How We Talk with Robots: Eliciting Minimally-Constrained Speech to Build Natural Language Interfaces and Capabilities

Kimberly A. Pollard, Stephanie M. Lukin, Matthew Marge, Ashley Foots, Susan G. Hill
U.S. Army Research Laboratory, Adelphi, MD 20783

Industry, military, and academia are showing increasing interest in collaborative human-robot teaming in a variety of task contexts. Designing effective user interfaces for human-robot interaction is an ongoing challenge, and a variety of single- and multiple-modality interfaces have been explored. Our work is to develop a bi-directional natural language interface for remote human-robot collaboration in physically situated tasks. When combined with a visual interface and audio cueing, we intend for the natural language interface to provide a naturalistic user experience that requires little training. Building the language portion of this interface requires first understanding how potential users would speak to the robot. In this paper, we describe our elicitation of minimally-constrained robot-directed language, observations about the users' language behavior, and future directions for constructing an automated robotic system that can accommodate these language needs.

INTRODUCTION

Human-robot teaming is of increasing interest in a range of contexts, such as learning (e.g., Saerbeck et al. 2010), health care (e.g., Moradi Dalvand et al. 2014), transportation (e.g., Bimbraw 2015), and military applications. For example, the U.S. Army has published a Robotic and Autonomous Systems Strategy (2017) that "seeks to achieve unity of effort in the integration of ground and aerial [Robotic and Autonomous Systems] capabilities into Army organizations." Similarly, in the recent Australian Department of Defence publication (2016), *Sharing Defence Science and Technology in the Land Domain 2016-2036*, one theme is "increased reliance on remotely controlled, automatic and autonomous systems." Effective human-robot teaming necessitates coordination between humans and robots, which requires the development of effective communication interfaces.

For a growing number of teaming tasks, simple or legacy interfaces are not ideal. There is a growing need for 1) bi-directionality of human-robot communication, 2) interfaces that require less training, and 3) interfaces that can accommodate physical tasks in an unpredictable physical environment. As robotic technology becomes more capable, with greater autonomous decision-making and action execution, both humans and robots will need to initiate communication, ask questions, seek clarification, or provide status updates using a variety of communication modalities. Systems will be built for use in dangerous, adversarial, or emergency situations. Time and budget may constrain costly training. As needs and constraints increase, we expect there will be ever more demand that an interface be intuitive for the user and require minimal training. Another important trend is the increasing demand for physically-embodied mobile robotic devices that must navigate through and interact with an often unpredictable physical environment. Human-robot communication interfaces that allow team performance of complex physical tasks in the real world must allow the humans and robots to establish sophisticated common-ground understanding with one another.

The focus of this paper is on our current work on the development of a natural language dialogue interface for human-robot communication, with a goal to find synergy among these diverse needs. Our goal is an interface that allows bi-directional communication, requires little training, and permits the flexible performance of tasks in a physical environment. Of the various communication modalities possible, natural language is particularly well-suited to these needs. Natural language dialogue can be an intuitive way for humans to interact with robots while also permitting a robot to ask questions of human teammates when uncertain. To allow flexibility to handle unpredictable physical environments, and in order to make the interface intuitive, we aim to move away from constrained language interfaces that confine users to a limited list of acceptable terms and phrases, i.e., we want to enable the human user to speak naturally, in a minimally-constrained fashion. To build such an interface, we must first understand what humans wish to say to their robot teammates, so an interface can be designed to accommodate the way humans want to, and are likely to, use the system.

A core contribution of our work has been the creation of a body (corpus) of minimally-constrained human-robot dialogue drawn from participant interactions with a robot teammate. Our approach is derived from human-virtual agent dialogue management strategies, but here it is used to handle dialogue necessary to communicate with an embodied robot operating in the physical world.

BACKGROUND

There is a large body of research literature on human-robot communication interfaces. A recent report on human-robot bi-directional communications provides an overview of various single- and multiple-modalities that have been documented (Hill, 2017).

