A REMOTE SYNTHETIC TESTBED FOR HUMAN-ROBOT TEAMING: AN ITERATIVE DESIGN PROCESS

Margaret Wong, Akudasuo Ezenyilimba, Alexandra Wolff, Tyrell Anderson, Erin Chiou, Mustafa Demir,
Nancy Cooke
Arizona State University

Urban Search and Rescue (USAR) missions often involve a need to complete tasks in hazardous environments. In such situations, human-robot teams (HRT) may be essential tools for future USAR missions. Transparency and explanation are two information exchange processes where transparency is real-time information exchange and explanation is not. For effective HRTs, certain levels of transparency and explanation must be met, but how can these modes of team communication be operationalized? During the COVID-19 pandemic, our approach to answering this question involved an iterative design process that factored in our research objectives as inputs and pilot studies with remote participants. Our final research testbed design resulted in converting an in-person task environment to a completely remote study and task environment. Changes to the study environment included: utilizing user-friendly video conferencing tools such as Zoom and a custom-built application for research administration tasks and improved modes of HRT communication that helped us avoid confounding our performance measures.

INTRODUCTION

Recent hardware and software advancements have increased the potential of autonomous agents (e.g., robots, Artificial Intelligence (AI), synthetic agents) to be more capable partners in complex dynamic task environments, such as Urban Search and Rescue (USAR). However, within such complex environments, the command and control of current robots remain challenging (Murphy et al., 2008). Therefore, it is important to study how humans and robots can team in these complex, dynamic situations through simulated task environments, allowing the study of specific characteristics of interest in the real-world dynamic environment (Gonzalez et al., 2005).

Autonomous agents can be described as systems that can intelligently respond to situations independent of human supervision. These responses, in turn, allow for the provision of interactive task processes (Cooke et al., 2020). A team task requires the coordination of communication between the team members to complete the common goal or task (Cannon-Bowers et al., 1995). Based on the requirements of teams, autonomous agents can be designed with heterogeneous roles that require interdependence among the teammates to complete the team task. In human-autonomy team (HAT) research, task-role interdependence and heterogeneity are important to consider in developing research scenarios capable of addressing the research questions in the context of the team task.

Depending on the research questions and testbed capability, several experimenters may be required to monitor the team task experiment. However, the experiment process has become even more challenging due to the limitations imposed by the coronavirus disease pandemic (COVID-19). With COVID-19, utilizing multiple experimenters at once is not always an option, especially with social distancing and other restrictions in place. To minimize the health risks of COVID-19, a team of human systems engineering researchers virtually collaborated over a 6 month period to design and implement a remote study for a USAR human-robot team (HRT) task. The current study aims to explain how a remote testbed for human subjects research in a USAR human-robot teaming (HRT) context was designed.

With that in mind, the Human Systems Engineering (HSE) researchers applied an iterative design process to evaluate and refine the testbed and the study design. Our process was informed by Gould and Lewis (1983), which determined three key principles to achieve a functional iterative design: (1) understanding the user and task, (2) utilizing qualitative and quantitative measurements, and (3) applying iterative design. While the iterative design process may be widely applied with consumer users in mind, we adapted this process to understand our research goals as the end "user". Ultimately, instead of meeting the needs of an actual person using a product, we wanted to iterate for a design that could successfully test our research question of transparency vs. explanation for information exchange in human-robot teams. To do so, we first motivate the need for a remote synthetic testbed given our research objectives. Then, with each iteration of the testbed, we utilize various qualitative and quantitative methods in order to constantly improve upon the testbed.

