2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

Robot Communication Via Motion: Closing the Underwater Human-Robot Interaction Loop

Michael Fulton[1], Chelsey Edge[2], Junaed Sattar[3]

Abstract— **In this paper, we propose a novel method for underwater robot-to-human communication using the motion of the robot as "body language". To evaluate this system, we develop simulated examples of the system's body language gestures, called kinemes, and compare them to a baseline system using flashing colored lights through a user study. Our work shows evidence that motion can be used as a successful communication vector which is accurate, easy to learn, and quick enough to be used, all without requiring any additional hardware to be added to our platform. We thus contribute to "closing the loop" for human-robot interaction underwater by proposing and testing this system, suggesting a library of possible body language gestures for underwater robots, and offering insight on the design of nonverbal robot-to-human communication methods.**

I. INTRODUCTION

In the United States of America as of May 2017, 3,280 people are employed as commercial divers, tasked with dangerous and difficult tasks underwater [1]. The presence of a capable Autonomous Underwater Vehicle (AUV) as a partner to assist in data collection, monitoring, search and rescue, or maintenance tasks has the potential to increase efficiency and ensure the safety of the diver [2]. In order for an AUV to be an effective partner to a human, it must be capable of accurate and efficient communication with its partner. A number of methods have been proposed to enable humans to communicate with robots underwater [3]–[9] but few have addressed the inverse problem of how the robot could communicate back. Underwater, the well-explored interaction vectors of voice interaction (through speech synthesis and text-to-speech systems) and text interaction (through keyboards and screens) are infeasible or less effective. It is therefore desirable to develop new modalities of *robot-to-human communication* for underwater robots to enable their use as workers, companions, and guides to divers.

Robot communication to humans underwater is quite challenging, as the two most common modalities used for human interaction are significantly limited. Sound is distorted, attenuated and can be masked by equipment noise. While vision is usually available, it is often degraded or altered [10]–[13]. In such a challenging environment, the *de facto* solution is to simply accept the AUV as a silent partner which listens but does not respond. This keeps AUVs from achieving their full potential as diver partners by offering relevant safety information, providing advice, and most importantly, engaging in dialogue with humans. Besides ignoring robot-to-human communication, the most common technique is the use of a display device such as an OLED or LCD display, integrated into the robot, as in the Aqua AUV [14] or the CADDY AUV [15]. However, such displays are typically very small and hard to read from a distance or at an angle, introduce additional weight and power requirements, and are susceptible to failure. The other primary method that is proposed and has been used is messaging via flashing lights [16]. This has the advantage of wider viewing angles and the fact that many AUVs already have some kind of lighting system. However, lights are not a natural communication vector and require divers to commit to memory a list of light codes and associated meanings. They also may have a limited number of possible meanings which can be communicated, since most built-in systems have only one light and are thus limited to varying the blinking rate of that single light.

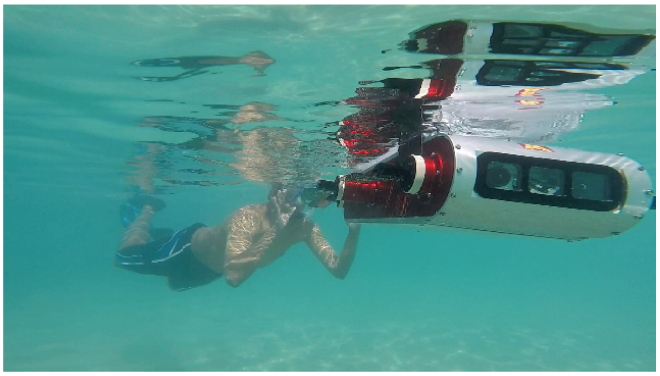


Fig. 1: An example of a swimmer interacting with the Aqua AUV, instructing the robot with free-form hand gestures.

In response to the drawbacks of the systems described above, we construct a list of desiderata. Underwater robot-to-human communication should: **a**) work from a distance and multiple viewing angles, **b**) require no additional hardware, **c**) be natural and easy to learn for the users, and **d**) allow for a large number of interaction phrases. To address these desiderata, we propose the use of robot motion as **kinemes** [17], a motion associated with a distinct meaning. We believe that a kineme communication system fulfills our desiderata, while remaining fast enough to be feasible for underwater interaction tasks.

To validate the use of motion as a communication technique, we conducted a study using simulated videos of the Aqua AUV to test the accuracy, efficiency, and ease of

The authors are from the University of Minnesota, Twin Cities, Minneapolis, MN, USA. {[1]fulto081, [2]edge0037, [3]junaed}@umn.edu

learning provided by such a system. Twenty-four participants tested the system against a baseline system comprised of colored lights flashing in codes, and the resultant data was analyzed to determine whether motion-based communication could be adequately accurate, efficient, and easy to learn for use in underwater robot-to-human communication.

