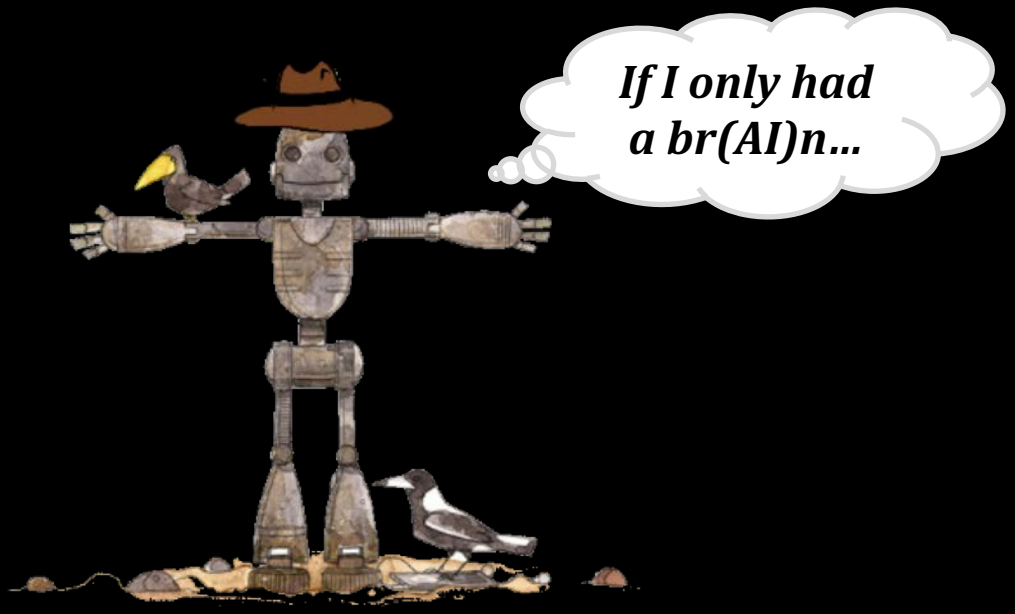


Towards Natural Dialogue with Robots

Matthew Marge

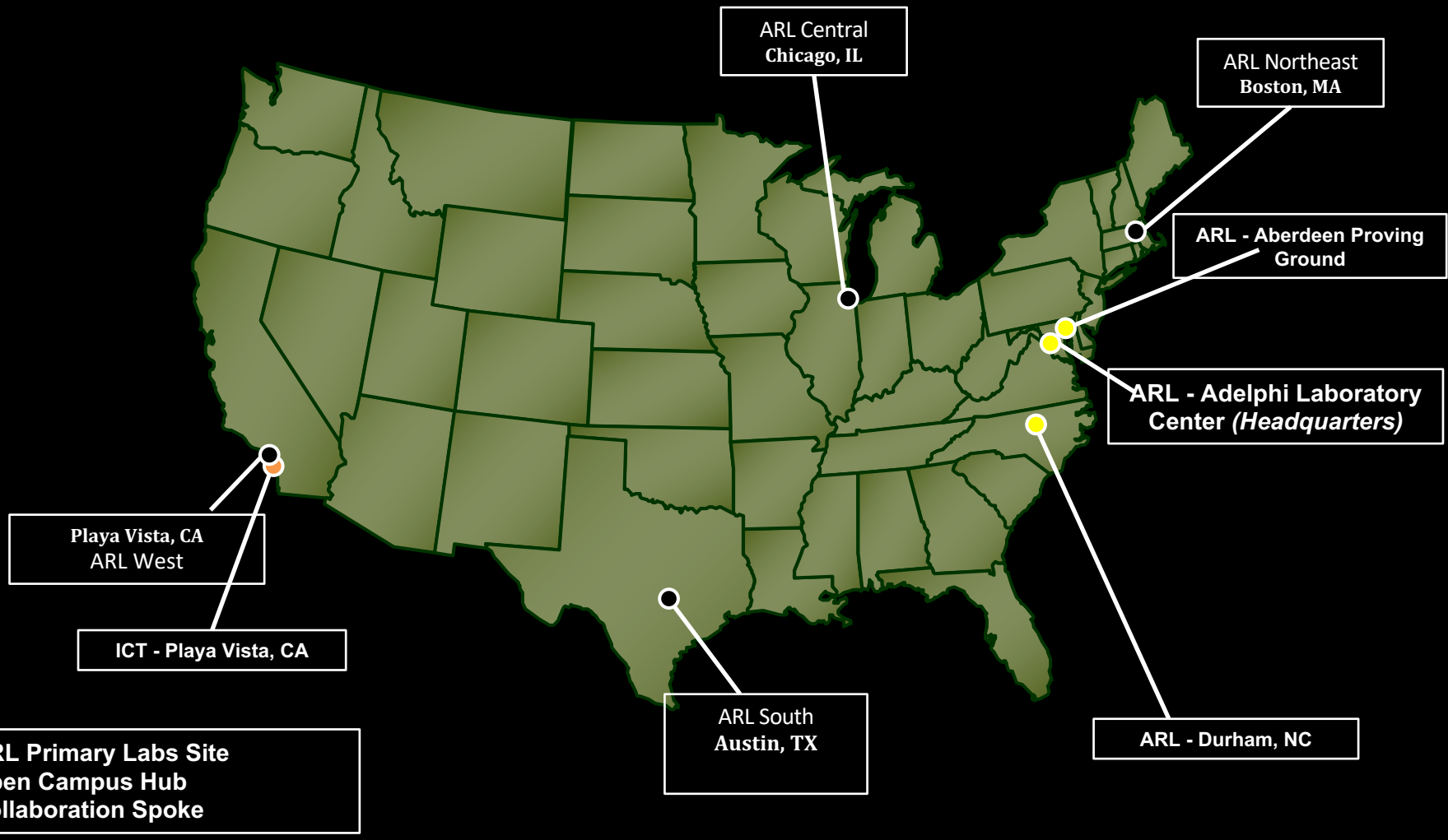
Army Research Lab

September 9, 2019

DEVCOM ARMY RESEARCH LABORATORY

USC Institute for Creative Technologies

Where is the Army Research Lab?



DEVCOM ARMY RESEARCH LABORATORY

Road Map

1. Motivation and Overview
2. Experiments Towards Natural Dialogue
3. Ongoing Work

Road Map

1. **Motivation and Overview**
2. Experiments Towards Natural Dialogue
3. Ongoing Work

Pulls together:

Robotics
Language
Computer Vision

into one AI system

Motivation

How to model dialogue?

What do you see in front of you?

I see a hole in a brick wall…

How can people build common ground with robots?

Research Question

How can we explore the natural diversity of communication strategies, while collecting language in a form that a robot could use?

This kind of autonomous system doesn't exist – could we start with humans?

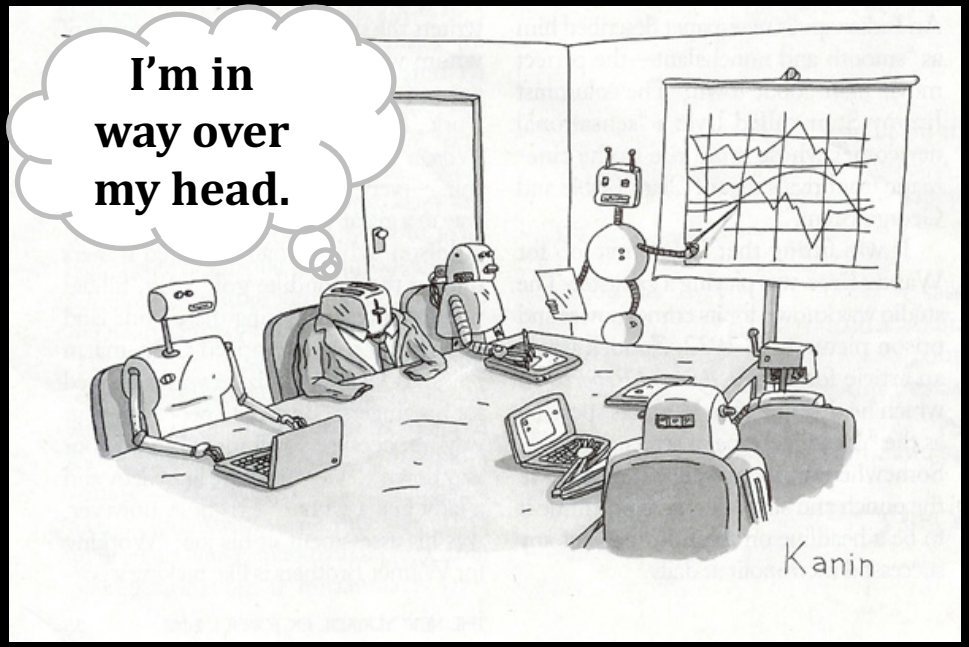
(Knepper et al., 2015)

Goal

Natural language understanding and generation to enable dialogue

- **Grounding** mechanisms like clarification strategies

Happens in everyday conversation, what about robots?



Background

Existing "Wizard of Oz" approaches to managing dialogue

- Supports low-development costs, with malleable system functionality
- Traditionally used in both dialogue system and human-robot interaction research communities (Riek, 2012; Gandhe and Traum, 2007; Green, et al. 2004)

Background

Existing "Wizard of Oz" approaches to managing dialogue

- Supports low-development costs, with malleable system functionality
- Traditionally used in both dialogue system and human-robot interaction research communities (Riek, 2012; Gandhe and Traum, 2007; Green, et al. 2004)

SimSensei Virtual Human

- Took multi-phase approach to build virtual human therapist
- Human "wizard" stood in for verbal communications during development

USC Institute for Creative Technologies Virtual Human (DeVault et al., 2014)

Towards Natural Dialogue

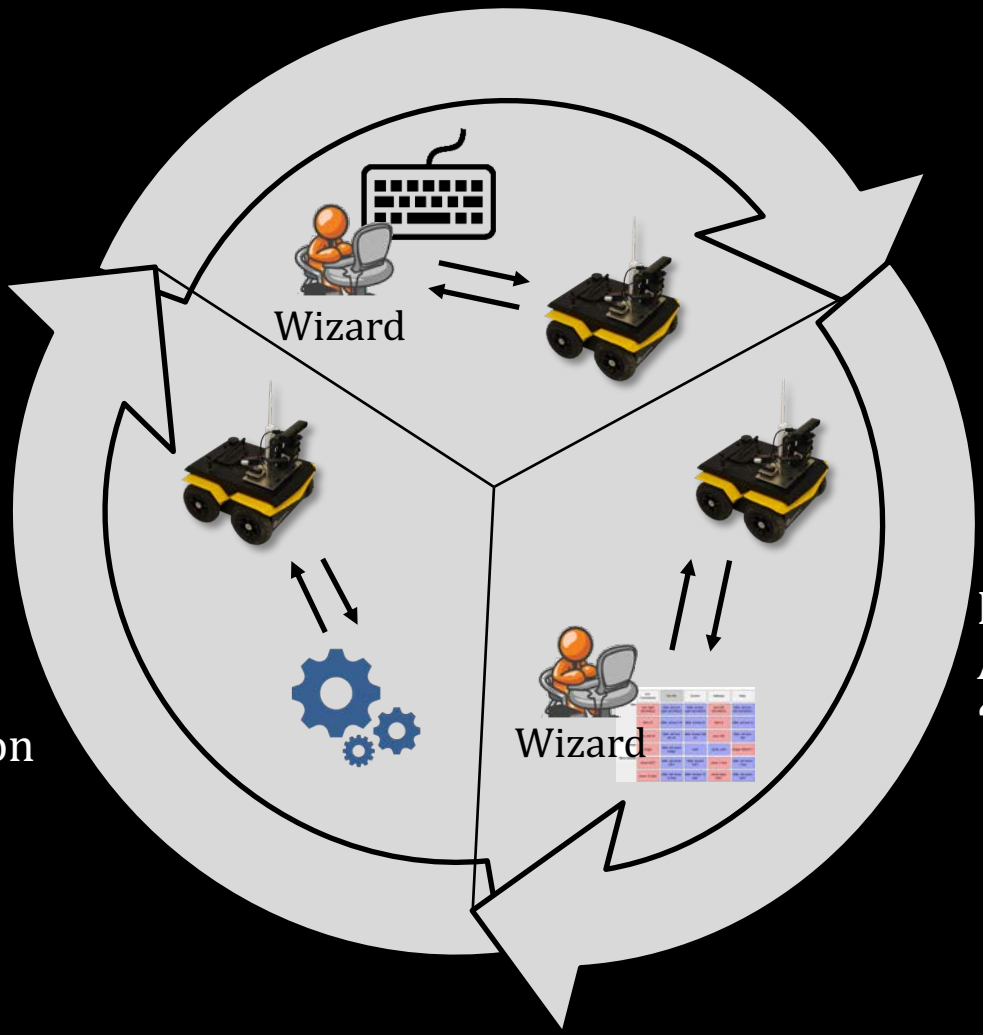
We extend and validate this approach to human-robot language communication

