Learning from Users: an Elicitation Study and Taxonomy for Communicating small Unmanned Aerial System States Through Gestures

Justin W. Firestone
Computer Science and Engineering
University of Nebraska, Lincoln
Lincoln, NE, USA
jfiresto@cse.unl.edu

Rubi Quiñones
Computer Science and Engineering
University of Nebraska, Lincoln
Lincoln, NE, USA
rubi.quinones@huskers.unl.edu

Brittany A. Duncan
Computer Science and Engineering
University of Nebraska, Lincoln
Lincoln, NE, USA
bduncan@unl.edu

***Abstract*—This paper presents a gesture set for communicating states to novice users from a small Unmanned Aerial System (sUAS) through an elicitation study comparing gestures created by participants recruited from the general public with varying levels of experience with an sUAS. Previous work in sUAS flight paths sought to communicate intent, destination, or emotion without focusing on concrete states such as Low Battery or Landing. This elicitation study uses a participatory design approach from human-computer interaction to understand how novice users would expect an sUAS to communicate states, and ultimately suggests flight paths and characteristics to indicate those states. We asked users from the general public (N=20) to create gestures for seven distinct sUAS states to provide insights for human-drone interactions and to present intuitive flight paths and characteristics with the expectation that the sUAS would have general commercial application for inexperienced users. The results indicate relatively strong agreement scores for three sUAS states: Landing (0.455), Area of Interest (0.265), and Low Battery (0.245). The other four states have lower agreement scores, however even they show some consensus for all seven states. The agreement scores and the associated gestures suggest guidance for engineers to develop a common set of flight paths and characteristics for an sUAS to communicate states to novice users.**

***Index Terms*—sUAS; Communication; Elicitation Study; User Design**

I. Introduction

As technology improves and companies refine their business models, the general public will increasingly encounter small Unmanned Aerial Systems (sUAS) in everyday life. Consumers might ask Amazon to deliver their packages via Amazon Prime Air [19] or have Alphabet deliver lunch through its experimental burrito service [12]. Because everyday users are not likely to be experts with an sUAS or even aeronautics generally, it will be important for everyday drones to communicate common states quickly and intuitively to bystanders.

As we move towards this ubiquity, not every sUAS will have hardware to communicate through sound or lights due to cost or battery limitations, but they should be able to indicate key states through motions in space (gestures). If an sUAS is about to land and drop off its payload, it is critical that bystanders interpret the intention to land so they can move away from the landing area. Similarly, users and bystanders will need to quickly understand when an sUAS has missed its target so they do not unnecessarily worry that their packages will be delivered to the wrong location.

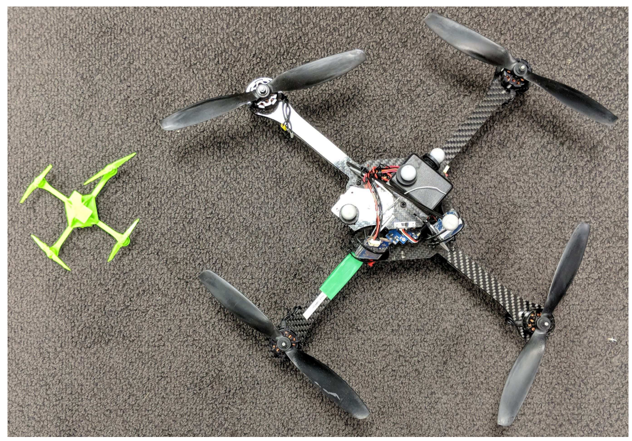


Fig. 1: For a side-by-side comparison, on the left is the palm-sized model given to participants and on the right is the Ascending Technologies Hummingbird used for demonstration flights in this study.

Furthermore, a well-defined set of gestures should improve sUAS user experiences and ultimately increase comfort with their greater prevalence in everyday life. Inexperienced users can be frightened or suspicious of a drone if they are confused by its intent, creating market barriers to adoption or innovation for new applications such as food delivery. As a result, a gesture set should be understood from various distances (even with partial occlusion), viewing angles, and qualities of lighting. These gestures should also not require specialized hardware so engineers can more easily incorporate them into pre-existing systems. Additionally, the limitation to movement alone allows this work to be better situated within exist-

This work was supported in part by NSF-1750750, NSF-1638099, and National Institute of Justice-GRF-STEM Fellowship 2016-R2-CX-0023.

ing human-human and human-robot gestural communications studies.

