Assessing Multimodal Interactions with Mixed-Initiative Teams

Daniel Barber(✉)

University of Central Florida, Institute for Simulation and Training, Orlando, FL 32826, USA
dbarber@ist.ucf.edu

Abstract. The state-of-the-art in robotics is advancing to support the warfighters' ability to project force and increase their reach across a variety of future missions. Seamless integration of robots with the warfighter will require advancing interfaces from teleoperation to collaboration. The current approach to meeting this requirement is to include human-to-human communication capabilities in tomorrow's robots using multimodal communication. Though advanced, today's robots do not yet come close to supporting teaming in dismounted military operations, and therefore simulation is required for developers to assess multimodal interfaces in complex multi-tasking scenarios. This paper describes existing and future simulations to support assessment of multimodal human-robot interaction in dismounted soldier-robot teams.

Keywords: Multimodal interfaces · Human-robot interaction · Simulation Tactile displays

1 Introduction

A desire to support the warfighters' ability to project force and increase their reach across a variety of future operations has resulted in a concerted push to advance the state-of-the-art in robotics. In current ground operations, robots are remote controlled assets supporting tasks where it is infeasible or unsafe for personnel to go (e.g. disposal of improvised explosive devices). These systems do not function collaboratively with human counterparts, requiring additional labor to not only manage the robot, but also provide force protection, which may instead create additional workload and reduce the controller's ability to perform secondary tasks, [1–3]. Although modernizations have taken place to make interfaces in dismounted applications more lightweight and portable, current interfaces for teleoperation focus on a one-to-one relationship where an operator observers sensor feeds (e.g. video) and manipulates hand controls, keeping their gaze heads-down, Fig. 1, [4].

Revolutionizing collaboration with robots will require a leap forward in their autonomy and equal development of robust interfaces. The current approach to meeting this goal is to design interfaces that model how human teammates interact today. Enabling a Soldier to use what is already familiar to them, such as speech, gestures, and vocabulary, will facilitate a seamless integration of robot counterparts. Building robot teammates that embed familiar communication methods will reduce the need for

S. Yamamoto and H. Mori (Eds.): HIMI 2018, LNCS 10904, pp. 175–184, 2018.
https://doi.org/10.1007/978-3-319-92043-6_15

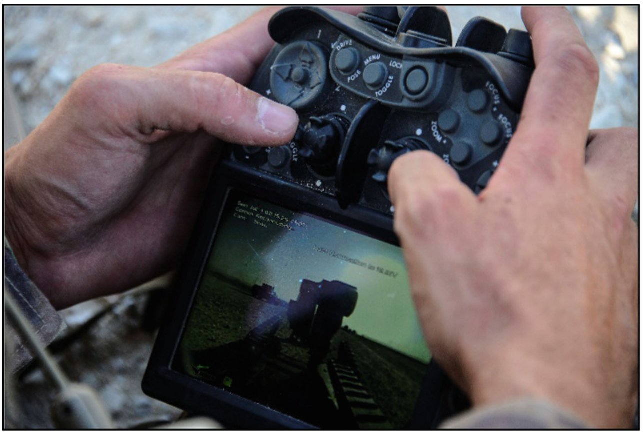


Fig. 1. Interface for teleoperation of a PacBot 310 robot, [5]

training to allow Soldiers to take advantage of these new assets, and lower demands on the Soldier. Incorporating these types of interactions will lead to the creation of collaborative mixed-initiative teams where Soldiers and robots take on different roles at different times to optimize the team's ability to accomplish mission objectives, [6].

