Towards Efficient Human-Robot Dialogue Collection: Moving Fido into the Virtual World

Cassidy Henry[1], Pooja Moolchandani[1], Kimberly A. Pollard[1], Claire Bonial[1], Ashley Foots[1], Cory Hayes[1], Ron Artstein[2], Clare R. Voss[1], David Traum[2] and Matthew Marge[1]

[1] U.S. Army Research Laboratory, Adelphi, MD 20783 USA
cassidy.r.henry.ctr@mail.mil
[2] USC Institute for Creative Technologies, Playa Vista, CA 90094 USA

***Abstract*—Our research aims to develop a natural dialogue interface between robots and humans. We describe two focused efforts to increase data collection efficiency towards this end: creation of an annotated corpus of interaction data, and a robot simulation, allowing greater flexibility in when and where we can run experiments.**

***Keywords*—NL dialogue, human-robot interaction, simulation**

I. Introduction

Effective human-robot teaming requires robots to engage in natural communication. Natural language (NL) dialogue allows bi-directional information exchange, with the benefit of familiarity and flexibility for humans. Our goal is to develop dialogue processing capabilities for an automated robot receiving instruction from a remote human teammate in a collaborative search-and-navigate task. The physical robot, affectionately named "Fido" by an experiment participant, is a Clearpath Robotics Jackal robot running ROS [1], featuring a PrimeSense RGB camera, an IMU, and a laser scanner.

The technology to support this kind of interaction does not yet exist; we take a multi-phase Wizard-of-Oz approach to data collection to bootstrap the robot's planned language capabilities [2]. The solution cannot simply decompose into autonomous robot control and language processing. Additionally, the robot cannot be fully autonomous because it must respond to and integrate instructions, questions, and information from human teammates into its action plans. Natural language processing (NLP) relies on situated interaction based on the dynamic state and robot action, perception, and inferential capabilities, that can be neither as simple as translation to a rigid command language nor as extensive as requiring the full range of human reasoning power and common sense knowledge. A two-wizard setup [3] (Figure 1) helps address this interdependence by allowing separate simulation of both NL interaction based on flexible but limited robot intelligence, and navigation controls. The *Dialogue Manager* (DM) listens to *Commander* (CMD) verbal instructions, and either types back dialogue replies, or types constrained action sets to the *Robot Navigator* (RN), who navigates the robot with a joystick.

The first experiment phase revealed several data collection challenges, described in the next section. In Section III, we describe efforts to annotate dialogue data, used to scope requirements for automated action and interaction, and provide training data for automating NLP components. In Section IV, we describe development of a robot simulator that provides the ability to collect additional training data more efficiently for both dialogue processing and robot navigation.

The effort described here is supported by the U.S. Army. The first two authors participated in this work as visiting research assistants funded by the U.S. Army Research Laboratory; the use of corporation or company names with regard to instrumentation and equipment used does not constitute a U.S. Army endorsement.

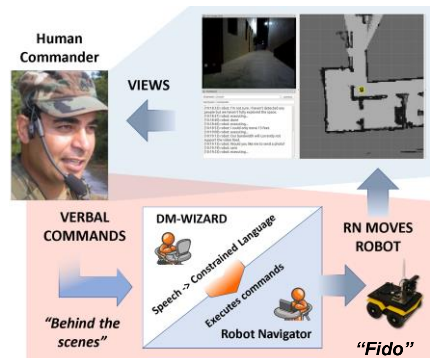

Figure 1: The Wizard-of-Oz experiment setup, where a human commander provides verbal instructions to a remote robot, given limited information from robot on screen (top). The robot is controlled by two wizards, one for handling NL and one for navigating the robot (bottom).

II. Challenges of Human-Robot Dialogue Collection

Substantial resources are required for data collection. With a single robot and test environment shared across projects, we could only run one participant at a time. In human labor, four researchers support each participant for a two-hour time block: a DM, RN, Experimenter (handles participant consent and instructions), and Session Coordinator (go-between to relay readiness to the robot's remote location). The location dependence restricts the participant pool and made recruiting a sufficient number of participants difficult. However, we found more data is necessary to both capture more natural participant language use variation, and collect sufficient training data to automate both dialogue processing and robot navigation capabilities.

Two focused efforts enable us to address these issues: development of an annotated corpus of language interactions (to be able to automate more aspects of the language process and reduce human wizard labor), and a virtual simulation that replicates our physical environment (allowing greater flexibility in when and where we can run experiments).

III. Annotation of Human-Robot Dialogue Data

We collected data from four different message streams (two speech, two typed), from three speakers (CMD, DM, RN) in two conversational floors as can be seen in Figure 2. This excerpt shows instructions, translations to the robot navigator

Commander	DM->CMD	DM->RN	Robot Navigator
	Which way should I turn?		
turn right forty five degrees			
		turn right 45, image	
	executing...		
			done
	sent		
go to the first orange cone on your right			
and send me a photo			
		move to the first orange cone on right, image	
	executing...		
			done
	sent		
are you able to move that orange cone in front of you			
	No, I'm not able to manipulate objects.		