In this paper, we make the following contributions:

- We propose a unique system: motion-based (kineme) communication for underwater robots which purely uses motion to communicate information to human collaborators.
- We show that there is a statistically significant improvement in the accuracy of communication when using kinemes compared to light codes.
- We show that kinemes outperform light codes in ease of learning and that given enough education, they can be nearly as quick to understand as light codes.

We also provide insight into the design of motion-based communication systems for other underwater robots, such as which kinds of information are best suited for kineme communication and discuss potential design implications for the implementation of our system using the Aqua AUV.

II. RELATED WORK

A. Underwater Human-Robot Interaction

1) Diver Communication: Underwater human-robot interaction (HRI) primarily focuses on the problem of human-to-robot communication. In this space, the emphasis has been on removing barriers to communication between the pair. A common approach is for a human operator on the surface to interpret the needs of the diver (possibly through hand signals from the diver) and teleoperate the AUV accordingly. This could be considered *direct control*, supported by high-speed tethered communication [18]. However, having the controller on the surface introduces latency and causes some issues by limiting situational awareness. In order to remove this latency, the use of waterproofed control devices at depth has been proposed [19], which essentially moves the controller down to the depth of the robot. This method requires additional hardware, however, which adds weight and complexity to the interaction scenario. To avoid this, onboard systems to enable human-to-robot communication are preferable. Examples of systems of communication which do not require an additional device include the use of fiducial tags [3] [20] or of hand gestures [5], [6], [8], [9], [16], which can be recognized and interpreted onboard a robot.

2) Robot Communication: As previously mentioned, the inverse communication problem of the robot communicating to the human has not been well explored. In many methods, the robot responds to the human's input via a small display [2], [5], [15], [21]. These displays are typically difficult to read from any distance or outside of optimal viewing angles. Larger screens, as in [22], would greatly reduce the depth rating and effective range of an AUV. Specific interaction devices generally are used for bidirectional communication, as in [19]. Once again, this requires the addition of other devices, which is costly and increases complexity. One of the more unique proposals is the use of an AUV's light system in [16], where illumination intensity is modulated to communicate simple ideas. This case study proved that a robot and a human could collaborate to achieve a task underwater, but the communication methods used were not validated by a multi-user study.

B. Nonverbal, Non-facial and Non-Humanoid HRI

Nonverbal methods form only a small portion of HRI, much of which focuses on displaying emotions in humanoid platforms, therefore, their results are only tangentially related to the problem of nonverbal underwater HRI, as our focus is non-humanoid robots displaying information rather than emotion. That said, there are a number of works which directly relate to our problem of robot-to-human communication using nonverbal, non-facial methods with a non-humanoid robot. These can be classified into categories based on their intended purpose. A useful source for these categories is the work of Mark L. Knapp [23], which defines five basic categories of body language:

1) Emblems, which have a particular linguistic meaning.
2) Illustrators, which provide emphasis to speech.
3) Affective displays, which represent emotional states.
4) Regulators, which control conversation flow.
5) Adapters, which convey non-dialogue information.

The bulk of nonverbal, non-facial, and non-humanoid HRI is concerned with affective display, while our work is primarily concerned with emblems.

1) Emblems: Emblems in nonverbal, non-facial HRI on non-humanoid platforms should be body movements which code directly to some linguistic meaning. This is the area to which our kineme communication system belongs, though it has little company here. The previously mentioned case study by DeMarco et al. [16] is one of the few attempts to communicate information rather than emotion using nonverbal methods, in this case via changes in the illumination of a light. Another example of emblems in this type of communication is the up-and-down tilt of a pan-tilt camera being used as a nod in [24], in which the robot simulates head gestures by controlling its camera's pan and tilt.

2) Affective Display: Affective display is by far the most explored realm of non-verbal communication with non-humanoid robots. The work of Bethel [25], particularly her doctoral dissertation [26], is a seminal work in this field, as it explores the use of position, orientation, motion, colored light, and sound to add affective display to appearance-constrained robots used for search and rescue. There has also been some work applied in this space to drones, such as adding a head-like appendage to an aerial robot [27]. This type of nonverbal communication has also been applied to a dog robot [28] and learned over time by an agent given the feedback of a human [29]. While many of these works focus on the actual display of the emotions and less on how they are generated, Novikova et al. [30] models an emotional framework to generate a robot's emotions and then display them, largely through body language.