Visual human-robot interfaces are traditionally computer displays which can convey images, charts, and status updates from a robotic system. Topics of appropriate scale and form factor are still investigated, e.g., smart watch vs. smart phone vs. tablet vs. larger screen sizes. Relevant human-computer interface design guidance continues to be developed and explored. Of note is the fact that visual modality interfaces are typically uni-directional, conveying information only from the autonomous system to the human. Control inputs are sent from the human to the robot via keyboard or other input devices, but do not easily provide for "dialogue" between the humans and autonomous systems.

Gestural interfaces can allow humans to use hand or body movements to communicate with robots/automation (e.g., Elliott, Hill and Barnes, 2016; LaViola, 2013). However, most described gestural interfaces are uni-directional, with less research on robots using gestures to communicate information to humans (e.g., Huang and Mutlu, 2013). *Tactile and haptic* interfaces have been used to describe force feedback to the human from the remote environment in which robotics technology is employed, for example haptic feedback from a remotely controlled surgical device (e.g., Moradi Dalvand et al, 2014). There has been considerable research on using tactile vibration signals as a means to provide cueing information to humans (e.g., Elliot et al 2015), but not in a dialogue. *Nonverbal auditory* signals can be used for getting human attention, cueing, reporting status, or can be mapped to a specific meaning, but require considerable training if complex information is to be exchanged.

Natural Language is one of the most intuitive modalities that humans use every day when teaming with other humans in person or remotely. A natural language interface for human-robot interaction carries the potential benefit of feeling natural and intuitive for users, and, when well-designed, can require minimal training for effective use. Unlike learning a visual layout of buttons and windows, or remembering the meaning of beeps, buzzes, vibrations, or new gestures, human users already know how to communicate verbally via natural language and simply need a robotic system that can understand and respond appropriately. Furthermore, natural language is one of the most commonly used modalities for human-human communication in physically grounded situations (e.g., "Do you see a brown box in the closet?"), so it seems plausible that a natural language human-robot interface could also work in such situations.

Natural language includes spoken and/or written language. Speech has been shown to be an effective way to command a robot (e.g., Atrash et al, 2009; Kollar et al., 2010; Poncela and Gallardo-Estrella, 2015). Text-based commands are also commonly used to interact with computerized and robotic systems. Robots can, in turn, respond to the human verbally using text or with speech generation and synthesis, making bi-directional natural language communication possible.

Bi-directional natural language interfaces are popular in many consumer-oriented applications, such as telephone reservation systems and conversational agent systems for the home such as Alexa and Siri. However, these systems are often highly constrained in the language they can process, and are limited in their ability to handle physically situated applications such as object manipulation, navigation, search, rescue, or reconnaissance. Alexa and Siri can perform specific task-based actions well, such as looking up a word, setting an alarm, or playing a song, but they are not geared toward referents in the physical world, nor can they move about and interact with the world physically.

APPROACH

Designing a natural language robot interface that can handle physically situated applications, while also being flexible to humans' natural everyday use of language, is a significant challenge. Natural dialogue will be more than just speech commands, but entails the ability 1) to have common ground and understanding about what is being discussed with respect to a specific physical environment, 2) to be bi-directional, with both (or more than two) participants interacting, and 3) to correct errors or seek clarification. The first fundamental question to address in advancing this research is: What do users want to say to the robot? While language corpora of human-human dialogue exist for physically-situated tasks (e.g., Stoia et al. 2008), robot-directed language differs from human-directed language. Humans have different perceptions and expectations of robots, and we need to understand how humans use language to talk with a robot. We therefore systematically collected robot-directed language in our study. Here we describe our elicitation of minimally-constrained robot-directed language during a physically-grounded task. We present observations about the users' language behavior and discuss future directions for constructing an automated robotic system that can accommodate this language.

To collect this information, we conducted a study to solicit human-robot dialogue in a collaborative search and navigation task. Because the dialogue interface system we wish to build does not yet exist, we could not rely on sampling dialogue from any existing system. Instead, we looked to methodology often used in the development of human-virtual human dialogue systems (see Gratch et al. 2015). Human-virtual human dialogue systems have been constructed using the Wizard of Oz (WoZ) methodology, wherein human users use natural language to interact with what they believe to be a virtual agent, not knowing that there are humans behind the scenes standing in for the future language capabilities of the eventually-automated system. This WoZ setup creates the experience of robot-directed language from the participant's perspective, allowing us to collect as close to the data we are after as possible. This methodology is important for gathering corpora of user language so that appropriate automated responses can then be built to accommodate the types of interactions and tasks actually observed. The process is iterative, with successive rounds of elicitation and experimentation using increasingly automated back-end language processing, until the human Wizards are eventually "automated away," yielding a fully-functional autonomous system. This methodology has been used in the construction of virtual human conversation agents such as SimCoach (Rizzo et al. 2011) and SimSensei (DeVault et al. 2014), which behave as virtual counselors, and New Dimensions in Testimony (Traum et al. 2015), where museum visitors can interactively interview Holocaust survivors.

