Field Assessment of Multimodal Communication for Dismounted Human-Robot Teams

Daniel J. Barber, Julian Abich IV, Elizabeth Phillips, Andrew B. Talone, Florian Jentsch
Institute for Simulation and Training (IST), University of Central Florida (UCF), Orlando, Florida

Susan G. Hill
U.S. Army Research Laboratory, Human Research and Engineering Directorate,
Aberdeen Proving Grounds, MD

A field assessment of multimodal communication (MMC) was conducted as part of a program integration demonstration to support and enable bi-directional communication between a dismounted Soldier and a robot teammate. In particular, the assessment was focused on utilizing auditory and visual/gesture based communications. The task involved commanding a robot using semantically-based MMC. Initial participant data indicates a positive experience with the multimodal interface (MMI) prototype. The results of the experiment inform recommendations for multimodal designers regarding perceived usability and functionality of the currently implemented MMI.

The Robotics Collaborative Technology Alliance (RCTA) is a consortium of government, academic, and industry robotics specialists initiated by the U.S. Army Research Laboratory (ARL) Autonomous Systems Enterprise to progress the state-of-the-art in robotics to enable mixed-initiative human-robot teams (HRTs; U.S. Army, 2015). A major area of interest within the program is investigating human-robot interaction (HRI), specifically communications among dismounted Soldiers and robot teammates. A dismounted Soldier refers to a Soldier that is conducting a military operation on foot, as opposed to a mounted Soldier whom is traveling in a vehicle (Beidel, 2011). Due to the dismounted Soldier's operational contexts and constraints, human-robot communication must support robust, flexible, natural, intuitive, and efficient forms of communication, hence the appeal of MMC (Oviatt, 2012).

Currently, teleoperation is the primary method of robot control by a Soldier, with communication dominated by visual displays. One goal of the RCTA is to advance current robotic systems "from tools to teammates" (Phillips, Ososky, Grove, & Jentsch, 2011) by developing communication interfaces that reflect the ways in which humans communicate with each other. This paradigm shift will drive future HRT communication technology, fostering a blend of modes to enhance team effectiveness while avoiding additional cognitive strain on the Soldier.

Multimodal communication

The song title from the band Journey sums up MMC quite nicely and poetically, "Any way you want it!" That is the principle behind MMC, to support the opportunity to communicate through various means of interaction that are suitable for any context. More formally, MMC is the ability to transmit and receive information through more than one communication method or "modality" with the emphasis on using all five senses (Dumas, Lalanne, & Oviatt, 2009). The use of MMC within dismounted HRTs has attracted much attention within the HRI community as a way of meeting operational demands (Barber, Lackey, Reinerman-Jones, & Hudson, 2013; Barber, Reinerman-Jones, & Matthews, 2014). Many systems

have been developed that integrate multiple input and output modes (e.g. smartphones, video game consoles), but less research has focused on assessing these systems in real-world applications.

For effective MMC to take place, the user, system, and type of interaction must be well understood (Jaimes & Sebe, 2007). It is therefore necessary to assess the abilities and limitations of the users, hardware and software capabilities and functionality, and how the end-users will utilize the MMC technology in addition to interaction preferences.

A user-centered design approach was applied to the development of the MMI and the various communication methods it supports. The Leventhal and Barnes (2008) usability model, adopted from previous models such as Eason's (1984), Shackel's (1986), and Nielsen's (1993), has an implicit causal structure meaning the usability of an interface is dependent on the interaction of many variables. The model identifies two main variables: 1) situational and 2) user-interface. Situational variables are decomposed into task and user variables. Task variables refer to the situational constraints, frequency of the task, and the rigidness of options available to complete the task. Dismounted Soldiers are exposed to operational conditions which impose external demands that challenge physical and cognitive abilities for processing and conveying information to and from teammates. MMC input and output devices must not only be sensitive enough to capture the information transmitted from the Soldier, but they must also be robust enough to withstand the continuum of military operations. Further, Soldiers and robots may not be co-located, and therefore will require methods that will support communication with no line of sight. Moreover, operational needs may limit the kinds of communications that can take place under some circumstances; for example, Soldiers may be required to be stealthy and limit sound and speech in their communications.