ITERATIVE DESIGN PROCESS

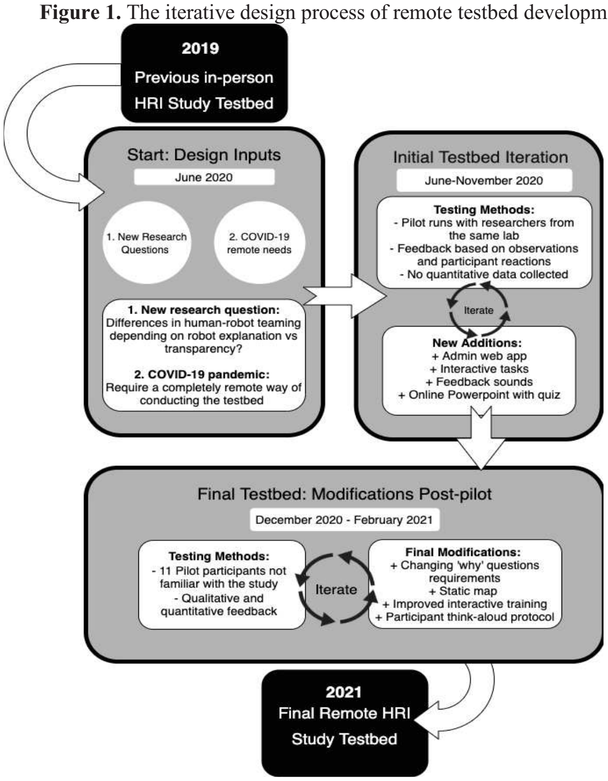
Between 2020 and 2021, we developed our remote testbed by applying an iterative design process to run the USAR-HRT study by considering three process steps which are depicted in Figure 1: (1) start - design inputs, (2) initial testbed iteration, and (3) final testbed - modifications post-pilot.

Design Inputs

Starting the development cycle, we conducted an extensive literature review on the basic USAR task, human-robot teaming, communication, and various situation awareness and performance indicators to guide the initial design. These different topics drove the main research question that we needed to test for: "What is the relative value of explanation versus transparency for user situation awareness and trust in the robot?" We also worked from our initial in-person USAR-HRT testbed design, which provided a starting point for the general task structure (Lematta et al., 2019). In the previous experimental testbed setup, we designed the study to be entirely in-person, located in a university campus lab. The testbed consisted of a Minecraft USAR-HRT task in which the participant viewed a robot played by a Wizard of Oz (WoZ) researcher navigating the environment (Kelley, 1984). The participant was able to view the environment through the robot's point of view through a laptop. They also communicated with the robot through a chat system through a separate laptop located adjacent to the main screen. As the robot went through the mission, the participant was tasked with marking victim locations on a physical map provided for them. Communication between the robot teammate was encouraged, and participants were able to direct robots through the environment if desired (Lematta et al., 2019). Going into this development process for the redesigned testbed, we had two major redesign guidelines to address:

(1) Due to the remote nature of the study, the testbed must be viewable synchronously by both researchers and participants. All materials involved in the study must be distributed through a virtual platform, including supplemental materials such as a virtual map for the participants.
(2) We needed to keep in mind the research questions of transparency versus explanations when creating the testbed interface and the environment.

Figure 1. The iterative design process of remote testbed development.



Initial Testbed Iteration

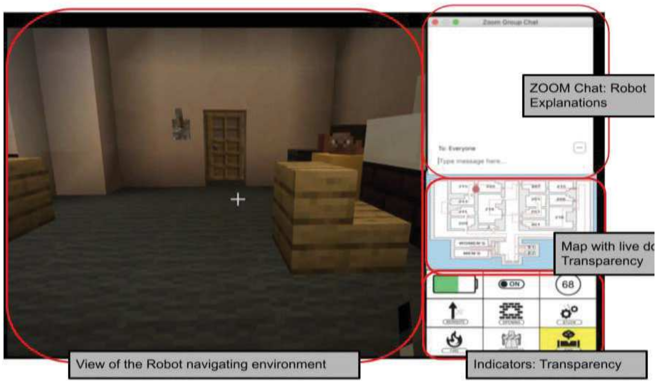
In this section, we first tested the inputs and then added new supplementary materials. After that, we iterated several times.

Testing Methods. Based on these design inputs, we needed to make adjustments from the previous in-person testbed design. We tested modifications to the testbed with other researchers in the lab, running them through the study and getting feedback from them. Feedback included direct inquiries, survey responses, and observations during runs. No quantitative data were collected during this portion of the design process. We mainly focused on adjusting the previous in-person testbed based on our design requirements.

