Designing a Virtual Reality Testbed for Direct Human-Swarm Interaction in Aquatic Species Monitoring

Arunim Bhattacharya and Sachit Butail *

* *Arunim Bhattacharya and Sachit Butail are with the Department of Mechanical Engineering, Northern Illinois University, DeKalb, Illinois, USA (email: sbutail@niu.edu)*

Abstract:
Swarm robotics is a natural candidate for monitoring large and complex environments. With the aid of a human who is directly involved in monitoring and search using a teleoperated robot, swarms of autonomous robots can solve problems related to navigation and changing objectives. In this context, understanding the role of nonverbal interactions have the potential to enable rapid and bidirectional communication between humans and robots. Quantifying such interactions however, requires a repeatable and engaging experimental setup that can track human action and cognition. This paper describes a desktop virtual reality testbed designed specifically to quantify human actions, perception, and cognitive load in real time during a monitoring mission. Motivated by recent deployment of underwater robots to monitor invasive species, the testbed is designed to mimic an underwater environment within the Great Lakes with five species of fish whose appearance, locomotion, and behavior are modeled based on videos from the field. Brain activity and pupillometry data are recorded synchronously in real time to aid in the measurement of cognitive load of the human operator. To quantify human perception, empirical data of visual acuity from the literature is used to model virtual object recognition. The capabilities of the testbed are evaluated in terms of frame rate achieved as a function of number of fish and robots in the environment, and demonstrated through two examples highlighting possible uses in human-swarm interaction studies.

Keywords: Human-swarm interaction, fish collective behavior, EEG, pupillometry, cognitive load, Great lakes

1. INTRODUCTION

Swarm robotics has applications in search and rescue, environmental monitoring, and surveillance (Bashyal and Venayagamoorthy (2008); Liu et al. (2019); Penders et al. (2011)). Despite these applications, owing to the complexity of monitoring possibly dynamic targets in a large and unstructured environment, an autonomous swarm alone may not constitute the most efficient monitoring strategy. Adding a human to the robot swarm has the potential to significantly increase the capability of the combined human-robot swarm. A human operator can for example influence the robot swarm by steering it in a specific direction, separating the swarm into subgroups, or by having the swarm exclusively monitor a particular region (Kolling et al. (2016), Dias et al. (2021)).

In human-swarm interaction (HSI), a human operator can interact with the swarm (Kolling et al. (2016)) indirectly by changing the environmental factors, or directly by steering one or more leader robots. With respect to direct control, depending on the application, the human may steer the swarm through an overhead interface (McLurkin et al. (2006)) or via a first person view (FPV) (Coucke et al. (2020)). Compared to an overhead interface, an FPV allows humans to directly investigate the environment and contribute in terms of improved detection and localization (Recchiuto et al. (2016); Chen et al. (2008)). With recent deployments of AUVs and drones for visual recognition of wildlife (Dawson and Allison (2021); Corcoran et al. (2021)), teleoperating a robot within an autonomous swarm may be the most relevant strategy to monitor dynamic targets. In this context, robot swarms that respond nonverbally to human cognition and perception in a direct HSI setup are expected to perform better than those operating under fixed autonomy (Hussein and Abbass (2018); Saunderson and Nejat (2019)).

This work enables cognitively and perceptually sensitive control strategies in direct HSI through: (i) the design of an experimental setup whereby a swarm can react to human cognitive load, (ii) quantifying human perception of the swarm to identify salient cues that are important in nonverbal interactions, and (iii) design and analysis of a realistic underwater virtual environment (VE) for swarm robot deployments. These contributions are realized in form of an experimental setup that integrates synchronous data streams of human cognitive load using electroencephalography (EEG) data, and gaze using eye tracking data, with control input. The use of virtual reality (VR) allows us to (a) overcome the challenges associated with

implementing large robot swarms and (b) create immersive environments that can mirror real-world setups.

To ground our approach in a real application that can elicit a variety of human responses in a monitoring task, we designed the VE inspired by the problem of monitoring aquatic invasive species (Wick et al. (2020)). The VE seeks to replicate underwater scenes within the Great Lakes (Larson and Schaetzl (2001)) with fish species that are native and invasive to the region. A mathematical model of collective behavior (Couzin et al. (2002)) is adapted for each fish species to match their natural behavior as observed in videos.

We present two examples demonstrating the capability of the experimental setup. First, to show the capability of designing swarm interaction strategies that can respond to human workload and movement, we demonstrate a proportional control strategy where the swarm follows the human controlled robot as a function of their cognitive load. Second, to show the capability of quantifying swarm perception, we use visual object recognition to identify a recognizable subset of the swarm.

