Development of a Squad Level Vocabulary for Human-Robot Interaction

Daniel Barber[1], Ryan W. Wohleber[1], Avonie Parchment[1], Florian Jentsch[1], and Linda Elliott[2]

[1] Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA
{dbarber,rwohlebe,aparchme}@ist.ucf.edu, florian.jentsch@ucf.edu
[2] Army Research Laboratory, Fort Benning, GA, USA
linda.r.elliott.civ@mail.mil

Abstract. Interaction with robots in military applications is trending away from teleoperation and towards collaboration. Enabling this transition requires technologies for natural and intuitive communication between Soldiers and robots. Automated Speech Recognition (ASR) systems designed using a well-defined lexicon are likely to be more robust to the challenges of dynamic and noisy environments inherent to military operations. To successfully apply this approach to ASR development, lexicons should involve an early focus on the target audience. To facilitate development a vocabulary focused at the squad level for Human Robot Interaction (HRI), 31 Soldiers from Officer Candidate School at Ft. Benning, GA provided hypothetical commands for directing an autonomous robot to perform a variety of spatial navigation and reconnaissance tasks. These commands were analyzed, using word frequency counts and heuristics, to determine the structure and word choice of commands. Results presented provide a baseline Squad Level Vocabulary (SLV) and a foundation for development of HRI technologies enabling multi-modal communications within mixed-initiative teams.

Keywords: Human-robot interaction, human-robot teaming, mixed-initiative teams, speech recognition.

1 Introduction

The U.S. military is increasingly looking to robots to perform complex and hazardous tasks such as reconnaissance and Explosive Ordinance Disposal (EOD). According to General Robert Cone of the U.S. Army Training and Doctrine Command (TRADOC), robots could replace one-fourth of combat Soldiers by 2030[1]. Today's robots bring advanced technology to the field, helping to keep Soldiers out of harm's way. As of 2011, the U.S. deployed over 2,000 ground robots to Afghanistan supporting EOD teams and entry control points [2]. However, state-of-the-art robotic ground systems currently serve as tools for Soldiers, not teammates. To enable unmanned systems operating in teams with Soldiers and manned vehicles as described by Department of

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part I, LNCS 8525, pp. 139–148, 2014.

Defense (DoD) Unmanned Systems Roadmap [3], advanced Human-Robot Interaction (HRI) capabilities are needed.

Current HRI paradigms utilize teleoperation, with control through physical input devices such as joysticks, track balls, and touch screens. These input devices constrain the robot operator in that they often require operation from stationary positions, hand eye coordination to map controls to output, dexterity to operate controls, and additional training and practice time [4]. Teleoperated robots also inherently add an additional step in the chain of communication, where a commander must direct a robot through a specially trained operator [5]. Advancements in robot perception, intelligence, and mobility will soon enable Soldiers to work with robots as collaborators rather than tools [6]. A key aspect of these advancements is communication using modalities Soldiers are already familiar with, facilitating Soldier-Robot (SR) collaborations without the addition of workload or training.

Speech provides a natural method of communication to robots. Research into speech for commanding robots clearly demonstrates it to be quicker, more intuitive, and superior to manual control for discrete tasks and Soldier multitasking and situation awareness [7], [4]. Ideally, continuous speech recognition systems would allow Soldiers to speak naturally to a robot collaborator as one might speak to another person. Unfortunately, full Natural Language Processing (NLP) and Automatic Speech Recognition (ASR) technology is limited [8]. Current systems have trouble with word classification accuracy and limited semantic understanding of continuous speech [9], [10]. Manual control is still more effective for continuous tasks than speech control [4]. A further challenge for true continuous speech NLP is working within the environment of military operations which includes noise from background sounds such as gunfire, echoes, and reverberations. Additional factors characteristic of this atmosphere including stress, fear, sickness, and pain may induce voice variability [11]. Moreover, special circumstances may change the way words are spoken [4], [11], [12]. Finally, common use of discrete words (e.g. commands) instead of normal speech in military environments may also inhibit speech understanding [7].

Due to current limitations in ASR technology and the use of discrete word commands typical of a military setting, programming systems to understand discrete commands using a limited lexicon is likely to show better overall performance and robustness in challenging environments. Such was the approach taken by Tellex et al. [8] who developed a body of natural language commands for operating a robotic forklift for the purposes of training and evaluating a robust robot speech recognition system. Applying a lexicon scoped to the application domain limited the NLP search space, resulting in the robotic forklift demonstrating an ability to competently handle discrete commands [8]. To successfully apply this same approach to develop ASR systems for SR teaming, development of lexicons should involve an early focus on the target audience, such as U.S. Soldiers [13], [4].

