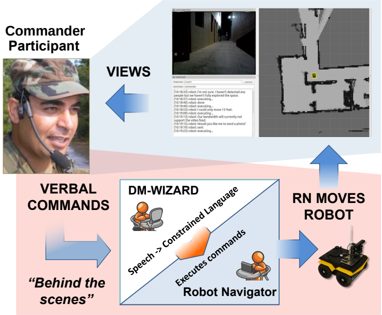
The Bot Language Project: Moving Towards Natural Dialogue with Robots

Cassidy Henry[1], Stephanie Lukin[1], Kimberly A. Pollard[1], Claire Bonial[1], Ashley Foots[1],
Ron Artstein[2], Clare R. Voss[1], David Traum[2], Matthew Marge[1], Cory J. Hayes[1], and Susan G. Hill[1]

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

Abstract

This paper describes an ongoing project investigating bidirectional human-robot NL dialogue with the goal of providing more natural ways for humans to interact with robots. We present the experiment's resulting corpus, current findings, and future work towards a fully automated system.



```
(10:18:33) robot: I'm not sure. I haven't detected any people but we haven't fully explored the space.
(10:18:37) robot: executing...
(10:18:40) robot: done
(10:18:46) robot: executing...
(10:18:53) robot: I could only move 13 feet.
(10:19:09) robot: executing...
(10:19:13) robot: Our bandwidth will currently not support the video feed.
(10:19:15) robot: Would you like me to send a photo?
(10:19:19) robot: sent
(10:19:23) robot: executing...
```

Figure 1: Illustrating the passage of information through the two wizard setup.

1 Motivation

Teams comprising both robots and humans working towards the same goals can leverage all participants' unique strengths, such as human abilities to see, reason, and command; and robot abilities to follow instructions, enter dangerous areas, and use various sensors. Coordinating the interplay of these strengths in the pursuit of a goal requires effective, detailed, and flexible interchange of information. Language is a main avenue through which humans collaborate and convey information about the world to one another, so this could be an efficient means for human-robot collaborative dialogue. The underlying research raises many challenges. It is not simply a problem of language understanding and generation —robots and humans must have a shared understanding of varied environments such that they can collaborate in rapidly changing and unique situations.

2 Experimental Setup

Because the current state of the art cannot yet facilitate interaction through fully flexible NL dialogue, employing the Wizard-of-Oz (WoZ) technique is beneficial. We present a multiphase experiment in a WoZ type setup, where there are two intermediary, non-copresent wizards, the Dialogue Manager (DM), and the Robot Navigator (RN). DM and RN seperation represents the separated processing modules for dialogue and navigation in future automated systems, respectively.

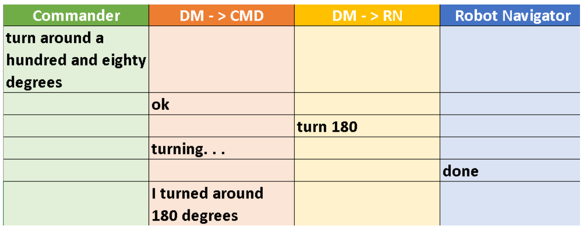
This approach has proven successful in both eliciting diverse dialogues, and task success overall (Marge et al., 2016). To complete a collaborative task, a human participant (Commander, or CMD) instructs a remote robot through a physical environment using spoken NL commands. Participant language use is unrestricted (no example utterances given). The DM passes the CMD instructions to the RN in typed constrained language, after which the RN executes commands by teleoperating the robot. The DM communicates back to the CMD through text replies. Data from the robot is provided, including still photos upon request from the robot's camera and a continuously-updating 2D map from the robot's LIDAR. The RN and CMD never communicate directly, instead speaking only to the DM via private audio streams.

3 Multiphase Approach

Our approach comprises multiple phases ('experiments') that successively automate system components until a fully automated system is reached, similar to the process in DeVault et al. 2014. In Experiment (Exp) 1, the DM freely typed responses, guided by a protocol. In Exp 2, DM responses were generated via a GUI, constructed from responses made by the DM in Exp 1 to further automate the process and to create a constrained response set for future training data (Bonial et al., 2017). Upcoming in Exp 3, we move from the physical world to simulation (Henry et al., 2017), which provides benefits such as reducing labor requirements, enabling rapid dialogue collection in parallel for multiple participants, and allowing dialogue elicitation across any conceivable situational domains, etc. Said simulation is utilizes on the same platform and virtual hardware as the physical robot. In planned subsequent experiments, Wizard roles will be fully automated; and in final experiments, the full system will be tested on-board an actual robot in the physical world.

4 The Data

We collect recorded speech data from the CMD and RN, text from the DM, still images that were requested by the CMD, and a wealth of information from the robot itself provided by its onboard OS. Our primary data is a growing corpus of human-robot interactions. In total across two completed phases, we have 20 participants, ~20 hours of participant audio containing 3,573 participant utterances (continuous speech) totaling 18,336 words, and 13,550 words from the DM in text. Corpus data is compiled into transcripts time-aligned to DM messages. A transcript contains two conversational floors, respectively: (1) CMD-DM, and (2) DM-RN.