New Additions. The remote testbed is implemented in a way that can be accessed remotely from any computer with the associated application and log-in information for the researcher's account. However, other previous related studies on USAR-HRT tasks have implemented a live interactive mission task through the use of applications such as Parsec (*Game, Work, and Play Together From Anywhere | Parsec*, n.d.), a web application that can allow remote control of games run on a separate computer, our remote study is conducted through a video recording (Cooke et al., 2020). This testbed achieves greater control through the use of pre-recorded missions. Participants undergo the study by viewing the video recordings of the robot navigating the environment, believing that the task is completed in real-time. During the debrief section, participants are fully informed of the nature of the pre-recorded study design. Unlike the previous in-person Minecraft testbed USAR-HRT study, our team needed to provide participants with additional materials due to the remote nature of the study, including separate hardware through pre-recorded videos for each mission.

Along with the view of the robot navigating the environment, the video required a side interface design that operationalized transparency and explanation through indicators, a map, and the Zoom chat system (*Zoom Chat*, n.d.), as shown in Figure 2. Accordingly, the view of the simulated Minecraft environment takes more than half of the screen (left). The rest of the screen is proportionally divided into three screens, including zoom chat (top right), map (either location-updating live or static – middle of the screen), and transparency indicators (bottom right). In addition to creating an integrated interface, we created a PowerPoint tutorial along with a quiz for participants to easily access through a link at the beginning of the study. This allowed us to ensure that all participants understood the task by utilizing an online quiz that reinforced the study task if a wrong answer was chosen.

Figure 2. Participant interface



We also developed our administrative website to control and manipulate the environment, allowing control over a point system, sound system, and timer throughout the mission. As the participant completes their mission tasks through the Zoom-enabled study interface, an experimenter is needed to moderate and control the timer, points, and start/pause the video, which is done through the administrative web application, see Figure 3. Prior to the addition of the administrative web application, monitoring the complex task environment was difficult due to the high cognitive load on the moderator. However, the addition of the administrative web app ensured greater control over the study, making the moderator's role easier to conduct (see Figure 3).

Figure 3. Admin application interface.

Operationalizing research objectives. Transparency is operationalized as a set of indicators on the interface that provide the remote human teammate access to information about the robot teammate and a location-updating live map of the robot's current location. For conditions without transparency, a static map without the robot's current location is displayed, and all of the indicator icons are off.

Explanations are operationalized as the robot teammate's answers to "why" questions asked by the human operator. Specifically, we planned to manipulate transparency and explanation for this current experiment by providing the human participant with constant access to the robot's information through indicators that update its status or by allowing them to ask questions regarding the robot's status. Robot explanations were available in conditions that allowed for explanations. In this case, robot communication through the WoZ occurred in real-time as participants asked "why" questions using the Zoom chat system. In all other conditions without explanations, the robot was only able to answer questions regarding location. To ensure that robot communications were consistent among conditions and participants, predetermined responses to commonly anticipated questions were utilized.

Final Testbed

Testing Methods. After the initial testbed re-design was finalized, the design still needed to be piloted by participants unfamiliar with our research question or procedure. These pilot participants were recruited through the university portal and with the researchers' family members. The research team utilized qualitative and quantitative methods to understand the issues with the testbed design and study protocol. With each pilot run, we recorded the Zoom sessions with the participant's consent. Recorded sessions allowed us to through the Zoom transcripts and text chat, giving greater insight into the communication between the human-robot team. We also gathered qualitative feedback on participant's overall reactions to undergoing the study and suggestions for improvements. Quantitative feedback involved post pilots' statistical analysis of the graded maps with points. The major problems identified in these pilot tests are highlighted below with our design modifications.

Final Modifications. In this process, we defined each problem and made appropriate modifications in response to the following problems.

"*Problem 1: Lack of team interdependence and communication*." There was a noticeable lack of team dynamics during the initial pilot runs. This was found through our observations of the Zoom chat system, which highlighted very limited human-robot teaming communication, even for participants in conditions allowing robot explanations. Additionally, it was noted that the act of reciprocity was absent; participants were lacking responsibility for their robot teammate, which is crucial in achieving a real team dynamic. Through feedback from the participants, as well as our researcher observations, previous iterations of the task design were limited in team interdependence. The addition of interactive tasks involving the human operator, robot, and an incident commander (mediates information) has been implemented. The interactive task is initiated through an unexpected event that requires the participant to utilize various methods to try and diagnose the robot's problem to continue the task. Participants also displayed significant frustration related to the limited dialogue within their team dynamic. Previously participants were only allowed to engage in communication via text chat with their robot teammate using 'why' questions. To rectify this dialogue in conditions that allow for explanations (Full and Explanation), participants can ask any and all questions via text chat that ends with a question mark. For conditions with limitations to explanation (Transparency and

Baseline), participants can ask questions related to location that ends with a question mark. The emphasis on punctuation is to actualize the implications of AI (artificial intelligence).

