A Human-Robot Speech Interface for an Autonomous Marine Teammate

Michael Novitzky(✉), Hugh R.R. Dougherty, and Michael R. Benjamin

Mechanical and Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
{novitzky,hughd,mikerb}@mit.edu
http://oceanai.mit.edu/pavlab/

Abstract. There is current interest in creating human-robot teams in which a human operator is in its own conveyance teaming up with several autonomous teammates. In this work we focus on human-robot teamwork in the marine environment as it is challenging and can serve as a surrogate for other environments. Marine elements such as wind speed, air temperature, water, obstacles, and ambient noise can have drastic implications for team performance. Our goal is to create a human-robot system that can join many humans and many robots together to cooperatively perform tasks in such challenging environments. In this paper, we present our human-robot speech dialog system and compare participant responses to having human versus autonomous vehicle teammates escorting and holding station at locations of interest.

Keyword: Autonomous teammate marine speech recognition dialog

1 Introduction

There is great interest in combining manned vehicles with autonomous vehicles to perform tasks in challenging environments. For example, the U.S. Army has the Manned Unmanned Teaming (MUM-T) program in which manned aircraft work with unmanned aerial systems (UAS) [10]. The manned aircraft are AH-64E Apache attack helicopters which are aided by Grey Eagle UAS. The crews are able to request various levels of control of the UAS from simply receiving its camera data to ordering waypoints for the UAS to visit. The U.S. Air Force's "Loyal Wingman" project is exploring manned-unmanned teaming in which an UAS and a manned aircraft work directly on missions such as air interdiction, attack on integrated air defense systems, and offensive counter air [1,6]. In this work we explore a similar manned unmanned teaming concept in the marine domain. The marine domain is more accessible for deploying autonomous vessels and yet still is challenging given the elements in the environment. Our manned vessel is a motorized kayak and our autonomous teammate is a surface vehicle (ASV). In particular we focus on preliminary user-centered design improvements

This work was supported by Battelle.

A. Agah et al. (Eds.): ICSR 2016, LNAI 9979, pp. 513–520, 2016.
DOI: 10.1007/978-3-319-47437-3_50

of our speech dialogue system in order to perform tasks of increasing complexity. In future experiments we wish to expand to coordinating with several teammembers while also recording physiological data of our participants. This will allow for the exploration of the interplay between operator load, robot autonomy, and human-robot trust. It is our goal to provide lessons learned from our platform in the marine domain to other challenging environments. Our design prinicpal is for novice users to begin using the system in a short period of time. This paper describes our manned unmanned teaming system and initial pilot study aimed at comparing human vs autonomous vehicle teammates.

2 Related Work

Uhrmann et al. [9] investigated the Manned-Unmanned Teaming (MUM-T) domain. The researchers simulated a full-scale military helicopter mission with the introduction of UAVs for route reconnaissance and for observing the designated landing sites prior to the approach of the manned helicopter. The mission was to have troops transported via manned helicopter to secure an object. Both manned and unmanned assets provided reconnaissance information and overwatch after troop delivery. Artificial Cognitive Units (ACUs) were implemented to interpret tasks with respect to the current mission and tactical situation and act upon those tasks in a situation-specific way. The goal of the ACUs was to allow a commander to communicate the task to the UAV and it take care of all the parameters just like a human pilot does. The human-robot interface varied depending on the focus of the task such as maps for spatial representations of a task and timetables or schedules for temporal representations. Speech interfaces were utilized when task representation involved a causal component, with previous or following tasks refereed in speech output or commands.

Draper et al. [3] investigated using speech input versus manual input for an unmanned aerial vehicle control station. The control station was designed to operate one vehicle at a time with multiple monitors and manual controls. They found that speech input was superior to manual input for flight/navigation and data entry tasks. Operators in this study indicated that speech provided a head up and hands-free advantage.