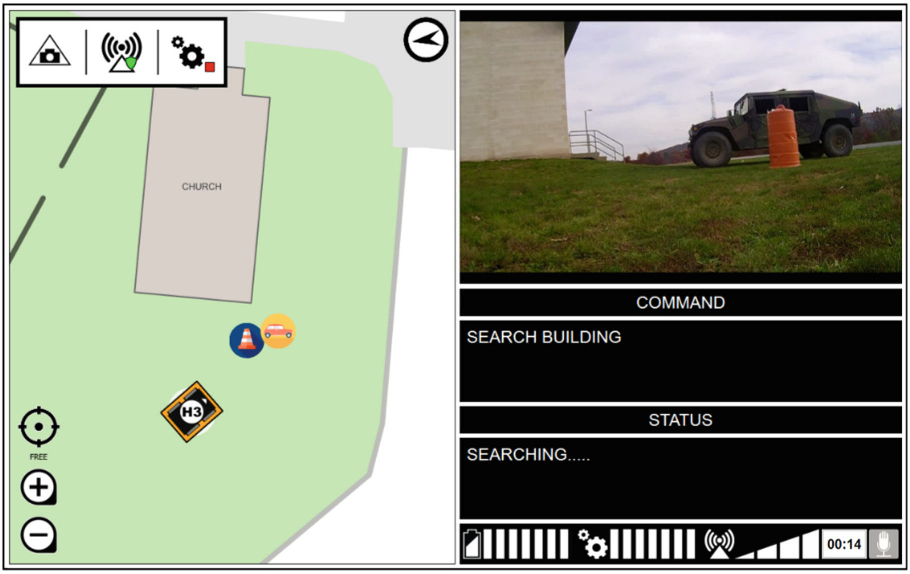
2 Multimodal Interaction

Developing advanced interfaces for human-robot collaboration that are modeled after human-to-human interactions will inherently require multimodal support. Throughout the literature, six common themes in multimodal communication efforts emerge: meaning, context, natural, efficiency, effectiveness, and flexibility. Numerous authors use multimodal communication to strive for meaning and context [7–9], more complex conveyance of information over multiple modes compared to single mode [10], and delivery of ideas redundantly (back up signals) and non-redundantly (multiple messages) [11, 12]. Ultimately, multimodal communication supports multiple levels of complexity [10].

In an effort to scope research efforts within the context of dismounted human-robot interaction (HRI), Lackey, et al. operationally defined multimodal communication as "the exchange of information through a flexible selection of explicit and implicit modalities that enables interactions and influences behavior, thoughts, and emotions," [13]. Leveraging this definition, explicit communication types from the literature for investigation within an HRI multimodal interface emerge and include speech, auditory cues, visual signals, and visual and tactile displays.

Future multimodal interfaces must support some or all of these explicit methods of communication to enable assessment of mixed-initiative team interactions. For example, to take full advantage of the auditory modality, interfaces must include functionality for both speech-to-text (STT), text-to-speech (TTS), natural language understanding (NLU), and other sound effects. Gestures are a common and natural form of communication among humans within the visual modality, and as such, robots must classify them. In addition to traditional visual displays (e.g. tablets), robots could also deliver their own gestures from manipulators and other body movements. Finally, an emerging field of research showing potential benefits is tactile displays. Tactile displays exist in many commercial-off-the-self products such as cell phones and smart watches that emit haptic cues for calls and text messages. In respect to dismounted operations, researchers using tactile belts have demonstrated improved navigation performance and wearer's ability to classify up to two-word phrases approaching the complexity of speech, [14, 15].

In an attempt to bring these individual technologies together, Barber et al. developed and assessed a prototype multimodal interface as part of the Robotics Collaborative Technology Alliance (RCTA), [16–18]. The RCTA's Multimodal Interface (MMI) supports multiple modalities for transactional communication with a robot teammate. With voice data captured on a Bluetooth headset, the Microsoft Speech Platform SDK version 11 classified speech commands to text, that were then converted into robot instructions using a natural language understanding module, [19]. For visual signals, a custom gesture glove captured arm and hand movement using an inertial-measurement unit, which a statistical model classified into gestures. Previous efforts using this glove have shown a capacity to classify 21 unique arm and hand signals, many from the Army Field Manual for Visual Signals, [20]. For robot-to-human communication, the MMI supports TTS, auditory cues, and a visual display. The MMI visual display contains all current mission information from the robot, including a semantic map, live video-feed, current command, and status, Fig. 2.



COMMAND

SEARCH BUILDING

STATUS

SEARCHING.....

Fig. 2. Multimodal Interface (MMI) visual display. Display supports three primary areas, semantic map (left), video feed (top right), and command/status information (bottom right).