"*Problem 2: Misrepresentation of appropriate levels of explanation and transparency within conditions.*" We found issues with the initial design of the task interface, namely the use of a dynamic map that presented the live location of the robot in the environment. This dynamic map was present in all conditions, resulting in an overreliance on this specific feature in the participant interface of the task. Initial pilot data revealed little differences in the conditions, which was detrimental to the research focus of understanding the impact of transparent vs. explainable robot information. Participants admitted that the map was by far the most useful tool for them, reducing their usage of robot explanations and indicators. Through this, we realized that this was a higher form of transparency that potentially masked the effectiveness of explainable behaviors. Our solution was to reduce the transparency of the dynamic map in explanation and baseline conditions, where transparency is not present. All else remaining the same (room numbers and intended route), the static map now lacks a live status update, displayed through the moving dot, of the real-time robot location. Although we were concerned with the task difficulty in the static map conditions, participants did not express frustration with the task when asked. Rather, they found that the task difficulty had increased at a manageable rate. This was especially so due to the modifications to the allotted communication (see lack of interdependence and communication).

"Problem 3: Participants' engagement in the task." In this study, participants had struggled to stay properly engaged in the task itself. There was initially very little feedback, and complaints about the lack of sound in the overall environment needed to be addressed. In the first iteration of the testbed, we implemented an associated sound system throughout the interactive tasks to create a sense of system feedback. Participants would now receive feedback through sound in response to their attempts at addressing unexpected robot behaviors or malfunctions. Participants would hear a positive chime for correct attempts, and for incorrect attempts, participants would hear a negative buzzer. Sounds were initiated by the moderator through the use of the Administrative Web Application. As a result, participants were less likely to be confused during interactive tasks, increasing their overall engagement in the task and performance.

The additional study protocol is introduced to participants during training and is reinforced throughout actual missions to address participants' engagement. Participants are now encouraged to think aloud about their current state of mind as they are completing the missions with the robot teammate. Researchers can subsequently take the recordings of the Zoom audio transcript and transcribe the audio into text. This approach not only encourages participants to stay engaged in the study throughout but also gives greater contextual clues for researchers during analysis. This audible data will be coded post missions to provide greater context to participants' decision-making and perceptions towards their robot teammate.

Lessons learned from the iterative design process are summarized in Table 1 below. In addition to our original study design centered around four conditions (baseline, explanation, transparency, and full) we were able to implement useful features to push our overarching goal of implementing a team dynamic in a remote USAR environment. Table 1 describes the various technologies and testbed insights implemented in the overall design.

CONCLUSION

We were met with many obstacles during the development of a practical USAR-HRT testbed design that we had to overcome as a team during the COVID-19 pandemic. However, the iterative design process, which involved collecting, synthesizing, and then acting on qualitative and quantitative feedback, allowed our research team to redesign the previous in-person testbed to achieve new research objectives during the COVID-19 pandemic. Overall, the final design of the testbed was able to achieve greater team interaction and communication between the human and robot teammates while conducting the study completely remotely.

The iterative nature of this design process was necessary because initial pilot testing of initial designs indicated that the task environment was not resulting in the type of teaming behaviors that we expected would naturally occur. There are still challenges to address when it comes to designing for HRT studies, especially for testbeds that are complex and dynamic enough to simulate real-world working conditions, but with enough experimental control to be able to make strong conclusions about the effects of our interface conditions, such as the many ways in which transparency could be represented.

Conducting pilot studies using a virtual subject pool enabled us to reach many interested study participants, but future studies should still consider the potential limited access to populations without the ease of internet access. Major requirements for conducting HRT research remotely and in real-time are Internet connectivity (for both the researchers and participants in their home environments) and participant engagement during the experiment. There are limited ways for researchers to intervene or assist participants experiencing technical issues or monitoring their focus on the task (e.g., checking cell phone notifications in the middle of the experiment, or having to attend to a child or pet). This is a concern for the reliability of the data generated from these studies. Our testbed redesign for remote data collection took advantage of pre-recorded videos to address difficulties of Internet connectivity or latency; however, this meant that certain aspects of teamwork, such as dynamic environments triggered by participant behaviors, and thus participant engagement were limited, compared to what was possible in previous in-person, Wizard of Oz experiments. Our redesign focused on adding back this interactivity and engagement through additional task features and interface modifications, noting that these additions were specific to the research objective that we were trying to achieve.