Meaning	Kineme	Light Codes	Human Eqiv?	Meaning Type
Yes	Head nod (pitch)	One solid green	Yes	Response
No	Head shake (yaw)	One solid red	Yes	Response
Maybe	Head bobble (roll)	One solid yellow	Yes	Response
Ascend	Ascend, look back, continue	Two solid yellow, one blinking green	No	Spatial
Descend	Descend, look back, continue	Two solid yellow, one blinking green	No	Spatial
Remain At Depth	Circle and barrel roll slowly	Two solid yellow, one blinking yellow	No	Spatial
Look At Me	Roll heavily and erratically	Three quick flashing green	No	Situation
Danger Nearby	Look around" then quick head shake	Three quick flashing red	No	Situation
Follow Me	Beckon with head, then swim away	Three blinking yellow	Yes	Spatial
Malfunction	Slowly roll over and pulse legs intermittently	Three solid red	No	State
Repeat Previous	"Cock an ear" to the human	One blinking yellow light	Yes	Response
Object of Interest	Orient toward object, look at human, proceed	Two solid green, one yellow blinking	Yes	Spatial
Battery Low	Small, slow loop-de-loop	One solid yellow, two blinking red	No	State
Battery Full	Large, fast loop-de-loop	One solid yellow, two blinking green	No	State
I'm Lost	Look from side to side slowly as if confused	Three solid yellow	No	State

TABLE I: Kinemes and light codes with their associated meanings.

3) Regulators and Adapters: In this final category, we mostly find work which attempts to communicate some non-emotional state, which we consider to be adapters. A simple example of an adapter in human body language would be standing with slumped shoulders due to tiredness. While adapters can frequently wander into the realm of affective display, we consider a robot's display of its state to be an adapter-like communication. Works exist which merge the two, such as Knight et al [31], which uses the motion of the robot to display the internal state and task state of the robot. A more straightforward example of an adapter, however, would be the work of Baraka et al. [32], which displays information such as intended path using an array of expressive lights around the robot's body. Regulators are not well explored in non-humanoid robots, but some works such as [33] address the problem of how to initiate conversations.

III. METHODOLOGY

In this section, we introduce the design and implementation of our kineme communication system using robot motion and the light codes system to which it is compared.

A. Kineme System

1) Guiding Principles: The development of meaningful motion is a somewhat out-of-scope problem for most computer scientists and engineers. It is most closely related to animation [34] in the way it must be approached, which is by considering how a given motion is likely to affect the viewer. In this particular case of informative display for a non-humanoid underwater robot, there is little previous work, so it is important to develop guiding principles for design.

We applied the following concepts to develop our kinemes:

- If a human analog for a gesture exists (such as nodding or shaking of the head), mimic it.
- Exaggerate motions so that they are clearly visible from distance.
- Exploit any humanoid-looking design elements of the AUV; *e.g.*, the front cameras of Aqua look somewhat like eyes, so "gazing" motions would likely work well.

2) Design Process: With these concepts in mind, we selected a set of appropriate meanings for the kinemes. We began by considering the types of information divers communicate with each other using hand signals as the basis for the information our robot should be able to communicate, identifying four primary categories: a) responses to queries, b) spatial information and commands, c) situational information and commands, and d) state information. With the possible meanings selected, we split them into those with obvious equivalent human gestures and those without.

The human equivalent group was developed by attempting to mimic the human gestures which existed. These kinemes were mostly in the Response category. For these kinemes, Aqua's front was viewed as a face, with the cameras serving as eyes. Then, by manipulating the motion of the whole robot, Aqua's "face" could be moved in a head-nodding motion for *Yes*, a head-shaking motion for *No*, etc.

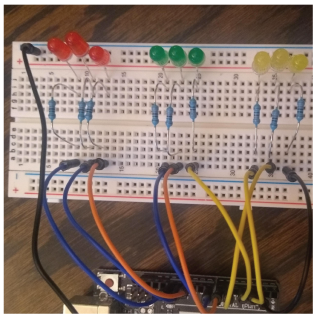
The group without human equivalent gestures was more challenging to design motion for. For these kinemes, if they were spatially oriented, the general approach was to orient Aqua's "face" to that location, move towards it, (*e.g.*, towards the surface to indicate *Ascend*), look back at the human, and then continue towards the location. For kinemes in the situation and state categories, design started by identifying a relevant emotion, such as fear for *Danger Nearby*, followed by developing a motion characteristic of that emotion. A complete list of kinemes along with their meanings can be found in Table I, as well as in the video submitted with this work.

3) Implementation: Implementation of kinemes was done using Epic Game's Unreal Engine™ to animate a 3D mesh of the Aqua AUV going through the selected motions. While physics simulation was not used to produce the kinemes, all kinemes were created by researchers familiar with the motion of the AUV in question and is achievable by the physical robot. The motion of Aqua's flippers was implemented as the forward swimming gait, regardless of the motion actually being executed. This is unlikely to have had an effect on participants, as none of them had experience with the Aqua robot, and almost all had never even seen it swim.

