drone.io: A Gestural and Visual Interface for Human-Drone Interaction

Jessica R. Cauchard *‡§, Alex Tamkin*, Cheng Yao Wang*,
Luke Vink, Michelle Park*, Tommy Fang* and James A. Landay*
*Stanford University, Dpt. of Computer Science, Stanford, USA
‡Interdisciplinary Center Herzliya, Dpt. of Computer Science, Herzliya, Israel
§Ben-Gurion University of the Negev, Dpt. of Industrial Engineering and Management, Beer-Sheva, Israel
Corresponding author: jcauchard@acm.org

Fig. 1. drone.io in use. (Left) Top view of the projected radial interface as seen by the user. (Middle) User extending their arm in a push gesture to navigate through the menu with the drone flying above. (Right) User selecting an item in the menu.

***Abstract*—Drones are becoming ubiquitous and offer support to people in various tasks, such as photography, in increasingly interactive social contexts. We introduce drone.io, a projected body-centric graphical user interface for human-drone interaction. Using two simple gestures, users can interact with a drone in a natural manner. drone.io is the first human-drone graphical user interface embedded on a drone to provide both input and output capabilities. This paper describes the design process of drone.io. We present a proof of concept, drone-based implementation, as well as a fully functional prototype for a drone tour-guide scenario. We report drone.io's evaluation in three user studies (N=27) and show that people were able to use the interface with little prior training. We contribute to the field of human-robot interaction and the growing field of human-drone interaction.**

***Index Terms*—Human-Drone Interaction, Robotics, UAV, Mobile Projector, Mid-Air Gestures, Radial Menus.**

I. Introduction

Small-sized drones are increasingly present in our lives, being used for photography, delivery, monitoring [1], [2], and search-and-rescue [3]. These flying robots are primarily used outdoors and typically operate in two modes: 1. *Autonomous*, where the drone is fully automated and follows a pre-determined path or uses sensors to adjust its path, and 2. *Manual*, where a pilot controls the drone in real-time, generally using a remote control or a phone. We find that these two modes of interaction are limiting, as they do not allow a person to interact on-the-fly or change control mode without prior knowledge.

To address this gap in the literature, we developed *drone.io*, a gestural and projected graphical user interface embedded on a drone (Figure 1). Integrating a projection-camera system on a drone has recently been shown as a possible way for people to communicate with a service drone [4]. It has the advantage of being embedded on the drone, so that users can walk to the drone or have the drone fly to them, and interact ad-hoc. Using the underlying drone.io infrastructure, people can interact with a drone, while focusing on their task, without the burden of control, by simply choosing options through a projected menu, and getting immediate feedback.

drone.io was iteratively designed as a body-centric graphical user interface that people can interact with using their hands. It allows for both single and multiple user interaction. The projector displays a radial menu around the user and the camera recognizes the position of the user and gestures for navigation and selection on the menu. Using this interface, we show that people can interact with a semi-autonomous drone without prior training or additional hardware. In addition to being the first direct Input-Output (I/O) projected interface for human-drone interaction, drone.io introduces a novel and unique perspective on how to interact with semi-autonomous devices in the future. It also allows the recognition of a range of gestures without inconvenient body-mounted sensors.

978-1-5386-8555-6/19/$31.00 ©2019 IEEE

This paper first discusses the related work before describing the design choices and implementation of a working drone.io proof-of-concept. The interaction is first evaluated indoor (N=15) in a fixed setting without the drone's constraints. It is then evaluated outdoor (N=6) with the system fully embedded on a drone. Lastly, a drone tour-guide application was developed to field test drone.io in realistic conditions (N=6). The paper concludes with design guidance on building I/O capable human-drone interfaces.

The primary contributions of this paper are as follow:

- drone.io: a gestural and visual user interface for human-drone interaction
- A novel drone platform enabling a variety of gesture- and foot-based applications.
- Three user studies (N=27) showing drone.io is easy to use, enjoyable, and highly reliable in close to real world conditions.

We contribute to the growing area of Human-Drone Interaction (HDI) in providing a menu-based interaction for drones.

II. Related Work

This section discusses prior work in HDI, mobile projected interfaces, menus and gesture-based interactions.

A. Human-Drone Interaction

Several techniques have been proposed to interact with a drone, using remote controls and phones [5]–[7], gestures and face poses [8]–[15], or even touch [16]–[18]. In terms of feedback, drones can adjust their flight path [19], [20], be fitted with screens [21], projectors [4], [22], or LEDs to communicate intent [23]. While most of the prior work focuses either on the input or the output of the drone, based on our earlier drone.io work, Brock et al. [4] implemented and evaluated an interface with both functionalities. Here, we describe a projected graphical user interface that, embedded on the drone, allows for full input and output of menu information.

B. Mobile Projected Interfaces

Mobile projected interfaces can be handheld, worn, or embedded in devices, such as robots and drones. They can be used to augment objects [24] or the world around the user, as for pedestrian navigation [25]. Sasai et al. [26] proposed a robot tour guide that projected an interface that the user could step onto to input their destination, before being guided by projected information along the way. TeleAdvisor [27] provides remote augmented reality assistance through a projector mounted on a robotic arm. This leaves the user free from carrying or wearing a device and gives more flexibility for interactions. In our setup, the projector is attached to the drone, and serves to project the menu around the user. To the best of our knowledge, there are no guidelines for designing graphical user interfaces that use mobile projectors, and projected interfaces are traditionally purpose-built.

C. Menus and Gesture-Based Interactions

Different types of menus exist in desktop, mobile, tabletop, and public display computing [28]. A comprehensive literature survey of menus [29] was recently published. drone.io was iteratively designed to use the most appropriate type of menu. We later explain the process, which led to radial and circular menus. This section therefore focuses on these two types.

