Robotics: Science and Systems 2022
New York City, NY, USA, June 27-July 1, 2022

Underwater Robot-To-Human Communication Via Motion: Implementation and Full-Loop Human Interface Evaluation

Michael Fulton, Muntaqim Mehtaz, Junaed Sattar
Deptartment of Computer Science & Engineering
Minnesota Robotics Institute
University of Minnesota – Twin Cities
Email: {fulto081,mehta216,junaed}@umn.edu

Owen Queeglay
Department of Electrical Engineering
Stanford University
Email: omqueeg@stanford.edu

***Abstract*—Autonomous underwater vehicles (AUVs) have long lagged behind other types of robots in supporting natural communication modes for human-robot interaction. Due to the limitations of the environment, most AUVs use digital displays or topside human-in-the-loop communications as their primary or only communication vectors. Natural methods for robot-to-human communication such as robot "gestures" have been proposed, but never evaluated on non-simulated AUVs. In this paper, we enhance, implement and evaluate a robot-to-human communication system for AUVs called Robot Communication Via Motion (RCVM), which utilizes explicit motion phrases (kinemes) to communicate with a dive partner. We present a small pilot study that shows our implementation to be reasonably effective in person followed by a large-population study, comparing the communication effectiveness of our RCVM implementation to three baseline systems. Our results establish RCVM as an effective method of robot-to-human communication underwater and reveal the differences with more traditional communication vectors in how accurately communication is achieved at different viewpoints and types of information payloads.**

I. INTRODUCTION

Despite the challenges inherent in the underwater environment, autonomous underwater vehicles (AUVs) have diversified both in form and applications over the last sixty years. AUVs now explore shipwrecks [10], chart biological habitats [30] and marine geology [31], observe the effects of climate change underwater [23], destroy subsea mines [24], inspect and repair undersea infrastructures such as pipelines or cables [21], aide in local water resource management [9], and help to control invasive species [1]. In many of these applications, AUVs could be deployed to work alongside humans, aiding them by carrying equipment, scanning the area for points of interest, maintaining dive safety, and guiding humans in their tasks underwater. To enable robots to work in these teams, methods for efficient, stable, and effective communication of information from robots to human partners are required. However, in the underwater environments where these diver-robot teams must work, standard robot-to-human communication methods suffer from a variety of challenges. Signals in the electromagnetic (EM) spectrum suffer from high attenuation and can only travel extremely short distances, the visual quality is often degraded due to turbidity, and spoken audio can be nearly incomprehensible due to the dampening effect of water.

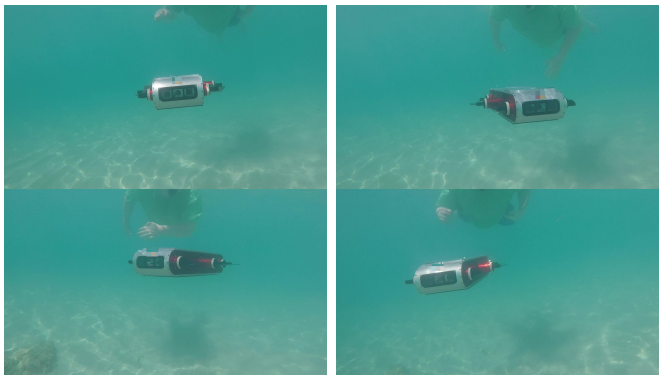


Figure 1: Clockwise from the top-left: An AUV from the Aqua family in the open ocean, producing a K_{No} kineme by shaking its "head".

The challenges of the underwater world limit the types of communication that can be applied there. Additionally, it is the opinion of the authors that methods of communication underwater should be natural and device-less, requiring little to no additional hardware, minimal training, and little mental effort on the part of the interactant. The reason for this is simple: diving is hazardous and mentally taxing. The diver must be aware of their environment, aware of their remaining oxygen, and remain focused on their task, all while carrying any equipment they require. Divers are therefore often at their physical and mental limits when underwater. If AUV communication adds to the diver's physical load by requiring them to carry an interaction-specific device, or if it adds to the diver's mental load by requiring intense focus, this will likely reduce the adoption of communicative AUV partners in underwater work and limit the effectiveness of the AUVs which are used. While the field of human-robot interaction research (HRI) in general contains a wide variety of works that propose methods of natural communication such as speech synthesis, text displayed on screens, and other options, such topics are less well explored underwater. Most AUVs tend to communicate with humans in limited and highly structured ways such as via small digital displays or wired controllers. These communication methods, while useful, fall far from the ease of use and simplicity of methods applied on terrestrial, humanoid robots, such as gestures and gaze cues, which have

not been fully implemented for AUVs or tested rigorously. Several ingenious methods for natural underwater human-robot communication have been proposed, such as implicit motion and light communication [7], which has only been evaluated in a single-participant case study. A form of "robot body language" [11] has also been proposed, but only evaluated in simulation. While the idea of using natural, human-like communication modes for AUVs is attractive, the question remains: can it be done?

To address this question, we implement an AUV "gesture" communication system which was previously proposed and evaluated in simulation [11], with modifications to the phrase-book which improve its viability in real-world operations. Called Robot Communication Via Motion (RCVM), this system maps meanings to specific, explicit robot motion phrases termed *kinemes* [19, 11] (noted in this paper as K_A for a kineme with the meaning "A"). An example of one of these kinemes is shown in Figure 1, where one can see the robot's back and forth yaw motions, which are mapped to the meaning "No", mimicing a human head-shake gesture.

In order to validate our implementation's effectiveness, we conduct a small pilot study (N=8, where N is the population size) which evaluates RCVM in a full-loop communication context. Following this, we perform a wide-reaching online study (N=121), comparing the RCVM system to three baseline communication systems: an audio text-to-speech (TTS) system, an LCD, and a system of blinking LEDs. In the results from this study, we compare the communication effectiveness of the kineme system to others in terms of the *viewpoint from which the interaction takes places* and the *content of the information communicated*. The results of this study demonstrate key strengths of RCVM over alternative communication methods: kinemes are easier to understand at challenging viewpoints and communicate simple concepts with human gesture equivalents such as "nodding" for a "Yes". In this paper, we make the following contributions:

- An improved set of RCVM kinemes, adaptable to any AUV with certain motion characteristics.
- The first implementation of RCVM, or any gestural system for an AUV, as well as the first studies of this implementation, including:
  - Results and analysis from an in-person pilot study testing kinemes in a full communication loop context.
  - Results and analysis from a wide-reaching online study comparing kinemes to three baseline systems.

