Exploring Variation of Natural Human Commands to a Robot in a Collaborative Navigation Task

U.S. ARMY RDECOM®

ARL U.S. ARMY RESEARCH LABORATORY 25 YEARS 1992-2017

USC Institute for Creative Technologies

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

1 U.S. Army Research Laboratory, Adelphi, MD 20783 USA   2 USC Institute for Creative Technologies, Playa Vista, CA 90094 USA   email: matthew.r.marge.civ@mail.mil

Introduction

- Challenge: Instruction-giving to robots depends on how people perceive them as conversational partners

- Experiment: Elicited robot-directed language through back-and-forth dialogue
  - 10 participants (8m, 2f), one hour of dialogue each

- Finding: Participant strategy in specifying endpoints during navigation changes over time, increasing in landmark references over metric units

Collaborative Navigation Task

- Goal: Collect dialogue data that is computationally tractable without sacrificing naturalness

- Focal task: Collaborative search-and-navigation with remote human teammate and on-location robot

- Method: Wizard-of-Oz with two human wizards to stand in as robot AI supports collecting data for training an initial system
  - **Dialogue Manager (DM):** intermediary, routes typed communications to Commander Participant and RN
  - **Robot Navigator (RN):** moves robot based on DM instruction

Commander Participant

VIEWS

VERBAL COMMANDS

Dialogue Manager

Speech -> Constrained Language

Executes commands

Robot Navigator
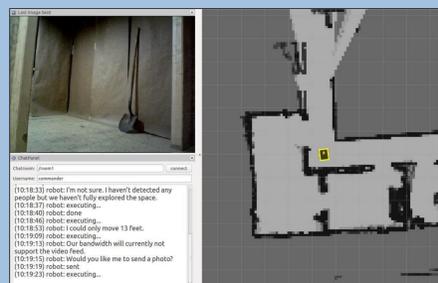
RN MOVES ROBOT

"Behind the scenes"
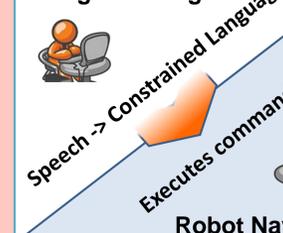
Two experimenters represent separable, automatable functions.

Approach: Eliciting Natural Language

- How can we collect natural communications, given that people may change strategies over time?

- DM followed guidelines to govern decisions
  - Minimal requirements: Clear action and endpoint
  - Guidelines provide response categories

> 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.`

Sample DM guideline for consistent dialogue behavior.

- To assess possible variation, annotated data for dialogue structure

- Four message streams (*two audio speech streams; two typed streams*) from Commander, DM, and RN

- Analysis focus: Parameters on motion commands
  - *Landmark:* Object references such as *doorway, table*
  - *Metric:* Specific distances such as *2 feet, 90 degrees*

	Commander (Audio Stream 1)	DM->Commander (Chat Room 1)	DM->RN (Chat Room 2)	RN (Audio Stream 2)
	face the **doorway** on your right		*Dialogue Move*	
	and take a picture		*Dialogue Move*	
		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		

Instruction Unit / *Time* / *Transaction Unit*

Annotation set marks dialogue moves, and structures such as instructions and transactions.

Results

Dialogue Moves

Dialogue Move	% Instruction Units (IU)
Command	94%
Send-Image	52%
Rotate	47%
Drive	42%
Stop	3%
Explore	1%
Request-Info	4%
Feedback	3%
Parameter	2%
Describe	1%
Total #IUs	858

Notable results:
- Send-Image appears in nearly half of all IUs
- Rotate and Drive also common instructions
- Other dialogue move usage based on assessment of robot capabilities

Landmark vs. Metric Usage in Dialogue Moves

TRAINING	27%	73%
MAIN TASK 1	30%	70%
MAIN TASK 2	36%	64%

$p < 0.05$

■ Landmark ■ Metric

Proportions of Landmark to Metric in Command:Rotate and Command:Drive moves.
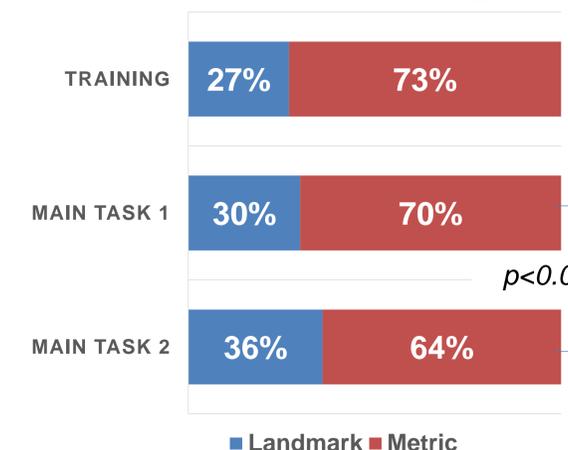
Notable results:
- Overall
  - 75% of IUs contained Metric mentions
  - 37% of IUs contained Landmark mentions
- Metric units initially dominant
- Subsided in favor of landmarks

Discussion & Conclusions

- Observed naturally occurring coordination efforts as Commanders gained experience with robot

- Effective language grounding will require interpretation of both metric and landmark usage

- Image requests very common due to Commander's limited situational awareness

- Dataset collected contains language and robot data, will be released in the next year

- Future work: Automate DM response generation with graphical interface