Radial menus exist in different shapes and forms, such as pie and hidden pie [30], [31], marking [31], and Frisbee [32] for large displays. Radial menus are mostly seen in desktop and mobile computing. Bailly et al. [33] propose free hand menus to interact at a distance with pie menus using mid-air gestures. Mid-air gestures are presented as part of a taxonomy of body centric interaction techniques [34]. They are classified in terms of relation to the body or as fixed in the world. Interestingly, with drone.io, the drone might be moving with the person, or both may be stopped at the same time, never truly fixed in the world. Velloso et al. [35] wrote a survey of foot-based interaction that is of relevance to this work.

Building on this earlier work, we now describe drone.io, its conceptual design, actual implementation, and its evaluation.

III. DRONE.IO

drone.io consists of a projector-camera system embedded on a drone. We first discuss the design process that resulted in drone.io and then describe its I/O capabilities. Finally, we describe the implementation of our hardware and software infrastructure.

A. Concept

drone.io projects a menu for a person to interact with the drone and was designed with the following goals:

Ease of use and Learnability: Users encountering the drone should be able to interact without prior knowledge.

Static Interaction: drone.io is designed for the moment when the user meets the drone and starts interacting with it. This interaction is static, meaning the drone and the person are not moving, as in [4]. This is equivalent to a person who interacts with their car navigation system before starting to drive. After the selection occurs, the drone can either project different content or fly and direct the user. How a drone should approach the person is out of the scope of this work and further discussed in [36], [37].

Spontaneous Short-Term Interaction: These interactions will be transitory and short. Spontaneous short-term interaction are "characteristic for robots that operate in public places (such as information kiosks, receptionists)" [38].

Drone Position: The user can take control of the drone by getting its attention. The drone then detects the user and starts hovering around their position (Figure 1 Middle).

Multiple Users: The projected interface allows for people to gather around the display and can foster conversations that would not happen if each person were looking at their own personal device. This vision pushes the boundary of the single user personal drone, which is traditionally remote controlled.

Fig. 2. (Left) User interacting with the drone.io menu during a campus tour. (Right) The drone.io concept allows for multiple projection spaces including the ground, a wall, or the user's body.

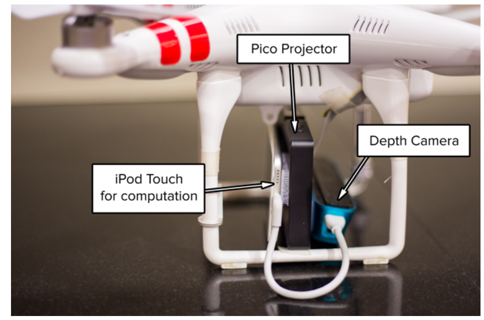

Fig. 3. Side picture of the drone with the embedded computing, projecting, and sensing infrastructure.

B. Gestural Input

We propose to embed the sensing, the projecting, and the computing infrastructure on the drone. This allows for impromptu interactions without additional hardware, trackers, or sensors on the user. The drone's flying ability helps support projector-camera alignment with the user in real-time.

Hands and Feet: With ground projection, we envision using both hands and feet to interact with the projected menu, to support a wide range of gestures.

Indirect Hand Interaction: drone.io is designed for stepping and mid-air gestures, which can be achieved using shadow [39] and in-the-air interaction above the projection area [40]. Controlling a shadow is unrealistic in an outdoor setting where there may be several light sources, such as street and buildings lights. As such, drone.io is designed for in-the-air interaction above the projection area. Constant feedback is necessary to show the hands' tracked position compared to the projected content.

Reference Frame: The user's body is used as a reference frame. The UI relies on human spatial memory and proprioception. We expect this to help with learnability of the system since within a few interactions the user should know how to position their hand and interact with the interface.

Stability: As the user is in a standing position, their hand and feet will be in stable positions. However, because of the drone's jitter, the gesture detection will have to be stabilized.

C. Projected Output

We present the defining elements of drone.io's output. These choices were made following many iterations and sketches helping us define the space within the realm of possibilities.

Projection Space: The position of the projection can change based on the user's needs at any given time, as has been done in the fixed [41] and mobile [42] projection literature. Figure 2 shows projections on the ground, a wall, or even a person's body to create a personal projection, as in [43]. We define the default position on the ground, as it is often flat, and always available, unlike other potential projection spaces.

Menu Projection: A radial menu [31] allows optimizing the space (over a linear menu) around the user, who can interact with the menu projected around their body. The menu needs to be within easy reach of hands and feet [44], [45]. Projecting all around the user was piloted. We found that users would rotate the menu rather than their body, so we chose to only project in front of the user. The front two-thirds of the radial menu is projected to maximize the menu size, while allowing participants to use their peripheral vision.

Menu Design: The radial menu contains one segment per menu item. The icons appear with the correct side up as the menu rotates around the user. The rotation speed is mapped to the current user's hand. The menu can rotate beyond 180 degrees. Two main actions are available: navigation through the menu and selection. A black and white color scheme is used for icons to maximize contrast and visibility, except when displaying color photographs of people. Selection is indicated by a blue segment increasing in size. Users should be able to go back to their previous choice. So, when navigating a sub-menu, a back button in the shape of a crescent appears in the top-right corner of the interface (Figure 4 Right). Unlike other menu items, the back button stays fixed as the user rotates the menu. The back button shows the last-selected menu item, indicating the position of the user in the menu hierarchy. To return to a previous menu, the user reaches the back button by extending their hand farther than the regular segment. When selected, the current menu fades and is replaced by the previous menu, which contracts inwards from outside.

D. Concept Video

As part of the iterative process, and to develop a vision of what the drone.io menu and interface would look like, we first designed a concept video[1]. This process helped us identify early on the challenges around privacy and multi-user interaction. Through the video editing process, many visual designs were fleshed out. We gathered feedback throughout the process until we were satisfied by the quality of the interface.

E. Implementation

This section describes the first implementation of drone.io.