Phase 1: Exploratory data collection of human-robot dialogue

Phase 2: Automate some of "Wizard's" labor

Phase 3: Automate "Wizard" entirely

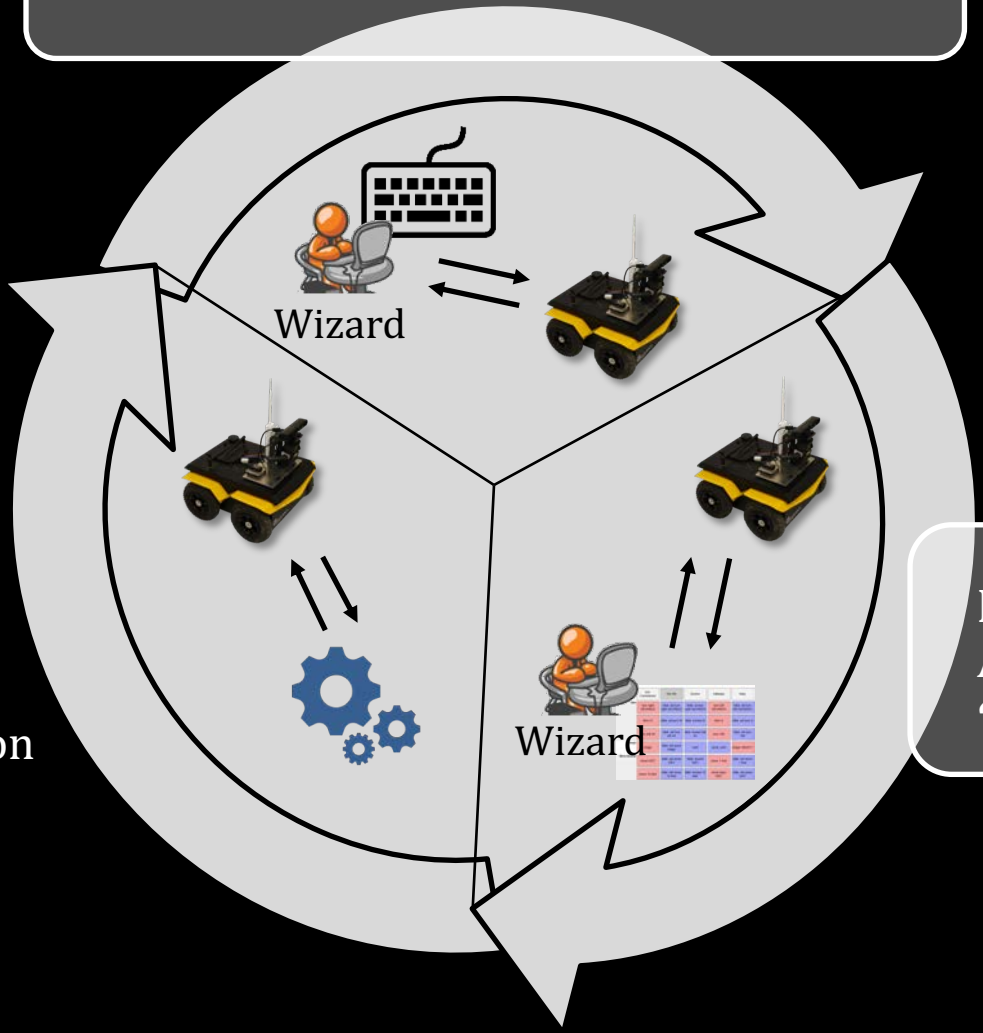
Phase 1 Exploratory Data Collection



Phase 2
Automate Some
"Wizard" Labor

Phase 3
Full Automation
Of "Wizard"

Phase 1 Exploratory Data Collection



Phase 2
Automate Some
"Wizard" Labor

Phase 3
Full Automation
Of "Wizard"

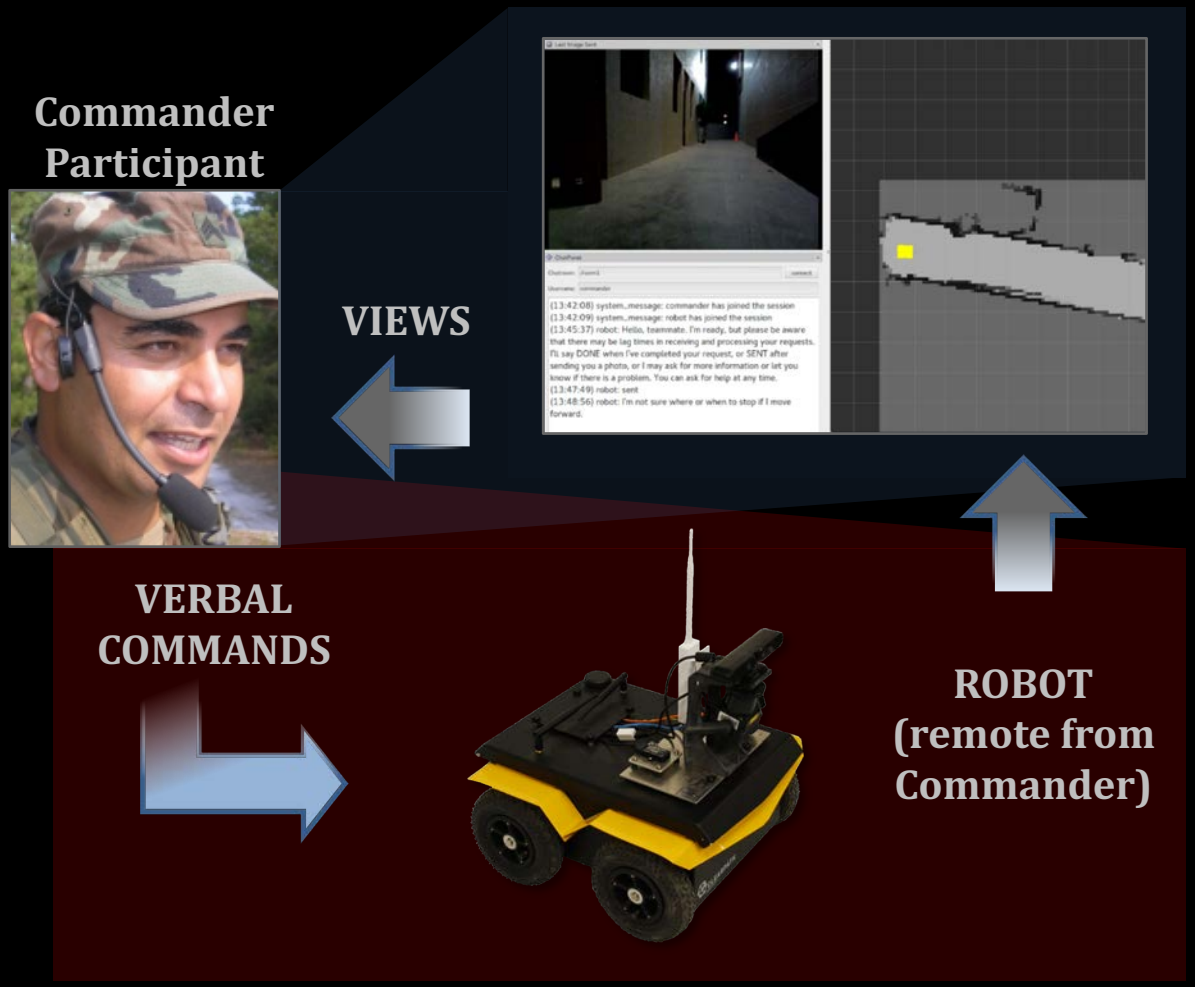
Phase 1 Exploratory Data Collection

Let's take a step back to see our end goal…

Phase 2
Automate Some "Wizard" Labor

Phase 3
Full Automation Of "Wizard"

Approach



(Marge et al., 2016, IEEE RO-MAN)

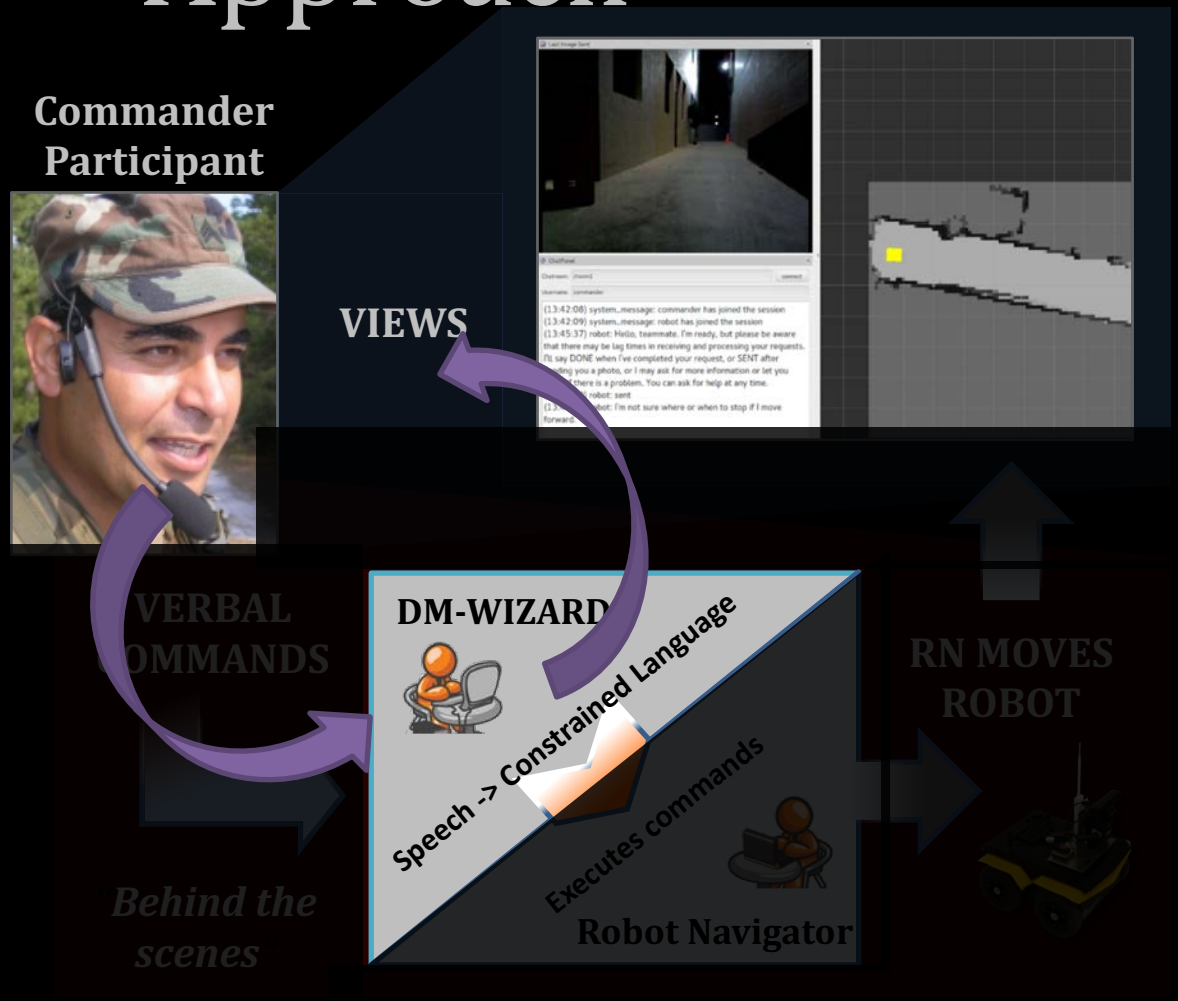
Approach

- Dialogue Manager (DM-Wizard) is the "brains" of the robot in natural language interactions