B. Lights System

1) Guiding Principles of Design: While the light system was designed as a baseline to compare to the kineme system, care was taken to make the light system as robust as possible.

(a) Unreal Engine^{TM} prototype.



(b) Arduino prototype.

Fig. 2: Experimental platforms for Kineme and Light Codes.

To be used as a baseline system, the same meanings for the kineme system were selected. To guide the development of the light codes, a number of principles were used, based on human perception of color [35] and blink frequency [36].

- Use natural color mappings (green = good, red = bad)
- The faster the blinking, the more time sensitive the communication.
- For related information, share a portion of the light code (*i.e.*, both battery info light codes have a single solid yellow light).

The list of light codes can also be found in Table I, as well as in the video submitted with this work.

2) Implementation: The light codes were chosen after the development of the kinemes and were implemented using an Arduino controlling 3 LEDs of each color. Blink frequencies were selected subjectively, with a duration of five seconds on for solid lights, 1 Hz blinking for five seconds for slow blink rates, and 5 Hz blinking for four seconds for fast blink rates.

IV. EXPERIMENTAL SETUP

To evaluate the kineme communication system as a method of robot-to-human communication, a user study was conducted using the color light system as a baseline. This section describes the hypothesis being tested, the population, design of the study, and experimental methods.

A. Hypotheses

The hypotheses we wish to test are simple: the accuracy of kinemes will be higher than that of light codes at all education levels (see Section IV-C) and the operational accuracy (accuracy of answers with confidence ≥ 3 on a scale of 1 to 5) of kinemes will be higher than light codes at all education levels. We also hypothesize that the confidence of participants will be higher with kinemes than with light codes and that the time-to-answer for kinemes will be significantly longer than light codes at low education, but eventually fall to approximately the same time as the education of participants increases.

B. Population

The population for this study was 24 participants (16 male, 8 female), largely undergraduate and graduate students at the University of Minnesota. The mean age of participants was 22 (*std_dev*=3.3). To ensure that the participants were representative of non-expert users, participants were asked to rate their experience with robots on a scale of one to five ($\mu = 1.54$, $\sigma = 0.5$), as well as their experience with nonverbal communication ($\mu = 1.33$, $\sigma = 0.7$).

Participants were split into three groups, based on the amount of preparation for the communication task they would receive. We designate these groups EDU0, EDU1, and EDU2, each with 8 participants. Within each group, the order in which the systems were displayed was alternated, so that half would see the kinemes first and half would see the light codes first.

C. Experimental Methods

After being enrolled in the study and providing basic demographic information from a survey, each participant was provided with the same basic information about the problem, namely that the purpose of the study was to develop underwater communication for robots. Next, participants received the appropriate level of education:

- EDU0: Participants were told the communication vector (motion, lights).
- EDU1: Participants were told the communication vector as well as the list of possible meanings.
- EDU2: Participants were told the communication vector and shown videos of each kineme and light code while being told the meaning.

Education was offered directly before testing each system. Once a participant was oriented and educated to a system, they were shown videos of the kinemes or light codes in a random order. The random order of the videos limits the order dependencies of the kinemes and light codes and produces independent measurements of each kineme.

For each video, three pieces of data were recorded: the user's understanding of what was communicated, the time taken from the start of the video to the start of the participant answering, and their confidence in their answer from 1 to 5, with five being positive. After-the-fact correctness for each answer was assigned according to some simple heuristics by an expert user.

D. Additional Modality

A third communication system employing messages display on an LCD screen was also tested in the study. The results from the LCD are not reported here, because it was not a realistic test of an LCD for underwater use. The purpose of the LCD in this study was simply to act as a control by which reaction times and confidence distributions for participants could be considered, as well as acting as a filter for participants who completely misunderstood the purpose of the experiment.