[1]Autonomous Wandering Interface: https://youtu.be/cqU_hR2_ILU

Hardware: We implemented a proof of concept of drone.io, using a DJI Phantom 2 drone, an iPod Touch (A8 processor), a Celluon laser pico-projector (always in focus with high contrast ratio), and a Structure Sensor [46] (Figure 3). The iPod Touch, projector, and depth camera were mounted on the drone, with a total payload <500 grams including cables and affixing material. In these conditions, the drone flies approx. 20 minutes for each battery charge[2]. In this first instantiation, the drone flies at around 4m. height above the user, in GPS hovering mode. A pilot stayed behind the participants with a remote control with the ability to go back to manual flight control as a safety backup. The current implementation requires an additional Android phone on the ground, due to the Structure Sensor SDK being available only on iOS (at the time of implementation), while the projector required a Miracast Wi-Fi connection (not iOS compatible).

Software: This first implementation focused on hand gestures with a calibration step required before each participant. The algorithm subtracts the background using Otsu thresholding [47], identifies the highest point of the depth image as the top of the head, checks whether the hand is visible and what its position and shape are. The gesture is then identified and the results are returned to adapt the projection accordingly. We define the head-hand angle to be the angle between the head-hand vector and the x-axis. We use this to determine the active segment in the projected menu.

Gestures: Two gestures were implemented: Hover, a flat open hand facing towards the UI for point and dwell, and Push, where the hand is rotated at 90 degrees vertically to navigate through the radial menu using a "turn the wheel" metaphor in both directions. When hovering, the selected segment changes color and size; when navigating, the menu rotates with the user's hand. Selection occurs after a 2sec. dwell time on a segment [28]. After the first trials outdoors, the algorithm was adapted to use an accumulation-decay model, where activating a segment would increase the dwell counter and leaving it inactivated would decrease the counter rather than resetting it.

IV. DRONE.IO USER STUDIES & EVALUATION

The following two user studies describe the evaluation of the drone.io interface: indoors using a fixed setup and outdoors in close to real world conditions.

A. Indoor User Study

We first tested drone.io in a fixed setup to evaluate the system's accuracy without having to account for additional factors such as the drone's drift, noise, or battery life. The computing, sensing, and projection components were attached to a ceiling at a height of 3.5m, similar to the drone's flying height. The study was run in low lighting to mimic the real world conditions. It took around 40 minutes per participant.

We recruited 15 volunteers from within our institution and local companies (7f, 8m), age 18-26 y.o. (μ=20.6, SD=2.1), who were compensated $20 for their participation. Two thirds of the participants had used mid-air gestures, such as with a Microsoft Kinect, before. After obtaining consent, we briefly demonstrated the interface to the participant. The study was constrained to using one-handed interaction for better comparison of the results. Participants were asked to keep their feet within the projected footsteps and told to select menu items as quickly and accurately as possible. To avoid anticipatory movements, the trial began when the target appeared and ended upon target selection on the deepest level of the menu.

[2]Although the Phantom 2's stated maximum payload capacity is under 400g with maximum flight time of 25 minutes, we were able to fly it with stronger propellers and safely carry a 500g payload for a 20 minute flight time.

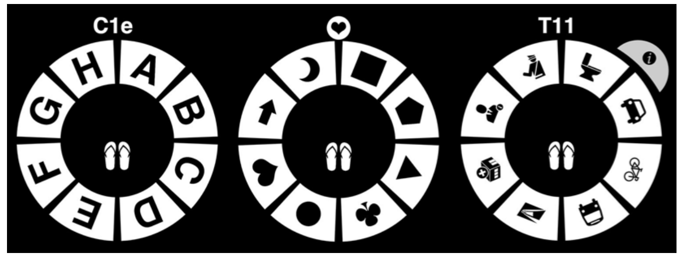

Fig. 4. Projected menus corresponding to the three visual search tasks: Task 1 (Left): Ordered letters and numbers, Task 2 (Middle): Non-ordered icons, and Task 3 (Right): Tour guide menu. The item at the top of each menu is the target. The back button is the crescent shaped button.

Three tasks were designed (Figure 4) and always presented in the same order, with increasing difficulty. All tasks involved visual search, requiring navigating between 1 and 3 menu levels to find the target. In **Task 1: Ordered numbers and letters**, users had to select a series of 3 characters: an uppercase letter, a number, and a lowercase letter, from ordered menus (e.g., F5a). There was one alphanumeric target per menu level. 15 trials were randomly generated. The first three trials were counted as practice and removed from the analysis. For **Task 2: Non-ordered icons**, users had to select one target icon from a non-ordered list. The icons were eight common 2D shapes such as hearts and clubs. The target and position of the items on the menu were randomly generated for each of the 12 trials. **Task 3: Tour guide menu** was specially designed for a tour guide scenario. Users were given each task verbally and had to select the corresponding item on the menu interface. For example, a trial included navigating to a "cactus garden" and another asked to "take a selfie". One task could be realized using only one level of menu navigation and the other eleven using 2 levels of menu hierarchy.

We calculate the minimum number of selection data points for Tasks 1 and 3. This number assumes the user gets the right answer on the first menu traversal for each task. The minimum selection data points across all three tasks and all 15 participants is: 15 x (36+12+23) = 1065 data points.

We recorded data for the following dependent variables:

- Task Completion Time (T_C): Time from the target's appearance to the end of the entire successful task (potentially multiple menus).
- Selection Time (T_S): Time from the menu appearance to the successful selection after dwell time (within a menu)

for all successful trials.

- Acquisition Time (T_A): Time measured from the end of the menu rotation to the successful selection after dwell time (single menu) for all successful trials.
- Selection Angle: Angle between the x-axis and the center of the selected segment.
- Selection Accuracy (A_C): An item-selection error was recorded when a wrong menu item was selected. The user would not be informed of the error and the interface would move on to the next menu level. A_C is the percentage of correct selections compared to the total number of items selected.
- Qualitative data: We used a NASA-TLX workload assessment after each task and conducted a post-study interview with each participant.