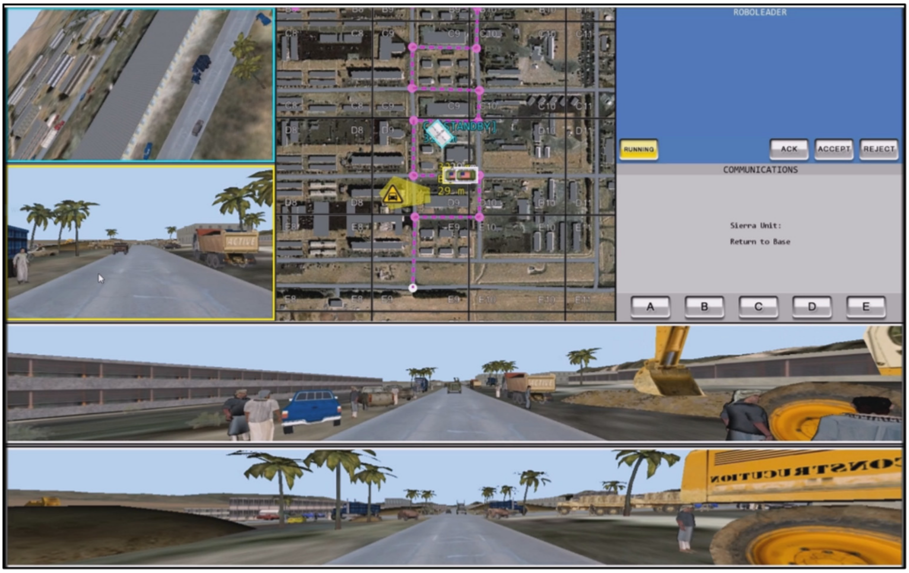
Through this combination of interfaces, users were able to give a complex speech command such as "screen the back of the building," receive confirmation, observe the robot execute, and receive feedback of mission completion without the need to look heads down at the visual display, [21]. Results from a field study revealed participants liked the ability to use multiple modalities and the interface form factor, but requested modifications to iconography, more intuitive gesture commands, and increased transparency into robot logic, [16]. Although successful in demonstrating a baseline level of human-robot interaction and multimodal assessment in a dismounted application, the types of mixed-initiative teaming available to researchers is limited to the capabilities of today's robots. The state-of-the-art in robot autonomy still does not come close to supporting human-teaming concepts such as back-up behaviors, shared mental models, and information prioritization. Without robots able to adequately perform in larger mixed-team scenarios and roles, they cannot drive adaptive multimodal interfaces with manipulation of modalities, information chunking, transparency, and dynamic reporting frequency for maintaining team situation awareness and performance.

3 Simulating Mixed-Initiative Teaming

There are several challenges when attempting to identify simulation environments for HRI experiments. For a given simulation to support mixed-initiative assessment, it must include tasks for participants to perform with robotic assets, however the majority of these environments focus on the engineering aspects of the robotic and not human interaction. For example, Gazebo is a robot simulation tool supporting Robot Operating System (ROS) users. Gazebo provides dynamics simulation, advanced 3D graphics, sensors with noise, and physics models of many commercial robots including the PR2, Pioneer2 DX, iRobot Create, and TurtleBot, [22]. Although an extremely powerful tool, it does not provide, and would be difficult to add, the human scenario elements needed to simulate mixed-initiative teaming for experimentation.

The Mixed-Initiative Experimental Testbed (MIX) is an open-source simulation designed up front for HRI, [23, 24]. MIX provides a research environment composed of two main applications: the Unmanned Systems Simulator (USSIM), and Operator Control Unit (OCU). USSIM simulates ground and air robots capable of autonomous navigation within a 3D environment. The OCU is a reconfigurable ground control station interface capable of managing one or more unmanned systems (both real and simulated) using the Joint Architecture for Unmanned Systems (JAUS), [25]. In addition to command and control of robots, the OCU simulates other relevant theoretically-driven mission tasks such as change detection and signal detection, [3]. Moreover, MIX generates a multitude of scripted events based on time or location triggers including: display visuals using imagery, injects into the 3d environment returned over robot video feeds, and audio events from sound files, Fig. 3.