This paper is organized as follows: Section 2 provides a background of the collective behavior model, cognitive load, and visual acuity. Section 3 presents models of fish appearance, locomotion, and behavior developed using observations from real data. Section 4 describes our approach to building an empirical model of visual acuity for a desktop VR setup. Section 5 evaluates the performance of the VE and presents the two examples demonstrating the capabilities of the experimental setup. We conclude in Section 6 with a summary of the work.

2. BACKGROUND

In this section, we review the mathematical model of collective behavior used to animate the fish and robots within our VE, measures of cognitive load, and an empirical model of human visual acuity.

2.1 Zonal model of collective behavior

The zonal model by Couzin et al. (2002) encodes motion of an agent within a collective in the form of three rules: (i) maintain a minimum distance (*zor*) from other agents to avoid collisions, (ii) orient in the same direction as those agents that are beyond *zor* but within a distance (*zoo*), and (iii) gravitate towards agents that are beyond *zoo* but less than *zoa* distance away. Other models of collective behavior include the boids (Reynolds (1987)) and the two-dimensional self-propelled particle model (Vicsek et al. (1995)). We select the zonal model for its predominant use to explain natural collective behavior as well as for simulating robot swarms (Tiwari et al. (2017)).

Within the zonal model, the position and velocity of i-th agent at time t is given by $\mathbf{r}_i(t) \in \mathbb{R}^3$ and $\mathbf{v}_i(t) \in \mathbb{R}^3$ respectively. The velocity of the $i-$th agent is updated as,

$$\mathbf{v}_i(t+\delta t) = \begin{cases} -s_i \mathbf{d}_{i,zor}(t) \text{ if } zor_i \neq \emptyset \\ 0.5 s_i \left(\mathbf{v}_{i,zoo}(t) + \mathbf{d}_{i,zoa}(t)\right) \text{ otherwise,} \end{cases} \quad (1)$$

where,$\mathbf{d}_{i,zor} = \sum_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|}$, $\mathbf{v}_{i,zoo} = \sum_j \frac{\mathbf{v}_j}{|\mathbf{v}_j|}$, and $\mathbf{d}_{i,zoa} = -\mathbf{d}_{i,zor}$; i and j represent focal agent and neighboring agent respectively ($i \neq j$) and δt denotes the simulation time step in seconds; $s_i \in \mathbb{R}$ denotes constant speed. Model parameters *zoo*, *zor*, and *zoa*, can be varied to produce different types of behavior (Couzin et al. (2002)). For example, a high value of *zoo* produces schooling behavior so that the agents move in the same direction; a relatively small value of *zoo* with a high *zoa* produces swarming behavior that looks like shoaling among fish.

2.2 Measuring cognitive load

Cognitive load is defined as a multi-dimensional construct that captures the amount of mental burden a human experiences as they accomplish a task (Paas et al. (2003)). Cognitive overload, when the task at hand reaches mental capacity, can affect performance (Jaeggi et al. (2007)). Cognitive load can be measured in a variety of ways including directly from EEG or pupillometry data, or indirectly from secondary task performance or workload survey responses (Sweller et al. (2011)). Within EEG data, cognitive load typically manifests in terms of synchronization and desynchronization within the α (7.5 - 12.5 Hz) and θ (4 - 7.5 Hz) frequency bands, are indicative of cognitive and memory performance (Anderson et al. (2011)).

Specifically, one measure of cognitive load is calculated in terms of energy shift at mean frequencies in both α and θ bands (Klimesch (1999)). The mean frequencies are calculated as a weighted mean of frequencies as (Klimesch (1999); Anderson et al. (2011))

$$\overline{f(\omega)} = \frac{\sum_{i=0}^{n-1} I_i f_i}{\sum_{i=0}^{n-1} I_i}, \quad (2)$$

where n is the number of bins in a frequency band ω, f_i is the frequency in bin i and I_i is the corresponding energy density. Cognitive load for a particular channel (location on the scalp) is then calculated as (Anderson et al. (2011))

$$L(t) = \Delta I(\alpha) \Delta \overline{f(\alpha)} - \Delta I(\theta) \Delta \overline{f(\theta)}, \quad (3)$$

where, $\Delta I(\omega)$ refers to the difference in energy density in the ω frequency band. The net cognitive load is a weighted sum of cognitive load of all channels.

2.3 Visual acuity

Human ability to perceive and recognize objects within the field of view varies with the distance of objects from the point of fixation of gaze. If the field of view is visualized as a cone projecting outward from a single point on the retina, the cone emanating at an angle 2.5° from the center of gaze corresponds to the fovea.

