A Platform for Studying Human-Machine Teaming on the Water with Physiological Sensors.

Michael Novitzky, Paul Robinette, Danielle K. Gleason, and Michael R. Benjamin
Laboratory for Autonomous Marine Sensing Systems (LAMSS)
Department of Mechanical and Ocean Engineering
Massachusetts Institute of Technology Cambridge, Massachusetts 02139
Email: novitzky, paulrobi, dgleason, and mikerb @mit.edu

***Abstract*—In this work we create human-robot teams in which humans pilot their own vehicle and team up with several autonomous vehicles to accomplish a common goal. We focus on human-robot teamwork in the marine environment as it is challenging and can serve as a surrogate for other environments. Humans use motorized kayaks to interact with autonomous surface vehicles. 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. We have previously presented our human-robot speech dialog system which has been through several rounds of user-centered design improvements. In this paper, we present our plan to integrate a human operator's physiological measurements so that an autonomous teammate can vary their communication and levels of control in response to the operator's cognitive load. We present initial results of using heart rate as a potential measure of cognitive load. Our goal is to perform further user-centered design improvements with both user feedback and physiological measures as guidance.**

I. Introduction

Modern systems are 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) [11]. 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, 5]. 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 (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. Our manned vessel is a motorized kayak and our autonomous teammate is an autonomous surface vehicle (ASV). In particular we focus on preliminary user-centered design improvements of our speech dialogue system in order to perform tasks of increasing complexity.

Fig. 1. Human operator in the motorized kayak, Mokai ES-Kape, and the autonomous robot teammate Arnold.

In future experiments we wish to expand to coordinating with several team-members while also recording physiological data of our participants. This will allow for 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 principal is for novice users to begin using the system in a short period of time. We have previously presented our manned unmanned teaming system along with user-centered design improvements and initial pilot study aimed at comparing human vs autonomous vehicle teammates [7] [8]. This paper describes our plan to integrate a human operator's physiological measurements so that an autonomous teammate can vary their interaction with the human based on their cognitive load. We present initial results of using heart rate as a possible measure of cognitive load for our motorized kayak operators.

II. Related Work

Uhrmann et al. [10] 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. 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 refered 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 vehicle performs its own planning and starts executing the plan. In this case, the operator can override any actions or plans of the vehicle. The autonomy in the surface vehicles in our work is higher than the direct control paradigm because it takes high level commands such as "Follow" instead of direct joystick inputs.

Solovey et al. [9] attempted classifying driver workload using physiological and driving performance data. Using machine learning techniques, they had classification accuracies by several algorithms of up to 90% for detecting elevated levels of cognitive load. These measures were collected in real-time and did not interfere with the primary task of driving the vehicle. The researchers used heart rate (HR) and skin conductance level (SCL) as the physiological measures. Driving speed, steering wheel position, and acceleration data were measures of vehicle control. Cognitive workload was increased using an "n-back" task which is an auditory presentation of digits followed by a verbal response by the participant delayed by n numbers. An important facet of this work is that the participants first drove in urban traffic for ten minutes to acclimate to the vehicle. They then drove twenty minutes on an interstate highway for additional familiarization. After familiarization, a single task of driving was used as a reference period. Then alternating task periods of "n-back" and rest and recovery were performed. The researchers found that HR was the most sensitive to cognitive load changes. They found Logistic regression and naive Bayes performed significantly better than all the other classifiers attempted.

III. Manned-Unmanned Teaming System

Our manned-unmanned teaming system is comprised of a human in a motorized kayak and autonomous surface vehicles which communicate through speech. The vehicle for human conveyance is an augmented Mokai ES-Kape motorized kayak. The ES-Kape weighs 88.45 kg, has a length of 3.63 meters, and is powered by a Subaru EX21 engine that can reach top speeds of 54 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.

The autonomous teammates are Clearpath Robotics Kingfisher M300s. The ASV is a mid-sized surface vessel with a weight of 28 kg and travels 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*.

A dual radio headset with dual push-to-talk (PTT) is used to mitigate the effects of wind and motor noise from the Mokai. The right speaker/PTT combination is connected to a 5-Watt waterproof handheld radio which is used to communicate with humans such as on shore or in a safety boat. 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 teammates.