- Robot Navigator (experimenter) navigates robot based on instructions from DM-Wizard

Approach

- Dialogue Manager (DM-Wizard) is the "brains" of the robot in natural language interactions



- Robot Navigator (experimenter) navigates robot based on instructions from DM-Wizard

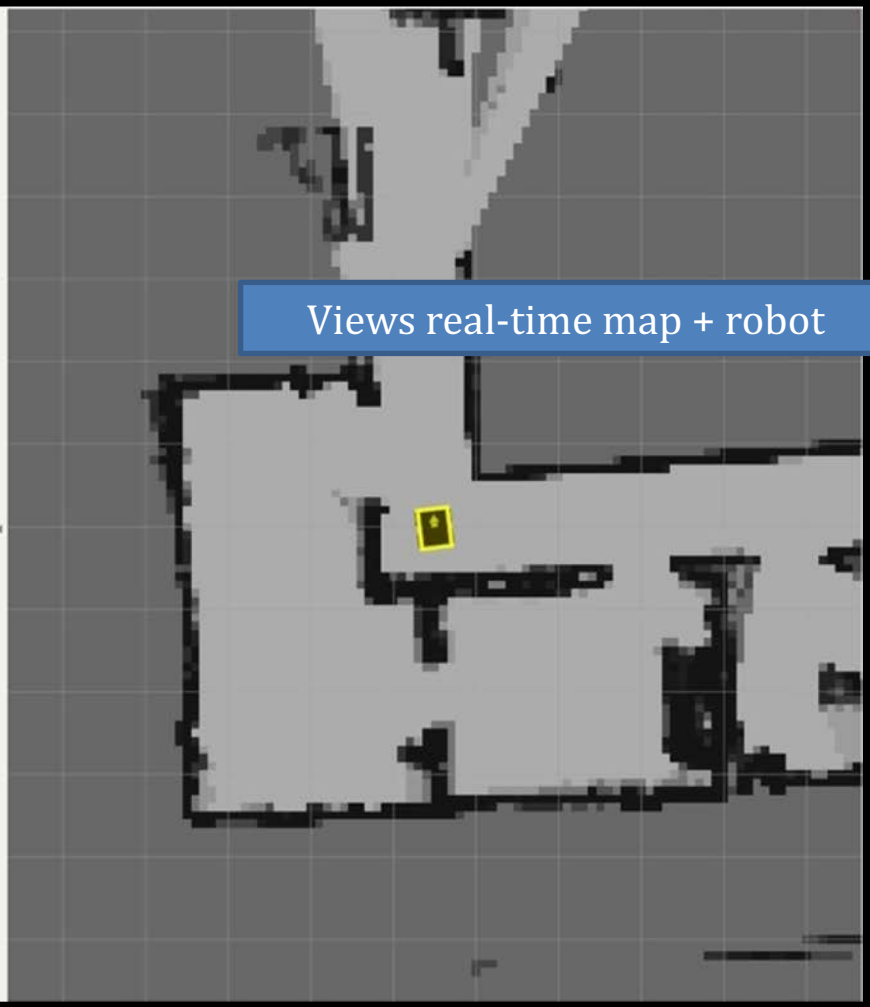
Commander View

Commander: Speaks into microphone

Requests and views photo from robot

Reads DM-Wizard dialogue responses

Views real-time map + robot

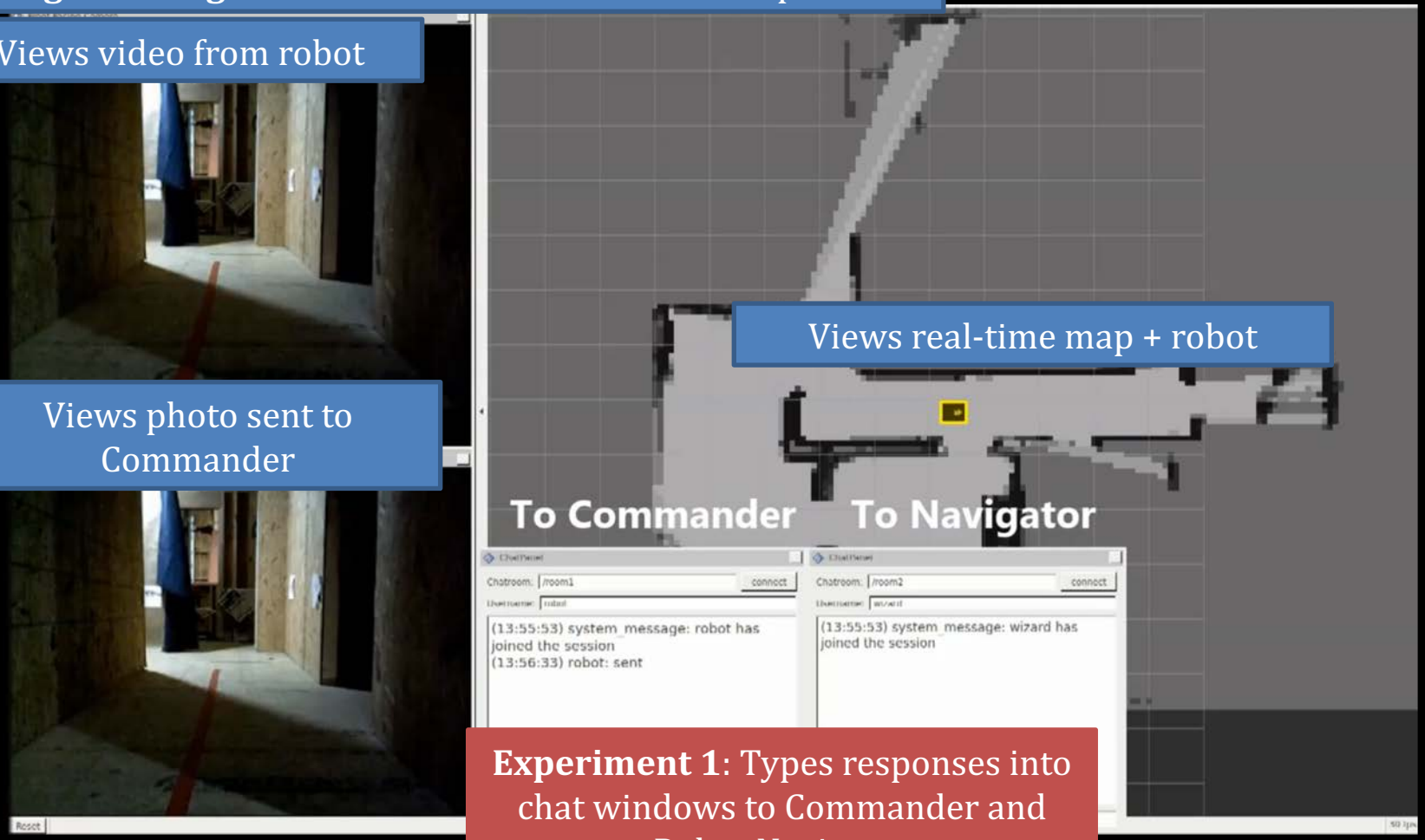


DM-Wizard View: Experiment 1

Dialogue Manager: Listens to Commander via headphones

Views video from robot

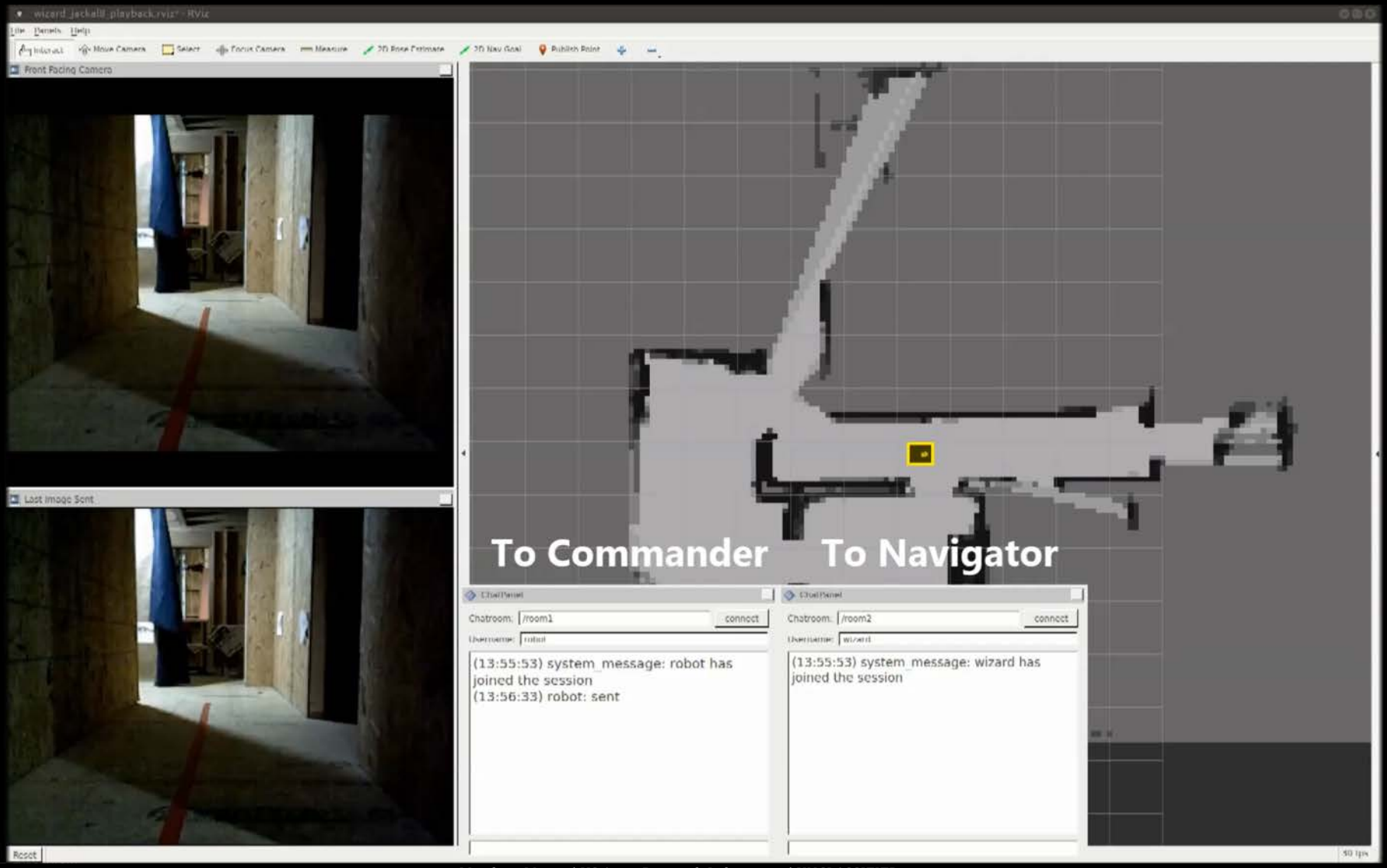
Views photo sent to Commander



Experiment 1: Types responses into chat windows to Commander and Robot Navigator

DEVCOM ARMY RESEARCH LABORATORY

Video: DM-Wizard in Experiment 1

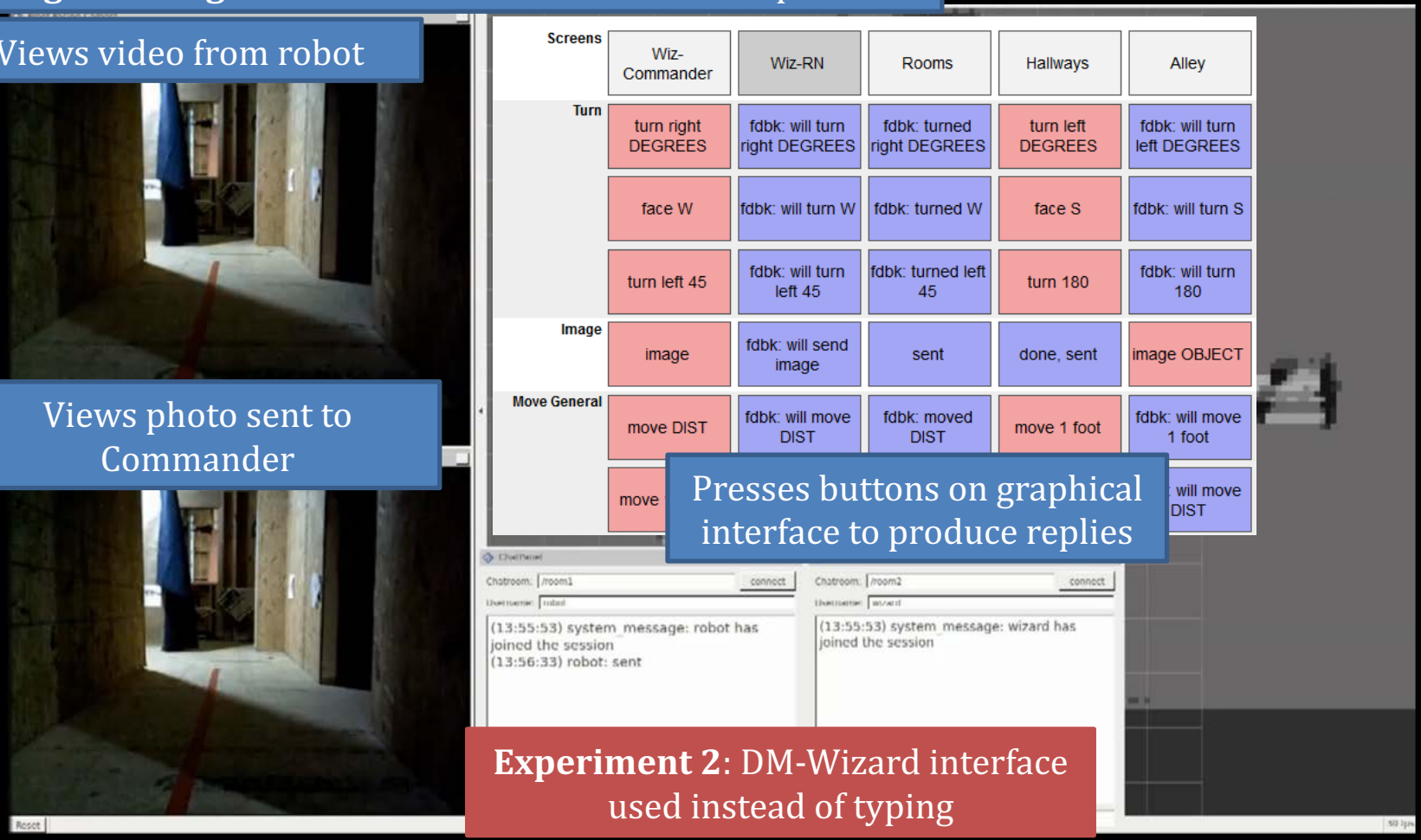


DM-Wizard View: Experiment 2

Dialogue Manager: Listens to Commander via headphones

Views video from robot

Views photo sent to Commander



Presses buttons on graphical interface to produce replies

Experiment 2: DM-Wizard interface used instead of typing

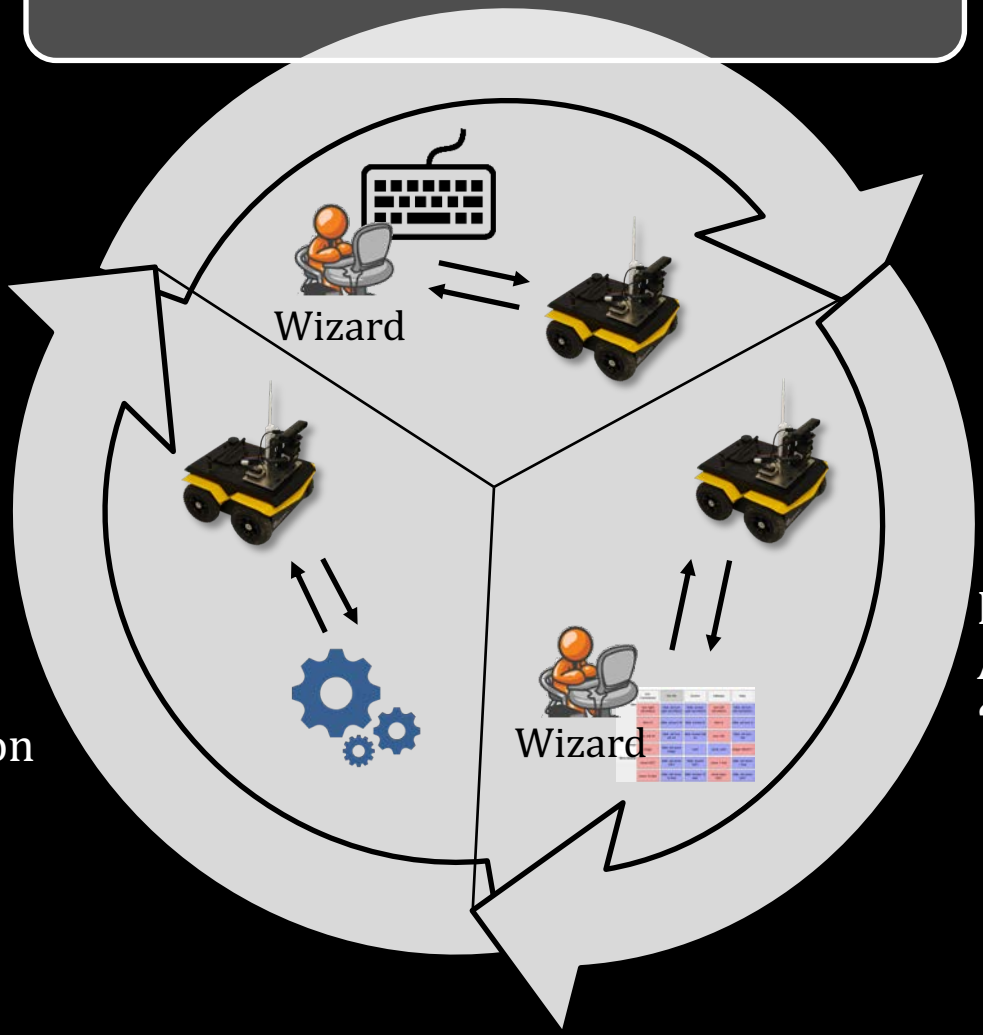
DEVCOM ARMY RESEARCH LABORATORY

Screens	Wiz-Commander	Wiz-RN	Rooms	Hallways	Alley
Turn	turn right DEGREES	fdbk: will turn right DEGREES	fdbk: turned right DEGREES	turn left DEGREES	fdbk: will turn left DEGREES
	face W	fdbk: will turn W	fdbk: turned W	face S	fdbk: will turn S
	turn left 45	fdbk: will turn left 45	fdbk: turned left 45	turn 180	fdbk: will turn 180
Image	image	fdbk: will send image	sent	done, sent	image OBJECT
Move General	move DIST	fdbk: will move DIST	fdbk: moved DIST	move 1 foot	fdbk: will move 1 foot
	move 10 feet	fdbk: will move 10 feet	fdbk: moved 10 feet	move back DIST	fdbk: will move DIST

Road Map

1. Motivation and Overview
2. **Experiments Towards Natural Dialogue**
3. Ongoing Work

Phase 1 Exploratory Data Collection



Phase 2
Automate Some
"Wizard" Labor

Phase 3
Full Automation
Of "Wizard"

Experiment 1 Setup

- Each experiment session: new Commander participant
  - 20 minutes of training with robot
  - 20 minutes in Trial 1 (first path)