We employed similar WoZ methods in our study. Behind the scenes, humans stood in for the robot's language (speech comprehension and text-production) and navigation (joysticking) capabilities. The WoZ approach is necessary to elicit realistic language from users, by having them interact with a system that resembles the final (but not yet existing) automated system. To our knowledge, these methods have not yet been successfully harnessed to generate a natural language interface that can handle flexible, often unpredictable, physically situated activities. Unlike SimSensei and SimCoach, Alexa or Siri, our goal is to develop a natural dialogue interface for use with an embodied, mobile, physically situated robot. Human-robot communication in this context requires establishment of physical common ground which is often not required in other human-agent dialogue applications. Acquiring a body of user language becomes even more critical when accommodating these needs.

Research efforts in the interface modalities described earlier continue for human-robot interaction. However, natural language dialogue may be one of the more challenging areas to understand well and implement effectively. We have several goals in conducting research toward natural dialogue. One is to allow humans to interact with autonomy, specifically physical embodied autonomy, (i.e., robots) much like they interact with other humans. Thus we do not constrain or restrict the vocabulary users may use when speaking to their robot teammate in our interface. Another is to inform both the development of a dialogue interface and the development of robotic language processing and future robotic intelligence capabilities. For example, how do people actually talk to robots when they are performing a joint task? What kinds of words do they use? How do they attempt to establish common understanding? If people express through their language the expectation of the robot to have certain capabilities, such as the ability to see and interpret colors, or the ability to move via cardinal directions (north, south, etc.) or other spatial relationships (to the right of, behind, etc.), then we need to develop robotic technology to support this.

METHODS

Our testbed was a search and navigate task wherein a user participant guided a remotely-located robot through a physical environment to assess characteristics about the environment (i.e., reconnaissance). The robot was a Clearpath Robotics Jackal, a four-wheeled robot the size of a small dog. The robot was fitted with LIDAR (light detection and ranging) sensors to detect obstacles, and a front-facing RGB camera to take still photos. The participant was seated at a control station (desktop computer with microphone) in a separate building and could speak to the robot via the microphone.

In order to elicit natural and relatively untrained and unconstrained language from human users, our participants were not given guidance on how to speak to the robot. Participants were provided with a list of the robot's basic physical capabilities, shown a photo of the robot, and told about the search and navigation task goals. Our natural language interface was part of a multi-modal interface that included text responses from the 'robot' (Wizard) to the user (natural language component); still images sent from the robot's front facing RGB camera when requested by the user, and a two-dimensional area map built from the robot's continuously updating LIDAR (visual components); and an audio "beep" when a new text-based response from the robot required the participant's attention (auditory component). The participant communicated to the robot solely through spoken natural language, and the 'robot' communicated back via its affordances of text, images, LIDAR, and audible cues. Figure 1 shows the participants' view in the study.

Twenty participants, 13 male, 7 female, participated in our study. Sixteen participants had zero years of experience in robot operation, development, design, or research; all twenty were naïve to the WoZ approach. We accepted participants with little robotics experience because such participants are unlikely to be familiar with the WoZ, method and because we ultimately want to design a system that can be used with minimal training. Ages ranged from 18 to 58, mean 43. For the first ten participants, the Wizard responsible for the production of language free-typed text responses to the participant, following a detailed response protocol. For the next ten participants, in an effort to increase automation of the robot language production, the Wizard used a Graphical User Interface with buttons that automatically generated text responses based on the response protocol. All user speech was audio-recorded and transcribed. For additional methodological details on study setup, refer to Bonial et al. (2017).