II. RELATED WORK

A. Underwater Robot-to-Human Interaction

In this paper, we focus on the robot-to-human direction of communication, which is not widely addressed in the literature, leaving the study of the complementary human-to-robot communication to other work [16, 6]. Underwater robots have typically relied on digital displays [26, 6], purpose-built interaction devices [27], and lights [7] for robot-to-human communication. These methods all have drawbacks, however. Digital displays are often hard to understand at distance or at an angle; dedicated interaction devices add complexity, failure points, and have limited range. Emitted light is a promising method of communication which has only been explored minimally for underwater HRI [7], but flashing lights and color codes used may be difficult to learn and remember [11]. Motion-based communication may be the answer, as all robots have the capability to move; motion is observable at challenging viewpoints, and is a natural method of communication for humans. The original paper which proposed the RCVM system [11] showed that motion might be a viable method of communication through simulated trials, but no implementation of motion-based communication for AUVs has ever been produced or tested on physical robots. Therefore, we must look to other sub-fields of human-robot interaction to find examples of how motion can be used for communication.

B. Motion-Based Human-Robot Interaction

The use of motion as a vector of communication for robots has always been a topic of interest, as human communication emphasizes motions (conscious and unconscious) as modalities of communication secondary only to spoken language. However, the majority of motion-based communication research has been done on the use of motion for communication of affect (not information) and has been focused on humanoid and indoor robots rather than non-humanoid field robots. In terms of non-humanoid motion interaction, Bethel's seminal thesis [3] is one of the most important works in the field, using the angle and motion of non-humanoid search and rescue robots to communicate affect from the robots to the humans they are helping. This work (along with Bethel's surveys of non-facial, no-verbal affective communication [2, 4]) helped to establish motion communication for field robots as a viable option for displaying affect. Non-affective motion communication has also been explored, from the use of a pan-tilt camera to simulate head nodding and other gestures [28] to using a digitally displayed virtual "head" to generate gaze cues in order to manage navigational conflicts with humans [13]. A large body of work by Dragan et al. has covered the topics of legible pointing [14], nonverbal communication for feedback in teaching by demonstration [15], the expressiveness of timing in manipulation motion [32], and the effect of different types of robot motion on human-robot collaboration outcomes [8] for terrestrial robots, mostly in the realm of manipulators. These works all have some similarity to our study of RCVM, though we focus on informative communication (unlike Bethel [3]) via explicit motion gestures (unlike Dragan [14]), and use the base motion of a non-humanoid robot instead of adding human features (unlike Hart [13]).

III. ROBOT COMMUNICATION VIA MOTION

In order to facilitate robot-to-human communication for AUVs, we present the first AUV implementation of the Robot Communication Via Motion system. This implementation of RCVM, previously proposed for a simulated AUV [11], refines the language design and brings the system to fruition by

Interaction Phrase	Meaning	Type
Affirmative	Yes	Conversational
Attention	Look at me	Conversational
Danger	Danger nearby	Status
Follow	Follow me	Directional
Indicate Motion(L/R)	Move to the left/right	Directional
Indicate Object(L/R)	Object to the left/right	Directional
Stay	Remain where you are	Directional
Lost	The robot is lost	Status
Malfunction	Something is wrong	Status
Negative	No	Conversational
Repeat Last	Repeat last instruction	Conversational
Report Battery	Battery level is...[level]	Status

Table I: The interaction phrases used in this paper.

providing physical AUV implementations. In this section, we present a brief discussion of kineme design, our modifications to the RCVM phrasebook, the implementation of RCVM, and the three communication systems which will serve as a baseline comparison to RCVM for our wide-reach study, discussed in Section V.

A. Design of Kinemes

The design process for kinemes is currently relatively unstructured. Once the set of phrases to be communicated is chosen, the designer or design team selects those with a direct human equivalent and attempts to mimic that motion with the AUV. Following this, kinemes with a directional component are designed, using the orientation and position of the AUV as a key indicator of the information. Lastly, any remaining kinemes are designed by appealing to emotional relations and attempting to evoke those emotions. For instance, K_{Danger} attempts to evoke a sense of fear, as fear in humans is related to the concept of danger. During the process of updating the RCVM phrasebook, we realized that the three types of kineme designed (human-equivalent, directional, and emotional) had strong groupings in terms of their content. We therefore began to group them as such, giving rise to the groups shown in Table I: **Conversational**, **Directional**, and **Status**.

B. Phrasebook Modifications

The library of phrases presented in the original RCVM proposal contained a total of fifteen interaction phrases, the content of which were drawn from the authors' experience with underwater robot operations. We modified the phrasebook somewhat to remove unnecessary kinemes (such as $K_{Possible}$), clarify the meaning of others, and condense others into a single kineme. Additionally, we modified the actual motion used for various kinemes to be more achievable for a physical robot, as the original phrasebook was designed for a simulated robot. These modifications to the RCVM phrasebook streamline the language and make it more appropriate for implementation on a physical AUV. Our final phrasebook is technically comprised of 12 kinemes (found in Table I), but for $K_{IndicateObject}$ and $K_{IndicateMotion}$ we tested two versions (left and right) in our studies, to ensure that the kinemes work in multiple directions and to test for confusion between directions. It would be beneficial to test these kinemes with a variety of indicated directions/orientations, to ensure that all possible 3D orientations can be effectively communicated. However, to allow direct comparison to other kinemes as well as comparison to other communication modalities not as capable of 3D orientation representation (such as the LED system in Section III-D3), we only test a left and right version of the $K_{IndicateObject}$ and $K_{IndicateMotion}$ kinemes.

C. Implementation

The process of implementing RCVM for an AUV is challenging due to the difficult environments AUVs operate in. The kinemes must satisfy three requirements:**a)** be performed faithfully, regardless of environmentally created motion such as currents, waves, or motion of other swimmers near the robot, **b)** be sufficiently expressive without sacrificing AUV stability, and **c)** not modify the robot's overall position unnecessarily. The Aqua AUV (our implementation platform) uses a Proportional-Integral-Derivative (PID) control [18, Chapter 9.3] controller for its motion planning system (referred to as the "autopilot" system)[12]. The autopilot (which is implemented using ROS [22]) uses PID controller feedback loop based on local motion estimation which allows the kineme designer to request movement to specific angles or at certain velocities.