V. RESULTS

A. Comparison Criteria and Methods

The kineme and light system are compared on the basis of the following criteria:

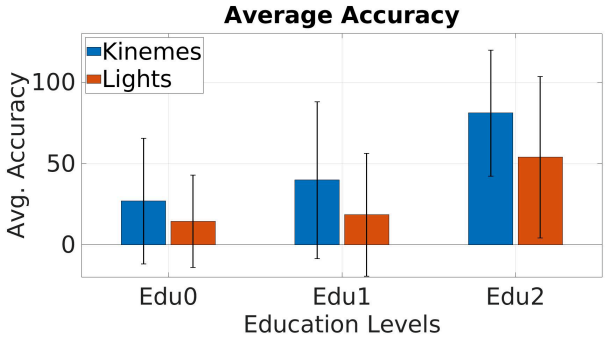


Fig. 3: Average Accuracy per education level

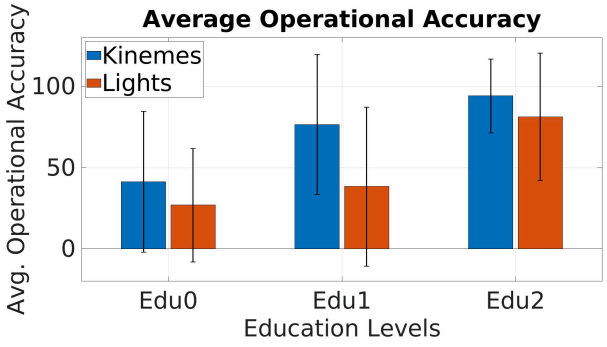


Fig. 4: Average Operational Accuracy per education level.

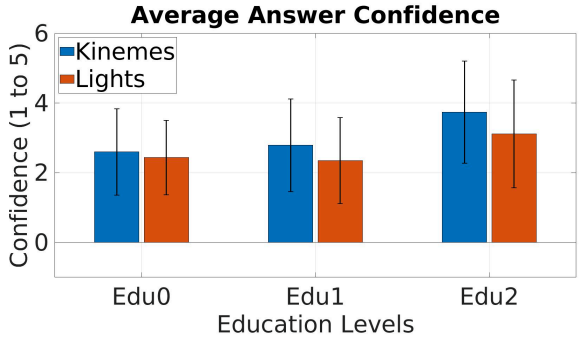


Fig. 5: Average Confidence per education level.

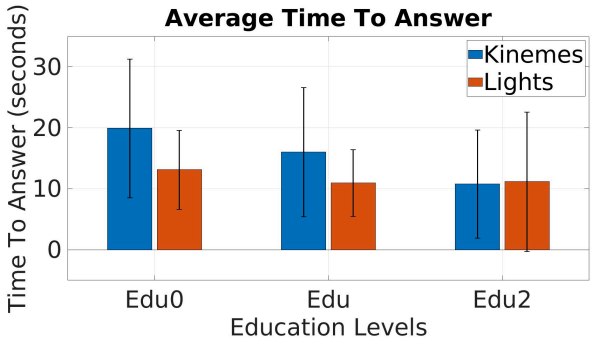


Fig. 6: Average Time To Answer per education level.

- **Accuracy** – The accuracy of a participant's understanding of a kineme or light code, rated from 0 to 10 in order of increasing accuracy.
- **Confidence** – The confidence a participant has in their understanding of a kineme or light code, rated from 1 to 5, in order of increasing confidence.
- **Operational Accuracy** – The same metric as accuracy, but only taking answers rated at a confidence level of 3 or higher, representing the answers that participants would be likely to act on.
- **Time To Answer** – The time it takes a participant to give the meaning of a kineme or light code, measured in seconds from the beginning of the signal to the beginning of their answer.

We use the Mann-Whitney test [37] to evaluate the hypotheses we set out in Section IV-A. We also use the Mann-Whitney test to validate that there is no statistically significant improvement in accuracy regardless of which system was shown to participants first and that there is no statistically significant difference between the accuracies of male and female participants.

The Mann-Whitney test is ideal for measuring the statistical effects of using the different systems in our trials, as it does not require normally-distributed data, while providing hypothesis testing capabilities. For each Mann-Whitney test, we report the p-value p, which is the probability under the null hypothesis of obtaining a result equal to or more extreme than what was observed. We also report the z-statistic z, which is used to calculate the approximate p-value. z is defined as

$$z = \frac{W - E(W)}{\sqrt{V(W)}}$$

where W is the Wilcoxon rank sum. Unless otherwise noted, all hypothesis tests are conducted at significance level $\alpha = 0.005$.

B. Statistical Results

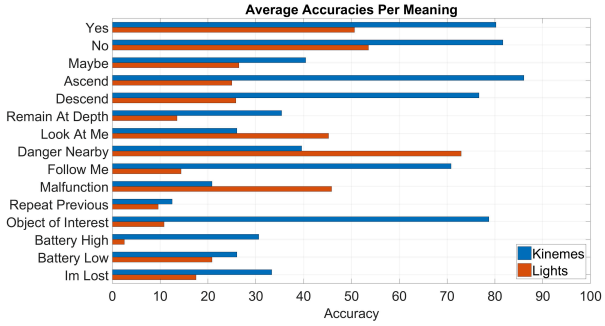
1) Accuracy and Operational Accuracy Between Education Levels: We compare kineme and light code accuracies at each education level, using a right-tailed Mann-Whitney test with this hypothesis:

$H_0 =$ Kineme accuracy does not have a higher median.