  - 20 minutes in Trial 2 (second path)

- Tasks:
  - Count doorways
  - Count objects of interest
  - Assess environment

- Ran 10 participants → ~10 hours of dialogue
  - 2 female, 8 male (age range: 28-58, mean = 44)

Dialogue Manager Guidelines

- DM followed guidelines to govern decisions
- **Clear action** & **endpoint**
  - Due to bandwidth limitations of scenario

Example command (speech): *Move forward.*

Communication problem: Open-ended action (no endpoint specified)

Relevant template: DESCRIBE PROBLEM + CAPABILITY

DM response to participant (text): **`How far? You can tell me to move to an object that you see or a distance.`**

(Traum et al., 2018; LREC)

Annotating Dialogue Structure

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing...		
			image sent
	sent		

Time

(Traum et al., 2018; LREC)

Annotating Dialogue Structure

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing…		
			image sent
	sent		

Time

Transaction Unit (TU)

(Traum et al., 2018; LREC)

Annotating Dialogue Structure

Instruction Unit (IU)

Time

Transaction Unit (TU)

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing...		
			image sent
	sent		

(Traum et al., 2018; LREC)

Annotating Dialogue Structure

Instruction Unit (IU)

Time

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing...		
			image sent
	sent		

Dialogue Move

Dialogue Move

Transaction Unit (TU)

Analysis: Dialogue Move

- Tabulated dialogue move types
  - Command (requests for the robot to do something)
  - Request-info (requests for information)
  - Feedback (acknowledgements, yes, no)
  - Describe (statements about scene or plan)
  - Reference type (landmark or metric)

- Focus of analysis: Reference type

Analysis: Reference Type

*"Move through the **doorway**"*

*"Move forward **two feet**"*

Landmark:
Object references

Metric:
Specific distances

2 feet

180 degrees

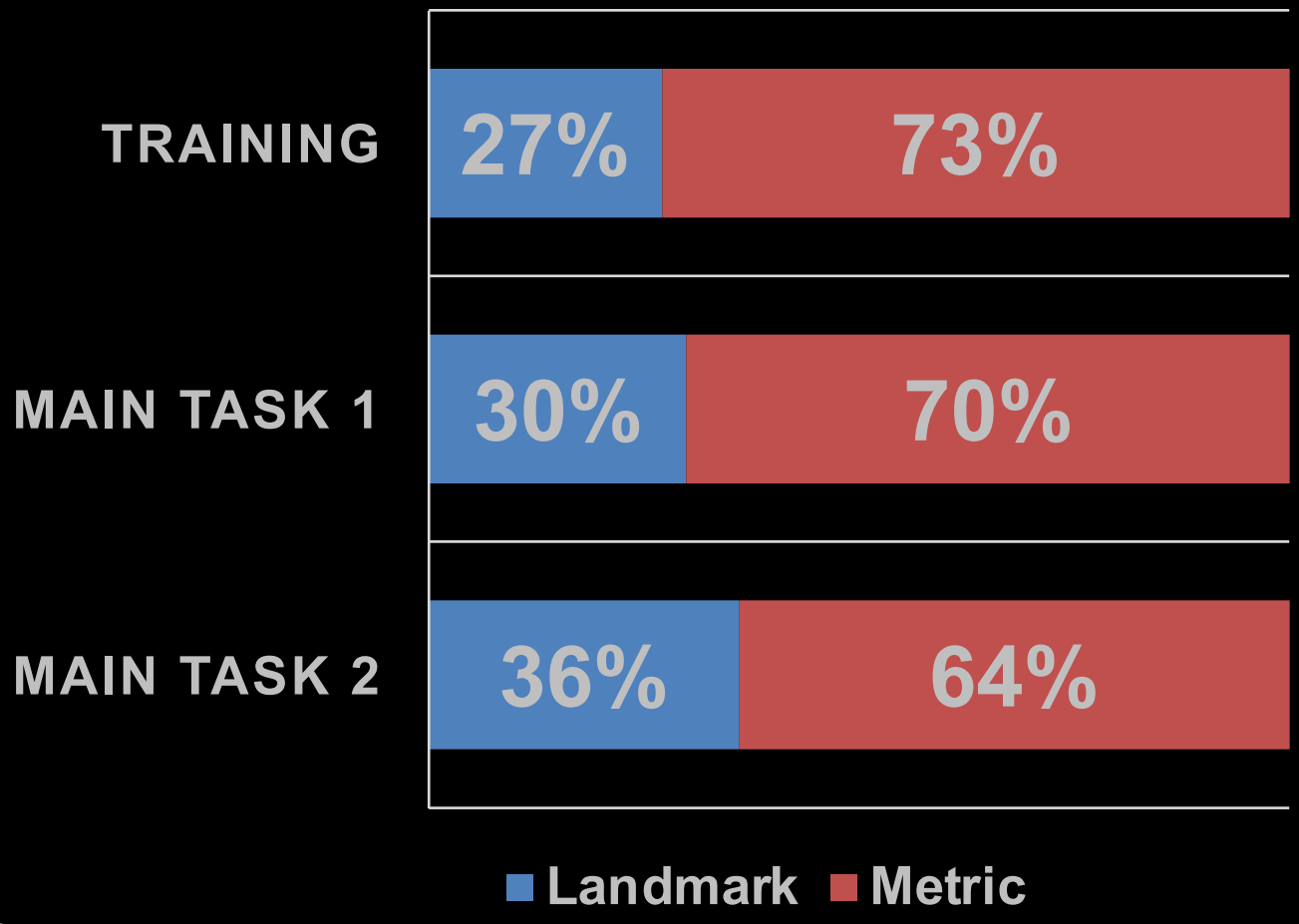
Annotation Results: Experiment 1

- 94% of instructions were commands
  - 52% had requests for images ("send a picture")
    - Situational awareness important
  - 47% had rotations ("turn right")
  - 42% had drive commands ("move to the doorway")