User variables refer to the level of expertise and motivation brought to the task by the Soldiers. Many Soldiers have little to no experience interacting with robots, but their motivation for mission and task success encourages a positive receptivity for HRI. Therefore, the usability of the MMI should

take advantage of their extensive military knowledge regarding communications and procedures to support a positive transfer of training. This also is applicable to the user interface variables which concentrates on items such as ease of use and learning, flexibility, user satisfaction, and task matching. MMC provides an interactive structure based on intuitive human communication exchanges that reduces the burden of extensive additional training because most of the commands programmed within the current MMI are already used for squad level commands or are commonly used during human-to-human conversations. The flexible nature of MMC allows Soldiers to choose the best method of communication for the task, resulting in strong user satisfaction.

Additionally, from a cognitive perspective, MMC facilitates time-sharing efficiency. Theoretically, MMC supports processing two or more streams of information through different modality channels to reduce cognitive overload (Wickens, 2008). Compared to unidimensional interfaces, MMIs have been shown to reduce tasking time and errors, as well as improve system reliability (Oviatt, 1997). User interface designers can use MMC to aid in developing tasks and configuring systems that are less demanding and more efficient in terms of interaction.

Experimental approach and aim

An RCTA Capstone Assessment was held in October 2014 at Fort Indiantown Gap, PA, to assess the integration of multiple aspects of the RCTA's program thrusts. As part of the Capstone Assessment, the goal for this particular field assessment was to explore the use of an MMI using speech, gesture, auditory cues, and a visual display within a field setting (i.e., outside of a laboratory; Hill, Barber, & Evans, 2015).

The aim of this paper is to report on the field assessment of a prototype MMI developed to support MMC for future dismounted Soldier-robot teams. This prototype was developed to assess elements of MMC, including the use of gesture and speech to communicate with a robot, as well as options for displaying information to a user (e.g., path of the robot, color, symbols). The results of this assessment will be used to inform designers on the ability of the device and its design principles to support effective MMC within HRTs, as well as user perceptions of the MMI and elements of MMC between the user and the robot.

METHODS

Participants

Ten individuals (9 males, 1 female, M_{Age} = 33.7, age range = 26-40 years) voluntarily participated in this field assessment held at the military training facility at Fort Indiantown GAP, PA. All participants had normal or corrected to normal (i.e., glasses or contacts) vision. Participants were also screened for colorblindness using the Ishihara Color Vision Test (Ishihara, 2014). One participant was identified as having a color deficiency, however, no major differences were identified between the responses provided by this participant and participants who did not have a color deficiency.

Materials

A variety of measures were used in this field assessment, however, this initial analysis only reports on two open-ended questionnaires. The first of these (Free Response Questionnaire) asked participants six questions in regards to the MMI that covered positive and negative aspects of their interaction with the device, and participants were asked to provide any suggestions for improvement.

The second questionnaire (Robot Reporting Questionnaire) asked participants to answer three questions related to the manner in which a robot might send reports to them in hypothetical situations.

Apparatus

Husky robot. The robot used for this field assessment was a Husky robot from Clearpath Robotics equipped with an XR LADAR for obstacle avoidance, semantic perception, and localization, an Adonis camera for semantic object and human detection, and the Hokuyo LADAR for near range obstacle avoidance while driving (Figure 1). The Husky was also equipped with the latest version of the autonomous capability software/hardware developed by RCTA researchers (an early version is described in Dean, 2013). The Husky had a top speed of 2.3 mph and was fitted with an emergency stop function for safety purposes.

Figure 1. (left) Participant's view of the robot navigating towards the identified object; (right) Participant interacting with the robot using the MMI

Multimodal interface (MMI). The multimodal interface (MMI) was created to instantiate bi-directional MMC between a human and a robot. The MMI consists of three integrated devices: a Panasonic Toughpad FZ-M1 tablet, a Plantronics M50 Bluetooth headset, and a gesture recognition glove (Figure 1; Barber et al., 2013). Through the MMI, users were able to interact with the robot through speech (e.g. using natural language interfaces; Howard et al., 2014), and/or gesture.

The MMI fosters bi-directional MMC with the robot by allowing users to (a) send commands to the robot (using speech or gesture), (b) receive feedback from the robot (visual and auditory), and (c) monitor the robot (via the tablet display). The tablet displays several types of information including a world map generated by the robot, camera feed from the robot, a window for displaying the command the robot was currently executing, and status information from the robot (Figure 2). The tablet display was also used as a means to help the robot when it requested assistance.