B. Outdoor User Study

In the outdoor study, drone.io was presented fully implemented with the drone flying above the user. We flew in the evening after dark, on non-cloudy and non-windy days.

1) Participants: We recruited six new right-handed participants (2f, 4m), age 19-36 y.o. (μ=25, SD=8.5) from our institution and local companies who were compensated \$20 for their participation. All but one participant had used mid-air gestures and two had controlled a drone before.

2) Methodology and Tasks: Participants performed 18 trials, 12 from Task 1, and 6 from Task 3. The experimental time was reduced from the indoor study to adapt to the drone's battery constraints. The study lasted about one hour and used the same methodology and dependent variables as in the indoor study. The minimum selection data points is: 6 x (12x3 + 6x2) = 288 data points.

C. User Study Results

This section presents and compares both studies' results.

1) Overall Use of the System: We find an increase in *Task Completion Time* (Table I) between the indoor and outdoor scenarios. This is consistent with our expectations since the outdoor conditions are more complex and the drone is not completely fixed and can drift or slightly change height during the interaction. We observe that on average for each selection across all tasks indoor, users spent approximately the same amount of time hovering (*Hover* = 31%) and selecting (Dwell = 35%), and less time navigating through the menu (*Push*=25%). The remaining time (9%) where no hand was detected was used to rest or when bringing one's arm towards the body to rotate the menu further.

These values are consistent indoor and outdoor (Figure 5). We see that Dwell time is reduced outdoors, which is due to the new implementation of the accumulation-decay dwell counter. Navigation (Push gesture) takes longer in Task 3 compared to Task 1, as expected, since users have to interpret the task and cannot immediately choose a visual target.

2) Selection Time: We calculate T_S for each task and condition (Table II-A), where a successful target was selected. T_S is slightly faster for Task 2 than for Task 1 where the

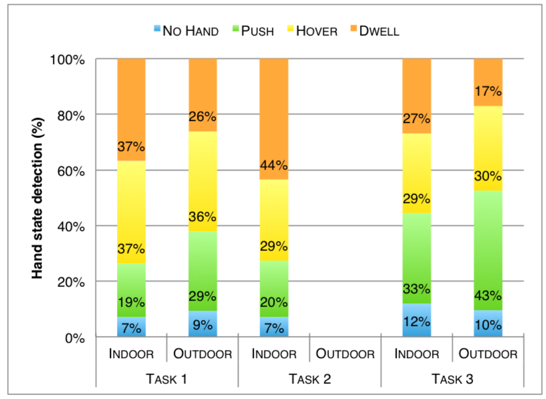

Fig. 5. Overall percentage of the hand state during the selection process across all successful trials.

menu was ordered. This might be due to fatigue because Task 1 requires 3 level of menu navigation for each trial. This is corroborated by the NASA TLX, where on a 7-point Likert scale, physical demand was 3.27 for Task 1 and 2.4 for Task 2 across all participants.

3) Acquisition Time: Despite the new dwell counter, item acquisition is slightly longer outdoors (Table II-B). This is likely due to the real-world study conditions, such as the drone's drift, making the selection more difficult than in the ideal fixed indoor setup.

4) Selection Angle: Most selections were made right in front of the user. When the segment appeared in front of the participant, they would directly select it without moving the menu. Surprisingly, outdoors most selections were shifted to the left of the user. The post-study interview revealed that participants were concerned with selecting the back button when the interface was not completely steady and chose to move the item slightly away to select it.

5) Selection Accuracy (A_C): Table II-C shows a high A_C of over 97% for Tasks 1 and 2, both indoors and outdoors. Outdoor A_C shows how well the recognition system works in close to real world conditions. We find high variation between users, with 60% making no errors indoors and 50% making no errors outdoors. Task 3 Indoor shows 92.8% overall accuracy in selecting the icon corresponding to the verbal query. This included 92.2% success on the 6 trials also used in the outdoor scenario. Task 3 Outdoor shows 88.9% successful

TABLE I

OVERALL AVERAGE *Task Completion Time* (T_C) FOR ALL PARTICIPANTS, ACROSS ALL TASKS AND CONDITIONS

	Task 1	Task 2	Task 3
Indoors	23.55s (SD=7.33)	6.64s (SD=1.16)	24.09s (SD=12.83)
Outdoors	30.82s (SD=8.31)	—	29.41s (SD=10.10)

Note: In all tables and graphs, Task 3 Indoors only shows the data of the 6 specific trials that were also run Outdoors.

TABLE II
TABLE A. AVERAGE *Selection Time* (T_S) OF AN ITEM WITHIN ONE MENU LEVEL FOR EACH SUCCESSFUL TRIAL ACROSS ALL TASKS AND CONDITIONS FOR ALL PARTICIPANTS. **TABLE B.** AVERAGE *Acquisition Time* (T_A) OF AN ITEM WITHIN ONE MENU LEVEL FOR EACH SUCCESSFUL TRIAL ACROSS ALL TASKS AND CONDITIONS FOR ALL PARTICIPANTS. **TABLE C.** SELECTION ACCURACY (A_C) FOR TASKS 1 AND 2 ACROSS ALL PARTICIPANTS

Table II-A.	Task 1	Task 2	Task 3
Indoor	6.86s (SD=0.31)	6.64s (SD=0.34)	9.24s (SD=0.49)
Outdoor	9.54s (SD=0.85)	—	13.88s (SD=0.64)

Table II-B.	Task 1	Task 2	Task 3
Indoor	3.66s (SD=0.12)	3.76s (SD=0.28)	3.72s (SD=0.16)
Outdoor	4.03s (SD=0.33)	—	4.6s (SD=0.28)

Table II-C.	Selections	Errors	Corrected Errors	Accuracy
Indoor T1	596	14	9 (12)	97.7%
Outdoor T1	242	5	3 (3)	97.9%
Indoor T2	180	2	—	98.9%

Note: Corrected errors shows the number of errors corrected compared to the total number of correctable errors.

trials for these same six trials. These errors were due to a participant misinterpreting the icon matching the verbal query and therefore selecting an incorrect icon. The results show that the designed icons can be used on their own without additional text for a fairly intuitive use in a tour guide context.