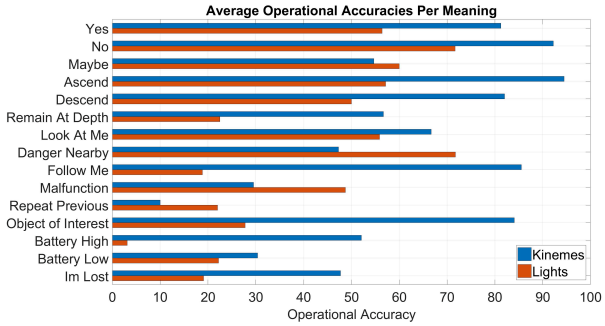
$H_a =$ Kineme accuracy has a higher median than lights.

When testing accuracy, we find statistically significant increases in accuracy when comparing kinemes to light codes for EDU0 ($p = 0.0113, z = 2.282$), EDU1 ($p = 0.0009, z = 3.114$), and EDU2 ($p = 0.000009, 4.27$). For operational accuracy, we again find statistically significant increases for EDU0 ($p = 0.029, z = 1.890$), EDU1 ($p = 0.007, z = 2.418$), and EDU2 ($p = 0.0006, z = 3.221$). In all of these cases, we can reject the null hypothesis in favor of the alternative. We can also see this visually in the plots of these statistics in Figures 3 and 4.

2) Confidence and Time-To-Answer Between Education Levels: We also test the median of confidence participants reported in their answers, and here we find that while there is no statistically significant increase in confidence in kinemes vs light codes at EDU0 ($p = 0.156, z = 1.01$), there is a statistically significant increase present at EDU1 ($p =$



(a) Average Accuracy per meaning.



(b) Average Operational Accuracy per meaning.

Fig. 7: Average Accuracy and Operational Accuracy per meaning.

$0.006, z = 2.496$) and EDU2 ($p = 0.0004, z = 3.297$). Finally, when considering the time to answer, we must slightly reformulate our test to be a left tailed test, testing the null hypothesis that the median time to answer for light codes is not lower than for kinemes, with the alternative being that the median for lights is lower. Here, we show that there is a statistically significant reduction in time when comparing lights to kinemes for EDU0 ($p = 0.0000002, z = -5.031$) and EDU1 ($p = 0.0003, z = -3.39$), but at EDU2 ($p = 0.4074, z = -0.828$) there is no significant reduction. This indicates that time to answer for kinemes drops with higher education, while time to answer for light codes remains approximately the entire length of the light code at all levels. These trends can be seen in Figures 5 and 6.

3) Comparison Between Specific Kinemes and Codes: For kineme-by-kineme comparison, we direct the reader to Figures 7a and 7b. We can see a particularly high accuracy for those kinemes in the spatial category, paired with low light-code accuracy for those same meanings. Conversely, we see that situation concepts such as *Danger Nearby* and *Malfunction* work better with flashing lights, likely due to a lifetime of being taught that flashing red lights mean danger. Operational accuracy figures are much closer between light codes and kinemes, but whether the kineme or light codes are the most accurate system does not change between accuracy and operational accuracy.

4) Internal Validation For Order and Gender: We validate these results by checking for a statistically significant bias based on system order. We find no statistically significant difference between the accuracy of kinemes when shown first or when shown second ($p = 0.449, z = -0.756$), nor do we find a statistically significant change in the accuracy of the light codes ($p = 0.748, z = -0.320$).

C. Opinion-Based Results

1) Participant Opinions: In their exit survey, participants were asked to rate the kineme and lights systems on a scale of 1 to 10 for several metrics. Participants rated kinemes easier to understand ($\mu = 5.6, \sigma = 2.2$) than lights ($\mu = 3.5, \sigma = 3.3$). They also considered kinemes easier to learn ($\mu = 7, \sigma = 2.1$) than lights ($\mu = 5.5, \sigma = 3.5$). When asked, 71.4% of participants also preferred the kineme system overall and 66.7% felt it would be most effective from a significant distance. Lastly when asked whether kinemes, light codes, or an LCD would be best for an underwater communication system, 45.6% preferred the kineme system, compared to 37.5% for lights and 16.7% for the LCD.

VI. CONCLUSION

In this paper, we proposed a unique motion-based communication for underwater robots, which we call kinemes, and implemented a version of these kinemes in Unreal Engine ^{TM} for testing. We evaluated the use of kinemes versus the use of colored light codes and found statistically significant superiorities in accuracy and operational accuracy, while remaining within an acceptable speed of recognition. Additionally, in our study, users preferred the kineme system over the light code system, and even over an LCD screen, especially when considering use at a distance or underwater.