Until new methods and algorithms, along with the hardware necessary for quickly handling large search spaces are developed to enable robots to fully interpret natural speech, a Squad Level Vocabulary (SLV) facilitating natural and intuitive communication within mixed-initiative teams is needed. A SLV scoped to commands Soldiers already use supports improved reliability of ASR technologies by reducing the input search space [14]. Further, by tailoring the speech commands to the preferences and conventions of the user, the language used for commands can be better retained and would allow for more efficient operation. Finally, this approach eliminates the need for additional training for robot specific communications. The goal for the present study was to inform the development of a SLV within spatial navigation and reconnaissance tasks.

2 Method

2.1 Participants

A total of thirty-one (N = 31; 25 male, 6 female, mean age = 24.9) Soldiers assigned to 11B (Infantry) or similar (e.g., Combat arms, Officer Candidate School) Military Occupational Specialty (MOS), located at Ft. Benning, GA participated in the study. Researchers informed participation in the study was voluntary and that they may withdraw at any time.

2.2 Spatial Navigation and Reconnaissance Survey

A Spatial Navigation and Reconnaissance (SNR) survey was created to begin development of an initial SLV within SNR tasks. In this survey, Soldiers were instructed to issue commands to a robot based on their level of experience and training. The survey consisted of four parts comprised of open ended and multiple choice questions regarding commands to give an autonomous robot. Commands for Parts 1 through 3 involved either (Check) reporting obstructions in lines of movement, (Move) moving to a location, or (Survey) performing surveillance at a location. Each command prompt included an image of a landscape with two buildings, the participant's avatar within the scene, and the robot. Participants were instructed to pretend they were in the scene and use the spatial information provided to issue one of the three commands. Language used within the survey training materials and questions was chosen to avoid words that may influence the responses from participants.

In Part 1, the images presented with each item included the location of the robot and the area of interest for commands. Participants were informed that the robot only has information about its current location, where buildings are, and where their avatar is, but not command marker in the image. The command markers in all images included a circle or arrow highlighting where the robot was in the scene and an 'X' representing the point of interest to command the robot to move to or observe. Command markers in the images complemented the questions and aided participants in understanding what the intent of each command they must give was while limiting use of words that may bias their responses (e.g. near building, right side). Fig. 1 illustrates an example Move command question with command markers from Part 1 of the survey.

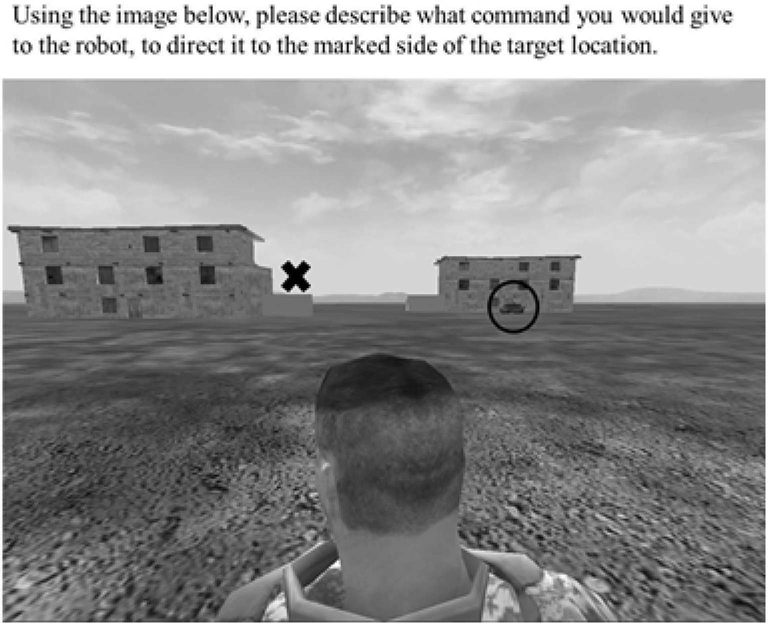


Fig. 1. Example question and image from Part 1 of the SNR Survey. In the image command markers include the 'X' that represents the location to command the robot (circle marker) to move to.

Part 2 of the survey augmented the images from Part 1 to include an aerial strip map with cardinal directions and building labels that could be used to help guide the robot. Similar to Part 1, participants were instructed that the robot had the same information as them (including the map), except for command markers in the scene and on the aerial strip map. Fig. 2 represents an example question from Part 2 of the survey.