6) Back Button Use: In Task 1, participants corrected 12 out of 15 correctable errors using the back button. In Task 3, they navigated through the different menu levels using the back button 40 times indoors (10% of the 12 trials selections) and 4 times outdoors (5% of the 6 trials selections). We therefore find a use for the back button both to correct oneself and also to navigate through the menu levels. The back button was selected by mistake twice indoors (<0.5% of selections) and 4 times outdoors (<2% of selections) in Task 1.

7) System Detection Errors: The detection was overall robust. We found, however, that the system struggled to recognize the participant's hand when the menu segment was on the far left. Moving the right hand across the body would bring it too close to the body and made it difficult to recognize.

8) Qualitative Data: This section describes the NASA TLX workload assessment and the post-study interviews results.

NASA TLX: When comparing the results of the workload assessment indoors and outdoors, we observe that while drone.io appears more mentally demanding outdoors in Task 1, it is less mentally demanding in Task 3. In terms of physical demand, all participants were comfortable using the system. Two complained about standing in the same position for an extended period of time, and only one mentioned that their arm was getting tired.

Post-Study Interviews: In post-study interviews, indoors participants highly agreed that the menu was easy to see (6.3 out of a 7-point Likert) and found that the system was easy to interact with (4.9/7). Two users found it hard to see items on the periphery and said they would mostly look at the three segments in front of them. Two others found it difficult to look at icons when the menu was rotating.

All participants[3] enjoyed using the system and commented: "interacting this way was really cool" (PI9), "really fun" (PI10), and PI1 mentioned "I felt like Tony Stark" in reference to Marvel's Iron Man. Participants found the interface straightforward, intuitive (PI8), responsive, accurate, easy to use (PI13), and user friendly (PI4). They liked its physicality, and the fact that the hands did not have to carry anything. PI8 liked the projection concept as "you can just go anywhere".

Participants mentioned two main difficulties when interacting. The first one is referred to as "flickering". If the system detected the hand as moving to a segment, it would start highlighting this new segment and restart the dwell counter. When the detected segment kept changing, it would give the impression of flickering. Participants quickly adjusted by keeping their hand in the middle of a segment, but this affected their interaction. PI6 describes: "the dwell felt out of my control because the selection could move at any moment". Indeed, the dwell time would be restarted on each new detection. The dwell counter implementation changes in the outdoor condition helped with this problem.

Another difficulty was with the navigation. PI6 mentions that "In beginning, it [the menu] spun out of control until I got a good grasp of it. After a few trials it didn't spin as much". The current menu rotation uses a Push gesture. If the user moves the hand back while in a Push gesture, the menu rotates back in the other direction, canceling the initial interaction. After a few trials people became comfortable with the navigation. Yet, Some felt that the detection system was too sensitive to their movements.

All participants felt that they improved over time. Some mentioned that they had a preferred location to select, such as in front of them, as it felt more natural.

Outdoors, participants paid less attention to their hand shadow, which was not always aligned with the highlighted area compared to during the indoor study. PO3 avoided selecting items next to the back button to not select it by mistake and preferred the top left 11 o'clock position instead. Participants found that the drone's drift made it hard to interact with the menu, as they sometimes had to readjust their gestures.

[3]PI and PO resp. correspond to the indoor and outdoor study participants.

V. Drone tour guide application

We then implemented an actual tour with an adapted version of the drone.io menu. We implemented this working scenario to validate the drone.io concept in the real world.

A. Implementation Changes

We replaced the DJI drone with a 3DR Solo, which SDK made it easier to pre-program a flight path. It also flew around 20 minutes with a total weight (drone + payload) of 1.9 kg (67oz). We implemented a pre-defined tour that was designed to last approximately 15 minutes, covering a distance of 263 meters (284 yards), which included a turnaround at the halfway mark. The tour presented four buildings, two on each side of a long walkway, and visitors had the option to pause and resume the tour at any point. The drone was flown at an altitude of 3.7 m. and a speed of 0.7m/sec. which was determined in a pilot study as appropriate for people to follow the drone at a comfortable walking pace.

Along the tour, the drone displays information inside a projected arrow pointing towards the building to the side. When the user raises their hand and gestures to the drone, drone.io pauses the tour and projects the menu (Figure 2 Left).

The tour guide menu includes the option for additional information about the current building, nearby places to get food or use the restroom, and continuing with the tour. Selecting the food or restroom option displays a map of the area with the relevant information overlaid on top of the map.

B. FIELD TRIAL

We ran a field trial on clear nights with no wind on the Stanford University campus.

1) Participants: Six new participants (2f/4m) age 19-36 y.o. (μ=22.8, SD=6.6) were recruited and divided into three pairs. None of the participants knew each other beforehand. All but one had seen a drone before, but none had significant experience using drones. Participant were compensated $15 for their participation.

2) Methodology: After signing the consent form, participants were given a scenario of taking a campus tour and were briefed on the user interface. They were to make their own decisions as to who controlled the interface and how control was handed off between them. An experimenter observed and recorded these interaction decisions. After the tour, participants were interviewed about their experience. The tour lasted approximately 15 minutes and the total field trial study time was around 45 minutes per group.

3) Observations: Participants began using the interface right away and did not ask the experimenters for help. They took turns interacting with the menu. They paused the drone, on average, 4.3 times per pair, mostly to get additional information on the four buildings. When a person raised their hand to pause the drone, they would interact with the menu and then either resume the drone flight or step out of the frame to let the other participant interact. When not actively interacting with the drone, participants would either observe their partner from a distance, or step in closer to read the projected text. Social interactions varied between pairs. Participants in one group were very talkative and traded control of the drone interface frequently, whereas in the other two groups, interaction was largely independent and more quiet. When one person raised their hand, the other would follow.