Franke et al. [4] describe systems for command and control for a single human operator to many autonomous vehicles. The authors note that using auditory cues and speech frees the operator's eyes and hands to observe other information and manipulate other tasks. They describe three primary control paradigms: direct control, management by consent, and management by exception. Direct control is when a human does all the decision making and information processing. This approach has a high workload as it requires the operator to constantly attend the controls. Management by consent has a lower operator workload as the vehicle performs planning but waits for the operator to approve it before proceeding. In management by exception the vehicles performs its own planning and starts executing the plan. In this case, the operator can override any actions or plans of the vehicle.

The work presented in this paper introduces a manned unmanned teaming mission using speech recognition for command and control. The autonomy in the following experiments is higher than the direct control paradigm because it takes high level commands such as "Follow" instead of direct joystick inputs. While the following experiments only have one teammate, future work will expand to having multiple teammates that recognize speech commands.

3 Experimental Setup

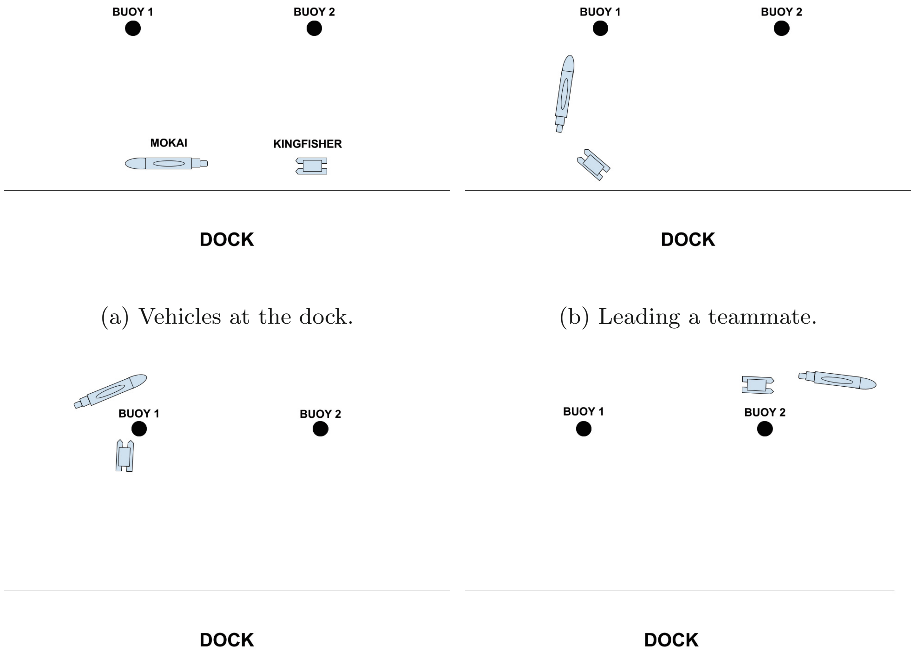
The experiments were conducted on the Charles River in Cambridge, Massachusetts where participants lead a teammate named Arnold to points of interest on the river. After each trial, the participants filled out a questionnaire. The questionnaire included the NASA Task Load Index (TLX) [5], Robot Liking, General Engagement, and Schaefer Trust Scale Items [8]. The TLX questions were rated on a likert scale between 1 and 7, very low load to very high load, respectively. The Robot Liking items were rated on a likert scale ranging between 1 and 10 where 1 was very poorly and 10 describes very well. The General Engagement items were on a likert scale between 1 and 7 where 1 was not at all enganged to 7 which was extremely enganged. The Schaefer Trust Scale items asked "what percentage of time will this teammate be ____". Participants responded on a likert scale between 0 % and 100 % with 10 % increments in between. The questionnaire included open ended questions about general experience and suggestions for improvements for user-centered design. The final portion of the questionnaire recorded demographic information. The 15 participants had a mean age of 29.5 years, are comfortable on the water, and had a comfort with robotics ranging from 2 to 6 on a 7 point likert scale.

3.1 Task Description

The human participant is asked to escort its teammate from the dock to two points of interest marked with buoys on the water, as seen in Fig. 1. They are instructed to have their teammate station as close as possible to each point of interest. Once all the points of interest have been visited they return back to the starting location at the dock.