To address the first requirement, we made some small modifications to the Aqua autopilot to allow some flexibility in how precisely angles were targeted and added a timeout for abandoning a motion request if it was attempted without success for long enough. These modifications prioritize execution of a "fuzzier" version of the kineme in a short period of time rather than a higher fidelity execution over a longer duration. The second requirement was dealt with by re-tuning the PID controller slightly to prioritize a fast approach to the target angles, allowing for some overshooting. Finally, our redesigns for certain kinemes prioritized keeping the AUV closer to its original position, to enable further communication more easily. Once we had completed these three refinements to improve robustness of field operations, we implemented RCVM as a ROS node which provided a number of services, endpoints which can be called by other nodes to trigger any of the kinemes in our phrasebook. Whenever it receives a request for a kineme execution, the RCVM node utilizes the modified autopilot API to produce the requested motions in the AUV. This modular design, somewhat decoupled from the autopilot code and utilizing services, makes it easy for RCVM to be ported to a new AUV (by updating the service callbacks) without modifying any ROS nodes which request RCVM services. The code comprising this implementation will be released for public use (not currently released for double blind review).

D. Baseline Systems

To provide an understanding of the place which RCVM occupies in underwater communication systems, we also developed three baseline systems for comparison. Demonstration

of each of the baseline systems, along with RCVM, can be found in the complementary video submitted with this paper.

1) TTS System: The text-to-speech (TTS) based communication system is quite simple: for each interaction phrase, a Bluetooth Speaker plays a short phrase that represents the interaction phrase's meaning, using English text-to-speech audio retrieved from Google's Text To Speech API. The speaker is mounted pointing up or down, not directly towards the interactant. We used a Kinps SoundCircular speaker, rated IPX8 for full immersion in water.

2) LCD System: The LCD communication system is similar to the TTS system in that the design simply involves expressing an English phrase on the designated device. In this case for each interaction phrase, a short text which represents the meaning is displayed for five seconds on a two-line, sixteen character-per-line backlit liquid crystal display, driven by an Arduino. The screen's display is white text on a blue background, which makes it particularly good for reading underwater compared to other displays, but it has poor viewing angles and is difficult to see from any distance.

3) LED System: The LED communication system was used as a baseline for kinemes in the original RCVM study [11]. In that work, a set of nine LEDs was illuminated in different ways to represent different interaction phrases. The set of light colors and timings used for each interaction phrase is termed a "light code" or "LED code". We improved upon this LED system by reducing the number of required LEDs to 3 and simplifying light codes. This system was implemented by using an Arduino to drive a series of three RGB LEDs.

Interaction Phrases Described In Appendix

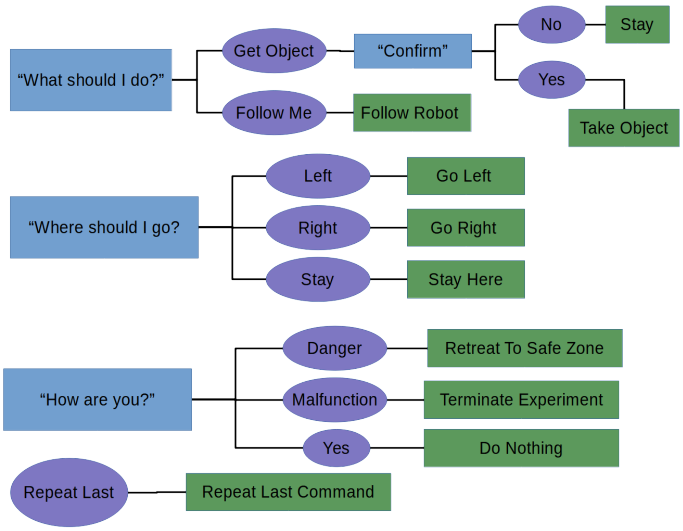
For brevity's sake, full descriptions of each system's interaction phrases are omitted from this paper. The Appendix provided with this article contains text descriptions of each kineme, LED code, LCD display, and audio phrase, along with URLs for video playlists of each system's interaction phrases.

IV. IN-PERSON PILOT STUDY

In order to validate this new physical implementation of RCVM, we performed a small pilot study with 8 participants, training them in the use of the RCVM system and then testing their ability to complete full-loop interactions. Given the small population size and complexity of the setup, this study was not intended as a definitive quantitative measurement of kineme effectiveness for communication. Rather, it serves as a confirmation that RCVM still works when translated to a physical robot, as well as when embedded in a full communication loop. While some quantitative measures are recorded, the study as a whole is treated as preliminary to more definitive experiments described in Section V. This study was determined to be Not Human Research, exempting it from IRB oversight (reference numbers: 00004699, 00004700).

A. Study Design

All participants were provided with educational material before the study began, to be completed at their own pace.



(a) Flow chart depicting the interaction process.

(b) Key: Boxes are gestures (blue) or actions (green) of the human, while the purple ovals are kinemes generated by the robot.

Figure 2: A flow chart of the pilot study's interaction loop, as described in Section IV-A.

The educational material familiarized them with the study layout and flow (using the image in Fig. 2), the gestures they would use for input, and videos of the kinemes on a simulated AUV (generated with the Gazebo simulator using our implementation). Participants were told to ask the Aqua AUV one of three questions, pay attention to the robot's response delivered via kineme, and take the appropriate action. The question asked was communicated using a set of gestures based on relevant American Sign Language signs for 'What should I do?', 'Where should I go?', and 'How are you?', with another sign for 'Confirm'. The users were told that Aqua would observe their gesture and automatically select a kineme to display. However, the kinemes were actually manually selected pseudo-randomly in secret by study staff. This selection was done manually in order to easily balance the number of each kineme shown throughout the study, and because gesture recognition in underwater environment is a challenging problem on its own. Therefore, the study is a Wizard-of-Oz study, where the study staff operate some aspect of the robot's function without the knowledge of the user. Once the robot displayed a kineme, the participant verbally identified the action they would take, appropriate to the response they received. If their selected action was correct, the interaction was marked as completed correctly. If the participant forgot the correct action, leading to them taking an incorrect action or simply refusing to take an action, the interaction was marked as a failure. Therefore, the recorded "Accuracy" of each kineme is more appropriately considered the success rate

System	Accuracy	Avg. Time	Avg. Conf.
Affirmative	**95.00**%	21.25	**9.00**
Negative	80.00%	**20.85**	8.10
Follow Me	50.00%	30.66	6.62
Indicate Motion (Left)	88.46%	24.77	7.54
Indicate Motion (Right)	86.36%	23.18	8.10
Indicate Object (Left)	57.89%	28.21	7.48
Indicate Object (Right)	52.63%	35.42	6.32
Indicate Stay	91.30%	23.78	8.60
Danger	64.29%	32.18	6.42
Malfunction	77.27%	29.09	7.00
Repeat Previous	25.00%	29.00	5.58
Total Avg. (Ours)	60.00%	34.75	7.34
Avg.@Highest Edu. Level ([11])	*85.00%*	*11.00*	*8.00*

Table II: Per-kineme results from the pilot study, along with the total average results compared to the two most appropriate education groups from [11].

of interactions including that kineme. After each interaction loop, users were asked for a confidence between one and ten (ten being high) on how accurately they had conducted their interaction. This confidence was recorded, along with the total time of the interaction loop and the sequence of question to kineme to action. For each robot, participants completed an interaction session composed of between ten and fifteen interactions. Two kinemes (K_{Lost} and $K_{ReportBattery}$) were not tested in this study, as they were under consideration for elimination at the time of the study.