Between interactions, participants would follow behind the drone at a distance of about 1.5 meters. Participants would first read the projected text and then look at the buildings until the text changed again.

4) Interview Results: Participants enjoyed their tour and having the opportunity to use a drone in an intuitive manner. They felt engaged with the tour and mentioned that "it was magic!" One participant commented that "the design itself was very nice" and "the simplest UI for what was needed." All participants found the scenario realistic. They had suggestions for improvements such as voice control and audio feedback, the option to ask for additional information as they would with a human guide, as well as to go on a different tour. Participants frequently compared the capabilities of the drone to a human tour guide and expected it to have a similar behavior.

Participants commented on the noise and jitter of the drone being disturbing, especially when reading text. One participant stated that the noise and wind generated by the propellers "made me feel less aware of my surroundings." Some participants were disturbed by the small height shifts of the drone. Participants' responses varied as to how fast they thought the drone ought to fly, with two participants stating the flight speed to be "ok", two wanting it to be faster, and two to be slower. These responses show the need for adapting the drone's speed to the users walking pace.

Participants felt comfortable using the drone in pairs, and said they would be happy to use it in groups of 2 to 12 people. They thought overcrowding around the projection and members of large groups stopping the drone too frequently were the limiting factors in determining how many people could use the interface. Participants stated that groups of families or close friends would be better able to use the interface than groups of strangers. One participant thought the drone interface would be best used one-on-one, such as on a personal tour.

When asked to rate on a scale of 1-5 how safe they felt (5 being the safest), four participants rated it a 4, and two a 3.5 (μ=3.8). Participants who did not feel as safe had no experience with drones and were unsure what would happen if the drone ran out of battery and fell down.

VI. Discussion

The section discusses the findings from the studies and reflections on the drone.io system.

A. Interaction Design

1) Input Technique: In the studies, participants were constrained to using one hand only and several would have preferred using both hands. Interaction will be faster when using both hands but this may also raise detection issues that need to be tested. A couple of people wanted to use feet to interact with the projection. A full implementation of the drone.io concept would indeed offer both hand and foot interaction. The combination will enable powerful interactions.

2) Drone's Position: Outdoor participants found interacting with the drone through drone.io to be natural. All felt safe, except one who felt neutral. While the above-the-head position of the drone is optimal for interacting with it using a projector-camera system, we had concerns that it may be uncomfortable. Yet, we found that the above the head drone position did not disturb the participants. Only one felt disturbed by the noise of the drone and none of them by the wind in the first study.

B. Interface Design

1) Navigational Ease: The fast selection times, low error rates, and qualitative data from both studies show that the menu is easy to navigate. In this design, participants could select items organized in a random order at least as quickly as ordered items.

2) Visual Feedback: All participants appreciated the visual feedback on the selected segment (blue highlight). Several mentioned that they wanted to use the shadow of their hand or that it should be mapped to the highlighted segment. While using the real shadow from the user is complex as it depends on uncontrolled factors, such as external light sources, future versions of drone.io could incorporate a mirroring technique where a cursor would represent the current hand position.

3) Back Button: This functionality was easily accessed with people correcting 80% of their correctable errors. Some participants complained about the flickering between the back button and the segment below it. Yet, the back button was only selected by mistake less than 0.5% of the time indoors, and less than 2% of the time outdoors. Its position is therefore adequate, although moving it further away from the segments could help reduce the flicker. In the tour guide field study, some participants mentioned that they would have liked the option to get back once they selected a segment or feedback confirming the selected icon. The back button was not present in this condition as the menu had only one level. We therefore recommend designing all menus with the back button.

C. Limitations

This section discusses the limitations of drone.io. We showed that the drone.io architecture of a drone paired with a projector-camera system works both indoors and outdoors under different constraints and limitations. The field study shows that the system is viable in close to real world settings. There are, however, restrictions in using drone.io in dim environments as mobile projectors do not currently display in bright daylight. While our current implementation does not allow for full interaction in the day, it corresponds to scenarios such as search and rescue, night navigation, and support of workers in anti-poaching missions. This first implementation of drone.io does not automatically correct image distortion, keystone and jitter. We minimized the distortion by projecting from above the user in a practically straight line and by only using part of the projection space, avoiding the edges. None of the participants seemed to notice any distortion. Our system was tested in the optimal condition of a single user interacting with the system. When the user is talking with friends or family while using the interface, we anticipate that conversational gestures may be recognized by the system and that false positives will occur.

VII. Legislation and Regulations

Since we started this project, we have seen legislation in the country, state, and on our campus evolving. As drones are becoming popular, legislators are being pushed to create new laws to regulate drones, trying to ensure safety first, but also to maintain privacy for all. While the administrators of some places, such as public parks and national monuments, have decided to ban recreational drones, others are finding ways to regulate the usage of drones.

Stanford University regulations ask the drone pilot to be nationally licensed and the flight to be registered and reviewed by a campus panel. We find that many rules have been created quickly because of a need to legislate and without a real understanding of all drone usage models and what it means to include drones into our environments. We believe that usage scenarios with slow-moving, low flying drone helpers are not properly captured by these rules and that part of the work in developing human-drone interaction will require discussion with legislators and other rule makers.

VIII. Future Work and Conclusions

Future work will include improvements to our algorithm, including stabilization of the drone and the projection, as well as leveraging head tracking to better position the drone compared to the user. It will also look at implementing new usage scenarios and testing them in real world settings.

In this paper we described drone.io, a novel input-output interaction system for human-drone interaction. drone.io is a fully mobile system that is embedded on a drone. It can recognize gestural input and adapt projected output in real time based on a user's movements. In three user studies with 27 participants, we show that the system is easy to use, enjoyable, and highly reliable when used outdoors in close to real world conditions. With drone.io, users can now walk to a helper drone and discover by themselves what the drone can do for them as well as make requests for help. drone.io is an intuitive interface that requires little prior training. Through drone.io, we have the opportunity to create a new generation of ambient and semi-public displays that are not limited by the infrastructure they are built in. Combining the projector and the drone gives flexibility in size, form, and positioning of the display at any given time with true interactivity, enabling a more natural Human-Drone Interaction.

