Integrating a Multi-Modal Interface in an On-Water Human-Robot Teaming Testbed

Michael Novitzky, Caileigh Fitzgerald, Danielle Gleason, Paul Robinette, and Michael R. Benjamin
Department of Mechanical and Ocean Engineering
Massachusetts Institute of Technology
Cambridge, USA
{novitzky, caileigh, dgleason, paulrobi, mikerb} @mit.edu

***Abstract*—Project Aquaticus is a testbed for manned-unmanned teaming that focuses on human-robot trust, human cognitive load, and robot teammate autonomy. It studies this intersection with a game of capture the flag with humans in motorized kayaks and small autonomous surface robot teammates on the water. Audio has been the main interaction modality between humans, their robot teammates, and the game. Humans give voice commands to the robots and receive audio responses from the robots. Game state information is conveyed over audio channels. During the summer of 2018, we were able to capture 13 of capture the flag games ranging from teams of 1 vs. 1 to 4 vs 4. Post-experiment questionnaires and interviews have revealed that participants had a very difficult time keeping situational awareness with just auditory cues. Our mitigation strategy to this problem is the conception of a visual icon display mounted in the cockpit of the motorized kayak. This vehicle state display (VSD) relays game state information to the user. In this paper, we present our survey results to find representative game state icons, the hardware and software implementation of our VSD, and our proposed experiments to determine if the VSD improves participants' situational awareness through a capture the flag game.**

I. Introduction

Modern systems are combining manned vehicles with autonomous vehicles to perform tasks in challenging environments. We have developed a manned-unmanned teaming concept in the marine domain called Project Aquaticus. The marine domain is more accessible for deploying autonomous vessels (no approval is required from government agencies and vehicles can be easily stopped on the water) and yet still challenging given the elements in the environment. As seen in Figure 1, our manned vessels are motorized kayaks and our autonomous teammates are autonomous surface vehicles (ASVs). The nominal composition of the experiment is to situate two humans and two robots on the same team competing in a game of capture the flag against a similarly situated team. Project Aquaticus has been designed to explore the interplay between human operator load, robot autonomy, and perceived human robot teammate trust. It is the goal of Project Aquaticus, not necessarily this paper, to provide lessons learned from our platform in the marine domain to other challenging environments.

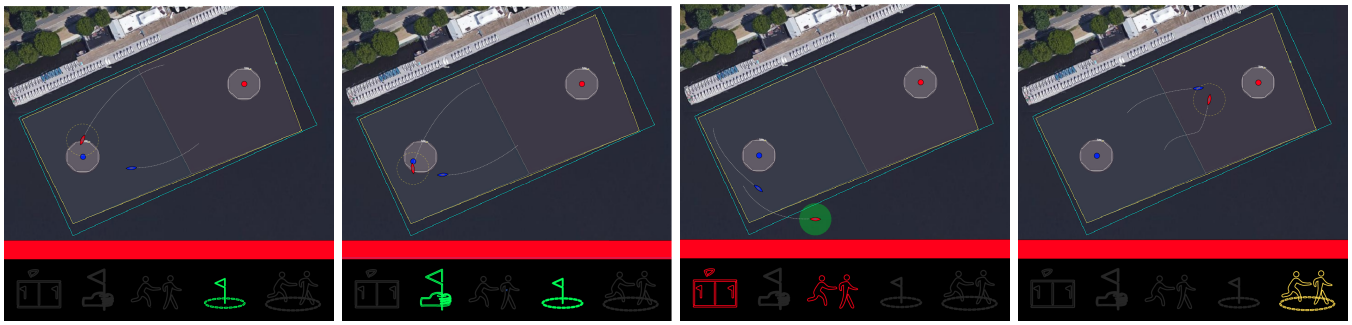
Project Aquaticus reached a major milestone this past summer of 2018. We were able to capture six full games of capture the flag between varying sizes of teams (2 vs 2, 3 vs 3, and 4 vs 4) [1]. These datasets included all of the vehicle track data, game state information along with the human-human and human-robot communications along with video of the participants. These datasets have been made publicly available at: http://oceanai.mit.edu/aquaticus/datasets

Fig. 1: Two human vehicles and two robot vehicles near the blue team's flag.

II. Project Aquaticus Testbed

Each Aquaticus game takes place in the area of the Charles River immediately adjacent to the MIT Sailing Pavilion. The field for the game extends along the entire 160m dock and 80m into the river. The capture the flag game mechanics we have implemented are based on scoring the most points in ten minutes. In order to score a point, the participant must go to their opponent's side of the field, virtually grab the flag and make it back to their home flag without being tagged or going out of bounds. Players can defend their side of the field through the use of tags. To tag an opponent, a request is made to the central game manager that checks if the requesting vehicle is in their home field side, that the opponent vehicle is both on the requesting vehicles side, within 10 meters, and the tagger is untagged. A tagged player must untag themselves by returning to their home flag zone. If a player is carrying a virtual flag and is tagged, then the flag is automatically reset to its original location.