As an initial step for developing a gesture set for seven important sUAS states requiring user interaction or bystander awareness, we present an elicitation study which gathered gestures from the general public (N=20). An "elicitation" study gives users the opportunity to develop gestures for a specific purpose within set parameters to understand whether there is a common agreement across participants. We then assessed their suggestions within a taxonomy and calculated agreement scores.

This work is inspired by techniques from the human-computer interaction (HCI) community [16], [24] which elicited user-generated gestures to better understand common characteristics for each state (Attract Attention, Sensor Lost, Low Battery, Signal Lost, Area of Interest, Missed Goal/Target, and Landing). Users were asked to: 1) draw a gesture based on the model sUAS shown in Fig. 1; 2) describe their gesture to an experimenter; 3) observe their gestures on a free-flying sUAS shown in Fig. 1 in a Vicon cage; and 4) confirm the free-flying sUAS gesture matched the gesture drawn and described for the model sUAS.

The concept of elicitation studies can be broadly applied within the social robotics community to understand naive assumptions common among users, classify them, and then make recommendations for candidates to evaluate in user studies. The elicitation study described here is similar to the methods used in [9], although our elicited gestures are not immediately interpreted by a participant.

II. Related Work

A. Human Gestural Communications

Researchers have studied human gestural communications to assess their effectiveness regarding how they are perceived and what kinds of information they communicate. Krauss, Morrel-Samuels, and Colasante [10] conducted a set of studies to understand how co-speech hand gestures are understood and found that although hand gestures can convey some information, they do not communicate as well as speech. Prati and Pietrantoni [14] investigated effectiveness of different types of hand gestures in conditions during which speech would be difficult, such as when firefighters are trying to communicate inside burning buildings. In both studies, the gestures were based on those with meanings already understood by participants.

B. Robot Gestural Communications

Gestural communications in robots can be split into ground robot gestural communications and sUAS gestural communications. While gestures have been examined in humanoid robots, that research has been limited to social gestures and collaborative gestures. The current state of the art with sUAS has been to communicate high-level state information or to use gestures to control vehicles.

1) Ground Robot Gestural Communications: Researchers have investigated social gestures for human-robot interaction (HRI) through studies similar to those used to investigate communicative hand gestures. Salem et al. [17] investigated the ability for co-speech gestures to enhance humanoid-robot communications. Huang and Mutlu [8] evaluated the use of gestures to improve recall from humanoid robot interactions. Ng, Luo, and Okita [13] developed a model to generate a set of gestures from text and manipulated specific parameters to convey excitement or expressiveness. Riek et al. [15] tested cooperative social gestures on a humanoid robot to understand the impact of gesture speed and viewing angle, finding that negative attitudes towards robots correlated with decreased ability to understand gestures. Overall, these works assessed understanding of gestures, but they focused on leveraging participants' pre-existing interpretations of human gestures.

More relevant to this paper are the collaborative gestures developed primarily for industrial applications as in [4], [7], but work in this area is limited by the assumed presence of a visible goal as reported in [22]. Dragan and Srinivasa [4] studied integration of an observer into motion planning for an industrial robot. Gleeson et al. [7] observed gestural communications between humans, derived terms and gestures for their robots, and implemented them to observe their communicative ability. These studies indicated gestures were more effective when they conveyed context *and* goal, which is more challenging for an sUAS.

2) sUAS Gestural Communications: Communications with sUAS can be split into communication *from* the sUAS and communication *to* the sUAS. Communication to the sUAS is outside the scope of this work, so will only be covered with respect to design-based approaches.

Flight paths have been investigated for their ability to communicate affective state [2], [18], intended destination [20], intended flight direction [21], and to influence interaction preferences [6]. These flight paths could enhance interaction with sUAS in collocated environments, but do not communicate actions or states to general bystanders.