Acknowledgments

We would like to thank Dr. Gilles Bailly and Prof. Terry Winograd for their time and feedback on this work.

REFERENCES

[1] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "Survey of Unmanned Aerial Vehicles (UAVs) for Traffic Monitoring," in *Handbook of Unmanned Aerial Vehicles*. Dordrecht: Springer Netherlands, 2015, pp. 2643–2666.

[2] A. Rango, A. Laliberte, J. E. Herrick, C. Winters, K. Havstad, C. Steele, and D. Browning, "Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management," *Journal of Applied Remote Sensing*, vol. 3, no. 1, p. 15, 8 2009. [Online]. Available: https://doi.org/10.1117/1.3216822

[3] P. Doherty and P. Rudol, "A UAV Search and Rescue Scenario with Human Body Detection and Geolocalization," in *AI 2007: Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1–13.

[4] A. M. Brock, J. Chatain, M. Park, T. Fang, M. Hachet, J. A. Landay, and J. R. Cauchard, "Flymap: Interacting with maps projected from a drone," in *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, ser. PerDis '18. New York, NY, USA: ACM, 2018, pp. 13:1–13:9. [Online]. Available: http://doi.acm.org/10.1145/3205873.3205877

[5] M. Lichtenstern, M. Angermann, and F. Martin, "IMU- and GNSS-Assisted Single-User Control of a MAV-Swarm for Multiple Perspective Observation of Outdoor Activities," 2011. [Online]. Available: http://elib.dlr.de/68957/

[6] M. Quigley, M. Goodrich, and R. Beard, "Semi-autonomous human-UAV interfaces for fixed-wing mini-UAVs," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3. IEEE, pp. 2457–2462. [Online]. Available: http://ieeexplore.ieee.org/document/1389777/

[7] I. Mota Group, "Lily camera," 2018. [Online]. Available: http://www.lily.camera/

[8] V. M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori, "HRI in the sky: Creating and commanding teams of UAVs with a vision-mediated gestural interface," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 11 2013, pp. 617–623. [Online]. Available: http://ieeexplore.ieee.org/document/6696415/

[9] J. Nagi, A. Giusti, G. A. Di Caro, and L. M. Gambardella, "Human Control of UAVs using Face Pose Estimates and Hand Gestures," 2014, pp. 252–253. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2559636.2559833

[10] T. Naseer, J. Sturm, and D. Cremers, "FollowMe: Person following and gesture recognition with a quadrocopter," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 11 2013, pp. 624–630. [Online]. Available: http://ieeexplore.ieee.org/document/6696416/

[11] W. S. Ng and E. Sharlin, "Collocated interaction with flying robots," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 143–149, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6005280

[12] A. Sanna, F. Lamberti, G. Paravati, and F. Manuri, "A Kinect-based natural interface for quadrotor control," *Entertainment Computing*, vol. 4, no. 3, pp. 179–186, 8 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1875952113000025

[13] J. Bruce, J. Perron, and R. Vaughan, "Ready Aim Fly! Hands-Free Face-Based HRI for 3D Trajectory Control of UAVs," in *2017 14th Conference on Computer and Robot Vision (CRV)*. IEEE, 5 2017, pp. 307–313. [Online]. Available: http://ieeexplore.ieee.org/document/8287709/

[14] J. R. Cauchard, J. L. E, K. Y. Zhai, and J. A. Landay, "Drone & me: an exploration into natural human-drone interaction," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. New York, New York, USA: ACM Press, 2015, pp. 361–365. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2750858.2805823

[15] J. L. E, I. L. E, J. A. Landay, and J. R. Cauchard, "Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. New York, New York, USA: ACM Press, 2017, pp. 6794–6799. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3025453.3025755

[16] A. Gomes, C. Rubens, S. Braley, and R. Vertegaal, "BitDrones: Towards Using 3D Nanocopter Displays as Interactive Self-Levitating Programmable Matter," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. New York, New York, USA: ACM Press, 2016, pp. 770–780. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2858036.2858519

[17] S. Rajappa, H. Bülthoff, and P. Stegagno, "Design and implementation of a novel architecture for physical human-UAV interaction," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 800–819, 6 2017. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364917708038

[18] P. Abtahi, D. Y. Zhao, J. L. E, and J. A. Landay, "Drone Near Me: Exploring Touch-Based Human-Drone Interaction," vol. 1, no. 3. ACM, 9 2017, pp. 1–8. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3139486.3130899

[19] D. Szafir, B. Mutlu, and T. Fong, "Communicating Directionality in Flying Robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*. New York, New York, USA: ACM Press, 2015, pp. 19–26. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2696454.2696475

[20] J. R. Cauchard, K. Y. Zhai, M. Spadafora, and J. A. Landay, "Emotion encoding in Human-Drone Interaction," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 3 2016, pp. 263–270. [Online]. Available: http://ieeexplore.ieee.org/document/7451761/

[21] S. Schneegass, F. Alt, J. Scheible, and A. Schmidt, "Midair Displays: Concept and First Experiences with Free-Floating Pervasive Displays," in *Proceedings of The International Symposium on Pervasive Displays - PerDis '14*. New York, New York, USA: ACM Press, 2014, pp. 27–31. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2611009.2611013

[22] P. Knierim, S. Maurer, K. Wolf, and M. Funk, "Quadcopter-Projected In-Situ Navigation Cues for Improved Location Awareness," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. New York, New York, USA: ACM Press, 2018, pp. 1–6. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3173574.3174007

[23] D. Szafir, B. Mutlu, and T. Fong, "Communication of intent in assistive free flyers," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*. New York, New York, USA: ACM Press, 2014, pp. 358–365. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2559636.2559672

[24] J. Schöning, M. Rohs, S. Kratz, M. Löchtefeld, and A. Krüger, "Map torchlight: a mobile augmented reality camera projector unit." in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. New York, New York, USA: ACM Press, 2009, p. 3841. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1520340.1520581

[25] A. J. Wecker, J. Lanir, T. Kuflik, and O. Stock, "Pathlight: Supporting navigation of small groups in the museum context," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. New York, New York, USA: ACM Press, 2011, p. 569. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2037373.2037462

[26] T. Sasai, Y. Takahashi, M. Kotani, and A. Nakamura, "Development of a guide robot interacting with the user using information projection basic system," in *2011 IEEE International Conference on Mechatronics and Automation*, Aug 2011, pp. 1297–1302.