Anstis (1974) conducted an experiment involving identification of letters and found that beyond the foveal region the tendency to resolve objects in high detail (and therefore identify them) tends to decrease with distance from the fixation point. In particular, he found a linear relationship between the angle subtended by a just resolvable object and its retinal eccentricity, defined as the angle subtended by the line joining the gaze fixation point

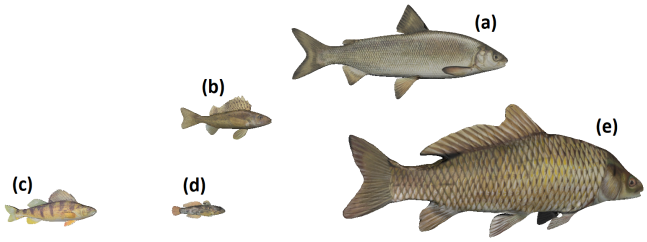

Fig. 1. Fish assets created for the virtual environment: (a) Lake whitefish, (b) Eurasian ruffe, (c) Yellow perch, (d) Round goby, (e) Common carp.

and object location (Read et al. (2009)). The reciprocal of the angle subtended by a resolvable object is called visual acuity. We utilize visual acuity to quantify the perception of the robot swarm by a human.

3. MODELING FISH APPEARANCE, LOCOMOTION, AND BEHAVIOR

This section describes the methods used to replicate appearance, locomotion, and behavior of various fish species within the VE. We populated the environment with five different species found in the Great Lakes. These include round goby (*Neogobius melanostomus*), lake whitefish (*Coregonus clupeaformis*), eurasian ruffe (*Gymnocephalus cernua*), yellow perch (*Perca flavescens*), and common carp (*Cyprinus carpio*).

3.1 Fish appearance and locomotion

Fish appearance was modeled to match the pictures in the aquatic species database from the US Geological Survey (USGS) (USGS (2022)). To create replicas, fish profiles were created from the lateral views. Specifically, lateral views were imported into Blender [1] modeling software as images, and vertices were placed to capture lateral profile, to match the fish silhouette (Fig. 1).

With respect to fish locomotion, we note that all the fish species that we modeled demonstrate carangiform motion, so that the fish body bent near the caudal regions during swimming (Winter (2007)). Accordingly, an animation was created in Blender by inserting two connected armatures within the fish body so that they joined near the caudal region. We then assigned two extreme positions corresponding to the tail beating locomotion and let the software interpolate to create a smooth tail beating motion.

Fish behavior was characterized on the basis of phenotypical observations from literature (Kováč (1998)) and videos available online. Videos were used because experimental studies in the literature were conducted in a laboratory environment, and therefore may not accurately represent fish behavior in the wild. A total of twenty-five videos, five per fish species, were used to conduct a detailed behavioral analysis. Videos were selected so that they: (i) were obtained from the field, (ii) had fish in view for at least 30 seconds, and (iii) did not contain environmental modification to lure fish (e.g. baits) or human activity of setting up cameras. Longer videos were cut to retain ninety seconds of continuous fish behavior. Fish in these videos were found to predominantly exhibit the following behaviors: (a) foraging, which involves searching for food as a group or independently (Cézilly (2008)) (b) shoaling, where the fish stayed but not necessarily moving in the same direction (Blakeslee and McRobert (2009)), and (c) schooling, where fish swam together in the same direction (Kasumyan and Pavlov (2018)).

Behaviors were encoded using the Behavioral observation research interactive software (BORIS) (Friard and Gamba (2016)). The frequency of each of the three behaviors was calculated across all five videos. Table 1 lists the percentage of time each species of fish was seen performing the three behaviors across all five videos. In particular common carp was found to forage most of the time, whereas eurasian ruffe and yellow perch exhibited shoaling more than 65% of the time. Lake whitefish was found to school most of the time. Round goby fish was seen in groups of two or three moving in a distinct stop and go pattern. All five fish were observed near the lake floor with common carp and round goby observed nearest the bottom of the lake. These same fish species were also observed in fewest numbers compared to other species.

3.2 Model of fish behavior

Motion of individual fish was simulated using model (1) albeit with different values of *zor*, *zoo* and *zoa* to qualitatively match the density and behavior as observed in the video analysis (Table 1). Inter-species interaction was limited to repulsion only so that fish from different species avoided collisions but did not align their motion or get attracted to each other. While schooling and shoaling behaviors for a fish species were simulated by varying the value of *zoo* between 0.48 – 1.872 virtual world units, foraging behavior was modeled for Common carp according to the model described next. Different behaviors were switched randomly every 90 seconds to match the frequency recorded in behavioral analysis. Round goby was simulated with large values of *zor* and *zoa* and to keep them apart yet within observing distance.