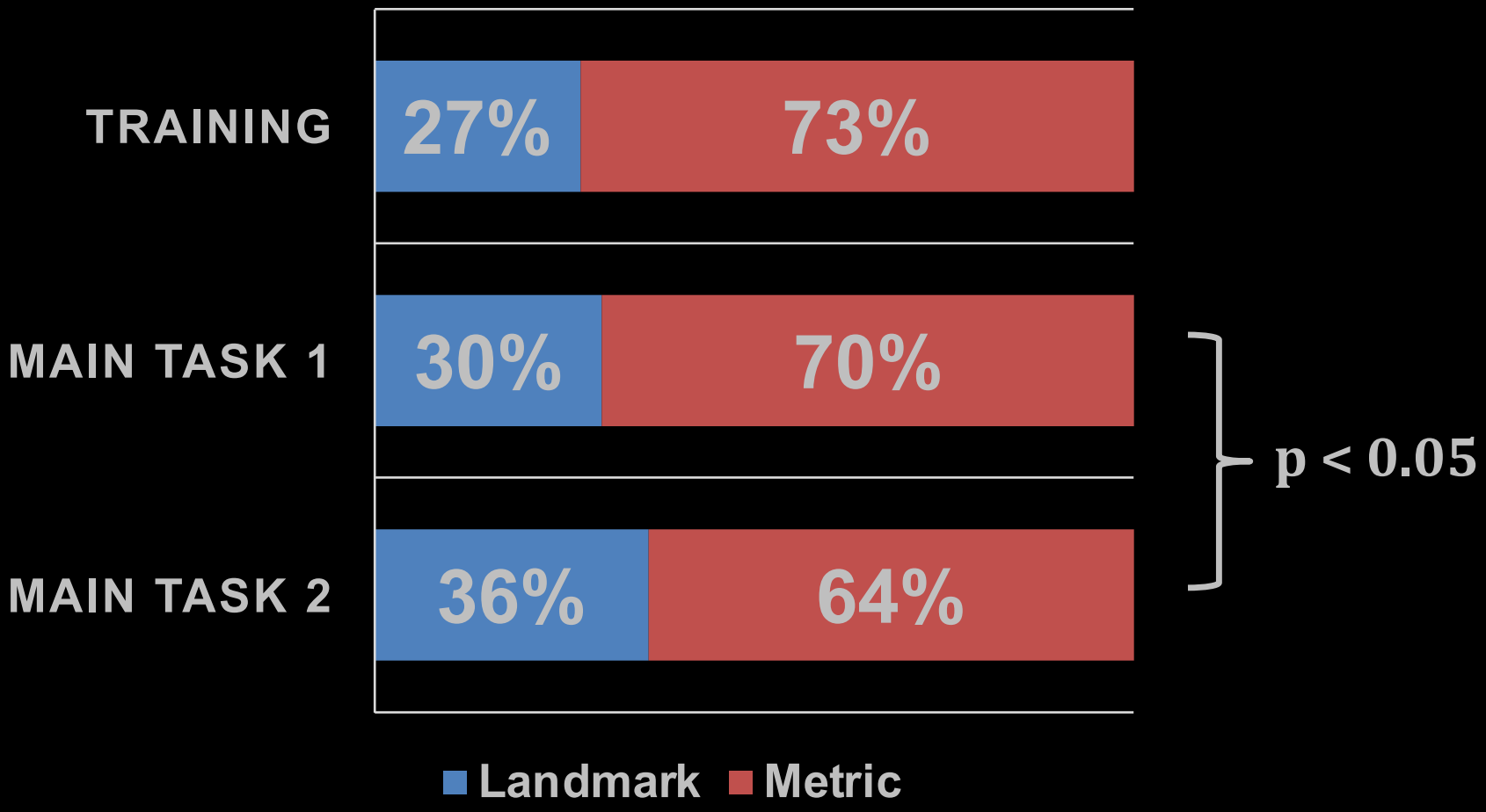
- Other dialogue moves based on how people assessed robot capabilities

(Marge et al., 2017; RoboNLP)

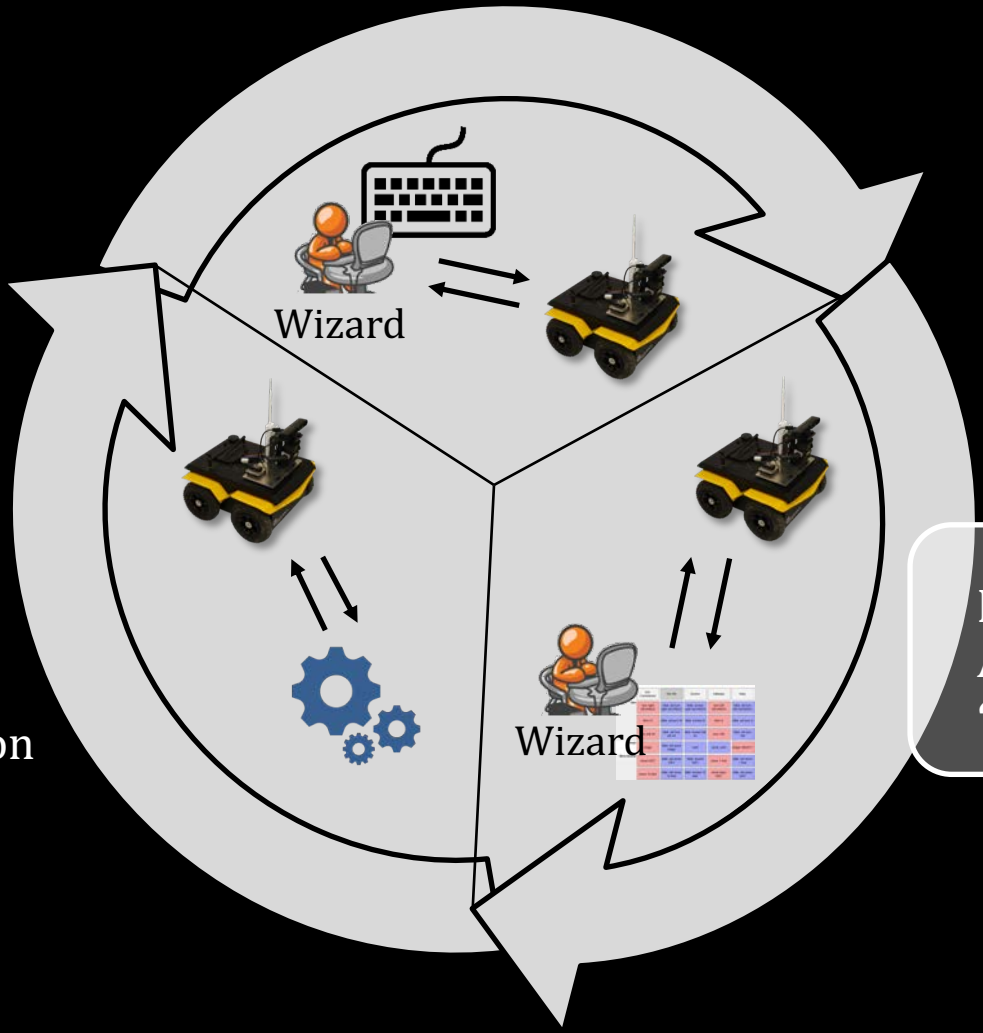
Landmark vs. Metric Results



Landmark vs. Metric Results



Phase 1 Exploratory Data Collection



Phase 2
Automate Some
"Wizard" Labor

Phase 3
Full Automation
Of "Wizard"

Transition to Experiment 2

- Observed naturally occurring coordination efforts in Experiment 1
- But…
  - Turn-taking was slow
  - Typed language had many variations
- With data collected in Experiment 1, developed a graphical interface for wizard to handle language

(Bonial et al., 2017; AAAI FSS)

Screens	Wiz-Commander	Wiz-RN	Rooms	Hallways	Alley
Turn	turn right DEGREES	fdbk: will turn right DEGREES	fdbk: turned right DEGREES	turn left DEGREES	fdbk: will turn left DEGREES
	face W	fdbk: will turn W	fdbk: turned W	face S	fdbk: will turn S
	turn left 45	fdbk: will turn left 45	fdbk: turned left 45	turn 180	fdbk: will turn 180
Image	image	fdbk: will send image	sent	done, sent	image OBJECT
Move General	move DIST	fdbk: will move DIST	fdbk: moved DIST	move 1 foot	fdbk: will move 1 foot
	move 10 feet	fdbk: will move 10 feet	fdbk: moved 10 feet	move back DIST	fdbk: will move DIST

DM-Wizard Interface: Experiment 2

What does it do?

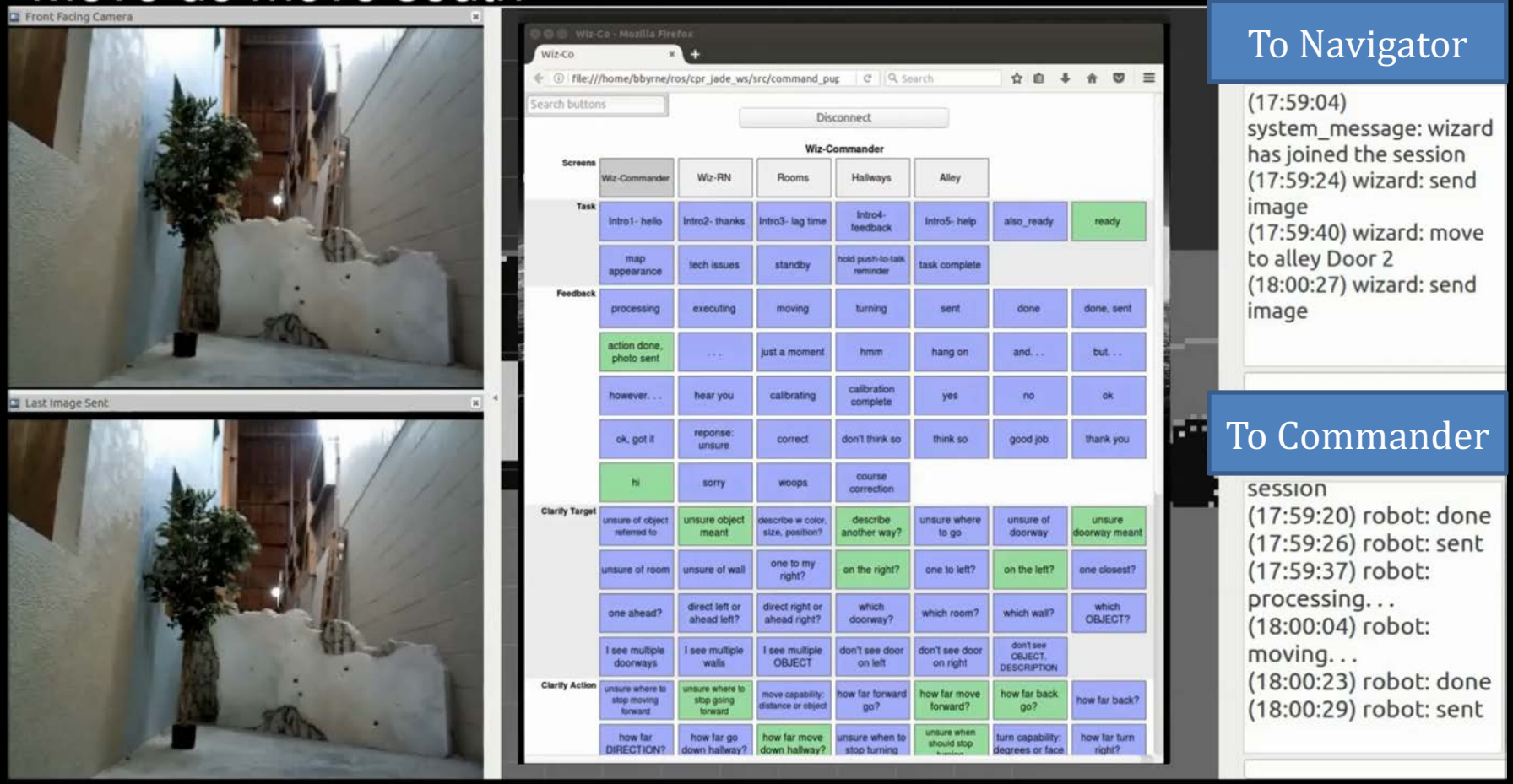
- Instead of typing (Exp 1) DM-Wizard presses a button in interface, which sends a text response to either the Commander or the Robot Navigator

Why is it important?