B. Results

The results from our pilot study (Table II) show that RCVM continues to function relatively well when implemented on a physical AUV and placed in a full-loop context. While our overall accuracy is less than that of the best-trained participants from [11], we see several high accuracies on certain kinemes, particularly $K_{Affirmative}$, $K_{Negative}$, K_{Stay}, and $K_{IndicateMotion}$. We note that average times increased, but since our measurement of duration included the entire interaction loop, this is of less significance. Given the small sample size, it is difficult to draw any statistically significant results from this data, but it serves to validate the in-person performance of our kineme system: kinemes implemented for the Aqua robot achieved similar levels of accuracy to the original simulated kineme system that we are implementing. In addition, RCVM operated effectively as a part of a full interaction loop. Further evaluation of RCVM in a full communication loop, including analysis of the cognitive load placed on interactants, would be interesting to explore. However, such an evaluation would be premature, without an understanding of how RCVM operates in comparison to other communication modalities, purely in terms of recognition accuracy.

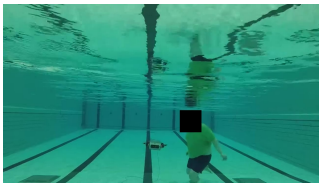
V. WIDE-REACHING ONLINE STUDY

With the kineme pilot study completed, validating the baseline performance of physically implemented AUV kinemes, we turn our attention to the task comparing RCVM to other viable underwater communication systems. Understanding the comparative performance of RCVM to other robot-to-human communication options is critical in determining how to utilize RCVM in actual interaction loop designs, for both further research and field applications. To achieve this understanding of RCVM's performance compared to other options, we designed a study comparing RCVM to the baseline systems described in Section III-D in terms of their efficacy from a variety of viewpoints and when communicating different types of information. Our participants, recruited using Amazon Mechanical Turk, were trained to use a randomly selected communication system, then tested on that system from a randomly selected viewpoint. This study design was submitted to the relevant Institutional Review Board for review and was determined to be Not Human Research, exempting it from IRB oversight (reference number: 00004695).

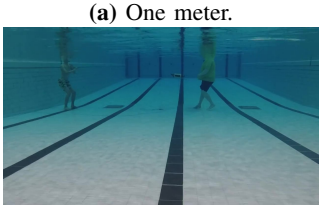
A. Viewpoints

To prepare videos for our survey, each communication system was recorded from a variety of viewpoints. These viewpoints (illustrated in Fig. 3) were 3 meters, 8 meters, 3 meters at an angle of 45 °, and 3 meters at an angle of 90 °. In addition, an ideal viewpoint for each communication system was recorded to be used for the training, as well as being a possible viewpoint condition for testing. **This viewpoint, referred to as the EDU viewpoint, is defined as the closest distance to the robot at which all portions of every interaction phrase are visible, minimum of 1 meter**. Therefore, the EDU viewpoint is a distance of 1 meter for all of the baseline systems, but a 5 meter distance for the kineme system, because some kinemes must be viewed from a slight distance in order to see all of the movement. As each viewpoint is not viable for all systems, some were not tested; *e.g.*, the LCD screen cannot be seen at all from a 90-° angle for instance, so accuracy should be close to 7%, as all attempts to identify the fourteen interaction phrases will essentially be random guesses. Additionally, the TTS systems should return similar results at the 3m viewpoints as the 45° and 90° viewpoints, as those viewpoints are also at a three meter distance and audio propagates evenly regardless of angle, given the position of our speaker. The viewpoint and robot/domain combinations which we tested can be seen in Table III along with the number of participants in each condition.

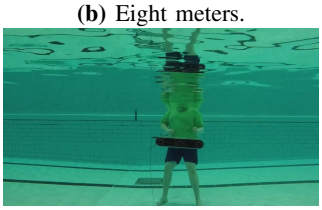
The viewpoints selected for this study represent a good sample of the possible distances and orientations viable for visual communication between an AUV and a diver. The orientations were selected by the assumption that divers and AUVs attempting to communicate with one another will be at least within a 90 degree orientation from one another. Distances were selected by considering the possible ranges of communication in the field. Visibility varies by levels of particulate matter, algae blooms, ambient light level, and depth. While visibility distances in the field range both higher and lower than the 1m-8m range tested in this study, this set of distances allows us to observe the effect of distance on communication, at distances that are realistic for a deployment (assuming that visibility is greater than 8m). Lower visibility



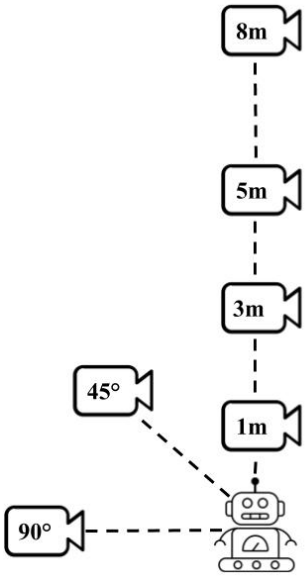
(a) One meter.



(b) Eight meters.



(c) Ninety degrees.



(d) Diagram of viewpoints

Figure 3: The viewpoints captured for this study, illustrated in a three examples(a-c) and a diagram(d).

distances than 8m would obviously reduce the effectiveness of all but the TTS system to nearly nothing past the distance of visibility, but diver-to-diver communication via hand signals would experience the same effect.

B. Video Recording

To capture videos of the systems at every viewpoint, several recording sessions were completed, using two GoPro™ HERO5 Black cameras, using a 1080p resolution in the linear aspect ratio at 24 frames per second. Due to scheduling constraints, the TTS system had to be recorded in a different pool than the rest of the Aqua systems. However, visual clips from the original pool were layered over the audio from the other pool to maintain a visual similarity. Due to this, the speaker is not visible in the video clips depicting the TTS system, but as the TTS system's communication is done over audio, this is not expected to impact results.