Figure 2: Excerpt of dialogue corpus, showing four message streams.

and feedback, but also clarification and question answering. Initial corpus processing included transcribing speech and aligning the four streams to enable analysis of the utterance relationships. We performed several annotations, including dialogue moves, structure, and relations.

Each typed utterance from the DM was labeled with a dialogue move, and then the categories of dialogue moves were analyzed and condensed to form a "best-representative reply" for each of these categories. That is, all DM utterances were clustered into a small set of specific utterances and utterance templates, which were used to build a GUI for the second phase that begins to automate the language production process, transforming a fully-generative task to one of selection and specifying a few details, speeding up response time and lowering the effort required to produce complex language instructions and feedback.

Utterance data is now being annotated and analyzed to automate language understanding and dialogue management modules, relying on observed behavior patterns to generate, tune, and evaluate policies for responding to the participant and translating instructions to the navigation component. These annotations, in conjunction with participant survey data collected, will support exploratory data analysis of individual differences in situated human-robot dialogue.

IV. Moving "Fido" into the Virtual World

Our simulation setup aims to reduce requirements to run the next phase of our research program and collect more dialogue data for human-robot interaction. The simulation is being developed on the same operating software as the robot, to facilitate comparisons and data transition between physical and virtual robots.

Under simulation development, we developed high-fidelity replications of the robot[1] and physical environment[2] using ROS [3] and Gazebo [4]. The virtual Jackal was equipped with the same sensors as the physical platform: it builds maps using SLAM, and features a virtual PrimeSense RGB camera. LIDAR allows for map population for participant view, while the camera provides images of the robot's point of view upon request. The CMD in our simulated setup will have the same

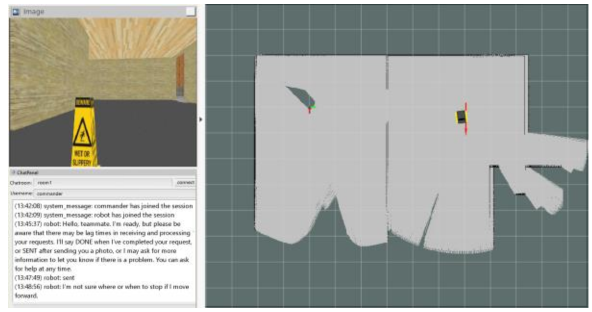

Figure 3: The Commander view in the simulated experiment setup with a Clearpath Jackal robot. The robot sends back images (top left), responds to instructions (bottom left), and displays a map (right).

tasks and see a screen (Figure 3) with the same layout as in earlier experiments (see Figure 1, top). A point-and-click navigation system was included alongside the virtual robot, which, along with the GUI, enables a single wizard to perform both dialogue management and navigation tasks.

Simulation thus gives us several advantages for data collection: first, it frees us from resource limitations, allowing parallel collection of data. Next, it allows experimentation without risk of damage to the physical robot or environment, and third, it reduces human labor by allowing simpler control. As the robot and simulation still operate on the same software and emulated hardware, we expect the experiment will smoothly transition back into a physical environment for validation purposes, after sufficient data collection.

V. Conclusion

Our overall program objective is to provide more natural ways for humans to interact and communicate with robots using language, using a sequence of multi-phase data collection experiments to incrementally automate the system, towards the ultimate goal of full automation. We highlighted two focused efforts to increase data collection efficiency via partial simulation: language corpus creation/GUI development effort and robot simulation. This corpus will help address many issues encountered in understanding and processing situated human-robot dialogue. The robot simulation replicates our physical environment while allowing greater flexibility in running experiments, and allows simulated results to be validated in a physical environment after completed data collection.

References

[1] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng. 2009. ROS: an open-source Robot Operating System. In *Proc. of ICRA workshop on open source software*.

[2] M. Marge, C. Bonial, K. Pollard, R. Artstein, B. Byrne, S. Hill, C. Voss, and D. Traum. 2016. Assessing Agreement in Human-Robot Dialogue Strategies: A Tale of Two Wizards. In *Proc. of IVA*.

[3] M. Marge, C. Bonial, B. Byrne, T. Cassidy, A. W. Evans, S. G. Hill and C. Voss. 2016. Applying the Wizard-of-Oz Technique to Multimodal Human-Robot Dialogue. In *Proc. of IEEE RO-MAN*.

[4] N. Koenig and A. Howard. 2004. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *Proc. of IROS*.

[1] Clearpath Robotics ROS libraries were used for this effort. `https://www.clearpathrobotics.com/`

[2] The authors are thankful to Darius Jefferson II and John G. Rogers III for creating a simulation of the physical environment.