3.3 Foraging

To simulate foraging exhibited by common carp, we introduced an external stimulus to act as food particles. The food particles are generated in the form of n_f clusters at random locations $\mathbf{r}_p \in \mathbb{R}^3, p = 1, \ldots, n_f$ at the start of simulation. We simulated foraging as an increased tendency by the fish to go towards the food source. This was accomplished by setting the strength of attraction to all food sources as

$$\mathbf{f}_i(t) = \sum_{p=1}^{n_f} \frac{C_p(t)(\mathbf{r}^p - \mathbf{r}^i(t))}{|\mathbf{r}^p - \mathbf{r}_i(t)|} e^{-|\mathbf{r}^p - \mathbf{r}_i(t)|}, \qquad (4)$$

where $C_p(t)$ denotes the concentration of the food source. Equation (4) models the attraction to a food source in proportion to the concentration and decreases exponentially with distance to the food source. To model reduction of food due to consumption, the concentration decreases with presence of fish in the vicinity as

$$C_p(t) = C_p(t) e^{\left(-0.2 N_p(t) t\right)}, \qquad (5)$$

[1] http://www.blender.org/

Table 1. Fish size, number, location and behaviors (foraging, schooling, and shoaling) as observed from video analysis. Round goby was never observed performing any of these behaviors. The corresponding parameters for the zonal model are also listed.

Fish	Size (m)	Number	Behavior (%)	Model Parameters		
			Foraging, Schooling, Shoaling	*zor*, *zoo*, *zoa* (school)	*zor*, *zoo*, *zoa* (shoal)	g_u, g_d
Common Carp	0.78	1-5	63, 13, 24	0.02 ,0.01, 0.1	0.06, 0.03, 0.31	0.1, 0.05
Eurasian Ruffe	0.2	50-100	0, 30, 70	0.008, 0.48, 0.50	0.016, 0.008, 0.08	0.1, 0.02
Yellow Perch	0.25	50	0, 32, 68	0.01 ,0.6, 0.63	0.02, 0.01, 0.1	0.1, 0.02
Round Goby	0.125	1-3	-	0.5, 0.7, 0.71	0.5, 0.7, 0.71	0.1, 0.02
Lake Whitefish	0.51	5-15	0, 71, 29	0.02, 1.22, 1.28	0.04, 0.02, 0.2	0.1, 0.02

where $N_p(t)$ is the number of fish that are within 1 m of the food source p. The tendency to forage is balanced with instantaneous velocity using a parameter α as

$$\mathbf{v}_i(t+\delta t) = \alpha \mathbf{v}_i(t) + (1-\alpha)\mathbf{f}_i(t), \tag{6}$$

where the value of $\alpha = 0.3$ to avoid fish colliding with each other as they forage.

3.4 Maintaining depth

Because each fish species was found to swim near the bottom of the lake, the zonal model (1) was updated to ensure that simulated fish maintained depth. Specifically, the vertical component of velocity $\mathbf{v}^l(t)$, identified as v_y^l, was updated as

$$v_y^l(t) = \begin{cases} g_d^l |v_y^l(t)| & \text{if } r_y^l(t) > r_{ymax}^l \\ -g_u^l |v_y^l(t)| & \text{if } r_y^l(t) < r_{ymin}^l, \end{cases} \tag{7}$$

where g_d^l and g_u^l are species-specific parameters (Table 1) that control the rate at which a fish returns to an assigned depth region and $r_{ymin}^l = 1$ m for Round goby, Common Carp and 3 m for other species and $r_{ymax}^l = 3$ m for Round goby, Common Carp and 6 m for other species.

4. VIRTUAL OBJECT RECOGNITION

In order to faithfully quantify swarm perception, we sought to determine the number of robots that the operator can see. This value may not be the same as the actual number of robots in the visible range and could depend on the ability to resolve an object in the human eye. More importantly, quantifying swarm perception can enable control strategies to emphasize certain features of the swarm such as cohesion and number of subgroups.

Accordingly, we fit a quadratic function from Anstis (1974) data to obtain the relationship for recognizable object as

$$\psi \geq 0.0008\phi^2 + 0.0234\phi + 0.123, \tag{8}$$

where ψ is the angle subtended by the object on the eye, ϕ is the retinal eccentricity and $1/\psi$ is the visual acuity. Although (8) uses the same data as in (Anstis (1974)), this nonlinear relationship can account for large retinal eccentricities in viewing objects within the field of view, and provides a basis for isolating members of a swarm based on whether or not they can be perceived.

To determine if the object is recognizable we first calculated ψ for virtual object on a viewer's eye, by finding the dimension at the object as it is displayed on screen, and then calculating the angle subtended by that image on the eye (Fig. 2). We calculated size of the object image on the screen using a perspective projection model so that given the vertical dimension l_o of an object in the virtual world, and distance d_{or} of the object from the first person camera view of user controlled robot, both in virtual world units, the size of the object on screen is $l_{scr} = \frac{l_o}{d_{or}} \frac{f_{\text{rov}}}{w_{chip}} w_{scr}$, where f_{rov} is the focal length of the camera mounted on the user controlled robot, w_{chip} is the width of camera sensor, and w_{scr} is the width of the screen; all variables are measured in virtual world units except the screen width which is measured in mm.