We have also found that certain concepts related to 3-dimensional space are especially easy to communicate through motion via perceived gaze directions, as are concepts with a direct human analog by mimicking that human analog. It is noted that to properly express these types of concepts, estimating the pose of the interactant will be a necessary capability of any AUV using this communication scheme.

Future work must explore further into the related control (how to make the motions look correct on physical robots) and perception (how to ground interactions by determining the interactant's pose and state) problems. We plan to extend this concept to other robotic systems and implement kinemes on the physical Aqua robot, further validating our findings by running studies involving fully closed-loop interaction tests and more participants. Testing our desiderata in more depth is another focus of future work, as interaction at distance and with different angles is not considered in this study. Furthermore, we plan to integrate light and sound alongside motion to create a communication system which fuses these communication vectors to effectively communicate information to human collaborators in underwater environments.

ACKNOLWEDGMENTS

The authors gratefully acknowledge the support of the MnDRIVE initiative on this research; Hannah Dubois, Marc Ho, and Jungseok Hong for their help in conducting studies; Md Jahidul Islam for his advice; The Mobile Robotics Lab at McGill University, Independent Robotics, and particularly Ian Karp for the use of their Unreal Engine Aqua robot model.

REFERENCES

[1] "Bureau of labor statistics, US department of labor, occupational employment statistics, commercial divers." [Online]. Available: https://www.bls.gov/oes/current/oes499092.htm#(3)

[2] J. Sattar, G. Dudek, O. Chiu, I. Rekleitis, P. Giguere, A. Mills, N. Plamondon, C. Prahacs, Y. Girdhar, M. Nahon, and J. Lobos, "Enabling autonomous capabilities in underwater robotics," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 3628–3634.

[3] J. Sattar, E. Bourque, P. Giguere, and G. Dudek, "Fourier tags: Smoothly degradable fiducial markers for use in human-robot interaction," in *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, May 2007, pp. 165–174.

[4] K. J. DeMarco, M. E. West, and A. M. Howard, "Sonar-Based Detection and Tracking of a Diver for Underwater Human-Robot Interaction Scenarios," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 2013, pp. 2378–2383.

[5] M. J. Islam, M. Ho, and J. Sattar, "Dynamic reconfiguration of mission parameters in underwater human-robot collaboration," pp. 1–8, May 2018.

[6] A. G. Chavez, C. A. Mueller, A. Birk, A. Babic, and N. Miskovic, "Stereo-vision based diver pose estimation using LSTM recurrent neural networks for AUV navigation guidance," in *OCEANS 2017 - Aberdeen*, Jun. 2017, pp. 1–7.

[7] M. J. Islam, J. Hong, and J. Sattar, "Person following by autonomous robots: A categorical overview," *CoRR*, vol. abs/1803.08202, 2018. [Online]. Available: http://arxiv.org/abs/1803.08202

[8] D. Chiarella, M. Bibuli, G. Bruzzone, M. Caccia, A. Ranieri, E. Zereik, L. Marconi, and P. Cutugno, "A Novel Gesture-Based Language for Underwater HumanRobot Interaction," *Journal of Marine Science and Engineering*, vol. 6, no. 3, p. 91, Sep. 2018. [Online]. Available: https://www.mdpi.com/2077-1312/6/3/91

[9] A. G. Chavez, C. A. Mueller, T. Doernbach, D. Chiarella, and A. Birk, "Robust Gesture-Based Communication for Underwater Human-Robot Interaction in the context of Search and Rescue Diver Missions," *arXiv:1810.07122 [cs]*, Oct. 2018, arXiv: 1810.07122. [Online]. Available: http://arxiv.org/abs/1810.07122

[10] A. Yamashita, M. Fujii, and T. Kaneko, "Color Registration of Underwater Images for Underwater Sensing with Consideration of Light Attenuation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, Apr. 2007, pp. 4570–4575.

[11] Z. Li, Z. Gu, H. Zheng, B. Zheng, and J. Liu, "Underwater image sharpness assessment based on selective attenuation of color in the water," in *OCEANS 2016 - Shanghai*, Apr. 2016, pp. 1–4.

[12] H. Lu, Y. Li, X. Xu, L. He, Y. Li, D. Dansereau, and S. Serikawa, "Underwater image descattering and quality assessment," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 1998–2002.

[13] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 7159–7165.