The speech recognition used in this project is provided by the open-source large vocabulary continuous speech recognition engine Julius [6]. 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 *uSpeechRec* and adapted for use in our system.

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 baseline commands are *Follow, Station, and Return*. 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 acknowledgment loop reduces error as the accuracy in speech recognition can be affected by wind, ambient noise, or user accents.

IV. Previous Work

We performed user-centered design experiments which were conducted on the Charles River in Cambridge, Massachusetts where participants lead a teammate named Arnold to points of interest on the river. User-centered design improvements were implemented based on post-experiment participant questionnaire and interview data.

The human participant was asked to escort its teammate to two points of interest marked with buoys on the water. They were instructed to have their teammate station as close as possible to each point of interest. Once all the points of interest were visited they returned back to the starting location at the dock.

Once the on-water orientations were completed the participant was briefed on their task, which was described above. After each experiment, the participant was asked to complete a questionnaire. After the participant completed the experiment and questionnaires, they were interviewed by an experimenter on their experiences.

Five participants used the baseline system. Based on the feedback in both questionnaire and interview data several improvements were implemented. The major common issue for the participants was a lack of feedback from their teammates, they were unsure of what task their teammate was performing. The command "Arnold Status" was added to the available sentences. Additionally, the *pBotDialog* application was created so that the autonomous teammate would respond with its current task upon receipt of the "Arnold Status" request. The IvP Helm autonomy was augmented so that upon switching to a new behavior it would inform the participant. Fifteen participants performed the experiment in this design iteration. Many of these participants used the "Arnold Status" command to verify their teammate was still performing the task they were commanded. Improvements in task performance between the initial iteration and this improved version are still being analyzed. Task performance measures include time to complete the task and distance a human operator can place their autonomous teammate to the buoy. Communication measures include the number of repeated commands to the robot. Other measures include participant responses on the TLX questionnaires.

V. Integrating Physiological and Alternative Measures

In previous versions of our system, if the participant errors in the command spoken or the system errors in speech recognition, the user may cancel a command. Previously, canceled commands were simply discarded. In a system trying to monitor the participant, repeated canceled commands can be an indicator of either system failure or participant overload. The *uDialogManager* can be adjusted to measure the number of consecutive participant command rejections and send a message to the autonomous teammate that there are communication problems.

Heart Rate (HR) will be measured using the Zephyr BioModule Device. Designed for athletes and high performance environments, the BioModule attaches to a strap that a participant can fit under their shirt. The BioModule uses a BlueTooth interface that connects directly to the human operator's vehicle Toughbook computer.

Measuring galvanic skin response is more difficult given the nature of our testing environment. The participants are close to the surface of the water while in their vehicle. Water sprays on them while their hands are on the hull of the kayak and their torso and face can be sprayed by water depending on their speed, water chop, and wind. Because participants use both hands, one for the joystick to control their vehicle and the other for communication, we are investigating sensors that attach on parts of their body not exposed to the water such as their feet.

Measuring the Mokai operator's head motion may be an additional piece of information to cognitive load or operator comfort. Initial attempts have been made with a head-mounted camera to observe optical flow of the image. A general sense of head turn can be seen in the optical flow results. However, when the vehicle turns or the operator looks to the water at the side of the vehicle, the optical flow produces large lateral movements in the image that do not correspond with the operator's head motion. Alternatively, we are exploring the use of a compass to track head turning along with the head mounted camera.