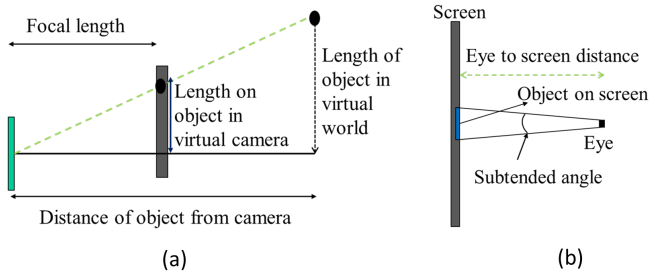

Fig. 2. (a) Projection of a virtual object on the virtual camera, and (b) angle subtended on the display.

With the screen d_{scr} mm away from the participant, the angle ψ_o subtended by an object of size l_{scr} is given by $\psi_o = 2\tan^{-1}\left(\frac{l_{scr}}{2d_{scr}}\right)$. The retinal eccentricity of an object located u_{scr} pixels away from the fixation point is then calculated as $\phi_o = 2\tan^{-1}\left(\frac{u_{scr}k_{scr}}{2d_{scr}}\right)$, where k_{scr} is a scaling factor equal to the ratio between the width of the display in mm to the wide resolution in pixels. Objects that subtend an angle ψ_o at a retinal eccentricity ϕ_o and satisfy the equation (8) are considered recognizable.

5. DEMONSTRATION OF CAPABILITIES

In this section we evaluate the performance of the VE in terms of refresh rate as a function of number of dynamic agents, and present two examples demonstrating the capability of the setup in utilizing and analyzing nonverbal interactions between a human and a robot swarm.

5.1 Experimental setup

The experimental setup consisted of a desktop computer with 16 GB memory, a 2.9 GHz i7 processor, an Nvidia GTX 1060 graphics card, a 1920 × 1080 pixel, 600 × 334 mm display, an EEG headset (Emotiv Epoc X, Emotiv Inc.) and an eye tracking device (Pupil Core, Pupil Labs

inc.). The EEG headset has 14 channels, with electrodes positioned according to the international 10-20 system with a sampling rate of 128 Hz. The eye tracking device can track gaze and pupil dilation at 200Hz with a resolution of 192 × 192 pixels per eye.

5.2 Virtual environment to mimic a lake

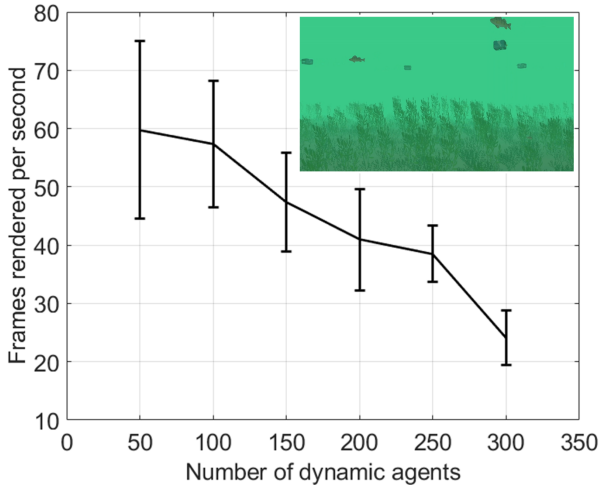

Fig. 3. Average ± standard deviation of fps as a function of total number of agents (fish and robot). Inset shows a scene from the experimental setup with yellow perch and autonomous robots (see video at `https://youtu.be/hCDTo5kRNxE`)

The virtual environment created in Unity software is a 1000 × 1000 × 50 m underwater region, authored to match the bottom of the Great Lakes region (Figure 3a). The environment is divided into five regions one for each species of fish, and features two kinds of submerged plants, chara and pondweed, found in the lake Michigan.

A swarm of autonomous underwater robots including one human operated robot are initialized near the center of the domain at the beginning of the simulation. All robots are designed after the Bluerobotics^{TM} remotely operated vehicle in terms of dimensions (0.25 × 0.45 × 0.65 m) and can move in six degrees of freedom. While these dynamics do not accurately capture those of an underwater robot, it serves to highlight the goal of this work, which is to capture cognitive and behavioral features of direct human swarm interaction. The human controlled robot can be maneuvered independently to move forward, yaw, and pitch using a keyboard.