The autonomy for the ASV is provided by the open source project MOOS-IvP [2]. The robot teammates in the Project Aquaticus testbed are fully autonomous: given a command, such as 'defend the flag' they perform the task unsupervised. The ASVs are Clearpath Robotics Heron M300s. The human



(a) In Flag Zone (b) Grabbed Flag and In Zone (c) Out of Bounds and Tagged (d) Within Tag Range

Fig. 2: Overhead simulation view along with the vehicle state display (VSD) for the red player. In a) the red player has entered the opposing team's flag zone. In b) the red player has grabbed the blue flag. In c) the red player has traveled out of bounds and is therefore tagged. In d) the red player is within tagging range of the blue player.

vehicles are Mokai ES-Kape motorized kayaks which are controlled via an on-board joystick and have been augmented with a semi-rugged laptop, compass, GPS, long range WiFi antenna, and a headset for voice interaction with the game. For further details please refer to [1].

All information to and interaction with the participant is piped through their headset in audio form. This is to better reflect what is encountered in tactical environments. The right speaker/push-to-talk (PTT) combination is connected to a 5-Watt waterproof handheld radio which is used to communicate with game referees and safety personnel. The left speaker/PTT combination is connected to an embedded computer. A button box connected to the embedded computer directs how the participant's speech is processed. The first button directs the speech to the MOOS-IvP modules that communicates with the autonomous robot teammate through speech recognition using the open-source large vocabulary continuous speech recognition engine Julius [3] and a basic dialogue manager. The second button directs the speech to a MOOS-IvP voice over IP (VOIP) application which integrates the open-source communication library Mumble [4]. Examples of game events are entering and exiting the flag zone which sounds like a bell and a buzzer. Events such as flag events and tagged events come over as text-to-speech (TTS). For example, when blue one grabs the red flag, the TTS comes over the headset saying "Blue One has the red flag." Similarly, when vehicles are tagged, it comes over TTS as "Blue One has been tagged." There is no current audio indication that you are within tag range of another vehicle.

Fig. 3: The prototype view of the actualized vehicle state display.

III. Multi-modal Interaction Studies

In post-experiment questionnaires and interviews from the summer of 2018, it became evident that participants reported a difficult time keeping track of game state information. In order to address issues that participants had with their situational awareness, we began looking into using icons for game state information to be displayed inside of their cockpit along with audio cues. The primary concern participants had, was how difficult it was to keep track of their game state. The states we chose to integrate into the vehicle state display (VSD) were, whether or not a participant was tagged, has the flag, is in a flag zone, or within range to tag an opponent. A series of icons were designed and put in a survey form and sent to 40 IRB approved potential participants. 17 participants responded. The icons for each category that were more accurately identified were those chosen for prototype implementation. The VSD, as seen in Figures 2 and 3, was created to display the chosen icons using an Arduino board, strip LEDs, and an acrylic encasement fitted to the cockpit of the motorized kayak. A new software module was created to run on the participant's computer which controlled the Arduino board, in turn, relaying game state.

IV. Upcoming Experiments

We have presented our Project Aquaticus testbed and difficulties our participants encountered with audio only interactions. Our goal with this work is to measurably demonstrate the improvement in participant situational awareness using a multi-modal approach over just using audio alone. As it is still in prototype state, we will only be able to test one participant at a time. We envision running a modified version of aquaticus with less players on the field to allow for a more controlled experiment. That way it will be more straight forward to compare participants in a multi-modal setup vs those in the unimodal (audio only) setup. We envision an experiment in which all participants experience the same series of events designed to create confusion. We will then query the participants as to their perceived game state so that we may compare the experimental conditions.

REFERENCES

[1] M. Novitzky, P. Robinette, M. R. Benjamin, C. Fitzgerald, and H. Schmidt, "Aquaticus: Publicly available datasets from a marine human-robot teaming testbed," in *ACM/IEEE International Conference on Human-Robot Interaction (Alt.HRI)*. Daegu, South Korea: ACM, March 2019.

[2] M. R. Benjamin, H. Schmidt, P. Newman, and J. J. Leonard, "Nested autonomy for unmanned marine vehicles with MOOS–IvP," *Journal of Field Robotics*, vol. 27, no. 6, pp. 834–875, 2010.

[3] A. Lee and T. Kawahara, "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*. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, 2009, pp. 131–137.

[4] "Mumble voice chat software," https://wiki.mumble.info/wiki/Main_Page, online; accessed 2018-05-30.