C. Population Recruitment and Statistics

The participants recruited on Mechanical Turk were required to be in the United States, have never taken the survey before, and have completed more than 5000 Human Intelligence Tasks (HITs) with an approval rating of 97% or greater. The HIT posted on Mechanical Turk paid participants $1.50 and was estimated to take an average of 12-15 minutes, meaning that users would be paid approximately minimum wage. Submissions were typically approved within 24-48 hours of completion. The only criteria for accepting Mechanical Turk work and paying the worker was that users spend an amount of time on the educational page equal to at least 25% of the duration of the education video. Users who spent less than a quarter of the video's duration before continuing were considered to have made a bad-faith effort and were rejected. They were, however, paid $0.75 (half-pay) for their time. The line for inclusion in the dataset was set higher: only users who spent at least 75% of the video duration on the page were included in the analyses.

		Viewpoint				
System	**EDU**	**3m**	**8m**	**45°**	**90 °**	**Total**
KINEME	7	8	7	7	7	36
TTS	8	10	10	0	0	28
LCD	7	8	0	6	0	21
LED	7	8	7	7	7	36
TOTAL	29	34	24	20	14	121

Table III: Study conditions marked by whether (green) or not (red) they are tested, with participant numbers for each.

The resultant population was relatively diverse. We surveyed 130 participants, with 9 of them being excluded from all analysis due to short education video watch times. The final population of 121 participants was 62% male, 38% female, middle-aged ($M = 37$, $SD = 21$), had a variety of education levels (42.1% had a bachelor's degree), came from 35 of the 50 states of the US, and were mostly employed (93.4%).

This study population is a decent sample of the US population, which was the target sample. Participants were not explicitly asked if they had diving experience. While divers are most likely to use the RCVM system and evaluation on a sample drawn specifically from experienced divers would be useful, all of the communication systems evaluated in this study are designed to be easily interpreted and understood with little training. Thus, studying the performance of our communication systems on a sample from the general population helps us to evaluate the performance of these systems with minimal effects of prior knowledge of diving and underwater communication on that performance.

D. Education and Distraction Procedure

Once participants entered the survey and passed a bot check, they were randomly assigned a condition and directed to an education procedure. The education procedure for each condition consisted of a video composed of the fourteen interaction phrases in a set order from the EDU (ideal) viewpoint of the robot and system of the participant's condition. Participants were asked to watch the entire video without skipping around and warned that their payment would depend on watching the entire video; however, they were permitted to leave the page at any time. As previously mentioned, only participants who spent at least 75% of the duration of their educational video on the page were included in the analysis. This education level falls somewhere between the two highest education levels described in [11](only meanings shown and meanings shown along with videos of kinemes), as the participant is shown both the communication system displaying an interaction phrase as well as the meaning of the phrase. However, since the participant was not being taught directly by study staff, the education provided in this study likely falls below the accuracy at the highest education level from [11]. Following

		Accuracy		Op. Accuracy		Confidence		Time (s)	
	System	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)	Mean (μ)	SD (σ)
OVERALL	KINEME	28.4	22.0	**43.7**	33.1	4.5	1.9	39.9	15.7
	TTS	21.2	24.8	29.4	40.2	4.4	2.6	**20.6**	9.3
	LCD	29.6	40.4	32.4	44.0	**4.9**	3.5	25.0	11.2
	LED	**35.3**	27.7	34.7	32.5	4.7	2.4	25.6	11.1
AT EDU	KINEME	41.8	13.3	77.7	18.9	4.5	1.7	44.7	10.0
	TTS	51.8	25.3	77.1	34.4	6.1	1.8	20.6	9.0
	LCD	**82.7**	21.8	**91.6**	15.2	**7.4**	1.9	**19.3**	2.6
	LED	70.4	23.8	61.6	36.8	6.9	2.7	30.8	15.4

Table IV: Mean and standard deviation of communication system metrics, averaged over all viewpoints and at the ideal viewpoint for each system. Bold values are the best (min/max respectively) mean for metric in group (overall or at EDU).

the education procedure, participants were asked to solve five ninth-grade level mathematics questions as a distraction procedure. Distraction procedures are a common method in psychology research used to induce forgetfulness in subjects. Some examples can be found in the 1950's memory research of Brown [5] and Peterson [20], or other more recent work on working memory by Waris et al. [29]. In our work, a distraction procedure is used to separate the training and testing phases of the study so that participants are less likely to be able to hold the entirety of the training they just completed in their short-term memory.

E. Testing and Evaluation

Each participant was shown videos of all 14 interaction phrases in a random order, using the communication system from the viewpoint indicated in their condition. Some received the EDU viewpoint as their testing viewpoint, so they had the same viewpoint for education and testing. For each video, the participant was shown a video hosted on YouTube™ and asked to select its meaning from a drop-down menu. Participants were also given the option to select "Unable to select meaning". If a meaning was selected, they were then asked what their confidence in their choice was, on an ordinal scale from 1 to 10 (10 being the most confident). Otherwise, the confidence question was not presented, and they were instead asked what made them unable to make a selection: forgetting the meaning, being unable to see the interaction phrase, the survey not displaying the video, or some other issue. The time from when the participant entered the webpage to when they left it was recorded, though participants were not told that it would be. Once participants had completed all of their videos, they were debriefed and given information on how to submit their responses for certification.

VI. RESULTS

A. Metrics

We use the same metrics to measure system efficacy that were used in the original RCVM study [11]: **accuracy**, **operational accuracy**, **confidence**, and **time to answer**. Accuracy is the correctness of a participant's answer, ranging from 0 to 100. Operational accuracy is the same metric but only considering answers also rated a 5 or higher in confidence, to simulate the answers that a user would be likely to act upon. Confidence and time to answer are simply the values recorded from the confidence question (0-10) and the time participants took to select a meaning for a video in seconds. Time to answer data was processed to remove outliers by discarding values greater than 150 seconds. This was set as the cutoff because for all interaction phrases, 95% of responses had a time to answer lower than 150 seconds (mean 95th percentile was 73.28 seconds). Durations of these outlier answers greatly exceeded 150 seconds (*e.g.*, 500 seconds or greater), which suggests that the webpage was left open while the participant briefly did something else.