Part 3 included multiple choice questions for Move commands using images from Parts 1 and 2. For each question, participants answered three multiple choice questions comprising a complete Move command sequence comprised of: a move command, position designation, and stop criteria. An example answer choice could result in the command, "Move to the South side of the left building." Part 3 was comprised of four images from Part 1 and four from Part 2 (with aerial strip map). Answer choices for the images from Part 1 used spatial phrases relative to the user (e.g. left, right, near, far), where images from Part 2 also included cardinal directions and building labels from the map (e.g. North, South, Building Alpha). Multiple choice answer options for Part 3 were derived from U.S. Army Field Manuals and the DoD Dictionary of Military Terms and were reviewed by a Non-Commissioned Officer (NCO) at Ft. Benning, GA. Fig. 2 represents an example question from Part 3 of the survey.

Using the image and strip map provided below, please describe what command you would give to the robot, to direct it to the marked side of the target location.

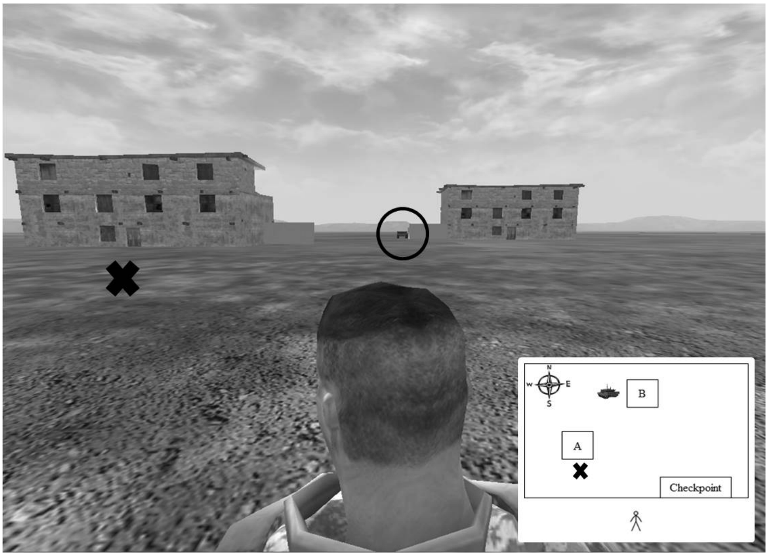
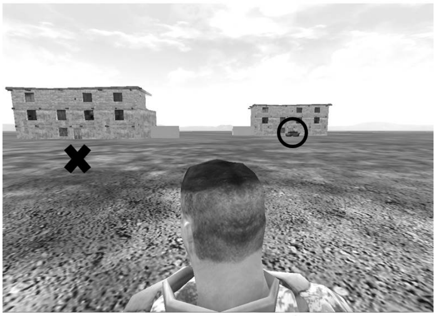


Fig. 2. Example question and image from Part 2 of the SNR Survey. In the image the ‘X’ represents the location to command the robot (circled) to move to



Using the provided figure, please build a speech command by circling the selections that you believe are the most intuitive, based on your experience and training, when commanding the robot to the marked side of the target location. Mark a choice for Column A, Column B, and Column C.

Column A	Column B	Column C
a) Travel to	a) the front side of	a) of the near building
b) Go to	b) the 6 o’clock side of	b) of my 11 o’clock
c) Drive to	c) the South side of	c) of the front building
d) Maneuver to	d) The near side of	d) of the left building
e) Move to		e) of the West building

Fig. 3. Example question and image from Part 3 of the SNR Survey. In the image the ‘X’ represents the location to command the robot (circled) to move to. Each column (A, B, and C) represents an answer choice to construct a command.

Part 4 presented no visual aids, and questions were different from the previous sections. At least one open-ended and one multiple-choice question addressed each of five new command types. Similar to Part 3, some multiple choice questions split the commands into parts to be answered separately, and multiple choice answer options were derived from U.S. Army Field Manuals and the DoD Dictionary of Military Terms and were reviewed by an NCO.

2.3 Procedure

After completing the informed consent, participants completed a demographics questionnaire. Participants were asked if they have any type of color blindness, and if so, or didn't know, the Ishihara Color Test, [15], was administered. Participants then completed the SNR Survey on paper. Participants completed one section of the survey at a time, returning each section upon completion to the experimenter. This prevented participants from changing their answers when presented with new information from later sections.

3 Results

3.1 Multiple Choice Question Analysis

Multiple-choice questions from Part 3 and Part 4 of the survey were analyzed using Chi-squared tests and word choice frequency counts. For Part 3, there are two categories (map, no map) of commands, each with three sub phrases: move command, position designation, and stop criteria. Table 1 lists overall recommended phrases comprising a full command based on word frequency counts, with significant values noted.