Sharma et al. [18] investigated the ability to communicate affective state via flight path with collocated users and found that to increase valence or arousal, paths should use space more indirectly and take less time. Szafir, Mutlu, and Fong [20] used both online and in-person interactions to explore the perception of animation principles applied to sUAS flight paths to improve communication of intent. Szafir, Mutlu, and Fong [21] next assessed the ability of a light ring to communicate direction of sUAS flight through in-person testing where participants made predictions regarding the end state of the vehicle. This work considered viewing angles, movement in multiple dimensions, occlusion, and ambient lighting. Duncan and Murphy [6] investigated whether the speed, cyclicity, and dimensionality of sUAS behaviors impacted the time, distance, and preference for interaction of users in a simulated interaction environment. However, none of these these works used a design-based approach to ask users to create their own gestures. Instead, they asked users to describe or react to predefined gestures.

C. Taxonomy Creation

Prior work on gesture classification has focused on creating taxonomies based on objective qualities to understand the relationship between humans and technology. Wobbrock et al. [24], constructed a taxonomy for tablet surface gestures to classify human hand movement and categorize different features of gestures on a 2-D plane. Wobbrock suggested surface computing has increased in prevalence compared to traditional input methods such as keyboards and mice because traditional methods limit users' motions. Wobbrock's participants were told a gesture's *effect*, and then they were asked to create a gesture that would *cause* that effect. This approach, according to Wobbrock, helped eliminate the "gulf of execution" between users and devices, because participants were told their answers were always correct and they stressed the importance of immediate usability to increase success rates for tasks [24]. The results from Wobbrock's taxonomy suggest tablet users prefer one-handed gestures, and also suggest a need for on-screen widgets to facilitate commands with low agreement scores.

Ruiz et al. [16] investigated motion gestures for invoking commands on a smart phone. Participants were given a list of commands and asked to construct a gesture that would execute each of the commands. Ruiz's taxonomy categorizes both the physical characteristics of the gesture as well as maps properties exhibited by each gesture. Their taxonomy indicates there is consensus among user-constructed commands in both physical characteristics and mappings.

Cauchard et al. [1] also foresee the possibility of drones becoming more prevalent in daily human activity. Therefore, they described interaction metaphors from participants. Participants were shown tasks which a drone could accomplish, then participants constructed gestures which users could perform to ask the drone to accomplish those tasks.

Well-defined taxonomies allow researchers to apply objective classifications to gestures from elicitation (or guessability) studies, leading to a useful calculation called an *agreement score*.

D. Agreement Score

After collecting user-generated gestures, a common practice is to calculate an agreement score for each task to evaluate consensus among participants, yielding objective data to compare against results from other studies. For this study, we adopted the agreement score calculation from Wobbrock [23] but do not convert scores into percentages or other ratios. Instead, our agreement scores conform to the standard used by Ruiz [16] and Cauchard [1]. An agreement score A_t, reflects the relative degree of consensus for a gesture among the participants. Wobbrock provides an equation to calculate an agreement score, where:

$$A_t = \sum_{P_i} \left(\left| \frac{P_i}{P_t} \right| \right)^2 \quad (1)$$

For Equation 1, t is one task from the set of all Tasks T. P_t is the set of proposed gestures for t, and P_i is the subset of identical gestures from P_t. For an example calculation, see [16], which implies an agreement score of 0.1 or higher indicates a minimum level of consensus. The possible range for A is [0, 1] and is not a percentage ratio.

III. Study and Methodology

The goal of our study was to have participants construct a preliminary gesture set for a list of sUAS states, all of which are important to communicate common conditions when operator interaction is needed to complete a task or to warn bystanders to avoid certain areas. Our study was heavily influenced by the work of Wobbrock and Ruiz [16], [24], who conducted elicitation (or guessability) studies to elicit gesture sets categorized based on a proposed taxonomy to calculate agreement scores.

A. Participants

Twenty English-speaking participants were recruited from the general public at a university in the Midwest U.S. through fliers and advertisements. As incentive for participation, successful completion of the study included entry into a random lottery for a $25 gift card. There were ten male and ten female participants with an age range of 19-79 (M=37.60, SD=18.36).

All participants were asked about their robot experience through broad questions to ensure they answered within their understanding of what classifies as a robot. Robot experience was assessed by asking: 1) whether participants had "ever interacted with a robot"; 2) the frequency of their interaction with robots; and 3) the type of robots with which they had interacted. Examples included consumer robots such as Roombas, pool-cleaning robots, Lego Mindstorms, Sony's Aibo, DJI Phantom, interactive robots in museums, or industrial robots. Seven of 20 participants reported some experience with such robots. Remote control (RC) experience was assessed by asking if they had ever owned or operated a "remote-controlled helicopter or airplane or an unmanned aerial system." Nine participants (three female, six male) reported prior interactions with RC aircraft.