- Represents the sum total of possible responses the "robot" can give to the Robot Navigator and Commander
  - Quality of automated system contingent upon interface design decisions
- Goal was domain coverage while balancing need to create an interface that is quick and easy to use

Interface Demo Video: Experiment 2

"Move do move south"

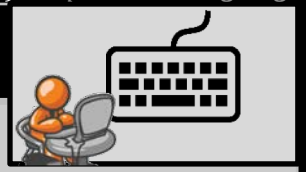


Experiment 2 Setup

- Equivalent to Experiment 1, except DM-Wizard used an interface

- Ran 10 different participants
  → ~10 hours of dialogue
  - 5 female, 5 male (age range 18-58, mean = 42)

Experiment 1: Free Response Mode



Participant: Move forward

Robot (DM-Wizard): You can tell me to move to something that you see.

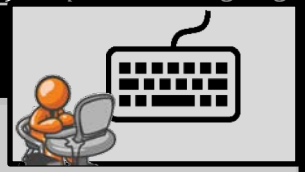
P: Go forward four feet

P: Continue to move down the hallway

Robot (DM-Wizard): How far should I continue down the hallway? Would you like me to reach something that you can see or go a certain distance?

P: Go down the end of the hallway where it appears to be a clock on the wall

Experiment 1: Free Response Mode



Participant: Move forward

Robot (DM-Wizard): You can tell me to move to something that you see.

P: Go forward four feet

P: Continue to move down the hallway

Robot (DM-Wizard): How far should I continue down the hallway? Would you like me to reach something that you can see or go a certain distance?

P: Go down the end of the hallway where it appears to be a clock on the wall

Experiment 2: Structured Response Mode

Participant: Robot go forward

Robot (DM-Wizard): I don't know where or when to stop moving forward. You can tell me to move a certain distance or to move to an object.

P: Robot move forward one feet

P: Can you move forward

Robot (DM-Wizard): I don't know where or when to stop moving forward. You can tell me to move a certain distance or to move to an object.

P: Can you stop at the second door

Analysis: Experiment 2 Coverage

- Analyzed number of times DM-Wizard would respond with a non-understanding
  - "I'm not sure."
- Across all dialogues, observed extremely **high coverage** using this measure
  - 11 of 926 total transactions had non-understanding
- 5 of 10 participants had at least transaction with a non-understanding

Novel request outside capabilities

Participant: how far away are you from the object in front of you?

Robot (DM-Wizard): I'm not sure.

Analysis: Experiment 1 vs. 2 Efficiency

- Analyzed **efficiency** within transaction units (TUs)

Analysis: Experiment 1 vs. 2

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing…		
			image sent
	sent		

Transaction Unit (TU)

Analysis: Experiment 1 vs. 2

Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
face the doorway on your right			
and take a picture			
	there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?		
the door ahead of you on the right			
		move to face the door ahead of you on the right, image	
	executing...		
			image sent
	sent		

Successful Interaction (SI)

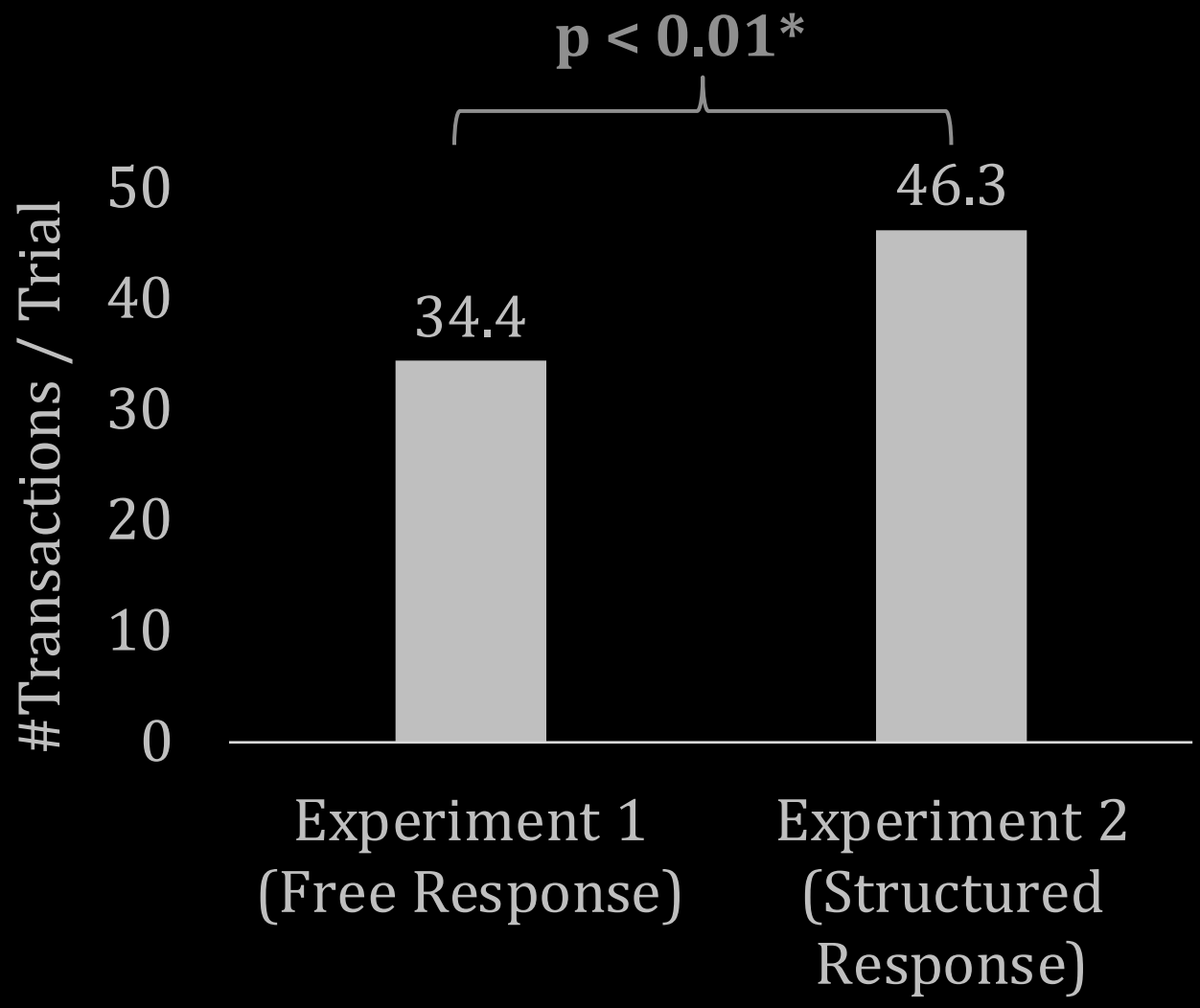
Transaction Unit (TU)

Analysis: Experiment 1 vs. 2

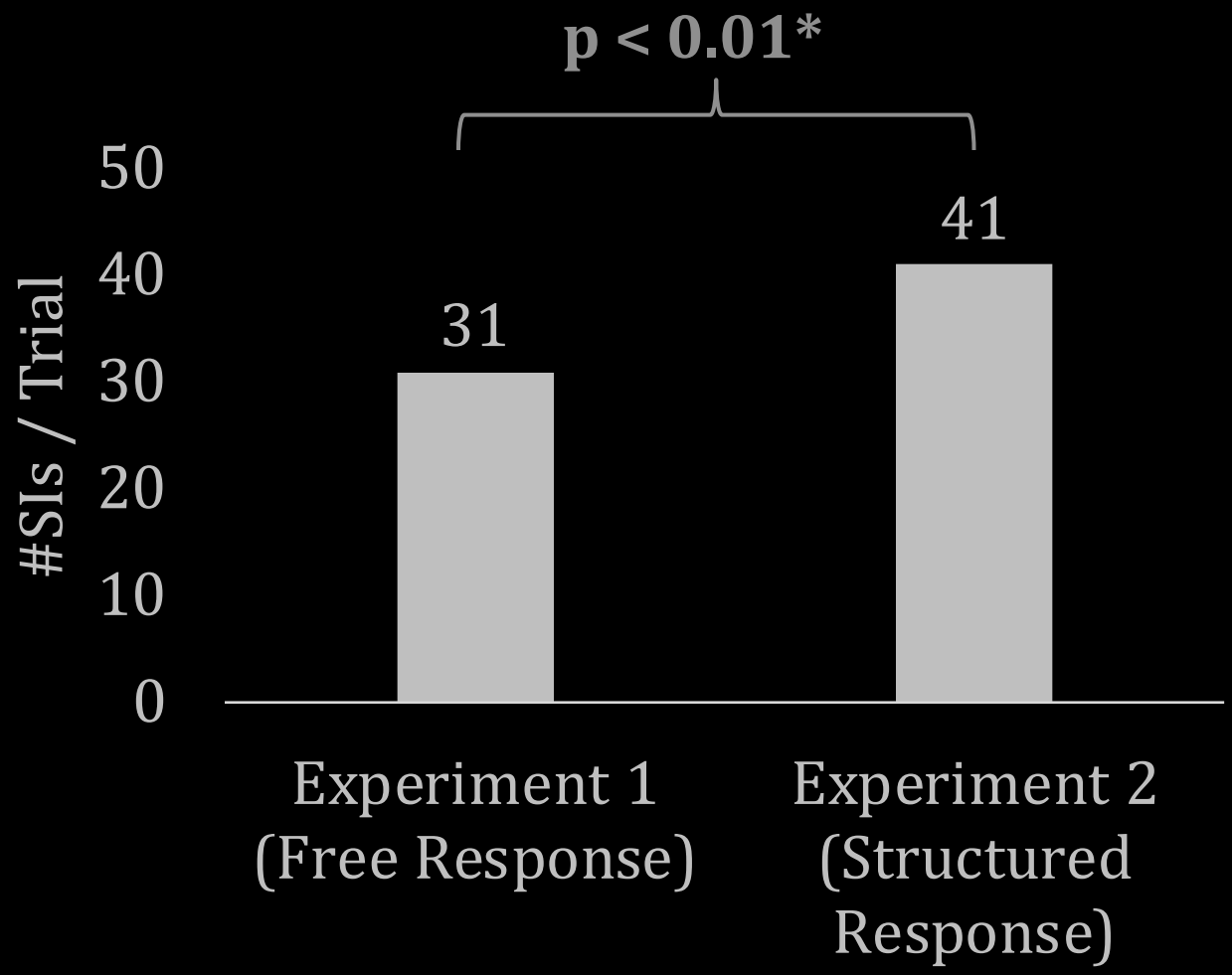
- Analyzed **efficiency** within transaction units (TUs) between
  - Experiment 1: Free Response Mode
  - Experiment 2: Structured Response Mode

- Per trial:
  - Number of TUs
  - Number of Successful Interactions (SIs)
  - Sum of utterances between Commander and DM-Wizard