Table 1. Recommended phrases for commanding a robot to a location from Part 3 of the survey

Map	Phrase 1	Phrase 2		Phrase 3
No	Move to*	the {North,South,East,West}	side of*	the {near, far} building
Yes	Move to*	the {North,South,East,West}	side of*	building {Alpha, Bravo}*

Note: Words within braces represent options within the phrase part.
*Significant for $p < .05$

For multiple choice questions in Part 4, each section was self-contained and a single recommendation is made for each question. Recommended commands are based on mean frequency across questions in each set analyzed, Table 2.

Table 2. Recommended phrases for commanding a robot from Part 4 of the survey

Command Question	Recommended Command
Follow a person of interest	Follow* I The target*
Retreat	(No clear recommendation)
Skip location	Avoid* I Location
Skip location	Skip landmark
Identify your location	Give I (No clear recommendation)
Identify your location	Identity your location

Note. Questions with two choices to build a full command are arranged with a divider between parts when necessary.
*Significant for $p < .05$

3.2 Free Response Question Analysis

Handwritten answers to open-ended questions were transcribed onto a spreadsheet. After standardizing words (variations of "building" included "bldng," "Building," and "bldg") and correcting typographical errors to reduce redundancies, a frequency count of each word for each Part of the SNR survey was generated. These words were then categorized by the parts of speech (noun, verb, adjective, etc.) to determine which words were used for each command response. These frequency counts ascertained the preferred word choice once the structure of each archetypal command was determined.

To determine the archetypal structure, it was first established heuristically whether the participant was verbally teleoperating the robot or assumed the robot had autonomous capabilities. Generally, it was assumed that the participants were trying to teleoperate if commands reached a certain level of detail where the responses indicated micromanaging the execution of the higher level task. If the Soldier gave direction or distance information in addition to the destination command, or if the Soldier walked the robot through the steps to get to the destination (e.g. "move North 75 meters. Move East 25 meters") the Soldier was assumed to be teleoperating the robot. Commands for teleoperating the robot were not consistent with automated robot capabilities and were therefore not included in final recommendations.

Commands given by participants who had assumed autonomous capabilities were broken down into phrases. These phrases were grouped into broad command categories such as move commands, position designation, and stop commands. Command categories were then divided into specific phrase types such as *zone-relative-object, zone-relative-self,* or *zone-relative-direction* for position designation category, or *stop-at-destination* or *stop-on-command* for the stop commands category, see Fig. 4.

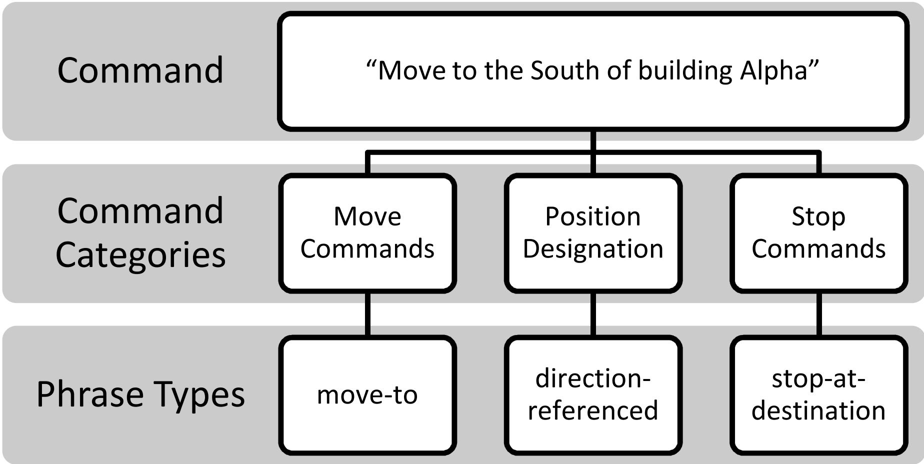


Fig. 4. Example of Move command mapped to command categories and then phrase types

The most frequently used phrase type was then chosen from each category to build an archetypal command phrase for each command type (Check, Move, and Survey). An example archetype command phrase consistent with "Move to the left of the building" would be *move-to* / *zone-relative-direction* / *stop-at-destination*. Word frequency counts from the data within chosen phrase types determined the words used for the command ultimately recommended. Final command recommendations for each command type are contained in Tables 3, 4, and 5.

Table 3. Recommendations for each command type (Check, Move, and Survey) for Part 1

Command Type	Recommended Command
Check	Report obstacles in your path to {left, right, front, back} side of building at my X o'clock
Move	Move to {left, right, front, rear, back} of building
Survey	Report recon Y meters to your {left, right, front, back}

Note: Words within braces represent options within the command. The variable X represents a unit of time on a clock [1-12] and variable Y a distance in meters.

