercial instruments are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

end result of a year's effort toward designing and documenting a purpose-built testbed. Throughout the design and build processes, we were guided by three principal design questions:

What applications could this testbed potentially be used for?
What research groups might use the resulting datasets?
What data will we need as a base requirement, and what data will be potentially useful for future studies by external researchers?

With the intended goal of supporting HRI research repeatability and dataset generation, we set out to design and document a testbed specifically intended for HRI teaming research, featuring portability and adjustability to accommodate the variability of humans and tasks. This testbed is modular to accommodate the rigid placement and localization of dedicated sensors and tools for human- and task-monitoring applications.

Additionally, the development and documentation processes are intended to address issues and limitations with many studies. Specifically:

- Construct the testbed based on the needs of the research question, rather than being limited by the equipment on hand;
- Clearly establishing the research space requirements for activities involving the general public;
- Focus on leveraging well-documented robots, equipment, controls, sensors, software, and test methods;
- Relegate custom-built solutions to those cases where established products are unavailable, unless said solutions are the focus of the study;
- Regardless, tolerances and uncertainties need to be known *a priori*, rather than measured and justified *post hoc*;
- Documentation of the design and functionality of testbeds frequently focuses only on those aspects that are distinct from preceding work, but design considerations that were inspired or derived from prior research are often omitted, effectively losing the context for design decisions.

To aid in this process, NIST collaborated with researchers at Michigan Technological University to design testbeds for HRI research in parallel. As components were integrated and data collected, design notes and solutions were shared to facilitate and accelerate technology transfer.

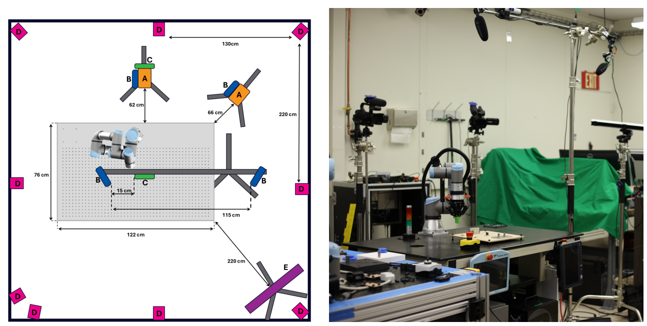
2.1 MASCAR Research Questions

As a first-pass basis of evaluation, we selected a robot-programming task to identify the impacts of different interfaces on the user's performance and satisfaction with the interfaces. The task itself consists of programming the robot to complete a simplified assembly task consisting of various small assembly components requiring different strategies to attach to a workpiece. Subjects were instructed to generate programs using lead-through teaching of the robot using different interfaces. The programs would then be run on the robot without any user input, and the performance of the assembly task would be quantitatively measured.

Additionally, throughout the training, programming, and evaluation processes, we wish to evaluate the following metrics:

- What does the HRI community need in terms of data collection?
- What are the common HRI devices, tools, and methods used?
- How do the subjects tend to interact with an interface?
- How do the interfaces impact task performance?
- How does the performance compare with human-only assembly, e.g., the metrics from Boothroyd & Dewhurst [1]?
- Which assembly tasks are impacted most by interface design?
- Which types of assembly tasks are the most challenging?
- What are the user's preferences in interface elements?
- How do the subjects prefer to interact with the task, when given multiple interface options?
- What are the impacts of the task on the user?
- What are the quantifiable actions taken by the user?
- How usable is the interface?

Whenever possible, survey tools and scales were selected from among the published literature to avoid creating competing or conflicting tools, and to avoid using measures that have not been verified or validated. To aid with this, we leveraged survey selection tools using the scoring criteria[2] from Saad *et al* [5].



(a). Sensor locations within testbed **(b). MASCAR Testbed Table**

Figure 1: The MASCAR testbed and sensor locations. More detailed diagrams are available within the software repository.[3]

3 Testbed Components

The physical testbed (Figure 1) largely consists of a 5m x 5m laboratory space with a 3m ceiling. This space can be extended as needed to accommodate a variety of robotic equipment and collaborative applications. In the current configuration, an adjustable-height, mobile cart with a custom-machined table top serves as a focal point for the human-robot collaboration. The table top was manufactured from 12.5mm-thick aluminum plate, with a regular grid–spaced 25mm apart–of drilled and M6-tapped holes for mounting artifacts, fixtures, and other task-relevant components. This material was chosen for its rigidity, strength, and durability.