3.2 Participant Conveyance

The vehicle for human conveyance is an augmented Mokai ES-Kape which is a motorized kayak, seen in Fig. 2. The ES-Kape weighs 88.45 kg, has a length of 3.63 m, and is powered by a Subaru EX21 engine that can reach top speeds of 54.39 km/h. In order to function as a vehicle on our network, the ES-Kape has been augmented with a semi-rugged laptop, compass, GPS, and long range wifi antenna.



(a) Vehicles at the dock.

(b) Leading a teammate.

(c) Stationing at buoy 1.

(d) Stationing at buoy 2.

Fig. 1. Example experiment (not to scale). The participant and teammate start at the dock, as seen in (a). The participant is asked to lead their teammate (a Kingfisher ASV in this example), as seen in (b), with the command "Arnold Follow" to a buoy where they use the command "Arnold Station" to have their teammate station as close as possible, as seen in (c). The participant will then escort their teammate to the second buoy, as seen in (d), with the same two commands. After visiting both buoys, the participant is asked to have Arnold return to the starting location at the dock. The participant can use the command "Arnold Follow" and lead their teammate to the dock or "Arnold Return" in which case the teammate will autonomously guide itself to the dock.

3.3 Autonomous Teammate

The autonomous teammate is a Clearpath Robotics Kingfisher M200, seen in Fig. 2, which is a surface vehicle. The ASV is a mid-sized surface vessel with dimensions of 1.35 m × 0.98 m × 0.32 m and a weight of 28 kg and operates at 1.5 m/s. The autonomy for the ASV is provided by MOOS-IvP [2]. MOOS is a robot middleware that utilizes a centralized database paradigm. The autonomy is provided by the IvP Helm behavior-based decision engine architecture. The IvP Helm behaviors used in this work are trail, station, and waypoint.

Fig. 2. Human operator in the motorized kayak, Mokai ES-Kape, and the autonomous robot teammate Arnold.

3.4 Human Teammate

As a control to the autonomous teammate, an alternate teammate is a human that operates a motorboat. The human teammate mimics the robot by traveling at 1.5 m/s and uses the exact same dialogue. The motorboat is augmented with a semi-rugged laptop, compass, GPS, and long range wifi antenna.

3.5 Speech Interaction

A dual radio headset with dual push-to-talk (PTT) is used to mitigate the effects of wind and motor noise from the ES-Kape. The right speaker/PTT combination is connected to a 5 W waterproof handheld radio which is used to communicate with humans. The left speaker/PTT combination is connected to a semi-rugged laptop which runs the speech dialog MOOS-IvP modules that communicate with the autonomous robot teammate.

The speech recognition used in this project is provided by the open-source large vocabulary continuous speech recognition engine Julius [7]. The engine allows for the specification of possible sentences and vocabulary to be recognized. The Julius engine has been encapsulated into the MOOS-IvP application called *uSpeechRec*.

A dialog manager was created called *uDialogManager*. Each command sentence recognized by *uSpeechRec* is acknowledged by asking the user "Did you mean, <command>?", where the <command>echos what the system believed to be uttered by the user. Possible commands are Follow, Station, Return and Status. The user can answer "No" in which case *uDialogManager* does nothing and responds with "Command Canceled" or the user can answer "Yes" in which case *uDialogManager* sends the appropriate command to the autonomous robot teammate and responds to the user with "Command Sent." The acknowledgement loop reduces error as the accuracy in speech recognition can be affected by wind, ambient noise, or user accents. If the teammate receives the command it will respond to the user with a text to speech message.

4 Experimental Protocol

We performed a within-subjects experiment in which the teammate is either an autonomous robot or a human (randomized order). Due to the experiments occuring on the water, each participant begins the experiment with a safety briefing and is briefed on how to operate the ES-Kape. They acclimate to the ES-Kape by driving it around a buoy and returning to the dock. The participant is then given a headset and is briefed on how to use the handheld radio to communicate with humans and the PTT for communicating with the autonomous teammate. Once the orientations are completed the participant is briefed on their task, which was described above in Sect. 3.1. After each treatment (where the teammate is human or autonomous) the participant is asked to complete a questionnaire. After the participant completes the two treatments and questionnaires, they are interviewed by an experimenter on their experiences.