[14] G. Dudek, P. Giguere, C. Prahacs, S. Saunderson, J. Sattar, L. Torres-Mendez, M. Jenkin, A. German, A. Hogue, A. Ripsman, J. Zacher, E. Milios, H. Liu, P. Zhang, M. Buehler, and C. Georgiades, "Aqua: An amphibious autonomous robot," *Computer*, vol. 40, no. 1, pp. 46–53, Jan 2007.

[15] N. Mišković, M. Bibuli, A. Birk, M. Caccia, M. Egi, K. Grammer, A. Marroni, J. Neasham, A. Pascoal, A. Vasilijević *et al.*, "Caddycognitive autonomous diving buddy: Two years of underwater human-robot interaction," *Marine Technology Society Journal*, vol. 50, no. 4, pp. 54–66, 2016.

[16] 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)*, Oct. 2014, pp. 3738–3743.

[17] W. Nöth, *Handbook of Semiotics*, ser. Advances in semiotics. Indiana University Press, 1995. [Online]. Available: https://books.google.com/books?id=rHA4KQcPeNgC

[18] T. Aoki, T. Murashima, Y. Asao, T. Nakae, and M. Yamaguchi, "Development of high-speed data transmission equipment for the full-depth remotely operated vehicle-kaiko"," in *Oceans '97. MTS/IEEE Conference Proceedings*, vol. 1, Oct 1997, pp. 87–92 vol.1.

[19] B. Verzijlenberg and M. Jenkin, "Swimming with robots: Human robot communication at depth," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 4023–4028.

[20] G. Dudek, J. Sattar, and A. Xu, "A visual language for robot control and programming: A human-interface study," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 2507–2513.

[21] J. Sattar and J. J. Little, "Ensuring safety in human-robot dialog a cost-directed approach," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 6660–6666.

[22] Y. Ukai and J. Rekimoto, "Swimoid: Interacting with an underwater buddy robot," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2013, pp. 423–423.

[23] M. Knapp, *Nonverbal Communication in Human Interaction*. Holt, Rinehart and Winston, 1972. [Online]. Available: https://books.google.com/books?id=_QFHAAAAMAAJ

[24] 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*, Sep. 2006, pp. 117–122.

[25] 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)*, vol. 38, no. 1, pp. 83–92, Jan. 2008.

[26] C. L. Bethel, "Robots without faces: Non-verbal social human-robot interaction," 2009.

[27] D. Arroyo, C. Lucho, S. J. Roncal, and F. Cuellar, "Daedalus: A sUAV for Human-robot Interaction," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '14. New York, NY, USA: ACM, 2014, pp. 116–117. [Online]. Available: http://doi.acm.org/10.1145/2559636.2563709

[28] L. Moshkina and R. C. Arkin, "Human perspective on affective robotic behavior: a longitudinal study," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug. 2005, pp. 1444–1451.

[29] T. Shimokawa and T. Sawaragi, "Acquiring communicative motor acts of social robot using interactive evolutionary computation," in *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, vol. 3, 2001, pp. 1396–1401 vol.3.

[30] J. Novikova and L. Watts, "A Design Model of Emotional Body Expressions in Non-humanoid Robots," in *Proceedings of the Second International Conference on Human-agent Interaction*, ser. HAI '14. New York, NY, USA: ACM, 2014, pp. 353–360. [Online]. Available: http://doi.acm.org/10.1145/2658861.2658892

[31] H. Knight and R. Simmons, "Expressive motion with x, y and theta: Laban Effort Features for mobile robots," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Aug. 2014, pp. 267–273.

[32] K. Baraka and M. M. Veloso, "Mobile Service Robot State Revealing Through Expressive Lights: Formalism, Design, and Evaluation," *International Journal of Social Robotics*, vol. 10, no. 1, pp. 65–92, Jan. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s12369-017-0431-x

[33] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to Approach Humans?: Strategies for Social Robots to Initiate Interaction," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '09. New York, NY, USA: ACM, 2009, pp. 109–116. [Online]. Available: http://doi.acm.org/10.1145/1514095.1514117

[34] T. Ribeiro and A. Paiva, "The Illusion of Robotic Life: Principles and Practices of Animation for Robots," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 383–390. [Online]. Available: http://doi.acm.org/10.1145/2157689.2157814

[35] M. Sokolova, A. Fernndez-Caballero, L. Ros, L. Fernndez-Aguilar, and J. Latorre, "Experimentation on emotion regulation with single-colored images," 12 2015, pp. 265–276.

[36] I. Langmuir and W. F. Westendorp, "A study of light signals in aviation and navigation," *Physics*, vol. 1, no. 5, pp. 273–317, 1931. [Online]. Available: https://doi.org/10.1063/1.1745009

[37] T. P. Hettmansperger, *Robust nonparametric statistical methods*, 2nd ed. CRC Press, 1998.