Commander	DM - > CMD	DM - > RN	Robot Navigator
turn around a hundred and eighty degrees			
	ok		
		turn 180	
	turning. . .		
			done
	I turned around 180 degrees		

Figure 2: Sample transcript of dialogue across multiple floors.

All transcripts are annotated to support various analyses and to serve as training data for a future automated system. We developed an extensive schema for annotating meso-level dialogue structure across multiple conversational floors (Traum et al., 2018) to track the flow of information across the floors and track task state. Dialogues are divided into groups called *transaction units*. A TU approximately comprises a single action or intent from initiation to completion. At the next level, each utterance's direct antecedent is labeled, and assigned a relation type to describe its dialogic function.

The data has been partially annotated for dialogue moves and accompanying parameters. The move set, which was inspired by a previous schema used for identifying dialogue moves in calls for artillery fire (Roque et al., 2006), was developed iterativelty to include the full spectrum of dialogue moves captured in the data. Marge et al. 2017 describes analyses conducted using information from both the dialogue structure and dialogue move annotations, where we identify an initial preference for participants to use metric information in commands ("move forward five feet"), which later shifts towards the use of landmark-based information ("move to the doorway on the left").

5 Ongoing & Future Work

Our project is working towards developing a fully automated system. This involves not just language generation and understanding techniques, but also a complex interaction of language and its situated context, requiring the robot to be grounded with its conversational partners.

Spoken command translation to executable actions present a challenge. Moolchandani et al. 2018 explores what people expect in terms of language to action realization (e.g. when a command "go to the doorway" is issued: the robot could maintain its current heading and simply move into the door's vicinity without facing it, or do so and then face it, or move into the doorway, etc.) There is no one best practice for interpreting these commands related to human expectations of execution, so further work is expected to investigate this further. Initial work investigates how to reduce need for clarification on behalf of the robot by exploring encoding human intentions (Hayes et al., 2018).

Via newly developed simulation, we hope to explore multiple situated domains in varied environments using this technology to elicit language that is as diverse as possible, to facilitate broad coverage of communications with robots.

Towards the goal of a fully automated, collaborative robotic system, our multi-phase approach deploying WoZ allows us collect valuable natural language data and center development around human expectations of communication with robotic systems in a manner that the current SOTA does not allow.

References

Claire Bonial, Matthew Marge, Ashley Foots, Felix Gervits, Cory J Hayes, Cassidy Henry, Susan G Hill, Anton Leuski, Stephanie M Lukin, Pooja Moolchandani, Kimberly A. Pollard, David Traum, and Clare R. Voss. 2017. Laying Down the Yellow Brick Road: Development of a Wizard-of-Oz Interface for Collecting Human-Robot Dialogue. *Proc. of AAAI Fall Symposium Series* .

David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proc. of AAMAS*.

Cory J. Hayes, Matthew Marge, Claire Bonial, Clare Voss, and Susan G. Hill. 2018. Team-centric motion planning in unfamiliar environments. In *To appear in the Proceedings of the SPIE Conference*.

Cassidy Henry, Pooja Moolchandani, Kimberly A. Pollard, Claire Bonial, Ashley Foots, Ron Artstein, Cory Hayes, Claire R. Voss, David Traum, and Matthew Marge. 2017. Towards Efficient Human-Robot Dialogue Collection: Moving Fido into the VirtualWorld. Vancouver, Canada.

Matthew Marge, Claire Bonial, Ashley Foots, Cory Hayes, Cassidy Henry, Kimberly A. Pollard, Ron Artstein, Clare R. Voss, and David Traum. 2017. Exploring Variation of Natural Human Commands to a Robot in a Collaborative Navigation Task. In *RoboNLP*.

Matthew Marge, Claire Bonial, Kimberly A. Pollard, Ron Artstein, Brendan Byrne, Susan G. Hill, Clare Voss, and David Traum. 2016. Assessing agreement in human-robot dialogue strategies: A tale of two wizards. In *Intelligent Virtual Agents*. Springer International Publishing, Cham, pages 484–488.

Pooja Moolchandani, Cory J Hayes, and Matthew Marge. 2018. Evaluating robot behavior in response to natural language. *To appear in the Companion Proceedings of the HRI Conference* .

Antonio Roque, Anton Leuski, Vivek Rangarajan, Susan Robinson, Ashish Vaswani, Shrikanth Narayanan, and David Traum. 2006. Radiobot-CFF: A Spoken Dialogue System for Military Training. In *Proc. of Interspeech*.

David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervits, Kimberly Pollard, Claire Bonial, Su Lei, Clare Voss, Matthew Marge, Cory Hayes, and Susan Hill. 2018. Dialogue Structure Annotation for Multi-Floor Interaction. In *Proc. of LREC*.