Figure 1. Visual and language components of robot-to-human interface: robot text responses (lower left), LIDAR-generated map (right), last still image requested from robot (upper left). The human-to-robot interface was a microphone for spoken natural language.

FINDINGS AND DISCUSSION

We report examples of minimally-constrained spoken language produced by our twenty participants when conducting a physical exploration task with a remotely located robot. We recorded approximately one hour of audio from each of our 20 participants, yielding a total of 18,336 robot-directed words. Here, we highlight and discuss several observed language patterns (Table 1) that we believe are instructive for understanding how users wish to talk with the robot during a physically-situated task and what will be required of the robot to accommodate these needs.

Observations

Users employ a range of referents for the same objects and concepts. Even a simple concept such as translating the robot forward in the environment yields several different phrasings, such as "move," "go," "drive," "proceed," "leave," "exit", and "return." When referring to objects or locations, the language becomes even more diverse. Sometimes users refer to objects directly, e.g., "move closer to the fire extinguisher." Other times, users may not be able to name an object and have to use visual adjectival properties, spatial relationships, or time referents to designate a particular object, e.g., "take a photo of the item on the right on the wall" (spatial), "go back out of the doorway you entered into" (time), or "face the left side of the orange cone" (visual adjective). This diversity of reference requires flexibility in the robot and is a main challenge in designing systems to handle physically situated dialogue. In addition, some referents are vague, e.g. "turn left and take a picture of the object." Handling this requires that the robot is able to disambiguate "object" in the context of the scene.

In general, the robot's language processing will include an expected vocabulary, or, preferably, be capable of inferring the referent and/or learning additional vocabulary. The robot needs to learn interactively; for example, if the robot does not understand the referent object, it can follow up with a question and ask for the participant to describe the object in another way using its color, size, or position. We follow this strategy in our studies and allow the robot to 'learn' new objects once adequately described by the user. 10% of robot-directed instructions we observed required this strategy.

Users vary in the workload they place on the robot within a single utterance/command. We annotated each instruction as to whether it contained one intent vs. many. Consider the packed, robot-directed instruction "go into that room and take a three hundred and sixty degree photo. Actually you know what, go into that room and take a photo every thirty degrees three hundred and sixty degrees in that room. Go back to where you are and wait for further instruction." Complex instructions such as these made up 29% of instructions in our corpus. The robot must be capable of parsing complex utterances into a series of executable commands, or alternatively must be capable of instructing the user in how to provide executable commands. We are pursuing the former strategy in our design, which seems generally more desirable, as a user could easily lose track of their intention if they are forced to break down instruction too atomically. Limiting instruction-giving style may also constrain the creativity of users' problem solving during the task, or induce frustration.

Users produce both higher-level and lower-level commands. Some users instruct the robot every step of the navigation task, e.g., "go forward five feet, turn left 90 degrees, go forward two feet, take a picture," while others assume the robot can parse the implicit steps, e.g., "go into the next room and take a picture," "robot take a better photo of the shoes," or "watch out for the crate on your left." These higher-level commands require decomposition by the robot, and are common in the natural language produced by users. Because our interface strives to be minimally-constrained, requiring that the user make simpler or more straightforward instructions is not ideal, as it threatens to turn the system into a training program and forces the user to constrain their language and thought process. Building a system that can accommodate these high level commands is an ongoing challenge.

Even some short commands require considerable inference and processing on the part of the robot, such as commands involving object referents (e.g., landmark-based commands). We annotated each movement-based instruction as to whether it used landmark vs. metric reference. Landmark references were used in 31% of movement-related commands in our study. While they are not as common as metric-based references (69% of movement commands in our study) such as "turn south thirty degrees" or "move forward eight feet," we cannot constrain the robotic system to solely process the more transparent metric-based instructions. Current work exists to disambiguate what users expect when issuing higher-level or landmark-based commands (Moolchandani et al. 2018).