Researchers using MIX have setup a variety of different HRI experiments for adaptive automation, supervisory control of multiple robots, and agent transparency, [3, 26, 27]. Another example HRI testbed is the Wingman SIL. The purpose for the Wingman program is to provide robotic technological advances and experimentation to increase the autonomous capabilities of mounted and unmanned combat support

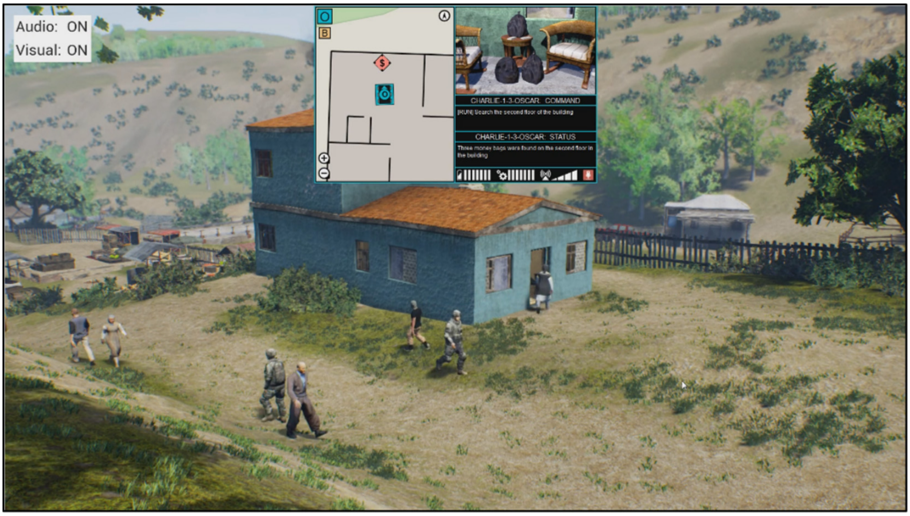


Sierra Unit:
Return to Base

Fig. 3. MIX Testbed Operator Control Unit (OCU) simulation. OCU includes video feeds from multiple robots (i.e. air, ground) and a 360 degree indirect vision display, map for route/mission planning, and dialogs for interaction with automated agents and command, [26]. Users configure OCU layout, graphics, content, and scenario using XML.

vehicles, [28]. The Wingman SIL includes Warfighter Machine Interfaces (WMI) for a Mobility Operator, Vehicle Commander, and Robotic Gunner. Combined together, this environment supports simulation of team combat exercises. Although capable of facilitating many research efforts, systems like MIX and Wingman SIL are focused on manned (i.e. in-vehicle) missions, (e.g. supervisory control for intelligence, reconnaissance, and surveillance), and are therefore not designed for dismounted teaming studies. Without extensive modification, they are not capable of modeling scenarios where researchers can have participants act as a squad with a robot, communicating with speech and gestures similar to interactions previously described using the RCTA MMI.

The Enhanced Dynamic Geo-Social Environment (EDGE) is a multiplayer, scalable, online training environment for first responders. Developed by the US Army Research Lab (ARL), Human Research Engineering Directorate (HRED), Simulation and training Technology Center (STTC) in partnership with TRADOC G2 and the Department of Homeland Security (DHS), EDGE is a government owned platform built using the Unreal Game Engine 4, [29, 30]. Designed for extension to other applications, researchers at the University of Central Florida working with ARL leveraged EDGE to support simulation of robots in dismounted operations. Under this effort, the RCTA MMI (henceforth referred to as MMI) was modified to communicate with a modified version of EDGE called the Visualization Testbed (VTB) that included a simulated robot. This virtual robot was capable behaviors emulating the semantic navigation capabilities driven from speech commands previously demonstrated with real platforms by Barber et al., [16, 19, 21]. For example, a user issues a speech command such as "go to the north side of the bridge," and receives acknowledgement and task status as robot executes the mission. In addition to integration with VTB, the MMI was further extended to support simulation of other content and tasks. These modifications enable using time or location-based events to trigger updates to the MMI text, images, color scheme, and map independent of content coming from VTB. Furthermore, when combined, the simulations support theory-based tasks (e.g. signal detection) and multi-tasking similar to the MIX testbed, but in dismounted human-robot teaming scenarios, Fig. 4.