For the initial setup, the cart hosts a Universal Robots e-series UR3 (a small7 degree-of-freedom industrial robot) with a Robotiq 2F-45 two-fingered, robotic gripper, a NIST assembly task board (ATB[4], [3]), and 3D-printed fixtures holding ATB parts. Attached to the gripper is a custom-built yoke handle to hand-guide the robot in freedrive mode. Each handle includes integrated buttons to actuate

[2] https://hriscaledatabase.psychology.gmu.edu/
[4] https://www.nist.gov/el/intelligent-systems-division-73500/robotic-grasping-and-manipulation-assembly/assembly

the gripper, enable and disable gravity compensation on the robot, and send additional digital signals to other onboard systems. The robot can be controlled using various interfaces, including 1) the manufacturer's default teach pendant, 2) the ForgeOS interface, 3) the NIST Collaborative Robot Programming Interface (CRPI)[5], and 4) custom-programmed augmented reality (AR) and virtual reality (VR) interfaces.

Networking, control, and recording equipment are mounted underneath the table surface, and include:

- **1 NETGEAR ProSafe G5728TP** gigabit network switch for inter-process communication,
- **1 CyberPower 1000W PFC Sinewave** rack-mounted uninterrupted power supply,
- **1 Universal Robots e-series controller** for UR3 control,
- **1 READY Robotics ForgeOS** commercial robot controller and teach pendant compatible with robots from multiple vendors,
- **1 Raspberry Pi 4B** microcomputer for controlling the Robotiq gripper and reporting gripper state,
- **1 Raspberry Pi 5** for time-synchronous logging of data from the UR3, gripper controller, and ForgeOS, and
- **1 Synology DiskStation DS418** 32TB network attached storage (NAS) for long-term storage of data logs.

4 Sensors and Data Logging

The principal function of the MASCAR testbed is to support the collection of high quality multimodal sensor data. In this section, we briefly discuss the current sensor configuration of the testbed.

4.1 Audio

The testbed consists of two standalone Sennheiser MKE 600 shotgun microphones, and two Senal MC24-ES short shotgun microphones (Fig. 1(a) - **B**) mounted on two Canon XA15 camcorders (Fig. 1(a) - **A** - see Section 4.2).

The polar pattern for the Sennheiser MKE 600 directional microphones perform best when recording directly in front of the target source, but will perform well within $\pm 30^o$ of the target source. For our application the MKE 600 microphones will be positioned overhead of the testbed, and will be situated to the left and right (not to exceed $\pm 30^o$) of the target subject, pointing at the subject, just out of frame of the front facing camcorder. The approximate distance from the subject standing normally will be determined following testing with the audio interface. However, distances from the subject should be large enough to ensure no collision with subjects and equipment for safety purposes.

The polar pattern for the MC24-ES microphone is optimal at 0^o and preferred to be within $\pm 30^o$ of the target source. However, since the MC24-ES microphones will be mounted on the vision camcorders–which will be positioned per the needs of the desired camera angle and field of view–the positioning and distance of the microphones relative to the subject are anticipated to be suboptimal for their polar patterns. As such, a small degradation in sample quality is expected. These sensors are intended to replicate a combination camera-microphone setup, and thus the optimal positioning aside from that for the camera itself is not of particular concern.

[5]https://github.com/usnistgov/CRPI

Each of the four microphones require phantom power and a preamplifier, and is connected to a Zoom f8n Pro audio recorder for data collection and synchronization. Audio files are stored in high-quality, multi-channel waveform audio files.

4.2 Visual

The testbed collects high resolution color video and depth data from multiple sources and angles. Video is captured using two Canon XA-15 camcorders (Fig. 1(a) - **A**) recording at Full HD (1080p), and two Intel RealSense D455 RGB-D sensors (Fig. 1(a) - **C**). One camcorder is positioned directly in front of the subject for face-on video collection, while the other should be positioned such that it is oblique to the subject. Both cameras are configured to stream their video feeds to a Blackmagic Design ATEM Mini Extreme ISO, which enables both feeds to be synchronized and centrally recorded in MP4 format.