Figure 2. Conceptual representation of the MMI

Scenarios

Participants completed two distinct interaction scenarios. Scenario 1 involved commanding the robot to navigate to one of three traffic barrels that were placed near the back of a building (Figure 3). One traffic barrel was next to a fire hydrant, one traffic barrel was located near the building, and one traffic barrel was located next to a vehicle. Participants instructed the robot to navigate to a particular traffic barrel using a speech command (e.g., "Husky navigate quickly to the traffic barrel near the fire hydrant"). After initiating the command, the robot began to autonomously complete the task. As the robot completed the task, participants could issue several commands including reorienting the robot, or pausing/aborting the command the robot was currently executing. To ensure participants had multiple opportunities to utilize both the speech and gesture commands, they were asked to command the robot to navigate to each traffic barrel three times. Thus, each participant had the opportunity to complete this interaction scenario as many as nine times.
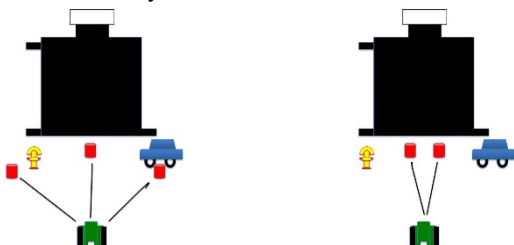

Figure 3. Diagrams of scenario 1 (left) and scenario 2 (right).

Scenario 2 involved commanding the robot to navigate to one of two traffic barrels placed equidistant from the back of the same building (Figure 3). After being given the command to "navigate to the traffic barrel near the building," if the robot could not perceive which barrel was closest, the robot would request assistance to help resolve the ambiguity in deciding which barrel was closest to the building. Participants assisted the robot using the tablet display by selecting the correct bar-

rel from the options provided. This scenario was completed by each participant once.

Procedure

Upon arriving at the field assessment, the informed consent procedure began. After reading and signing the informed consent, participants were administered a demographics questionnaire and the Ishihara Color Vision Test. Following completion of these pre-scenario measures, each participant was trained on how to use the MMI and briefed on safety regarding the robot (e.g., maintaining an appropriate distance between oneself and the robot). Participants were trained on all commands and had to successfully perform each to have the system understand both their verbal and gestural responses. Participants could practice as long as they wanted before performing the task. During the experiment runs, there was not a time constraint associated with the task, therefore participants could take as long as they needed.

Once training was complete, participants completed as many as 10 interaction scenarios. Following completion of the interaction scenarios, participants were asked questions to assess their perception of the robot's performance by completing the Free Response and Robot Reporting Questionnaires.

RESULTS

Open-ended responses to the Free Response and Robot Reporting Questionnaires were obtained from each participant and were then aggregated for each item. Two independent raters then conducted an evaluation of the data for common themes across participant responses to each item. This method represented an inductive approach to organizing common ideas, patterns, and/or motifs within the data. The raters were given clear guidelines for rating the free response questions administered. The free response questions were fairly straight forward and did not require much interpretation on the part of the raters (e.g. rate the positive aspects of MMI device). The raters have worked together on many projects and are very aware of their rating style which helped to ensure similar judgments were made. Raters began by organizing the textual responses into common ideas and then comparing the frequency with which certain themes occurred in the text for each item. Common motifs that were identified as being mentioned by at least three respondents for each item were retained. The two assessments were then compared for overlapping patterns identified by both raters.

Common themes

Across participants, several themes were identified for each item. Tables 1 and 2, present the identified themes for each item as well as examples of prototypical responses that were grouped into common themes.

Table 1
Table of common themes among participant responses to the Free Response Questionnaire

	Free Response Questionnaire	
Item/Question	**Theme**	**Prototypical responses**
Positive aspects of the device	(a) Multiple Modalities (b) Portability	(a) The ability to use multiple modalities including speech was helpful. (b) The device is compact, easy to use, light-weight, and works well in sunlight.
Negative aspects of the device	Trouble with gesture recognition	Gesture commands were not completely intuitive, required excessive precision, or would take additional training to master.
Object identification	Mismatch of object representation	Displayed objects did not match real-world representations, did not represent the real-world shape of the object, or were too small.
Navigation information	Support for robot's projection of its path	The robot displayed its path clearly, it was helpful to see its projected path, and fostered understanding of where the robot was navigating.
Suggested improvements	Improve gesture recognition	The gesture commands did not always work well, required too much precision, but could be improved with additional training.