The two metrics upon which we will perform statistical analysis are accuracy and operational accuracy. Shapiro-Wilk [25] tests were performed for accuracy $W = 0.84$, $p < .001$ and operational accuracy $W = 0.84$, $p < .001$, both finding evidence that data was not normally distributed, and direct observation of the data confirmed this. Due to this finding, we will perform the following hypothesis testing using the Kruskal-Wallis H-test [17]. Kruskal-Wallis is also referred to as one-way ANOVA on ranks and is a non-parametric equivalent to one-way ANOVA which does not assume a normal distribution of data. Tests were run at a confidence level of 99%, a significance of $\alpha = 0.01$.

B. Internal Validity

The Kruskal-Wallis H-test found no significant relationship between the percentage of their education video that a participant watched and their average accuracy in the testing phase $H(5) = 2.89$, $p = .717$. Further, we found no significant relationship between accuracy and gender $H(1) = 0.10$, $p = .750$. A correlation test using Spearman's method detected no significant correlation between accuracy and age $r(119) = -0.145$, $p = .112$ No threats to internal validity were detected, but participant recognition accuracy was considerably lower for all systems than expected.

The kinemes in the pilot study achieved 60% accuracy, while the highest RCVM accuracy achieved in this study was 41.8%, at the ideal viewpoint. This may be due to the fact that the study was conducted online via video, or due to low education absorption. However, the performance of the kinemes and LED systems at the EDU viewpoint is generally between the accuracy level reported at education levels 1 and 2 in [11], which is consistent with our education

	Viewpoint	Accuracy		Op. Accuracy		Confidence		Time (s)	
		Mean (μ)	**SD (σ)**	**Mean (μ)**	**SD (σ)**	**Mean (μ)**	**SD (σ)**	**Mean (μ)**	**SD (σ)**
KINEME	EDU	**41.8**	13.3	**77.7**	18.9	4.5	1.7	44.7	10.0
	3m	25.9	25.3	38.4	38.0	**5.4**	2.4	45.8	16.6
	8m	35.7	26.4	44.4	31.9	4.3	2.3	41.8	8.1
	45°	12.2	7.9	20.3	25.7	4.1	1.3	34.4	16.5
	90°	26.5	23.2	38.3	25.1	3.7	1.4	**32.0**	22.2
TTS	EDU	**51.8**	25.3	**77.1**	34.4	**6.1**	1.8	20.6	9.0
	3m	12.1	10.7	8.5	10.5	4.6	2.8	**19.9**	10.4
	8m	5.7	6.6	12.1	31.2	2.9	2.2	21.4	9.4
LCD	EDU	**82.7**	21.8	**91.6**	15.2	**7.4**	1.9	**19.3**	2.6
	3m	2.7	3.7	3.0	6.1	3.0	3.1	33.0	10.8
	45°	3.6	6.0	2.4	5.8	4.6	3.9	21.1	12.8
LED	EDU	**70.4**	23.8	**61.6**	36.8	**6.9**	2.7	30.8	15.4
	3m	43.8	18.9	41.1	30.6	4.6	2.0	22.8	10.0
	8m	14.3	17.0	14.9	19.5	4.3	2.1	27.3	10.0
	45°	33.7	20.9	40.0	26.6	5.4	1.7	27.8	10.1
	90°	13.3	12.7	15.0	27.8	2.4	1.7	**19.9**	8.8

Table V: Mean and standard deviation of communication system metrics, for all evaluated viewpoints. Bold values are the best (min/max respectively) mean for metric in system group.

procedure's expected success (as mentioned in Section V-D). Higher accuracy at all viewpoints could likely be achieved by administering an education procedure which would train users until a certain competency level is reached. However, we believe the general trend of results to be correct, and statistically large effects should persist in in-person testing, as our Pilot study demonstrates similar accuracy levels to [11] for in-person testing.

C. Overall Results

When considering differences between the four communication systems we tested, we are most interested in the effects on accuracy and operational accuracy. When considered over all viewpoints, none of the systems tested have statistically significant differences in accuracy $H(3) = 7.60$, $p = .055$ or operational accuracy $H(3) = 4.27$, $p = .234$. Due to the challenging nature of the underwater environment and the viewpoints at which they were tested, none of these systems have achieved high accuracy overall. However, when we consider the accuracy of these systems at different viewpoints, we see significant differences, despite the overall low accuracy, which indicate how RCVM performs compared to other communication options underwater (and often outperforms them).

D. Viewpoint Comparisons

The effect that viewpoint has on the accuracy of tested systems can be found in Table V. While we see RCVM accuracy is the lowest of any system at the EDU viewpoint, once we move to the more challenging viewpoints, the TTS and LCD systems become entirely non-competitive, with accuracies near to that of a random guess (7%).

Kruskal-Wallis tests show that viewpoint has a statistically significant effect on every communication system, **with the exception of the kineme system** $H(4) = 8.94$, $p = .063$. The effect is most significant with the TTS $H(2) = 14.38$, $p < .001$ and LCD $H(2) = 14.71$, $p < .001$ systems, but is also present for the LED system $H(4) = 20.99$, $p < .001$. Considering the values shown in Table V, it is apparent that non-EDU viewpoints reduce accuracy significantly for each of these systems. We also test operational accuracy with Kruskal-Wallis tests, which show that viewpoint affects operational accuracy for the LCD system $H(2) = 15.87$, $p < .001$ and the TTS system $H(2) = 10.75$, $p = .005$. However, there is no statistically significant difference in operational accuracy by viewpoint for the kineme system $H(4) = 10.78$, $p = .029$ or the LED system $H(4) = 10.33$, $p = .035$.

To summarize, through statistical testing we find that TTS and LCD communication begin to fail quickly at any challenging interaction viewpoint, while **LED and Kineme communication are more viewpoint-invariant, particularly kinemes**. Since the accuracy of the kineme system is above that of a random guess (7% accuracy) at challenging viewpoints, we suggest that this shows that kinemes are more viewpoint invariant than other systems, though the LED system is a strong competitor.