Table 4. Recommendations for each command type (Check, Move, and Survey) for Part 2

Command Type	Recommended Command
Check	Report obstacles from your location to {North, South, East West} side of building {Alpha, Bravo}
Move	Move {North, South, East West} to building {Alpha, Bravo}
Survey	Report recon {North, South, East, West} of current location

Note: Words within braces represent options within the command. Alpha and Bravo building names were within provided maps and may be substituted for other labels.

Table 5. Recommendations for each command type from Part 4

Command Objective	Recommended Command
Follow Person	Follow target
Retreat	Retreat back to point
Skip specific building	Move past building
Return to starting point	Return back to start point
Identify current location	Report location

4 Conclusion

The phrases produced by this heuristic-driven process are useful for future research and design of robotic command languages and interfaces, but they should not be mistaken for conclusive or exact recommendations, with the possible exception of some of the multiple-choice based recommendations. However, consistent with the exploratory goal for this effort, these phrases serve as a necessary first step for a better understanding of preferred word choice and ordering for Soldiers commanding a robot within a reconnaissance and surveillance task. Moreover, this baseline SLV provides a foundation for the development of HRI technologies enabling true multi-modal communication within mixed-initiative teams using a well-known lexicon that is intuitive to Soldiers.

Acknowledgements. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the author's and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. CBS News: U.S. Army general says robots could replace one-fourth of combat soldiers by 2030. In: CBS News, http://www.cbsnews.com/news/robotic-soldiers-by-2030-us-army-general-says-robots-may-replace-combat-soldiers/ (accessed January 23, 2014)
2. Magnuson, S.: "Robot Army" in Afghanistan surgest past 2,000 units. In: National Defense Magazine, http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=300 (accessed February 2, 2011)
3. Office of the Secretary of Defense: Unmanned Systems Integrated Roadmap: FY2013-2038., Washington D.C (2013)
4. Redden, E., Carstens, C., Pettit, R.: Intuitive speech-based robotic control. Aberdeen Proving Ground, MD (2010)
5. Cosenzo, K., Capstick, E., Pomranky, R., Dungrani, S., Johnson, T.: Soldier Machine Interface for Vehicle Formations: Interface Design and an Approach Evaluation and Experimentation. Technical Report ATRL-TR-4678, Aberdeen Proving Ground (2009)
6. Ososky, S., Schuster, D., Jentsch, F., Fiore, S., Shumaker, R., Lebiere, C., Kurup, U., Oh, J., Stentz, A.: The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In: Proceedings of the 2012 SPIE Defense Security & Sensing Symposium (Unmanned Systems Technology XIV Conference DS114), Baltimore, MD (2012)
7. Pettitt, R.A., Redden, E., Carsten, C.B.: Scalablity of Robotic Controllers: Speech-based Robotic Controller Evaluation (ARL-TR-4858). Aberdeen Proving Ground, MD: US Army Research Laboratory, 1-46 (2009)
8. Tellex, S., Kollar, T., Dickerson, S., Walter, M., Banerjee, A., Teller, S., Roy, N.: Understanding natural language commands for robotic navigation and mobile manipulation. In: Proceedings of the National Conference on Artificial Intelligence (2011)
9. Anusuya, M., Katti, S.: Speech recognition by machine: A review. International Journal of Speech Technology 6(3), 181–205 (2009)
10. Yucong, D., Cruz, C.: Formalizing semantic of natural language through conceptualization from existence. International Journal of Innovation, Management and Technology 2(1), 37–42 (2011)
11. Pigeon, S., Swail, C., Geoffrois, E., Bruckner, G., Van Leeuwen, D., Teixeira, C., Orman, O., Collins, P., Anderson, T., Grieco, J., Zissman, M.: Use of Speech and Language Technology in Military Environments, North Atlantic Treaty Organization, Montreal, Canada (2005)
12. Williamson, D., Barry, T., Liggett, K.: Flight Test Results of ITT VRS-1290 in NASA OV-10, Wright-Patterson AFB, OH (1996)
13. Karis, D., Dobroth, K.: Psychological and Human Factors Issues in the Design of Speech Recognition Systems. In: Syrdal, A., Bennett, R., Greenspan, S. (eds.) Applied Speech Technology, pp. 359–388. CRC Press, Ann Arbor (1995)
14. Phillips, E., Rivera, J., Jentsch, F.: Developing a tactical language for future robotic teammates. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Santa Monica, CA (2013)
15. Ishihara, S.: Tests for color-blindness. Hongo Harukicho, Handaya (1917)