Towards a Conversational Agent for Remote Robot-Human Teaming

José Lopes David A. Robb Muneeb Ahmad Xingkun Liu Katrin Lohan Helen Hastie

Heriot-Watt University

Edinburgh, United Kingdom

jd.lopes@hw.ac.uk

***Abstract*—There are many challenges when it comes to deploying robots remotely including lack of operator situation awareness and decreased trust. Here, we present a conversational agent embodied in a Furhat robot that can help with the deployment of such remote robots by facilitating teaming with varying levels of operator control.**

***Index Terms*—remote robots; autonomous systems; conversational agent**

I. Introduction

Robotic and autonomous systems are being deployed in remote, hazardous locations such as agriculture farms, offshore wind farms, oil and gas platforms, and for space/defense applications. In these scenarios, human operators need to be kept continuously in the loop with the remote robots for safety reasons and in case they need to step in and take control. Keeping situation awareness high is particularly challenging with robots that are out-of-sight, indeed studies have shown that remote robots instill less trust than local robots for this reason [1].

In this paper, we describe late breaking work of a voice-enabled automated assistant to help manage these types of scenarios, where robots are doing tasks autonomously and remotely. The goal is to create an assistant who manages multi-robot activity, specifically emergency response and inspection on an offshore oil-rig, facilitating teaming by acting as a go-between for the remote robots and the human. Here, we embody the voice-enabled assistant as a FurHat Robot [2] (see Block 5 in Figure 1). Key research questions include: 1) how much control should the human operator have? 2) how much information do they need to manage the situation without being overloaded, and 3) how does this affect operator trust in the conversational agent assistant's ability to get the job done?

This paper describes the first step towards building such an assistant through a Wizard of Oz (WoZ) data collection. The data collected can be used to inform the design of such an assistant and bootstrap data-driven and hybrid interactive systems, such as the one described in [3]. Here, we describe the WoZ data collection and provide a short summary of the dataset features and example dialogs with varying levels of user control.

II. Wizard of Oz Data Collection

The semi-autonomous web-based Wizard interface used for this data collection is shown in Figure 1. The interface was inspired by work reported in [4]. However, some differences were introduced to deal with the unique aspects of robot control and situation-awareness required by the scenario.

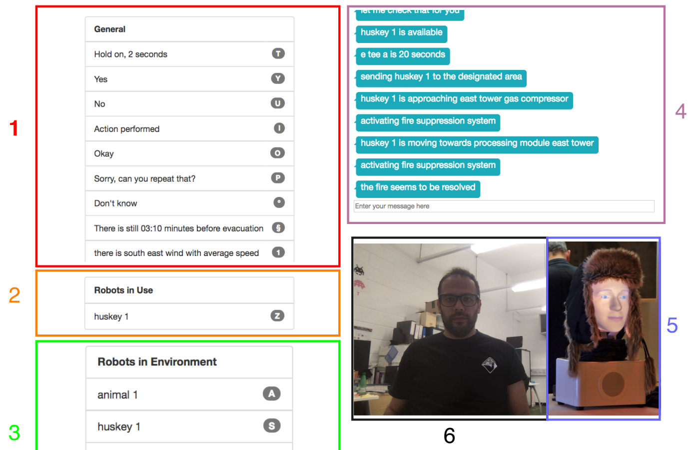

Fig. 1: The WoZ interface, with an insert of the Furhat robot (5). (1) Prompts available, (2) active robots, (3) available robots, (4) chat window and (6) view of the participant.

This scenario involves an emergency on an offshore oil rig, where there are 3 types of robots active: ANYmal, Husky and Quadcopter Unmanned Aerial Vehicles (UAVs). Specifically, participants are given a limited time (3 minutes) to extinguish a fire and assess the damage. If they do not complete these tasks within the time, they are told to evacuate the rig. Participants were informed that they should avoid evacuation at all costs. The relevant information to run the emergency scenario was scripted in a markup language file. This information is initially loaded into the system's *Situation Knowledge Base* and is dynamically updated during the course of the dialog. In addition, a database with the robots and their capabilities is also loaded into the *Situation Knowledge Base*. During the interaction, the Wizard has two live lists for the robots (Blocks 2 and 3 in Figure 1): one listing the robots currently in use and one listing available robots. Robots are activated by pressing the corresponding key. When a robot is active, an utterance is available to report the current status of that robot.

The goal of the WoZ interface is to keep a natural pace in the

High User Control	Low User Control
A: Which robot should we send to put out the fire? **P**: The Husky please **A**: Moving Husky to processing module East Tower. **P**: Can you send a Quadcopter to the North Tower as well? **A**: Okay.	**A**: I'll check which robots are available to resolve the fire. **A**: Husky 1 is available. **A**: Should Husky 1 be sent to processing module East Tower? **P**: Yes. **A**: Okay.