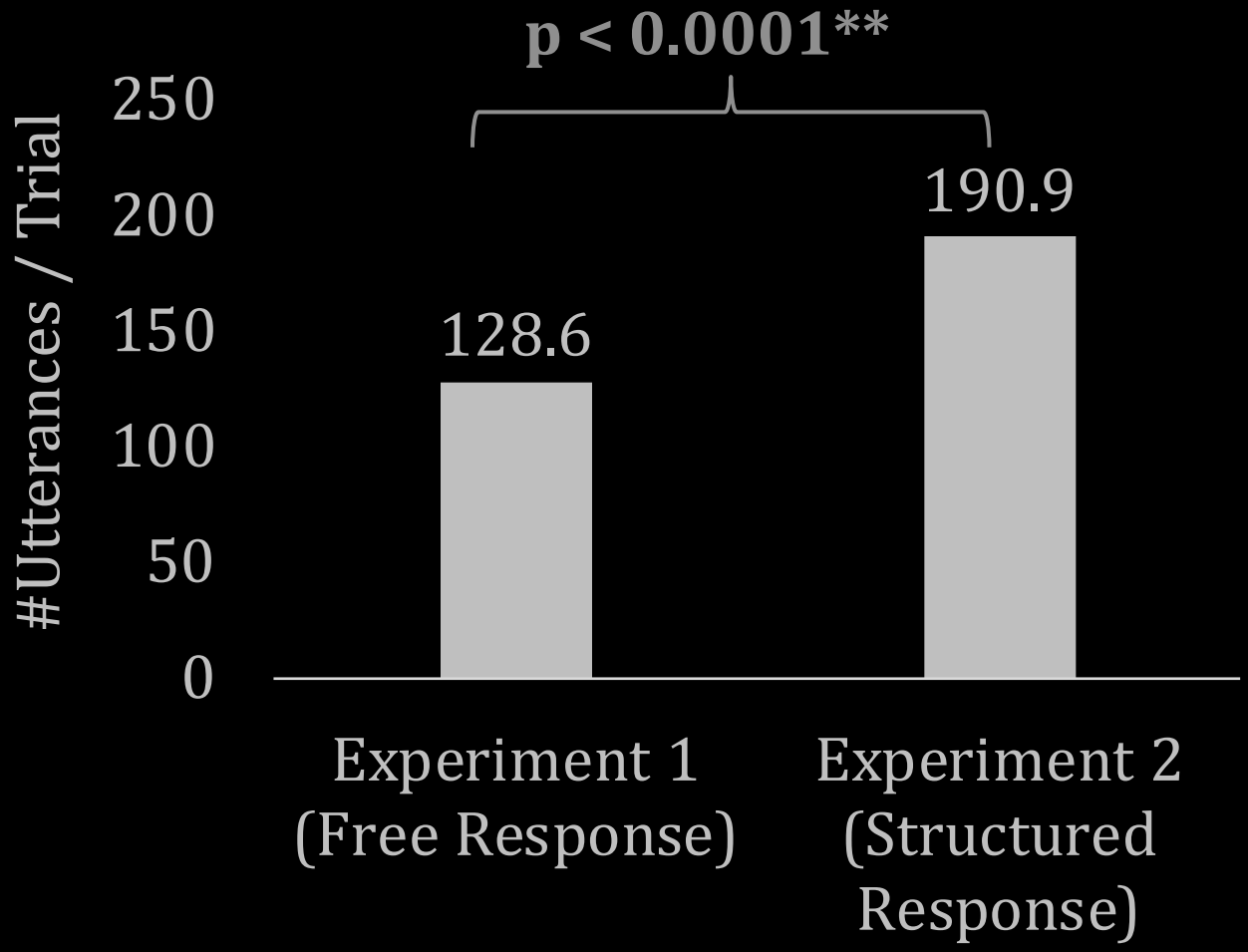
Results: Transaction Units (TUs)



Results: Successful Interactions (SIs)



Results: Total Utterances between Commander and DM-Wizard



Discussion

- Maintained sustained quality of instruction handling & coverage
- Responses provide natural classification of corresponding participant utterances

Discussion

- Approach holds promise for collecting **efficient** dialogue data

- Structured Response Mode (Experiment 2) with the interface supports generating dialogue
  - Enables participants to issue more instructions
  - Balances efficiency with naturalness
  - Dialogue now easier to incorporate in training dataset

Lessons Learned

- Speed and responsiveness at processing dialogue is crucial
- Simple messages (*"processing…"*) provide transparency & allow robot to "hold the floor"

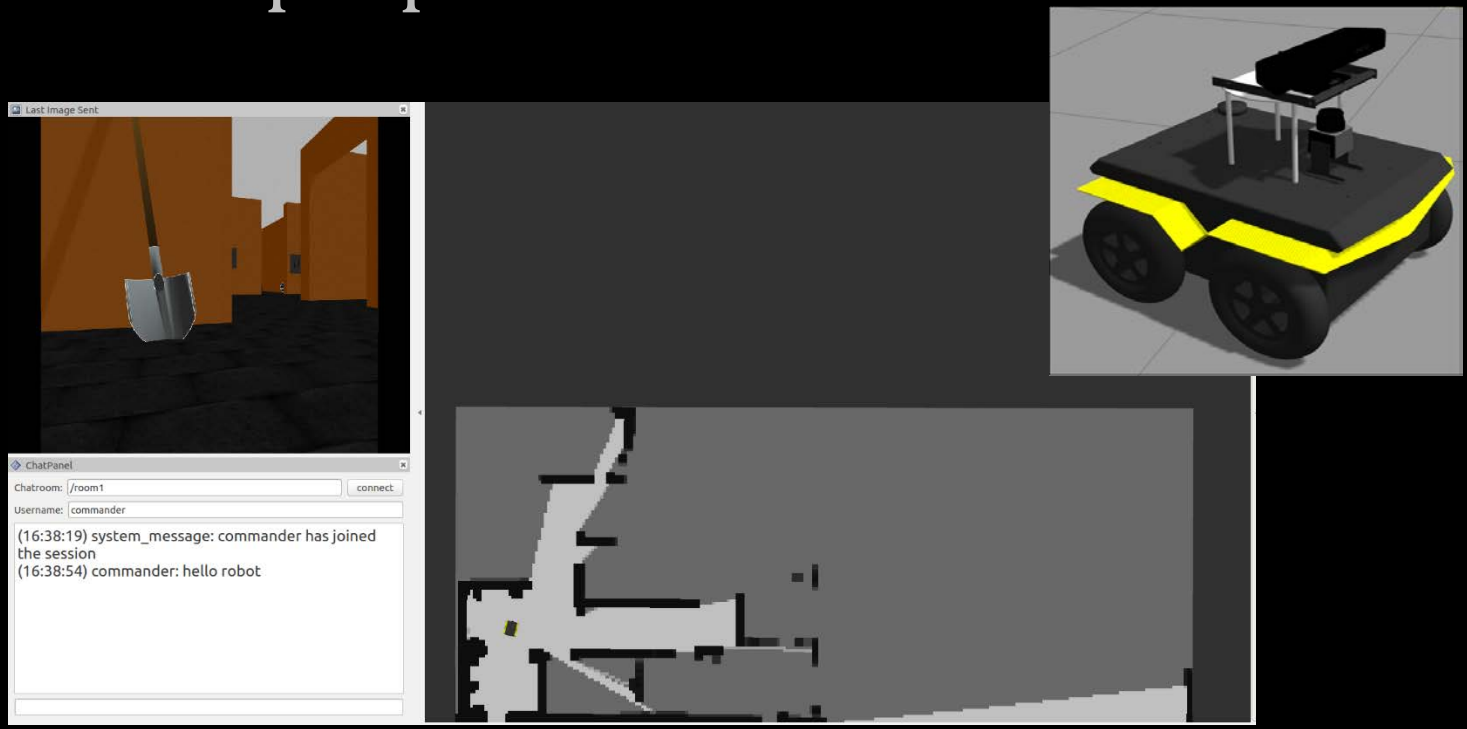
- Essential responses can be categorized:
  - Common status updates and clarifications
  - Slightly generalized buttons ("which one" over "which cone")
  - Flexible templates for uncommon referents ("I see <…>")
  - Very general non-understanding ("I'm not sure.")

Road Map

1. Motivation and Overview
2. Experiments Towards Natural Dialogue
3. **Ongoing Work**

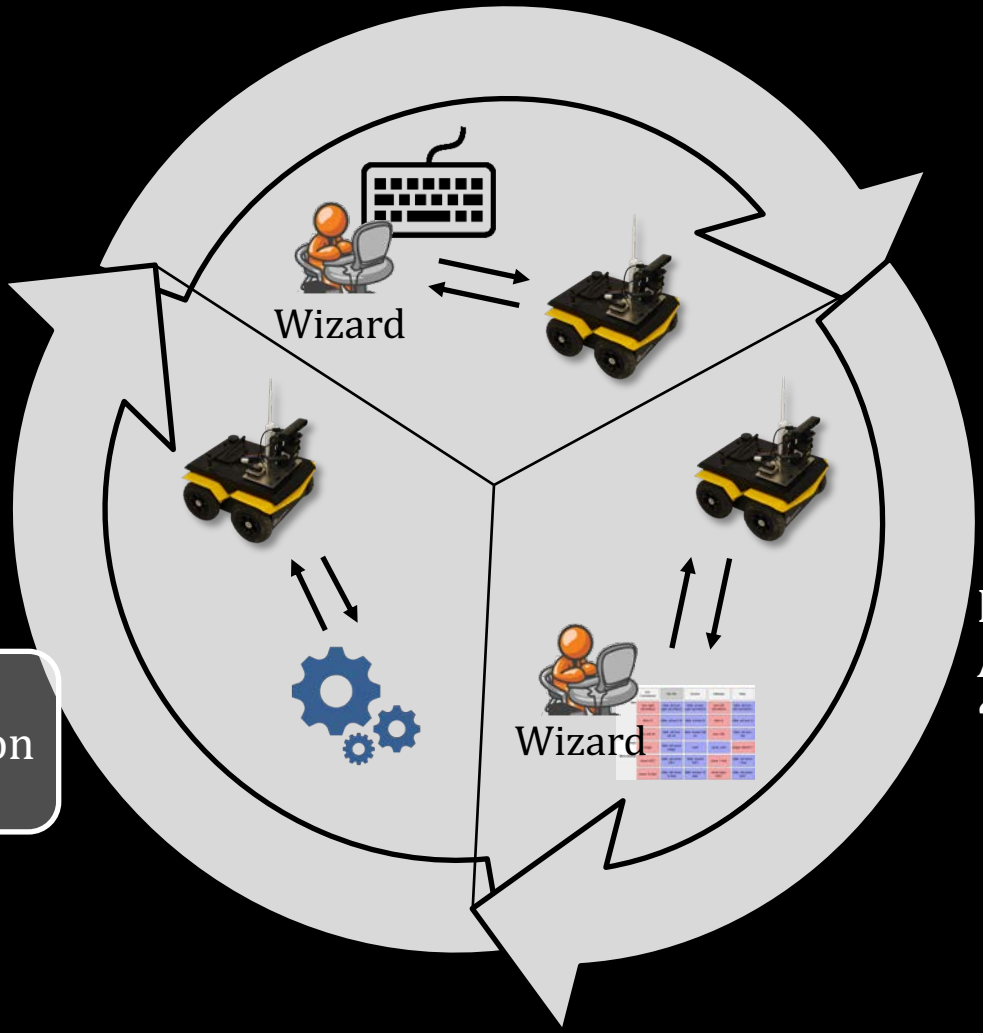
Transition to Simulation

- Moved to simulation to collect more data from more people



(Henry et al., 2017; WiNLP)

Phase 1 Exploratory Data Collection



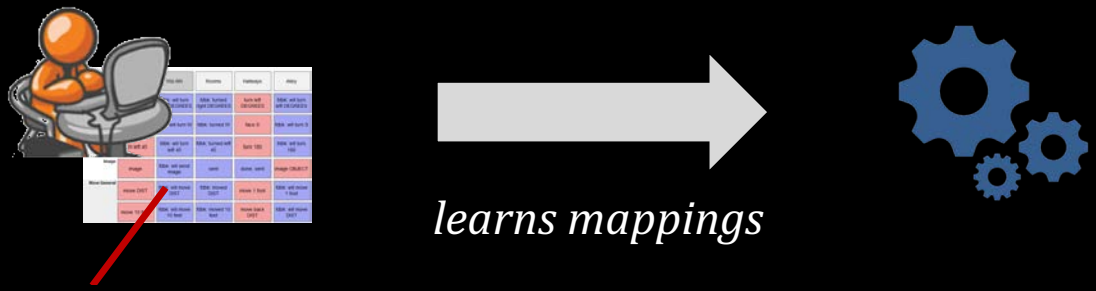
Phase 2
Automate Some
"Wizard" Labor

Phase 3
Full Automation
Of "Wizard"