VI. Preliminary Heart Rate Experiments

In order to determine if heart rate (HR) could be an indicator of cognitive load we performed explorative experiments similar in fashion to Solovey et al. [9] and their use of the "n-back" procedure. We established a baseline for the heart rate of the Mokai operator by setting up a slalom course on the water. The slalom course provided for a consistent challenge for the operator rather than the operator driving into areas that were relatively calm or full of traffic which would alter their heart rate. The Mokai operator steered repeatedly in-between buoys for 10 minutes. We then attempted to induce higher cognitive load by giving the Mokai operator math equations to solve via the radio from shore while still driving through the slalom course. The types of math problems included simple addition and subtraction in addition to a few more difficult problems, namely the multiplication of 2 "larger" numbers and an integral. As can be seen in Figure 2, the HR for the slalom only modality was relatively consistent except for when a stray paddleboarder entered the slalom and the operator had to figure out how to maneuver in order to avoid a collision. Figure 3 contains the modality in which the Mokai operator navigates a slalom course and performs math over the radio. Spikes in HR can be seen while the operator is processing the more difficult problems. During just the slalom phase the average HR was 116.94 with a standard deviation of 5.15 while in the slalom with math modality the average HR was 122.67 with a standard deviation of 8.11. The average and standard deviation for the slalom modality were calculated after removing an episode of the stray paddleboarder entering the slalom, described above.

Based on these results, we will continue to investigate using HR as a measure for cognitive load of our Mokai operators. In particular, observing spikes in HR during the difficult math problems gives an indication that it is a measure of cognitive load. A difficulty with our current experimental setup is that cognitive load was not maintained at an elevated level throughout the ten minute time span. This indicates that the math problems may have been too easy for the operator. Additionally, radio reception was not always clear and some time was needed to clarify and resend both questions and

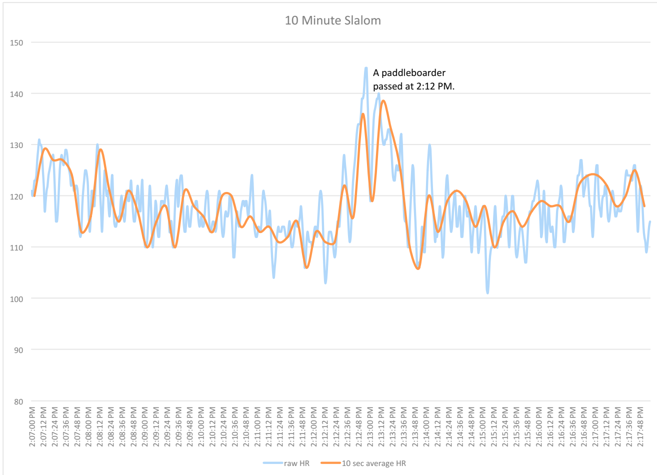

Fig. 2. Heart rate (in beats per minute) of the Mokai operator over a 10 minute period performing just a slalom course.

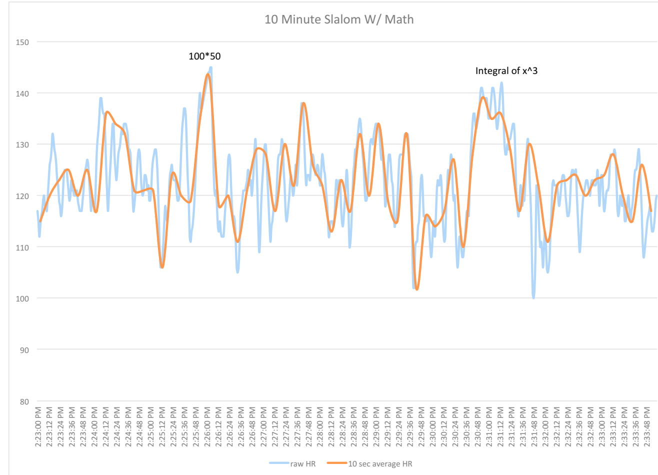

Fig. 3. Heart rate (in beats per minute) sof the Mokai operator over a 10 minute period performing a slalom and math problems over the radio.

answers. Further exploration is required to design a modality that maintains higher levels of cognitive load while the Mokai operator is navigating the slalom.

VII. Autonomy Adaptation Further Human-centered Design Experiments

The first new set of human-centered design experiments will begin with the improvements indicated by the participants in the last iteration of experiments described above. This will allow for an establishment of comfort with respect to the physiological measuring instruments and data analysis, along with any machine learning classifier creation, for the researchers. It will also allow for exploration to determine if the previous experiment of points of interest visitation actually involves a high cognitive load, and if so, at which point it occurs. Additionally, it will allow for comparison against an augmented system in which the autonomy adapts to the participant's cognitive load.