Similar to fish, the motion of autonomous robots are modeled to swarm according to the zonal model, zoo = 4 m, zor = 5 m and zoa = 8 m at a constant depth. With respect to the human controlled robot, the velocity of autonomous robots in (1) is updated with a gain $k_a \mathbf{d}_{i,zoa}(t)$ so that a high value of k_a would result in following-type behavior. Although the virtual fish are repelled by the robots if they come close, robots are only repelled by each other to avoid collisions.

5.3 Rendering Performance

To measure performance with increase of complexity in the scene we varied the total number of fish and robots and measured the frame rendering rate. In particular, we increased the total number of fish and robots from a baseline value of 50 agents (38 fish and 12 robots) to 300 agents (260 fish and 40 robots). During this time, the human controlled robot was maneuvered to move in circles. In each scenario, we recorded the frame rate for 35 seconds. Although the frame rate can be locked during actual experiments, we measure the performance of our setup in terms of unlocked frame rate to capture the lower bounds with respect to different environments.

Figure 3b shows the average number of frames per second (fps) as a function of total number of mobile interacting agents (robots and fish). For comparison, human perception of quality of video drops significantly as the frame rate goes below 15 fps (Ou et al. (2008)). For a virtual environment requiring human interaction, Claypool and Claypool (2007) contend that frame rate significantly impacts playability, with frame rates higher than 30 fps not contributing significantly towards increased playability.

The large standard deviations in frame rate is likely because of mismatched timing of incoming EEG and pupillometry data stream as frame renderer waits for new data sample to arrive. This issue can partially be addressed by locking the rendering rate to refresh rate of the screen.

5.4 Example I: Closing the loop on human cognitive load

In this example, we show the system capability in modulating swarm autonomy with respect to human cognitive load measured from raw EEG data. While a meaningful response to cognitive load must be such that the operator feels aided and not annoyed (Chen and Barnes (2014)), here we demonstrate a simple proportional control strategy to demonstrate closing the loop. The environment consisted of 11 robots including the human teleoperated robot. The number of fish for each species were 70 yellow perch and eurasian ruffe, 7 common carp, 10 lake whitefish and 5 round goby. These fish were placed at random locations within the environment so that the experimenter did not know where to look.

Cognitive load was calculated as $L_{avg}(t) = 1/8 \sum_{i=1}^{8} L_i(t)$ where, $L_i(t)$ is cognitive load of ith channel located in the frontal part of the scalp. Baseline cognitive load was calculated using EEG data recorded during the first 60 seconds when the experimenter simply looked at the screen without navigating the robot. As the environment is explored, each swarm robot member is able to measure human cognitive load with additive noise as $\hat{L}(t) = L_w(t) + \eta_c$, where $L_w(t) = \frac{1}{N} \sum_{\tau=t-w}^{t} L_{avg}(\tau)$ is the average cognitive load over a window of past $w = 1$ seconds and N is the number of samples captured over the window and $\eta_c \sim N(0, 0.01)$. In an underwater environment, such observations could be relayed by using light emitting diodes (LEDs) (Berlinger et al. (2021)) or acoustic communication. The cognitive load within a certain range is then used to proportionally update the attraction gain as

$$k_a = \begin{cases} k_{low} \text{ if } \hat{L}(t) < 0 \\ k_{cl}(\hat{L}(t) + 1) \text{ if } 0 < \hat{L}(t) < L_{max} \\ k_{high} \text{ if } \hat{L}(t) > L_{max} \end{cases} \quad (9)$$