Audio: ON
Visual: ON

CHARLIE-1-3-OSCAR: COMMAND

[RUN] Search the second floor of the building

CHARLIE-1-3-OSCAR: STATUS

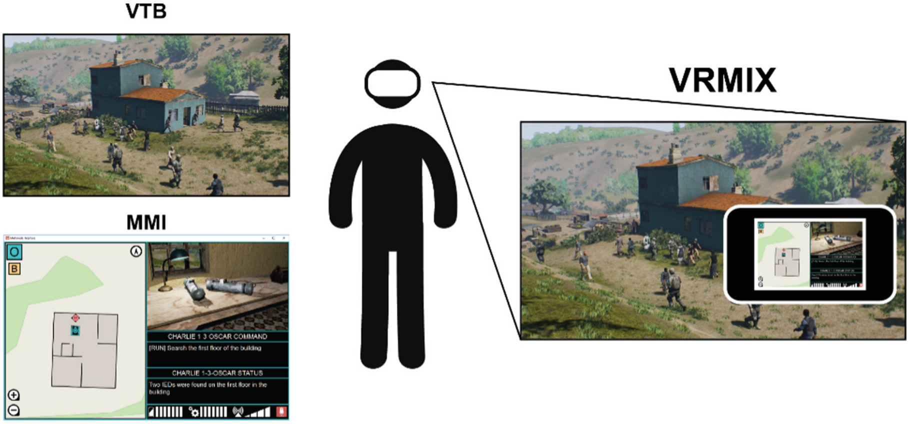
Three money bags were found on the second floor in the building

Fig. 4. VTB and MMI Simulation for dismounted human-robot teaming. VTB generates virtual world images, content, and simulated robot, with the MMI (overlaid top-middle) supporting interactions with the robot and other simulated communications and tasks. Characters moving in the environment support a signal detection task.

In addition to support for different interaction types, any assessment in dismounted scenarios also requires an ability for researchers to model relevant missions. One of the most frequent operations a Soldier may perform in a team is cordon and search. Cordon and search is complex in nature, with reconnaissance, enemy isolation and capture, and weapons and materials seizures, [31]. This combination of tasks makes cordon and search ideal for multimodal HRI experimentation. The VTB/MMI environment is capable of supporting cordon and search and other dismounted missions for human-in-the-loop studies. Using VTB to simulate robot teammates and an outer cordon task from the Soldier's perspective and the MMI for robot reports and commands, one can explore a variety of use-cases. In a recent example of this, in an effort to investigate adaptive multimodal communication, researchers conducted an experiment to assess recall of different robot reports using single (visual, auditory) and dual (visual and auditory) communication modalities under different environmental demands for a cordon and search mission, [32]. Although promising, there were limits to representing a dismounted mission, in that the task was performed on a desktop workstation, which does not provide the type of immersion or demands that may be needed to translate findings to the real world. Moreover, without the ability to provide some semblance of robot presence with participants, researchers cannot study implicit communication (e.g. social distance) or anthropomorphic affects in HRI.

4 Virtual Reality for Dismounted HRI

With recent advancements in commercial-off-the-shelf virtual reality (VR) displays, the cost associated with immersing someone in a virtual world is dramatically reduced; making incorporation of VR into human-in-the-loop experiments approachable to researchers. The HTC VIVE VR system is an example of this, with a cost of $600 and direct integration support for multiple game engines, including Unreal Engine 4 (UE4), [33, 34]. Using VR, one can address the gaps associated with the desktop-based VTB/MMI simulation for enhanced empirical validity and to cover a broader range of research. To meet this goal, this paper proposes a new simulation platform called VRMIX, which combines the UE4-based VTB simulation, MMI, and HTC VIVE to produce an immersive virtual world for exploration of multimodal interactions with mixed-initiative teams. With direct support for the HTC VIVE, developers can update VTB cordon and search scenarios for use cases where participants are “physically” present with characters and robots in the scene. The MMI, previously integrated with VTB for sharing of data, only requires integration of any visual display elements within VTB, as speech and audio modalities are supported with existing hardware (e.g. microphone, speakers), Fig. 5.



CHARLIE 1 3 OSCAR COMMAND

[RUN] Search the first floor of the building

CHARLIE 1-3-OSCAR STATUS

Two IEDs were found on the first floor in the building

Fig. 5. VRMIX concept. Through the combination of the UE4-based VTB simulation and MMI software (left), and virtual reality headset, participants are immersed within a dismounted mission where they may interact using speech, gestures, and visual display in game (right).