The development of testbeds aims to create a controlled environment that can also capture the complexity in the work environment that affect key performance measures in human-robot teaming research (Gonzalez, 2005). In this case, our research team was specifically interested in how transparency and robot explanations impacted human-robot team performance. These "design inputs" guided our choices for the redesign of the testbed for a remote study. Our modifications after pilot testing focused on *increasing the teaming aspect (i.e., interdependent interactions)* as well as *task engagement (i.e., participants' focus on the study as a whole)*. We hope that the insights gained from the iterative process of redesigning a previous in-person testbed to a remote testbed is beneficial for future research studying the complexities involved in HRT tasks.

Table 1. Task-Specific Insights for USAR-HRT task environment

Feature	Advantages	Disadvantages
Zoom Web-enabled video telephony and online chat application	• Design structure: simple, portable, convenient, highly compatible with most computers. • Allow for USAR simulation through screen sharing of pre-recorded videos • Minimal working parts	• Heavily dependent on internet connectivity; screen sharing lags • Still new to many users; unfamiliarity with features and functions
Administrative Web Application Coded experimenter application	• Simulates an active ongoing mission despite being a pre-recorded mission • Places time urgency on participants through the use of a timer • Allows participants to self-assess with the introduction of a point system	• Tracking interactive task points can be difficult.
PowerPoint Instruction Presentation program for basic task instruction	• Monitored training via Zoom screen share • Pre-screening for colorblindness • Self-guided quiz: reinforces PowerPoint materials with immediate correction if needed.	• Participants with lesser computer proficiency struggled more to screen share. • Screen sharing may risk participants' privacy.
Experimenter-guided Training Video Separately pre-recorded training mission viewable when loaded into Admin App	• Enables all participants, regardless of condition, to be thoroughly trained in the mission tasks, reinforcing the Powerpoint information.	• Participants without the Full condition may require more time to adjust to their actual experimental condition.
Interactive Tasks Experimenter planned "unexpected events" embedded in missions	• Allows the participant become involved in an unexpected event and take greater responsibility for robot teammate; encourages team dynamic through an interdependent task	• Participants may automatize interactive task processes instead of communicating.

REFERENCES

Cannon-Bowers, J., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies: Implications for training requirements and strategies—Google Scholar. In *Team Effectiveness and Decision Making in Organizations* (pp. 333–380). http://scholar.google.com/scholar?q=Defining+team+competencies%3A+Implications+for+training+requirements+and+strategies&btnG=&hl=en&as_sdt=0%2C3

Cooke, N. J., Demir, M., & Huang, L. (2020). A Framework For Human-Autonomy Teaming Research. *22nd International Conference on Human-Computer Interaction*. INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION, Copenhagen, Denmark.

Game, Work, and Play Together From Anywhere | Parsec. (n.d.). Parsec Gaming. Retrieved March 19, 2021, from https://parsec.app

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. https://doi.org/10.1016/j.chb.2004.02.014

Gould, J. D., & Lewis, C. (1983). Designing for usability: Key principles and what designers think. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 50–53. https://doi.org/10.1145/800045.801579

Kelly, J. F. (1984) An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems*, *2*(1), 26-41. https://doi.org/10.1145/357417.357420

Lematta, G. J., Coleman, P. B., Bhatti, S. A., Chiou, E. K., McNeese, N. J., Demir, M., & Cooke, N. J. (2019). Developing Human-Robot Team Interdependence in a Synthetic Task Environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1503–1507. https://doi.org/10.1177/1071181319631433

Murphy, R. R., Tadokoro, S., Nardi, D., Jacoff, A., Fiorini, P., Choset, H., & Erkmen, A. M. (2008). Search and Rescue Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1151–1173). Springer. https://doi.org/10.1007/978-3-540-30301-5_51

Zoom Chat. (n.d.). Zoom. Retrieved March 19, 2021, from https://zoom.us/