Way Forward: Learning

- Robot's natural language generation capabilities can learn from DM-Wizard selections & responses

Participant: *"Move forward"*



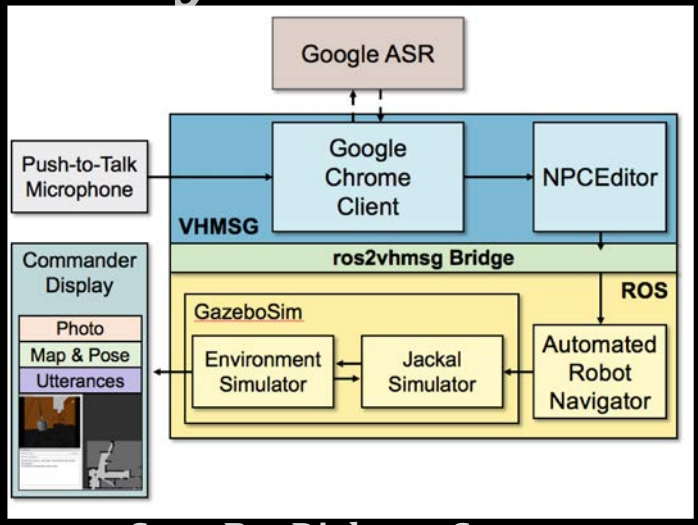
DM-Wizard: *"I'm unsure when or where to stop…"*
Auto DM: Learns selections from DM-Wizard

ScoutBot Dialogue System

- Created ScoutBot dialogue system incorporating ICT Virtual Human Toolkit
- Supports rapid creation of new domains
- Uses "intent retrieval" technique to select responses to user's utterance by matching response to training data

$$r = response\left(\max_{t \in C} \frac{q^T t}{|q||t|}\right)$$

Given user utterance q and corpus C, retrieve utterance turn t in C that is most similar to q and return response to q



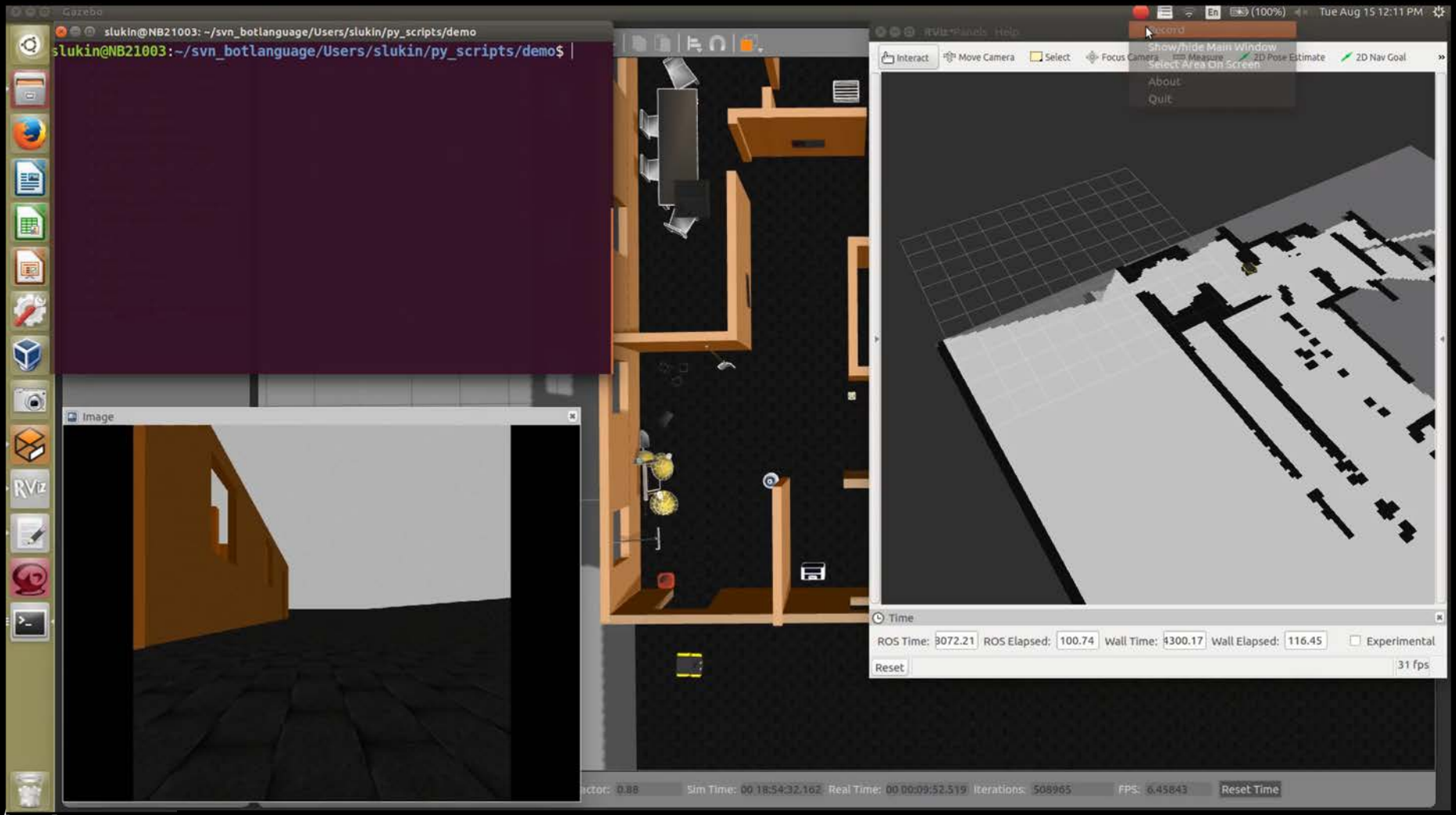
ScoutBot Dialogue System (Lukin et al., ACL 2018)

- Constructed mobile simulation platform using ROS that enables rapid dialogue collection
- Maintains sensory data similar to physical platform
- High-fidelity simulations of indoor/outdoor environments

Outdoor Simulation

DEVCOM ARMY RESEARCH LABORATORY

Video Demo: Autonomous Dialogue Manager



Video recorded by Stephanie Lukin (ARL) and Felix Gervits (USC/ICT intern)

Planning in Response to Language

- How should dialogue work with access to full situation?
  - History
  - Environment
  - Uncertainty

- How should robot behave in response to natural language?

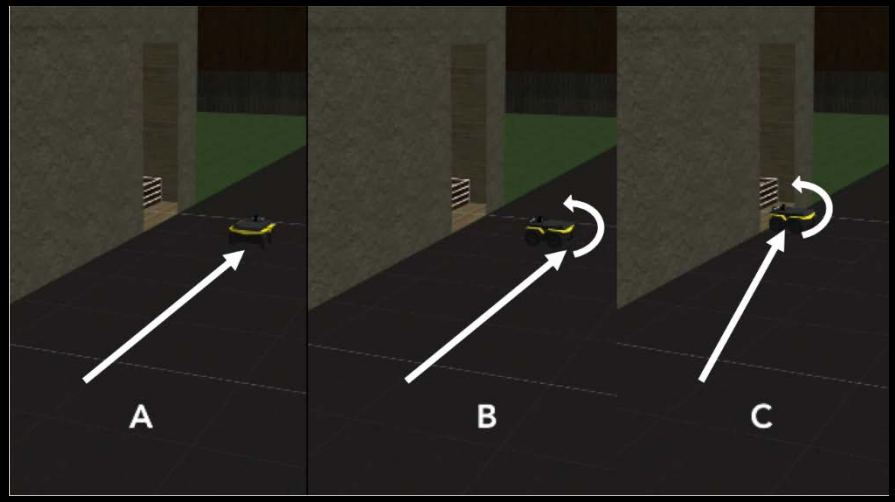
Evaluating Robot Behavior in Response to Natural Language

Problem: Given a command to navigate, how should a robot execute the command?

- Movement should match human expectation for task efficiency and naturalness
- Many possible variations

Progress: Web study with 21 ARL volunteers

- Analyzed in-house natural language navigation corpus to uncover ambiguous instructions
- Identified set of instruction classes with similar intents
- Incorporated HRI parameters for "natural behavior"
- Participants evaluated robot movement from videos



Example command: "Go to the Doorway"

Preliminary findings:

- Robot movement more accurately meets user expectation when:
  - it navigates with an awareness of its environment
  - demonstrates a sense of self-safety

USC

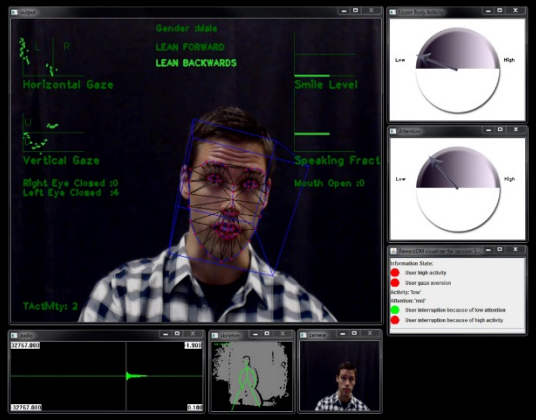
Pooja Moolchandani
(USC undergraduate)

(Moolchandani et al., 2018; HRI LBR)

DEVCOM ARMY RESEARCH LABORATORY

Future Directions

- Scale to other domains
- Multimodal information processing



- Return to the physical platform

Conclusions

Methodology for supporting natural communication with robots

- Observed Commanders adapting use of metric and landmark references as they gained experience with robot
- Need to handle both metric and landmark

- Graphical interface automating wizard labor balances efficiency of dialogue collection with coverage
- Dataset collected contains language and robot data, will be released in next year

Collaborators

Project Members at ARL

Claire Bonial	Linguistics	(Adelphi)
Ashley Foots	Audiology	(APG)
Cory Hayes	Human-Robot Interaction	(Adelphi)
Susan Hill	Human-Robot Interaction	(APG)
Stephanie Lukin	Computational Linguistics	(ARL West)
Matthew Marge	Computational Linguistics	(Adelphi)
Kimberly Pollard	Biology	(ARL West)
Clare Voss	Computer Sci., Linguistics	(Adelphi)
Cassidy Henry	Linguistics	(SMART Scholar)

USC Institute for Creative Technologies

Project Members at USC/Institute for Creative Technologies

Ron Artstein	Linguistics
Anton Leuski	Computer Science
David Traum	Computational Linguistics

And a host of interns!

DEVCOM ARMY RESEARCH LABORATORY

Thank you!

Questions?

Email: matthew.r.marge.civ@mail.mil

We are hiring! Contact me if interested in postdoctoral or internship positions.

DEVCOM ARMY RESEARCH LABORATORY

USC Institute for Creative Technologies

References

- Matthew Marge, Claire Bonial, Brendan Byrne, Taylor Cassidy, A.William Evans, Susan G. Hill, and Clare Voss. 2016. Applying the Wizard-of-Oz Technique to Multimodal Human-Robot Dialogue. In Proc. of IEEE RO-MAN.
- Matthew Marge, Claire Bonial, Ashley Foots, Cory Hayes, Cassidy Henry, Kimberly Pollard, Ron Artstein, Clare Voss, and David Traum. "Exploring Variation of Natural Human Commands to a Robot in a Collaborative Navigation Task." In Proceedings of the First Workshop on Language Grounding for Robotics (RoboNLP), pp. 58-66. 2017.
- Claire Bonial, Matthew Marge, Ron Artstein, 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”. In Proc. of the AAAI Fall Symposium Series: Natural Communication for Human-Robot Collaboration.
- Cassidy Henry, Pooja Moolchandani, Kimberly Pollard, Claire Bonial, Ashley Foots, Cory Hayes, Ron Artstein, Clare Voss, David Traum, and Matthew Marge. 2017. Towards Efficient Human-Robot Dialogue Collection: Moving Fido into the Virtual World. In Proc. of the SoCal Robotics Symposium and the Workshop on Women and Underrepresented Minorities in Natural Language Processing.
- Pooja Moolchandani, Cory J. Hayes, and Matthew Marge. 2018. Evaluating Robot Behavior in Response to Natural Language. In HRI’18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion.
- 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. Dialogue Structure Annotation for Multi-Floor Interaction. International Conference on Language Resources and Evaluation (LREC), 2018.