In order for users to access the visual display of the MMI within VR, VTB is modified to perform screen captures (i.e. frame grabbing) of the actual MMI display software and render it in game. Thus, users are able to perform a visual search as they would in the real world by looking around with their head, command with speech and gestures, receive auditory cues, TTS, and simulated radio chatter, and access a visual display within a complex multi-tasking scenario. Furthermore, the tactile modality can be supported using the haptic channel of the HTC VIVE controllers or with integration of a tactile display. Moreover, the first-person nature of VR lends itself well to the egocentric spatial characteristics of tactile belts, [14]. Thus, VRMIX has the potential to provide a means of investigating all modalities in a laboratory setting with maximum fidelity and experimental control.

5 Conclusion

The goal for this paper is to discuss robotics research and technologies for advancing human robot collaboration, and what is needed to assess these future mixed-initiative teams. There is a clear demand to improve soldier-robot teaming established in congressional mandate and Department of Defense (DoD) funded research programs, [17, 35]. However, the state of the art in artificial intelligence and the cost to emulate the complexity of real-world mission scenarios (e.g. cordon and search) requires researchers to rely heavily on simulation. Simulation provides a means to explore future robot capabilities, keep costs low, and enable experimental control for assessment of multimodal communication. Many robotics simulation environments focus on simulation of sensors and physics for robotics development, with few platforms supporting human robot interaction. In order to drive future requirements, interface capabilities, and understand the human factors of multimodal communication, a new simulation environment called VRMIX was presented. VRMIX will provide researchers the necessary tools to assess multimodal interaction in relevant dismounted military missions with robot capabilities yet to come.

Acknowledgement. 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. Amazon.: HTC VIVE Virtual Reality System (HTC), 08 February 2018. https://www.amazon.com/HTC-VIVE-Virtual-Reality-System-pc/dp/B00VF5NT4I?th=1. Accessed 02 Aug 2018
2. Barber, D.J., Leontyev, S., Sun, B., Davis, L., Nicholson, D., Chen, J.Y.: The mixed initiative experimental (MIX) Testbed for collaborative human robot interactions. In: Army Science Conference. DTIC, Orlando (2008)

3. Barber, D.J., Reinerman-Jones, L.E., Matthews, G.: Toward a tactile language for human-robot interaction: two studies of tacton learning performance. Hum. Factors **57**(3), 471–490 (2014). https://doi.org/10.1177/0018720814548063
4. Barber, D., Abich IV, J., Phillips, E., Talone, A., Jentsch, F., Hill, S.: Field assessment of multimodal communication for dismounted human-robot teams. In: The Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, vol. 59, pp. 921–925. SAGE Publications (2015)
5. Barber, D., Carter, A., Harris, J., Reinerman-Jones, L.: Feasibility of wearable fitness trackers for adapting multimodal communication. In: Yamamoto, S. (ed.) HIMI 2017. LNCS, vol. 10273, pp. 504–516. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58521-5_39
6. Barber, D., Howard, T., Walter, T.: A multimodal interface for real-time soldier-robot teaming. In: SPIE Defense, Security, and Sensing - Unmanned Systems Technology, Baltimore, Maryland USA (2016)
7. Barber, D., Lackey, S., Reinerman-Jones, L., Hudson, I.: Visual and tactile interfaces for bi-directional human robot communication. In: SPIE Defense, Security, and Sensing - Unmanned Systems Technology. Baltimore, Maryland USA (2013)
8. Bischoff, R., Graefe, V.: Dependable multimodal communication and interaction with robotic assistants. In: 11th IEEE International Workshop on Robot and Human Interactive Communication, pp. 300–305. IEEE (2002)
9. Chen, J.Y., Barnes, M.J., Qu, Z.: RoboLeader: an agent for supervisory control of multiple robots. In: Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction (HRI 2010), pp. 81–82 (2010)
10. Chen, J., Joyner, C.: Concurrent performance in gunner's and robotic tasks and effects of cueing in a simulated multi-tasking environment. In: Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, pp. 237–241 (2009)
11. Childers, M., Lennon, C., Bodt, B., Pusey, J., Hill, S., Camden, R., Navarro, S.: US army research laboratory (ARL) robotics collaborative technology alliance 2014 capstone experiment. Army Research Laboratory, Aberdeen Proving Ground (2016)
12. Cosenzo, K., Chen, J., Reinerman-Jones, L., Barnes, M., Nicholson, D.: Adaptive automation effects on operator performance during a reconnaissance mission with an unmanned ground vehicle. In: Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, Los Angeles, CA, pp. 2135–2139 (2010)
13. Elliot, L.R., Duistermaat, M., Redden, E., Van Erp, J.: Multimodal Guidance for Land Navigation. U.S. Army Research Laboratory, Aberdeen Proving Ground (2007)
14. Endeavor Robotics. (2018). Endeavor Robotics Products (uPOINT). (Endeavor Robotics). http://endeavorrobotics.com/products. Accessed 02 May 2018
15. EPIC.: Setting up UE4 to work with SteamVR, 08 February 2018 (EPIC). https://docs.unrealengine.com/latest/INT/Platforms/SteamVR/QuickStart/2/. Accessed 02 Aug 2018
16. Glass, D.R.: Taking Training to the EDGE, 14 March 2014 (Orlando Marketing & PR Firm Capital Communications). http://www.teamorlando.org/taking-training-to-the-edge/. Accessed 02 Feb 2018
17. Griffith, T., Ablanedo, J., Dwyer, T.: Leveraging a Virtual Environment to Prepare for School Shootings. In: Lackey, S., Chen, J. (eds.) VAMR 2017. LNCS, vol. 10280, pp. 325–338. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57987-0_26
18. Hearst, M., Allen, J., Guinn, C., Horvitz, E.: Mixed-initiative interaction: trends and controversies. IEEE Intell. Syst. **14**, 14–23 (1999)
19. Kvale, K., Wrakagoda, N., Knudsen, J.: Speech centric multimodal interfaces for mobile communication. Telektronikk **2**, 104–117 (2003)