E. Content Comparisons

We chose to study the effect of message content because message content is known *a priori*, meaning that if a communication system shows an affinity for communicating certain types of messages, we can autonomously switch to using that system to get the message across most effectively. For this experiment, we consider three categories of our interaction phrases (see Section III-A and Table I): Conversational, Directional, and Status phrases. In Table VI, we can see that kinemes and LEDs have similar accuracies for Conversational and Status interaction phrases, but the accuracy of LEDs is higher for Directional phrases. This is unexpected given the spatial nature of kinemes, but analysis of kineme identifications suggests that participants may have been confused as to

		Accuracy		Op_Acc.		Confidence		Time (s)	
	Phrase Content	**Mean (μ)**	**SD (σ)**	**Mean (μ)**	**SD (σ)**	**Mean (μ)**	**SD (σ)**	**Mean (μ)**	**SD (σ)**
KINEME	Conversational	**35.0**	28.8	**45.5**	38.8	**4.8**	2.2	**37.5**	18.2
	Directional	23.3	27.3	29.3	36.3	4.3	2.2	41.7	20.7
	Status	23.9	21.3	42.2	41.3	4.1	2.1	42.0	17.2
TTS	Conversational	17.9	27.4	26.2	42.2	4.2	2.6	20.3	12.3
	Directional	**23.6**	27.2	**30.5**	41.7	**4.6**	2.9	**19.7**	11.0
	Status	22.9	30.2	29.4	43.1	4.5	2.7	21.7	11.9
LCD	Conversational	28.6	42.7	31.0	45.3	**5.0**	3.5	**22.2**	10.4
	Directional	**32.4**	38.7	**34.5**	42.8	**5.0**	3.6	28.0	14.6
	Status	28.6	42.7	32.4	47.1	4.7	3.5	25.5	14.1
LED	Conversational	**40.6**	31.9	38.8	39.3	4.8	2.6	28.7	16.1
	Directional	**40.6**	33.6	**39.8**	38.0	**5.0**	2.7	**24.0**	13.2
	Status	25.0	26.3	23.0	33.1	4.5	2.6	24.6	11.6

Table VI: Mean and standard deviation of communication system metrics, separated by phrase content. Bold values are the best (min/max respectively) mean for metric in system group.

whether "left" referred to their left or the robot's left, and vice versa for right. A Kruskal-Wallis test does not show a significant effect on the accuracy of Directional phrases based on which communication system is used $H(3) = 6.40$, $p = .094$, however. Simply observing the higher accuracy of the LED system suggests that LED codes may be more effective in expressing directional information than kinemes, but this difference is not statistically significant.

No communication system is found to have statistically higher accuracy than others when considering the accuracy of Status phrases $H(3) = 1.48$, $p = .686$, but system type does have a significant effect on the accuracy Conversational phrases $H(3) = 11.61$, $p = .009$. Post-hoc analysis with Dunn tests using a Bonferroni-adjusted alpha level of 0.0017 $(0.01/6)$ was used to compare pairs of systems. No comparisons were significant after Bonferroni adjustment (all ps $> .012$), but the TTS system under-performs both the kineme and LED systems on Conversational phrases. This may be due to the fact that Conversational phrases are typically short for the TTS system, meaning that they might be entirely missed or easily misidentified at challenging viewpoints, while the longer phrases of the kineme and LED systems provide more opportunity to understand the phrase.

To summarize, while we had hoped that Directional accuracy of the kineme system would be high, no statistically significant effect is detected for Directional accuracy based on system type. Not effect is detected for Status phrases either, but Conversational accuracy is affected by system type (though no system is shown to be better than others in post-hoc analysis).

VII. CONCLUSION

In this paper, we presented the first physical implementation of RCVM, an explicit motion-based human-robot communication system for AUVs along with a short evaluation of its effectiveness in a full communication loop and an extensive analysis of performance compared to other communication systems. After modifying and refining the phrasebook from the original proposal of RCVM (which contained some unrealistic kinemes), we implemented the system using ROS and a modified version of the Aqua AUV's PID motion controller. To test our implementation, we performed a small in-person pilot study, in which we tested RCVM in a full-loop communication scenario, and found it to be sufficiently effective for further study. Following this we performed a large online study, comparing RCVM to three baseline systems we produced in terms of their efficacy at different viewpoints. Our results suggest that while RCVM is not the most accurate in the ideal conditions for any given system, it does perform better at challenging viewpoints than our LCD and TTS systems, and competes well with our new LED system (improved from the baseline system in [11]). Additionally, we find that RCVM does not outperform other systems in communicating directional information, as initially expected. RCVM and the LED systems are more accurate than the TTS and LCD systems for Conversational information, though not to the point of statistical significance. Based on our results, it seems that while RCVM would likely be insufficient for robust communication on its own, it has strengths that other underwater communication options do not have, and should be integrated with other communication systems. In future work, we plan to do just that, integrating RCVM with a system which is capable of autonomously selecting a communication strategy based on information about the interaction context it is in. Kinemes will be a key part of this system, to be used to communicate information when interaction viewpoints are challenging, as this is when RCVM thrives.

ACKNOWLEDGEMENT

This work was supported by the US National Science Foundation awards IIS-#184536 & #00074041, and the MnRI Seed Grant. This research was conducted when Owen Queeglay served as a volunteer at the UMN CS&E and Interactive Robotics and Vision Lab. The authors wish to thank all members of the Interactive Robotics and Vision Lab, particularly Khiem Vuong, Chelsey Edge, and Jungseok Hong for their input.

References

[1] Mark Allison, Heather Dawson, and Grant Rusin. Towards an AUV swarm based mobile underwater sensor network for invasive species data acquisition. In *2018 4th International Conference on Universal Village (UV)*, pages 1–4, 2018. doi: 10.1109/UV.2018.8642151.

[2] C. L. Bethel and R. R. Murphy. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(1):83–92, January 2008. ISSN 1094-6977. doi: 10.1109/TSMCC.2007.905845.

[3] Cindy L Bethel. Robots without faces: Non-verbal social human-robot interaction. PhD Thesis, University of South Florida, 2009.

[4] Cindy L. Bethel and Robin R. Murphy. Non-facial and non-verbal affective expression for appearance-constrained robots used in victim management*. *Paladyn, Journal of Behavioral Robotics*, 1(4):219–230, 2011. ISSN 2081-4836. doi: 10.2478/s13230-011-0009-5.

[5] John Brown. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10(1):12–21, 1958. doi: 10.1080/17470215808416249.

[6] Arturo Gomez Chavez, Christian A. Mueller, Tobias Doernbach, Davide Chiarella, and Andreas Birk. Robust gesture-based communication for underwater human-robot interaction in the context of search and rescue diver missions. In *Workshop on Human-Aiding Robots, International Conference on Intelligent Robots and Systems (IROS)*, October 2018.

[7] K. J. DeMarco, M. E. West, and A. M. Howard. Underwater human-robot communication: A case study with human divers. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3738–3743, October 2014. doi: 10.1109/SMC.2014.6974512.