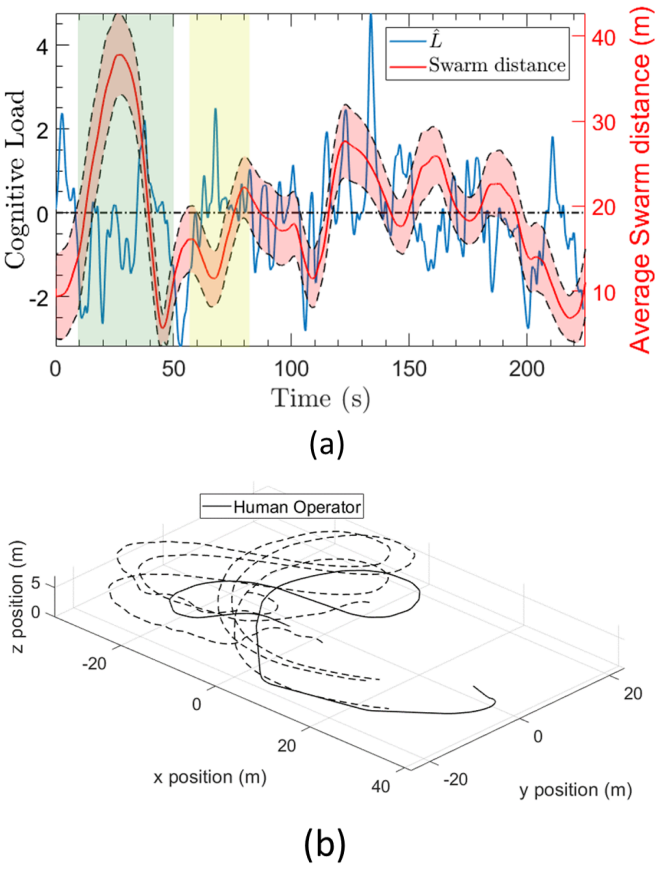

Fig. 4. (a) Swarm reaction to human cognitive load in terms of average distance to the human controlled robot; envelope denotes standard deviation. Regions of interest are highlighted with the swarm moving away (10–30 s) in response to low cognitive load and moving closer (55–80 s) as the cognitive load gets higher, (b) Three-dimensional trajectories of the human operator with three (out of eleven) autonomous robots (dashed lines) for the first 130 seconds.

where $k_{low} = 1, k_{high} = 5, k_{cl} = 3$, and $L_{max} = 1.61$ set on the basis of maximum cognitive load observed in an interpretation task (Anderson et al. (2011)).

Figure 4b shows sample trajectories of robots in response to human cognitive load. Times during which the robot swarm responded to human cognitive load by moving towards or away from it are highlighted. Although measuring cognitive load using EEG is difficult in real world setups due to movement artifacts that may manifest in the data stream, here it provides the basis for identifying cognitively responsive HSI strategies.

5.5 Example II: perception of the robot swarm by human

This example was designed to show that swarm perception depends on gaze location and distance. The experimenter's calibrated gaze is used to identify retinal eccentricity and the angle subtended by robots that are in view for two minutes as they explore the virtual environment.

Figure 5 compares the total number of robots (out of 51 robots) in the view with the those that are recognizable

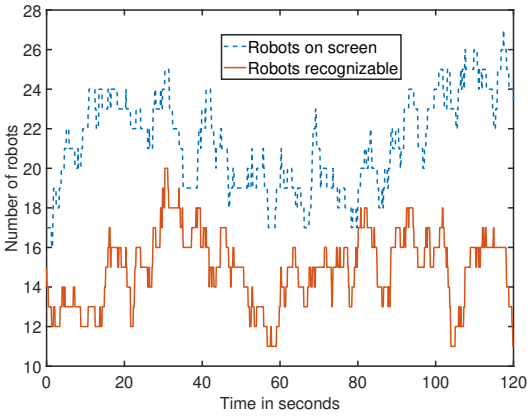

Fig. 5. Comparison of the number robots on screen and the number of robots perceived by human.

based on equation (8). We note that the recognizable portion of the swarm is nearly always less than the total number of robots in view. In light of this result, it is possible that a subset of the robot swarm may exhibit a collective behavior different from the one that is perceived by the human operator.

6. CONCLUSION

In this paper we describe the design of an experimental testbed where a human's cognitive load and perception are tracked as they interact with a robot swarm in an underwater monitoring mission. To achieve high levels of immersiveness and ease of realizing robot swarms, an underwater VE was designed and populated with replicas of five fish species whose appearance and behavior was modeled to match their real counterpart. Performance of the VE was evaluated in terms of frame rate achieved as a function of swarm size. The capability of the setup in terms of testing cognitively responsive HSI strategies were then demonstrated through two examples carried out using the experimental setup. This setup will be used in future studies to evaluate the performance of mixed initiative strategies in HSI that respond to human cognitive load and swarm perception.

ACKNOWLEDGEMENTS

This research was supported by National Science Foundation under grant # IIS-2033918.

REFERENCES

Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A., and Silva, C.T. (2011). A User Study of Visualization Effectiveness Using EEG and Cognitive Load. *Computer Graphics Forum*, 30(3), 791–800.

Anstis, S. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14(7), 589–592.

Bashyal, S. and Venayagamoorthy, G.K. (2008). Human swarm interaction for radiation source search and localization. *IEEE Swarm Intelligence Symposium*, 1–8.

Berlinger, F., Gauci, M., and Nagpal, R. (2021). Implicit coordination for 3D underwater collective behaviors in a fish-inspired robot swarm. *Science Robotics*, 6(50).

Blakeslee, C. and McRobert, S. (2009). Shoaling behavior in fish. *Handbook of Social Interactions in the 21st Century*, 137–158.

Chen, J.Y., Durlach, P.J., Sloan, J.A., and Bowens, L.D. (2008). Human–Robot interaction in the context of Simulated Route reconnaissance missions. *Military Psychology*, 20(3), 135–149.

Chen, J.Y. and Barnes, M.J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44(1), 13–29.

Claypool, K.T. and Claypool, M. (2007). On frame rate and player performance in first person shooter games. *Multimedia systems*, 13(1), 3–17.