Table 2
Table of common themes among participant responses to the Robot Reporting Questionnaire

	Robot Reporting Questionnaire			
Item	**Theme(s)**	**Theme(s)**	**Theme(s)**	**Prototypical responses**
Status updates on Terrain and people	Traversability, type, and obstacle	Friendly/Hostile, objects, location, classification	Objects in environment	User would expect notification of whether terrain is difficult to maneuver, if insurgents are present, the location of people, people with weapons, and objects that have been previously identified as important.
Status updates on completion of tasks	Success/ failure	Distance, time	Object identification	User would expect notification of success or failure of task/ subtask completion, how long it took to complete, its distance to the goal, estimated time to completion, and any objects it identifies along the way.
When would robot provide status updates	Completed	Failure	Periodically	User would expect notification once the robot has completed its task or immediately beforehand, as well as if it has failed during task completion, and periodically if it's a long mission.

DISCUSSION/RECOMMENDATIONS

The results of this field assessment begin to uncover the factors that impact efficient MMC within dismounted HRTs. The presented data gathered from the subjective, free-response measures highlight aspects of MMC to expand upon and improve in the future. Below is a discussion of the results with suggested recommendations for future MMC and MMIs.

Flexible communication. Participants had positive comments on the use of MMC for HRTs. As indicated, having the option to choose the form of communication to interact with the robot was rated as a positive feature. Other modes of communication (e.g. tactile) might be included for improved flexibility for conveying information to and from a robot teammate across tasks and environments.

Intuitive communication. The language used to communicate with a robot is important for effective MMC to take place. Although participants did prefer the ability to communicate with the robot using various modalities, the gestures needs to be more intuitive, require less training, and expanded to include vocabulary or commands phrases that Soldiers would prefer to use to interact with a robot teammate.

Display. The physical aspects of the MMI are important for use across various operational environments. Participants noted that the display was clearly visible in direct sunlight, a critical feature for dismounted Soldiers. In making the shift from large, traditional display-based interactions to more intuitive and multimodal interactions, the exact form factors will need to consider such human factors concerns as display visibility and well-designed display "real-estate."

Content features. Participants indicated a level of ambiguity between the objects displayed in the robot's world map and the actual objects in the real-world. For this task, there were very few objects present in the environment and it might not have been as critical compared to a more visually complex environment However, participants stated that they would much prefer the objects in the robot's world map (displayed as colored geometry) to resemble the actual real-world objects (e.g. icons). Additionally, some participants suggested that it would be beneficial for developing a better mental model to have the option to touch an icon on the screen in order to display more detailed information about the object of interest (e.g. click on a traffic barrel and get information about what it is, its size, its geographic location, etc.).

The projected navigation path of the robot was displayed prior to executing the given command. Participants preferred having this navigation information available on the display. This feature helped build a shared mental model among the HRTs because participants were able to see where the robot was planning to navigate and it allowed for early intervention and correction if necessary.

MMC hardware/software. Speech and gesture recognition systems must be robust enough to capture the commands transmitted from any participant, regardless of varying demographic features. One concern participants had with the gesture recognition system was that it required precise arm and hand movements. This is a feature that requires further investigation. If the system is too lax then unintentional messages might be sent, but on the other hand, if it is too strict, then it will be difficult to transmit any signal.

Portability/ruggedness. Technological advances should not equate to bulky, power-consuming equipment. Soldiers already must carry over 100lbs of gear (Beidel, 2011), and therefore, adding to that burden should be avoided. The current MMI is a small, light-weight, low-energy consumption device that could easily fit into a Soldier's pack. However, ultimately the interest is to examine even smaller form factors, such as smartphones, that Soldiers can drop in their pockets for use only when needed. The concentration is on minimizing the need for a "device" by using speech and gesture interfaces as possible solutions.

Robot reports. Based on the results of the Robot Reporting Questionnaire, participants provided descriptions of what information they would expect to receive when a robot provides an update report (Table 2). In general, participants indicated that they would expect information on terrain, location, the presence of people in the environment, if people were suspicious (perhaps because of actions or objects), and objects. This type of information would help Soldiers make important decisions about the robot's actions and their own actions. This type of information would also assist in the development of appropriate mental models of the operational environment.

When reporting a status update regarding task completion, participants mentioned they would prefer the robot to indicate whether the task was completed successfully or not, the duration of time estimated to complete the task, and how far it already traveled and will travel to reach the goal. All of this information will enhance the HRTs shared mental model and support decision making.