TABLE I: Example dialog excerpts with varying control.

Feature	Mean (Standard Deviation)
Time on Task (sec)	228.2 (28.3)
Number of User turns	9.78 (7.67)
Time between User Turns (sec)	25.5 (9.08)
Number of System turns	43.6 (4.76)
Time between System Turns (sec)	5.38 (0.62)
Avg Turn Length (words)	3.02 (1.59)
% typed utterances	2 (3)
Overall Trust (score out of 100)	79.2 (13.4)

TABLE II: Mean (standard deviation) values for interaction features.

interaction by limiting the utterances available to the Wizard and therefore facilitating more rapid responses. This behavior is facilitated through the design of interaction flow based on a Finite State Automaton (FSA), where each state represents a dialog state. For each state a number of hand-crafted prompts are uploaded and selected by the Wizard by pressing the corresponding keyboard key. These prompts contain slots corresponding to items that are fetched during the interaction from the *Situation Knowledge Database*, such as robot status. Utterances could also include tags with instructions to change Furhat's non-verbal behaviour. In addition to the utterances generated from the flow, the Wizard has at his disposal a number of predefined utterances ("Yes", "No", "Okay", "Hold on, 2 seconds", "Action performed", "Sorry, can you repeat that?" and "I don't have that information at the moment") and backchannels. In addition, there are also utterances with weather status, with the time remaining before evacuation, with the risk of the emergency spreading to different areas of the rig and a repeat last utterance key. These utterances would appear in Block 1 of the interface shown in Figure 1.

The WoZ system was implemented using the FARMI framework [5], which provides additional tools for performing synchronized recordings of various data streams, using the timestamps from the computer, which acts as a central hub of the framework. In this setup, participant audio was recorded, as well as all keystrokes from the Wizard (for utterance and robot selection) and the sequence of states from the dialog flow. Finally, the Wizard has the possibility to type utterances in a text box below the chat window (Block 4 in Figure 1). The chat window shows the most recent utterances spoken by the Furhat robot. Block 6 in Figure 1 shows the webcam used to observe the participants. The interface could be extended to show speech recognition output, as well as streaming video from the remote robots' perspective during live trials.

III. The Interaction Data

25 participants were recruited from a pool of University students from a variety of disciplines. Table I gives example dialog. In the scenario, the participants (**P**) were instructed to partake in two types of dialog: one where the assistant (**A**) takes more control of the planning and execution of the task (2nd column of the table) and the another, where the participant is encouraged to take more control (1st column of the table).

Descriptive statistics are given in Table II. As only 2% (on average) of the prompts needed to be typed, this reflects that the FSA and predetermined dialog acts cover the dialog phenomena well. The difference in the time between turns for system and user may be attributed to the varying roles in the interaction, as exemplified by the dialog in Table I. Further detailed results of this study and analysis of the data will be presented in a future publication.

IV. Discussion

Slower-than-human interaction pace is a recognized disadvantage of these types of WoZ studies. Our qualitative analysis from the post-test questionnaire indicates that the interaction pace was indeed something that could be improved. Whilst the WoZ interface allows for a high level of automation, there do still seem to be delays in cases where the participant requests the Wizard's opinion, and thus where automation is not necessarily available.

Finally, the trust scores were gathered using Schaefer's model [6] and are given in Table II, showing that overall participants' interaction with the robotic assistant resulted in a level of high trust. This is encouraging in terms of potential adoption of both remote robotic and autonomous systems combined with a voice-enabled assistant.

V. Conclusion and Future Work

This paper describes an initial step of creating a powerful, semi-automatic interface for collecting data to inform the design and development of a voice-enabled assistant for remote robot control. Future work includes iteratively replacing the Wizard decisions with automatic decisions, through methods such as Reinforcement Learning.

References

[1] W. Bainbridge, J. Hart, E. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *Proc. of ROMAN*, 2008.

[2] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granstrom, "Furhat: A back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems*. Springer.

[3] J. D. Williams, K. Asadi, and G. Zweig, "Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning," *arXiv preprint arXiv:1702.03274*, 2017.

[4] J. Lopes, O. Engwall, and G. Skantze, "A first visit to the robot language café," in *Proc. of the ISCA workshop on Speech and Language Technology in Education*, 2017.

[5] P. Jonell, M. Bystedt, P. Fallgren, D. Kontogiorgos, J. Lopes, Z. Malisz, S. Mascarenhas, C. Oertel, E. Raveh, and T. Shore, "FARMI: a framework for recording multi-modal interactions," in *Proc. of LREC*, 2018, pp. 3969–3974.

[6] K. Schaefer, "The perception and measurement of human-robot trust," Ph.D. dissertation, 2013.