Utterance	Category	Identified Robot Needs
take a photo of *the item on the right on the wall*	spatial reference	understand spatial relations
go back out of *the doorway you entered into*	time reference	recall past behavior
face the left side of *the orange cone*	visual adjective reference	understand visual adjectives such as color
turn left and take a picture of *the object*	vague reference	determine the target object
***go into the next room* and take a picture**	higher-level command	infer what precise actions are expected
robot, *take a better photo* of the shoes	higher-level command	infer what precise actions are expected
***watch out for* the crate on your left**	higher-level command	infer what precise actions are expected
turn right one foot	impossible command	infer intent and/or query user
***how many* yellow helmets do you see?**	higher-level assessment	count objects
robot, *what is in front of you*	higher-level assessment	identify objects or provide scene descriptions
***are you sure* there's no way to get into the first doorway?**	higher-level assessment	infer what to reassess and then reassess
are you a male or a female?	social interaction	handle an appropriate level of chit-chat

Table 1. Selected subset of illustrative utterances from our human-robot dialogue corpus, with categories and needed capabilities to address them.

Users unintentionally give impossible commands. Due to situational awareness disparity or simply by mistake, users may give impossible instructions. Some are relatively straightforward to resolve, such as when a participant asks the robot to drive forward a distance that would run it into a wall. It is more difficult to interpret the implied intent in commands like "turn right one foot." The robot may be able to infer what the user meant and follow up with "did I do what you wanted me to?" or can ask for clarification. Thirteen participants gave instructions that needed this form of error-handling, which is critical to build into the robot's language capabilities to avoid a potential communication breakdown.

Users request higher-level assessments from the robot. Users often request higher-level assessments and interpretations from the robot, requests which transcend simple inquiries as to the robot's physical capabilities. The robot is asked to identify and count objects ("is that a map in front of you?", "how many yellow helmets do you see?"), describe scenes ("robot what do you see?", "robot what is in front of you"), or even to reevaluate or provide its own feedback ("are you sure there's no way to get into the first doorway?", "is there anything that indicates that the environment has been recently occupied?") Users also occasionally initiated friendly interaction with the robot, such as asking personal questions (i.e., "are you a male or a female?") or giving positive feedback to the robot (i.e., "good job"). This may suggest a desire for a social relationship with the robot teammate, and along with requesting higher-level assessments, suggests a tendency or desire to treat the robot less as a tool and more as a teammate, even in what may be expected to be a rather impersonal task. This highlights the benefits of our WoZ approach for eliciting natural language. This method allows us to uncover needs for dialogue features and robot intelligence capabilities that may not have been foreseen *a priori*.

PATH FORWARD

Our future work will help develop robot language processing capabilities as well as identify users' needs and expectations of robot intelligence, as revealed in our minimally-constrained robot-directed language corpus. In our next iteration, we will record the user participants on video and use the software MultiSense to analyze facial expressions (Stratou and Morency, 2017), thus adding an additional human-to-robot communication modality to our interface. Acoustic analyses are also planned to improve the robot's capability to determine user affect, attention, or other relevant user state variables from vocal tones and nonverbal utterances (e.g., laughs, sighs). We are developing error handling techniques to address a broad taxonomy of communication error types observed in our dialogue, while also developing algorithms for handling higher-level commands.

In our next iterations, we will perform the same study on a simulated environment platform. This will allow us to make changes to the "physical" environment to determine how this alters the language use and the concomitant capabilities that will need to be incorporated into the robotic system.

The results of these efforts will help build a natural language dialogue interface that is bi-directional, requires minimal training, and allows for performance in physically situated environments. The modality of natural language lends itself to these needs, but much work is left to be done.

ACKNOWLEDGEMENTS

We acknowledge our participants and all project team members who have contributed to planning, data collection, analysis, and documentation of past and future work.

REFERENCES

Atrash, A., Kaplow, R., Villemure, J., West, R., Yamani, H., and Pineau, J. (2009) Development and validation of a robust speech interface for improved human-robot interaction. *International Journal of Social Robotics* 1(4) 345-356.

Australian Department of Defence (2016). *Shaping Defence Science and Technology in the Land Domain (2016-2036).*

Bimbraw, K. (2015). Autonomous Cars: Past, Present and Future - A Review of the Developments in the Last Century, the Present Scenario and the Expected Future of Autonomous Vehicle Technology. In *Proceedings of the 12th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2015), 191-198. DOI: 10.5220/0005540501910198.