When asked about frequency of reports, participants stated they would prefer reports when the task was complete, when the robot or task failed, or periodically, if the task was long in duration. It is important to not overwhelm Soldiers with information constantly throughout a mission. Better understanding of human-robot trust will help determine how often Soldiers and robots need to report their status and how confident Soldiers are that an autonomous robot will complete a task without requiring frequent monitoring.

CONCLUSION

Field experiments provide an opportunity to evaluate the validity of laboratory results within a relevant operational setting. The RCTA's Capstone Assessment showcased the successful integration of many aspects of the program, with the focus of this reported experiment on multimodal communications. Results of this experiment suggests a positive response to the user-centered design of the MMI and the use of MMC for HRTs. Additionally, it generated foundational data to use for version improvements and subsequent research.

Future studies will expand upon the results gathered from this experiment to further modify the MMI and to investigate other questions regarding MMC within dismounted HRTs. More specifically, future field and laboratory studies must begin to incorporate environmental factors that may impact performance, such as low visibility or noise congestion

REFERENCES

Barber, D., Lackey, S., Reinerman-Jones, L., & Hudson. I. (2013). Visual and tactile interfaces for bi-directional human-robot communication. In *SPIE Defense, Security, and Sensing* (pp. 87410U-87410U). International Society for Optics and Photonics.

Barber, D., Reinerman-Jones, L., & Matthews, G. (2014). Toward a tactile language for human-robot interaction: Two studies of tacton learning and performance. *Human Factors,* 1-20. doi: 10.1177/0018720814548063.

Beidel, E. (2011). Army shifts focus to dismounted Soldiers. Retrieved on January 14, 2015 from NationalDefenseMagazine.org: http://www.nationaldefensemagazine.org/archive/ 2011/April/Pages/ArmyShiftsFocustoDismountedSoldiers.aspx.

Dean, R. M. S. (2013, May). Common world model for unmanned systems. In *SPIE Defense, Security, and Sensing* (pp. 87410O-87410O). International Society for Optics and Photonics.

Dumas, B., Lalanne, D., & Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In D. Lalanne & J. Kohlas (Eds.), *Human machine in-teraction: Lecture notes in computer science* (pp. 3-26). Berlin: Springer Berlin Heidel-berg.

Eason, K.D. (1984). Towards the experimental study of usability. *Behaviour and Information Technology, 3*(2), 133-143

Hill, S., Barber, D., & Evans, A.W., III. (2015). Achieving the vision of effective Soldier-robot teaming: Recent work in multimodal communication. *Proceedings from Human Robot Interaction: Extended Abstracts*.

Howard, T., Chung, I., Propp, O., Walter, M., & Roy, N. (2014). Efficient natural language interfaces for assistive robots. In *Proceedings of the Workshop on Rehabilitation and Assistive Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Ishihara, S. (2014). Ishihara's tests for colour deficiency: Concise edition. Tokyo: Kanehara Trading Inc.

Jaimes, A. & Sebe. N. (2007). Multimodal human computer interaction: A survey. *Computer Vision and Image Understanding, 108*(1), 116-134.

Leventhal, L., & Barnes, J. (2008). *Usability engineering: Process, products, and examples.* Upper Saddle River, NJ: Pearson Prentice Hall.

Nielsen, J. (1993). *Usability engineering.* Boston: Academic Press.

Oviatt, S.L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction 12*, 93–129.

Oviatt, S. (2012). Multimodal interfaces. In J. Jacko (Ed.), *Handbook of Human-Computer Interaction (3rd ed.).* New Jersey: Lawrence Erlbaum.

Phillips, E., Ososky, S., Grove, J., & Jenstch, F. (2011). From tools to teammates: Toward the development of appropriate mental models for intelligent robots. *Proceedings from the Human Factors and Ergonomics Society 55th Annual Meeting*, 55(1), 1491-1495.

Shackel, B (1986). Ergonomics in design for usability. In M.D. Harrison & A.F. Monk (Eds.), *People and computers: Designing for usability, Proceedings of the HCI'86 Conference on People and Computers II 1986.* Cambridge, UK: Cambridge University Press, 44-64.

U.S. Army. (2015). Robotics Collaborative Technology Alliance. Retrieved on January 24, 2015 from http://www.arl.army.mil/www/default.cfm?page =392.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.