[8] Anca D. Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S. Srinivasa. Effects of robot motion on human-robot collaboration. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 51–58, Portland, Oregon, 2015. ACM/IEEE.

[9] Mike Eichhorn, Christoph Ament, Marco Jacobi, Torsten Pfuetzenreuter, Divas Karimanzira, Kornelia Bley, Michael Boer, and Henning Wehde. Modular AUV system with integrated real-time water quality analysis. *Sensors*, 18, 2018. ISSN 1424-8220. doi: 10.3390/s18061837.

[10] Brendan P. Foley, Katerina Dellaporta, Dimitris Sakellariou, Brian S. Bingham, Richard Camilli, Ryan M. Eustice, Dionysis Evagelistis, Vicki Lynn Ferrini, Kostas Katsaros, Dimitris Kourkoumelis, Aggelos Mallios, Paraskevi Micha, David A. Mindell, Christopher Roman, Hanumant Singh, David S. Switzer, and Theotokis Theodoulou. The 2005 Chios Ancient Shipwreck Survey: New Methods for Underwater Archaeology. *Hesperia: The Journal of the American School of Classical Studies at Athens*, 78(2):269–305, 2009. ISSN 0018098X, 15535622.

[11] M. Fulton, C. Edge, and J. Sattar. Robot communication via motion: Closing the underwater human-robot interaction loop. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4660–4666, May 2019. doi: 10.1109/ICRA.2019.8793491.

[12] P. Giguere, Y. Girdhar, and G. Dudek. Wide-speed autopilot system for a swimming hexapod robot. In *2013 International Conference on Computer and Robot Vision*, pages 9–15, 2013. doi: 10.1109/CRV.2013.13.

[13] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. Using human-inspired signals to disambiguate navigational intentions. In *Proceedings of the 12th International Conference on Social Robotics (ICSR)*, Golden, Colorado, November 2020. Springer Publishing.

[14] Rachel Holladay, Anca Dragan, and Siddhartha Srinivasa. Legible robot pointing. In *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication*, volume 2014. IEEE, August 2014. doi: 10.1109/ROMAN.2014.6926256.

[15] I Huang, R Pandya, and AD Dragan. Nonverbal robot feedback for human teachers. In *Conference on Robot Learning (CoRL)*, 2019.

[16] M. J. Islam, M. Ho, and J. Sattar. Dynamic reconfiguration of mission parameters in underwater human-robot collaboration. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8, May 2018. doi: 10.1109/ICRA.2018.8461197.

[17] William H. Kruskal and W. Allen Wallis. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952. ISSN 0162-1459. doi: 10.2307/2280779. URL https://www.jstor.org/stable/2280779. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[18] Norman S. Nise. *Control Systems Engineering*. Wiley, Hoboken, New Jersey, 7 edition, 2015. ISBN 1118170512,9781118170519.

[19] W. Nöth. *Handbook of Semiotics*. Advances in semiotics. Indiana University Press, 1995. ISBN 978-0-253-20959-7.

[20] Lloyd Peterson and Margaret Jean Peterson. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3):193–198, 1959. ISSN 0022-1015(Print). doi: 10.1037/h0049234. Place: US Publisher: American Psychological Association.

[21] Yvan R. Petillot, Scott R. Reed, and Judith M. Bell. Real time AUV pipeline detection and tracking using side scan sonar and multi-beam echo-sounder. In *OCEANS '02 MTS/IEEE*, pages 217–222 vol.1, Oct 2002. doi: 10.1109/OCEANS.2002.1193275.

[22] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y

Ng. ROS: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009.

[23] Kim R. Reisenbichler, Mark R. Chaffey, Francois Cazenave, Robert S. McEwen, Richard G. Henthorn, Robert E. Sherlock, and Bruce H. Robison. Automating MBARI's midwater time-series video surveys: The transition from ROV to AUV. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–9, 2016. doi: 10.1109/OCEANS.2016.7761499.

[24] Sanem Sariel, Tucker Balch, and Jason Stack. Distributed multi-auv coordination in naval mine countermeasure missions. *Tech Report – Georgia Institute of Technology*, 2006.

[25] S. S. Shapiro and M. B. Wilk. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52 (3/4):591–611, 1965. ISSN 0006-3444. doi: 10.2307/2333709. URL https://www.jstor.org/stable/2333709. Publisher: [Oxford University Press, Biometrika Trust].

[26] Y. Ukai and J. Rekimoto. Swimoid: Interacting with an underwater buddy robot. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 423–423, March 2013. doi: 10.1109/HRI.2013.6483628.

[27] Bart Verzijlenberg and Michael Jenkin. Swimming with robots: Human robot communication at depth. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4023–4028, Taipei, Taiwan, 2010. IEEE/RSJ. doi: 10.1109/IROS.2010.5652751.

[28] J. Wainer, D. J. Feil-seifer, D. A. Shell, and M. J. Mataric. The role of physical embodiment in human-robot interaction. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 117–122, September 2006. doi: 10.1109/ROMAN.2006.314404.

[29] Otto Waris, Anna Soveri, Miikka Ahti, Russell Hoffing, Daniel Ventus, Susanne Jaeggi, Aaron Seitz, and Matti Laine. A latent factor analysis of working memory measures using large-scale data. *Frontiers in Psychology*, 8, 06 2017. doi: 10.3389/fpsyg.2017.01062.

[30] Stefan B. Williams, Oscar Pizarro, Michael Jakuba, and Neville Barrett. AUV Benthic Habitat Mapping in South Eastern Tasmania. In Andrew Howard, Karl Iagnemma, and Alonzo Kelly, editors, *Field and Service Robotics*, pages 275–284, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-13408-1.

[31] Russell B. Wynn, Veerle A.I. Huvenne, Timothy P. Le Bas, Bramley J. Murton, Douglas P. Connelly, Brian J. Bett, Henry A. Ruhl, Kirsty J. Morris, Jeffrey Peakall, Daniel R. Parsons, Esther J. Sumner, Stephen E. Darby, Robert M. Dorrell, and James E. Hunt. Autonomous underwater vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience. *Marine Geology*, 352:451–468, 2014. ISSN 0025-3227. doi: 10.1016/j.margeo.2014.03.012.

[32] Allan Zhou, Dylan Hadfield-Menell, Anusha Nagabandi, and Anca D. Dragan. Expressive robot motion timing. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, pages 22–31. Association for Computing Machinery, 2017. ISBN 978-1-4503-4336-7. doi: 10.1145/2909824.3020221.