Table 1. Questions and Wilcoxon signed-rank test

Question	Median human	Median robot	W	p value
How successful were you in accomplishing what you were asked to do?	6	5	34	0.0312
The experience caused real feelings and emotions for me	2	3	9	0.0371
Reliable	80	70	47	0.0488
Led astray by unexpected changes in the environment	20	40	8	0.0234
Protect people	50	40	32.5	0.0469
Malfunction	10	30	17	0.0464
Warn people of potential risks in the environment	70	30	57	0.0312
Provide appropriate information	80	80	66	0.0371
Make sensible decisions	70	60	65.5	0.0405
A good teammate*	80	70	70.5	0.0088
Performs task better than a novice human user*	70	40	103	0.0004
Possess adequate decision-making capability*	50	40	72	0.0078
Meets the needs of the mission*	90	60	76	0.002

*$p < 0.01$

5 Results

Wilcoxon signed-rank test was performed on 59 likert scale responses. The questions with p values less than 0.05 are listed in Table 1. In Table 1 the column labeled W are the Wilcoxon signed-rank test values. In the TLX section, question "How successful were you in accomplishing what you were asked to do?" had a p value of 0.0312. In the General Engagement section of the questionnaire the question "The experience caused real feelings and emotions for me" had a p value of 0.0371. The rest of the questions listed in Table 1 were from the Schaefer Trust Scale Items. The questions "A good teammate", "Performs task better than a novice human user", "Possess adequate decision-making capability", and "Meets the needs of the mission" had p values below 0.01.

6 Conclusions and Future Work

The experiments in this pilot study demonstrate our systems' initial capability for a human-robot team to perform tasks in a challenging environment. Overall, it seems that human teammates were viewed in a more favorable light than their robot counterparts. Even though their dialog was designed to be the same. Future work will include recruiting a larger number of participants along with gathering physiological data to characterize cognitive load. Additionally, analysis of quantitative data such as the task duration or distance teammates were stationed from points of interest may reveal differences in performance between teammates or improvements in the system. Further user-centered design iterations and increasing the team size will aid in finding lessons learned for application to other challenging environments.

Acknowledgments. The authors would like to acknowledge Alon Yaari for his technical support and the MIT Sailing Pavilion staff for their professionalism and support in designing and maintaining our safety protocols. We thank Paul Robinette and Alan Wagner for feedback and guidance throughout this work.

References

1. United States Air Force unmanned aircraft systems flight plan 2009–2047. Headquarters, U.S. Air Force, p. 34, May 2009
2. Benjamin, M.R., Schmidt, H., Newman, P., Leonard, J.J.: Nested autonomy for unmanned marine vehicles with MOOS-IvP. J. Field Robot. **27**(6), 834–875 (2010)
3. Draper, M., Calhoun, G., Ruff, H., Williamson, D., Barry, T.: Manual versus speech input for unmanned aerial vehicle control station operations. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 47, pp. 109–113. SAGE Publications (2003)
4. Franke, J.L., Zaychik, V., Spura, T.M., Alves, E.E.: Inverting the operator/vehicle ratio: approaches to next generation UAV command and control. In: Proceedings of AUVSI Unmanned Systems North America 2005 (2005)

5. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
6. Kearns, K.: RFI: autonomy for loyal wingman. Air Force Research Laboratory (AFRL), July 2015
7. Lee, A., Kawahara, T.: Recent development of open-source speech recognition engine Julius. In: Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, pp. 131–137. Asia-Pacific Signal and Information Processing Association, International Organizing Committee (2009)
8. Schaefer, K.E.: The Perception and measuerment of human-robot trust. Ph.D. thesis, University of Central Florida, August 2013
9. Uhrmann, J., Strenzke, R., Schulte, A.: Task-based guidance of multiple detached unmanned sensor platforms in military helicopter operations. In: COGIS, Crawley (2010)
10. Whittle, R.: MUM-T is the word for AH-64E: Helos fly, use drones. Breaking Defense, January 2015