Similarly, one Intel RealSense sensor is intended to be positioned below the front facing camcorder, while the other is mounted overhead such that the testbed table is in the center of its field of view. The first RealSense captures information regarding the subject, while the other records information regarding the collaborative assembly task. The data from both sensors is stored in `rosbag` format.

4.3 Object Tracking

Three motion capture (MoCap) systems are set up to track rigid-body objects within the testbed volume. These systems use passive, retro-reflective markers arranged in unique patterns to provide real-time tracking of objects. Data from these motion capture systems can be recorded locally on their host computers, or streamed via Ethernet to dedicated data-logging computers.

The first MoCap system consists of eight Vicon Bonita 10.1-megapixel, power over Ethernet (PoE) infrared cameras rigidly mounted to the walls and ceiling on the periphery of the testbed (Fig. 1(a) - **D**). This system is intended to track task-relevant objects as they move throughout the testbed, and records object and marker information at 200 Hz.

The second and third MoCaps are OptiTrack V120 Trio standalone cameras, and are used to measure and record the extrinsic configuration of the audio and vision sensors within the MASCAR testbed volume such that their locations can be registered to one another. The OptiTracks can be repositioned to provide optimum visibility of the other sensors (Fig. 1(a) - **E**), and are registered with the Vicon and Captury systems (see Section 4.4).

4.4 Human Tracking

In addition to rigid body and sensor tracking, a dedicated human-specific motion capture system is also used in the testbed to track subject positions and poses. A nine-camera (1.3 megapixel, PoE color cameras), artificial intelligence (AI)-based motion capture system from Captury is installed along the upper walls of the testbed area (Fig. 1(a) - **D**), stacked vertically below the Vicon system. The system is configured to record subject poses based on trained models, and does not require participants to wear special clothes, markers, sensors, or other fiducials. The intention is to encourage more natural interactions between the subjects and the robot system. Logs consisting of raw video files (in AVI format) and subject pose data from the Captury are stored locally on the host computer.

4.5 Interface Usage

One of the principal aspects of the planned study is the evaluation of the use and impacts of various interfaces. Inputs from the yoke gripper are constantly monitored and recorded, so we always know which buttons have been pressed and when. Further, custom-built AR and VR interfaces have been set up to record the time and nature of different interface interactions (e.g., pushing a virtual button, selecting and editing saved poses, and direction / duration of eye gaze).

Unfortunately, capturing how subjects interact with physical interfaces is neither straight-forward nor simple. A video feed of the UR3's teach pendant (a touch screen device) can be recorded using an HDMI cable routed from the robot's controller to the Blackmagic interface discussed in Section 4.2, or via a screen-recorded virtual network control (VNC) viewer. Screen feeds of the UR3 teach pendant (TP) show highlights when elements of the GUI are interacted with, and the TP can also be controlled externally via VNC, such that more granular usage tracking can be implemented.

The ForgeOS interface, in contrast, cannot share the screen of its teach pendant. Instead, a camera must be mounted to the teach pendant to stream video of the pendant screen and finger motions of the subject. Currently, three different camera systems are undergoing initial trials: an ESP32-CAM (a microcontroller with attached camera module), a Raspberry Pi Zero-W with an Adafruit 5389 camera module, and a GoPro Hero 7 Black.

All videos of the interfaces are stored in MP4 format. Reporting specific user interface usage currently requires manual annotation, which may introduce a source of human error, and will be streamlined in future work.

4.6 Robot Data

Robot state and configuration information is streamed via the Universal Robots Real-Time Data Exchange (RTDE)[6], and is recorded on a dedicated Raspberry Pi 5. Similarly, the robotic gripper state and configuration information is shared with the logging Raspberry Pi over the network such that it can be recorded and timestamped using the same system clock.

4.7 Time Synchronization

While providing a rich set of information, having varied sources of data with different update rates and formats presents a challenge for synchronization. Moreover, a large percentage of the data recording systems do not support networked time protocols for system clock alignment. For the purpose of data synchronization and alignment, the MASCAR testbed leverages Deity TC-1 timecode generators that produce audio signals with embedded timestamp information. Each timecode generator interfaces with a single recording device via audio inputs, and is synchronized with a centrally-controlled application via Bluetooth.