20. Laboratory, U.A., Schaefer, K.E., Brewer, R.W., Pursel, R.E., Zimmermann, A., Cerame, E., Briggs, K.: Outcomes from the first wingman software-in-the-loop integration event: January 2017. US Army Research Laboratory (2017)
21. Lackey, S. J., Barber, D. J., Reinerman-Jones, L., Badler, N., Hudson, I.: Defining next-generation multi-modal communication in human-robot interaction. In: Human Factors and ERgonomics Society Conference. Las Vegas: HFES (2011)
22. Nigay, L., Coutaz, J. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In: INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 172–178 (1993)
23. Oh, J., et al.: Integrated intelligence for human-robot teams. In: Kulić, D., Nakamura, Y., Khatib, O., Venture, G. (eds.) ISER 2016. SPAR, vol. 1, pp. 309–322. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50115-4_28
24. Open Source Robotics Foundation, 25 January 2018. Gazebo. http://gazebosim.org/. Accessed 14 Feb 2018
25. Parr, L.: Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. Anim. Cogn. **7**, 171–178 (2004)
26. Partan, S., Marler, P.: Communication goes multimodal. Science **283**(5406), 1272–1273 (1999)
27. Raisamo, R.: Multimodal Human-Computer Interaction: A Constructive and Empirical Study. University of Tampere, Tampere (1999)
28. Reinerman-Jones, L., Taylor, G., Sprouse, K., Barber, D., Hudson, I.: Adaptive automation as a task switching and task congruence challenge. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. vol. 55, pp. 197–201. Sage Publications (2011)
29. Sutherland, J., Baillergeon, R., McKane, T.: Cordon and search operations: a deadly game of hide and seek. Air Land Sea Bull. Cordon Search, pp. 4–10 (2010)
30. U.S. Air Force: EOD craftsment balances family, mission, 24 May 2016. from http://www.af.mil/News/Article-Display/Article/779650/eod-craftsman-balances-family-mission/. Accessed 7 Feb 2018
31. U.S. Army Research Laboratory, 17 March 2017. Robotics. U.S. Army Research Laboratory: http://www.arl.army.mil/www/default.cfm?page=392. Accessed 7 Feb 2018
32. U.S. Congress: National Defense Authorization Act for Fiscal Year 2001, Washington, D.C (2001)
33. University of Central Florida.: Mixed Initiative Experimental (MIX) Testbed, 23 July 2013. http://active-ist.sourceforge.net/mix.php?menu=mix. Accessed 02 Sept 2018
34. US Army Research Laboratory Aberdeen Proving Ground United States.: Agent Reasoning Transparency: The Influence of Information Level on Automation Induced Complacency. US Army Research Laboratory Aberdeen Proving Ground United States (2017)
35. Wikipedia: JAUS, 06 July 2017. https://en.wikipedia.org/wiki/JAUS. Accessed 02 Sept 2018