There are several adaptations an autonomous teammate can perform in response to perceived operator load. The first is to modify the frequency and content of its messages. An observed behavior by the participants in our experiments is the constant need to look over their shoulders to verify the autonomous robot teammate is performing the commanded action. A new behavior in robot communication to address this will be the continual communication of the robot's status without being prompted by the operator. Having the autonomous teammate adjust the frequency of this status message may be regulated by the perceived operator load, reducing the frequency when the load is high. The second autonomy adaptation is to change the way in which they are controlled based on operator cognitve load: direct control, management by consent, or management by exception. As described above, the current implementation of the autonomous teammate is better than direct control because the robots respond to commands such as "follow". In order for these autonomous teammates to be capable of management by consent or management by exception, they need a method of determining their possible actions or plans based on the mission. This may then be leveraged by the autonomous teammate to help reduce the human's cognitive load by switching itself from requiring explicit mode commands to being managed by consent where it can request permission to change modes based on context.

VIII. Conclusions and Future Work

Our previous pilot study experiments demonstrated our systems' initial capability for a human-robot team to perform tasks in a challenging environment. Improvements in the dialog system were implemented based on feedback from participants in the baseline experiments. It is our goal to implement improvements to the autonomy that can adapt to the perceived operator cognitive load. Our initial heart rate (HR) experimental results do indicate that increased cognitive load created by more difficult math problems does increase the HR of the Mokai operator. Our next steps in the HR evaluation for measuring cognitive load is to identify modalities that elicit consistent cognitive load for extended periods of time. We will also investigate whether direct ECG or heart rate variability are more indicative of cognitive load. Future work will include gathering physiological data to characterize cognitive load. Our plan is to continue investigating galvanic skin response and operator head motion. Further user-centered design iterations and increased team size will aid in finding lessons learned for application to other challenging environments.

Acknowledgments

This work was supported by Battelle. The authors would like to thank the MIT Sailing Pavilion staff for their professionalism and support in designing and maintaining our safety protocol. We thank Nick DePalma and Alan Wagner for their feedback and guidance during early portions of this work.

References

[1] United States Air Force unmanned aircraft systems flight plan 2009-2047. *Headquarters, U.S. Air Force*, page 34, May 2009.

[2] Michael R. Benjamin, Henrik Schmidt, Paul Newman, and John J. Leonard. Nested autonomy for unmanned marine vehicles with MOOS-IvP. *Journal of Field Robotics*, 27(6):834–875, 2010.

[3] Mark Draper, Gloria Calhoun, Heath Ruff, David Williamson, and Timothy Barry. Manual versus speech input for unmanned aerial vehicle control station operations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 109–113. SAGE Publications, 2003.

[4] Jerry L Franke, Vera Zaychik, Thomas M Spura, and Erin E Alves. Inverting the operator/vehicle ratio: Approaches to next generation uav command and control. *Proceedings of AUVSI Unmanned Systems North America*, 2005, 2005.

[5] Kirsten Kearns. RFI: Autonomy for Loyal Wingman. *Air Force Research Laboratory (AFRL)*, July 2015.

[6] Akinobu Lee and Tatsuya 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*, pages 131–137. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, 2009.

[7] Michael Novitzky, Hugh R. R. Dougherty, and Michael R. Benjamin. Towards a human-robot speech interface for an autonomous teammate. In *Robot and Human Interactive Communication, 2016. Proceedings. ROMAN 2016. The IEEE International Symposium on*. IEEE, 2016.

[8] Michael Novitzky, Hugh RR Dougherty, and Michael R Benjamin. A Human-Robot Speech Interface for an Autonomous Marine Teammate. In *International Conference on Social Robotics*, pages 513–520. Springer, 2016.

[9] Erin T Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. Classifying driver workload using physiological and driving performance data: two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4057–4066. ACM, 2014.

[10] Johann Uhrmann, Ruben Strenzke, and Axel Schulte. Task-based guidance of multiple detached unmanned sensor platforms in military helicopter operations. *COGIS, Crawley*, 2010.

[11] R Whittle. MUM-T is the word for AH-64E: Helos fly, use drones. *Breaking Defense*, January 2015.