[27] P. Gurevich, J. Lanir, B. Cohen, and R. Stone, "TeleAdvisor," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. New York, New York, USA: ACM Press, 2012, p. 619. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2207676.2207763

[28] R. Walter, G. Bailly, N. Valkanova, and J. Müller, "Cuenesics: using mid-air gestures to select items on interactive public displays," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. New York, New York, USA: ACM Press, 2014, pp. 299–308. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2628363.2628368

[29] G. Bailly, E. Lecolinet, and L. Nigay, "Visual menu techniques," *ACM Comput. Surv.*, vol. 49, no. 4, pp. 60:1–60:41, Dec. 2016. [Online]. Available: http://doi.acm.org/10.1145/3002171

[30] J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman, "An empirical comparison of pie vs. linear menus," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*. New York, New York, USA: ACM Press, 1988, pp. 95–100. [Online]. Available: http://portal.acm.org/citation.cfm?doid=57167.57182

[31] G. Kurtenbach, A. Sellen, and W. Buxton, "An Empirical Evaluation of Some Articulatory and Cognitive Aspects of Marking Menus," *Human-Computer Interaction*, vol. 8, no. 1, pp. 1–23, 3 1993. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1207/s15327051hci0801_1

[32] A. Khan, G. Fitzmaurice, D. Almeida, N. Burtnyk, and G. Kurtenbach, "A remote control interface for large displays," in *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*. New York, New York, USA: ACM Press, 2004, pp. 127–136. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1029632.1029655

[33] G. Bailly, R. Walter, J. Müller, T. Ning, and E. Lecolinet, "Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus," in *Human-Computer Interaction – INTERACT 2011*, P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 248–262.

[34] J. Wagner, M. Nancel, S. G. Gustafson, S. Huot, and W. E. Mackay, "Body-centric design space for multi-surface interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. New York, New York, USA: ACM Press, 2013, p. 1299. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2470654.2466170

[35] E. Velloso, D. Schmidt, J. Alexander, H. Gellersen, and A. Bulling, "The feet in human–computer interaction: A survey of foot-based interaction," *ACM Comput. Surv.*, vol. 48, no. 2, pp. 21:1–21:35, Sep. 2015. [Online]. Available: http://doi.acm.org/10.1145/2816455

[36] A. Wojciechowska, J. Frey, S. Sass, R. Shafir, and J. R. Cauchard, "Collocated human-drone interaction: Methodology and approach strategy," in *Accepted for Publication at the 2019 ACM/IEEE international conference on Human-robot interaction - HRI '19*. New York, New York, USA: ACM/IEEE, 2019, pp. 1–8.

[37] B. A. Duncan and R. R. Murphy, "Effects of speed, cyclicity, and dimensionality on distancing, time, and preference in human-aerial vehicle interactions," *ACM Trans. Interact. Intell. Syst.*, vol. 7, no. 3, pp. 13:1–13:27, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/2983927

[38] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: a second-generation museum tour-guide robot," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, vol. 3. IEEE, pp. 1999–2005. [Online]. Available: http://ieeexplore.ieee.org/document/770401/

[39] D. Molyneaux, S. Izadi, D. Kim, O. Hilliges, S. Hodges, X. Cao, A. Butler, and H. Gellersen, "Interactive Environment-Aware Handheld Projectors for Pervasive Computing Spaces." Springer, Berlin, Heidelberg, 2012, pp. 197–215. [Online]. Available: http://link.springer.com/10.1007/978-3-642-31205-2_13

[40] J. Zhao, H. Liu, C. Zhang, and Z. Zhang, "MobileSurface: interaction in the air for mobile computing," in *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. New York, New York, USA: ACM Press, 2010, p. 459. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1866218.1866270

[41] C. S. Pinhanez, "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces," in *Proceedings of the 3rd international conference on Ubiquitous Computing*. Springer, 2001, pp. 315–331. [Online]. Available: https://dl.acm.org/citation.cfm?id=741324

[42] J. R. Cauchard, M. Fraser, T. Han, and S. Subramanian, "Steerable projection: exploring alignment in interactive mobile displays," *Personal and Ubiquitous Computing*, vol. 16, no. 1, pp. 27–37, 1 2012. [Online]. Available: http://link.springer.com/10.1007/s00779-011-0375-3

[43] C. Harrison, S. Ramamurthy, and S. E. Hudson, "On-body interaction: armed and dangerous," in *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction - TEI '12*. New York, New York, USA: ACM Press, 2012, p. 69. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2148131.2148148

[44] N. P. Holmes and C. Spence, "The body schema and the multisensory representation(s) of peripersonal space." *Cognitive processing*, vol. 5, no. 2, pp. 94–105, 6 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16467906 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1350799

[45] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth, "Body-centric interaction techniques for very large wall displays," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10*. New York, New York, USA: ACM Press, 2010, p. 463. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1868914.1868967

[46] Occipital Inc., "Structure Sensor," 2018. [Online]. Available: https://structure.io/

[47] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1 1979. [Online]. Available: http://ieeexplore.ieee.org/document/4310076/