Corcoran, E., Winsen, M., Sudholz, A., and Hamilton, G. (2021). Automated detection of wildlife using drones: Synthesis, opportunities and constraints. *Methods in Ecology and Evolution*, 12(6), 1103–1114.

Coucke, N., Heinrich, M.K., Cleeremans, A., and Dorigo, M. (2020). Hugos: A multi-user virtual environment for studying human–human swarm intelligence. In *International Conference on Swarm Intelligence*, 161–175. Springer.

Couzin, I.D., Krause, J., James, R., Ruxton, G.D., and Franks, N.R. (2002). Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology*, 218(1), 1–11.

Cézilly, F. (2008). An evolutionary perspective on behaviour. *Behavioural Ecology*.

Dawson, H.A. and Allison, M. (2021). Requirements for autonomous underwater vehicles (auvs) for scientific data collection in the Laurentian Great Lakes: A questionnaire survey. *Journal of Great Lakes Research*, 47(1), 259–265.

Dias, P.G., Silva, M.C., Rocha Filho, G.P., Vargas, P.A., Cota, L.P., and Pessin, G. (2021). Swarm robotics: A perspective on the Latest Reviewed concepts and applications. *Sensors*, 21(6).

Friard, O. and Gamba, M. (2016). BORIS : a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325–1330.

Hussein, A. and Abbass, H. (2018). Mixed Initiative Systems for human-swarm interaction: Opportunities and challenges. *Systems Modelling Conference (SMC)*.

Jaeggi, S.M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W.J., and Nirkko, A.C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience*, 7(2), 75–89.

Kasumyan, A. and Pavlov, D. (2018). Evolution of Schooling Behavior in Fish. *Journal of Ichthyology*, 58, 670–678.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3), 169–195.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., and Lewis, M. (2016). Human Interaction With Robot Swarms: A Survey. *IEEE Transactions on Human-Machine Systems*, 46(1), 9–26.

Kováč, V. (1998). Biology of Eurasian Ruffe from Slovakia and Adjacent Central European Countries. *Journal of Great Lakes Research*, 24(2), 205–216.

Larson, G. and Schaetzl, R. (2001). Origin and evolution of the Great Lakes. *Journal of Great Lakes Research*, 27(4), 518–546.

Liu, R., Cai, Z., Lewis, M., Lyons, J., and Sycara, K. (2019). Trust Repair in Human-Swarm Teams. *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–6.

McLurkin, J., Smith, J.and Frankel, J., Sotkowitz, D., Blau, D., and Schmidt, B. (2006). Speaking Swarmish: Human-Robot Interface Design for Large Swarms of Autonomous Mobile Robots. *Human-Robot Interaction*.

Ou, Y.F., Liu, T., Zhao, Z., Ma, Z., and Wang, Y. (2008). Modeling the impact of frame rate on perceptual quality of video. In *2008 15th IEEE International Conference on Image Processing*, 689–692. IEEE.

Paas, F., Tuovinen, J.E., Tabbers, H., and Van Gerven, P.W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71.

Penders, J., Alboul, L., Witkowski, U., Naghsh, A., Saez-Pons, J., Herbrechtsmeier, S., and El-Habbal, M. (2011). A Robot Swarm Assisting a Human Fire-Fighter. *Advanced Robotics*, 25(1-2), 93–117.

Read, J.C.A., Phillipson, G.P., and Glennerster, A. (2009). Latitude and longitude vertical disparities. *Journal of Vision*, 9(13), 265–278.

Recchiuto, C.T., Sgorbissa, A., and Zaccaria, R. (2016). Visual feedback with multiple cameras in a uavs human–swarm interface. *Robotics and Autonomous Systems*, 80, 43–54.

Reynolds, C.W. (1987). Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4), 25–34.

Saunderson, S. and Nejat, G. (2019). How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. *International Journal of Social Robotics*, 11(4), 575–608.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). Measuring Cognitive Load. In *Cognitive Load Theory*, 71–85. Springer New York, New York, NY.

Tiwari, R., Jain, P., Butail, S., Baliyarasimhuni, S.P., and Goodrich, M.A. (2017). Effect of leader placement on robotic swarm control. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 1387–1394.

USGS (2022). Nonindigenous Aquatic Species. `https://nas.er.usgs.gov/taxgroup/fish/default.aspx`.

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6), 1226–1229.

Wick, M.J., Angradi, T.R., Pawlowski, M.B., Bolgrien, D., Debbout, R., Launspach, J., and Nord, M. (2020). Deep Lake Explorer: A web application for crowdsourcing the classification of benthic underwater video from the Laurentian Great Lakes. *Journal of Great Lakes Research*, 46(5), 1469–1478.

Winter, H.V. (2007). *A fisheye view on fishways*. PhD Thesis, Wageningen University.