Bonial C., Marge, M., Artstein, R., Foots, A., Gervits, F., Hayes, C.J., Henry, C., Hill, S.G., Leuski, A, Lukin, S.M., Moolchandani, P., Pollard, K.A., Traum, D., Voss, C.R. (2017) Laying Down the Yellow Brick Road: Development of a Wizard-of-Oz Interface for Collecting Human-Robot Dialogue. In *Proc. of AAAI Fall*.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M. et al. (2014) SimSensei Kiosk: A virtual human interviewer for healthcare decision support. *In Proc. Of AAMAS.*

Elliott, L., Mortimer, B., Cholewiak, R., et al., (2015) *Salience of tactile cues: An examination of tactor actuator and tactile cue characteristics*, ARL-TR-7392, US Army Research Laboratory, Aberdeen Proving Ground, MD.

Elliott, L., Hill, S. and Barnes, M. (2016) *Gesture-Based Controls for Robots: Overview and Implications for Use by Soldiers,* ARL-TR-7715, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

Gratch, J., Hill, S., Morency, L-P., Pynadath, D., and Traum, D. (2015) Exploring the implications of virtual human research for human-robot teams. In *International Conference on Virtual, Augmented and Mixed Reality.* Springer, 186–196.

Harris, J. & Barber, D. (2014). Speech and gesture interfaces for squad-level human-robot teaming. *Proceedings Volume 9084, Unmanned Systems Technology XVI; 90840B*; doi: 10.1117/12.2052961. SPIE Defense + Security, 2014, Baltimore, Maryland, United States

Hill, SG (2017). *Human-robot bi-directional communications literature overview*. DST-Group-TR-3416, Australian Dept of Defence Science and Technology Group, Fishermans Bend, VIC, AU.

Huang, C. M. and Mutlu, B. (2013, June) Modeling and evaluating narrative gestures for humanlike robots. *Robotics: Science and Systems* 57-64.

Kollar, T., Tellex, S., Roy, D. and Roy, N. (2010) Toward understanding natural language directions. In *Proceeding of the 5th ACM/IEE.*

LaViola Jr., J.J. (2013) 3D gestural interaction: the state of the field. *International Scholarly Research Notices Artificial Intelligence* Vol. 2013 Article ID 514641.

Mehlmann G, Janowski K, Baur T, Markus Häring EA, Gebhard P (2014) Exploring a model of gaze for grounding in HRI. In: *Multimodal Interaction, ICMI '14*, pp 247–254.

Moolchandani, P., Hayes, C.J., and Marge, M. (2018) Evaluating Robot Behavior in Response to Natural Language. In *HRI'18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion*.

Moradi Dalvand, M., Shirinzadeh, B., Nahavandi, S. and Smith, J. (2014) Effects of realistic force feedback in a robotic assisted minimally invasive surgery system. *Minimally Invasive Therapy and Allied Technologies* 23(3) 127-135.

Poncela, A. and Gallardo-Estrella, L. (2015) Command-based voice teleoperation of a mobile robot via a human-robot interface. *Robotica* 33(1) 1-18.

Rizzo, A., Sagae ,K, Forbell, E, Kim, J., Lange, B, Buckwalter, J., Williams, J., Parsons, T., Kenny,P., Traum ,D., Difede ,J. and Rothbaum, B.. (2011) SimCoach: An Intelligent Virtual Human System for Providing Healthcare Information and Support. In *I/ITSEC 2011*.

Saerbeck, M, Schut, T., Bartneck, C., and Janse, M.D. (2010) Expressive Robots in Education. Paper presented at the *CHI 2010: Classroom Technologies*, Atlanta, GA, USA, April 10-15, 2010.

Stoia, L., Shockley, D.M., Byron, D.K., and Fosler-Lussier, E. (2008) SCARE: A situated corpus with annotated referring expressions. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*.

Stratou, G., Morency, L-P. (2017) MultiSense—Context-Aware Nonverbal Behavior Analysis Framework: A Psychological Distress Use Case. In *IEEE Transactions on Affective Computing*, 8:190-203.

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., Jungblut, K., Leuski, A., Smith, S., and Swartout, W. (2015) New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling. In *Interactive Storytelling.* Vol. 9445. Springer International Publishing, Copenhagen, Denmark, 269–281.

U.S. Army Capabilities Integration Center (2017) *U.S. Army Robotic and Autonomous Systems Strategy*.