The audio coming from the timecode generators are higher-voltage, line-level signals, and require "line-in" interfaces on the devices taking the timecode data. The f8n and Blackmagic systems have audio inputs that natively connect to and record these audio signals, however the computers used for data logging require separate USB-based audio interfaces with line-in ports to read and record the the timecode signals.

5 Iterative Design Processes

During the design, construction, and integration phases of the MASCAR testbed development, almost every decision and configuration underwent an iterative change management process to ensure continued functionality and optimum performance. For example, the testbed tabletop discussed in Section 3 went though four iterations in which layouts and material characteristics (e.g., color and reflectivity) were altered to ensure subject comfort (e.g., with robot and task handling) and sensor compatibility. Similarly, the physical handle interface changed significantly over five iterations while undergoing pilot studies.

While the finalized designs are important for replicability and technology transfer, the iterative processes need to be fully documented to capture the methodology and rationale behind design decisions. Recording system performance and capturing the reasons for failure is critical for understanding, technology evolution, and sharing knowledge. As such, each design decision, prototype, evaluation, and change was thoroughly documented.

6 Testbed Design Documentation

Throughout the MASCAR testbed design and construction processes, all steps, decisions, design factors and iterations, and evaluations were documented. Such documentation includes digital and physical models and prototypes, standard operating procedures for assembly and testing (such as installing and configuring software, troubleshooting issues, building or modifying physical components, and positioning and integrating equipment), risk assessments for safety verification and ethics board review, bills of materials[7], and custom source code.

All materials that can be disseminated by NIST (i.e., materials created by NIST or were created by third parties and that NIST has been given permission to share publicly) are publicly available at https://github.com/usnistgov/MASCAR. Such materials include 3D models of component parts, ATB designs, design and instruction documents, and source code. Future human subject data will be made publicly available using separate archival systems to include International Standard Book Number (ISBN) identifiers for traceability. This data will be collected under NIST Institutional Review Board approval.

7 Conclusion & Ongoing Work

To address the issues of research repeatability and reproducibility a new reporting methodology is being developed by NIST to address testbed replication, data scalability, and result generalization. This paper introduces the design and configuration of a new testbed created to develop this methodology through the collection of content- and context-rich data. As NIST conducts new HRI research, the data logging capabilities of this testbed will be leveraged to provide publicly disseminated assets, metadata, and datasets in support of technology transfer and research reproducibility.

[6] https://docs.universal-robots.com/tutorials/communication-protocol-tutorials/rtde-guide.html

[7] Such bills of material include specific products, links, and pricing of materials and equipment used in the construction of the MASCAR testbed. These bills of material are provided for informational purposes only and may not be in-date.

References

[1] Geoffrey Boothroyd, Peter Dewhurst, and Winston A Knight. 2010. *Product design for manufacture and assembly*. CRC press.

[2] Hatice Gunes, Frank Broz, Chris S. Crawford, Astrid Rosenthal-von der Pütten, Megan Strait, and Laurel Riek. 2022. Reproducibility in Human-Robot Interaction: Furthering the Science of HRI. *Current Robotics Reports* 3, 4 (Dec. 2022), 281–292. doi:10.1007/s43154-022-00094-5

[3] Kenneth Kimble, Justin Albrecht, Megan Zimmerman, and Joe Falco. 2022. Performance measures to benchmark the grasping, manipulation, and assembly of deformable objects typical to manufacturing applications. *Frontiers in Robotics and AI* 9 (2022), 999348.

[4] Benedikt Leichtmann, Verena Nitsch, and Martina Mara. 2022. Crisis ahead? Why human-robot interaction user studies may have replicability problems and directions for improvement. *Frontiers in Robotics and AI* 9 (2022), 1–15.

[5] Laura Saad, Eileen Roesler, Elizabeth K Phillips, and J Gregory Trafton. 2025. Choosing the "perfect" scale: a primer to evaluate existing scales in HRI. *ACM Transactions on Human-Robot Interaction* (2025).

Received 2025-12-08; accepted 2026-01-12