B. Experiment Materials

The sUAS was an Ascending Technologies (AscTec) Hummingbird (see Fig. 1) weighing 365 grams (0.81lbs) with a diameter of 0.54m (21in). A palm-sized model of the Hummingbird (as shown in Fig. 1) was used for ease of interaction during the participants' time authoring gestures. The palm-sized drone was roughly one-third the size of the Hummingbird. The flights were controlled by a ROS script coordinated with a Vicon motion-capture system, and all flight paths were fully autonomous. A backup pilot was present in all study runs to take over control of the vehicle if needed.

C. Experiment Procedure

The study took approximately one hour to complete three parts: 1) Pre-interaction; 2) Flight Path Design; 3) and Flight

Path Observation. The first part was scheduled for 15 minutes and the final two parts were scheduled to consume the last 45 minutes. Each part included a survey and the experiment concluded with an interview.

1) Pre-interaction: Each participant was escorted to a room occupied by no other people except for one researcher. The researcher read them the consent form including a description of the objective of the study: to design gestures which a drone could execute to communicate seven specific states. After signing the consent form, the participants were given a pre-questionnaire to collect data about their general background.

2) Flight Path Design: The seven sUAS states were chosen because they are representative of some of the most common states the general public might encounter increasingly in every-day life. The states require either operator intervention, such as re-establishing communication, or awareness of bystanders, such as avoiding a landing site. The most common states for bystander avoidance would likely be landing, low battery, and signal lost. Bystanders need to know whether an UAS intends to land so they can adapt their behavior accordingly. Operators need to know about lost sensors, area of interest, or a missed goal in order to intervene. These states were primarily picked due to the motivating scenarios that these drones will be performing missions and need to communicate with operators who are otherwise engaged. We expected some of the states, specifically vehicle-centric states like Landing or Low Battery, to be intuitively understood by participants because they would be familiar with them from prior observations of hobbyist drones or even more traditional aircraft like airplanes and helicopters. We expected other states, like Lost Sensor or Lost Signal, to be similar to those same states users had already experienced with mobile phones or tablets. Lastly, we avoided choosing states that might need specific hardware or software, such as Take Picture or Deploy Sensor.

Each participant was given a sheet of paper that listed the seven states (Attract Attention, Sensor Lost, Low Battery, Signal Lost, Area of Interest, Missed Goal/Target, and Landing) along with extra whitespace for supplemental notes. This sheet included instructions at the top stating they would have 15 minutes to brainstorm and requesting they "design an appropriate gesture that the drone may take to communicate that state using the drone model provided within the defined area. Area is defined as between: (i) The top of the table to the top of your head, and (ii) Your chest to the tip of your fingers when your arm is extended. It is highly recommended that you provide a gesture for each task. It is also highly recommended you write notes in the space provided for each task." The page then had a space for each of the seven states for the participant to use when brainstorming.

Participants were given 15 minutes to design a set of appropriate gestures, but they were also welcome to request more time. Two participants completed this task in less than five minutes, ten participants needed five to ten minutes, seven participants needed ten to 15 minutes, and one participant requested an additional five minutes (20 minutes total) to complete the task. Two of these participants designed gestures with lights and sounds in less than five minutes, but then requested time to redesign their flight paths after clarification. These participants ended with one in the five-to-ten and one in the ten-to-15 minutes groups. Each participant was told that after the first 15 minutes lapsed, a researcher would come to announce each state to prompt them to demonstrate their gestures using a palm-sized model drone (see Fig. 1). They were told their gestures were always correct, even if they conflicted with the physical capabilities for the sUAS. For example, some participants thought the sUAS should ascend quickly to indicate Low Battery despite ascension being an energy-intensive flight path.

Participants were told they had to demonstrate their gestures within a bounded area to restrict participants from wandering around or making more complicated gestures that would be difficult for an sUAS to perform, and the bounded space was roughly to scale with respect to the room and drone used in the next part of the study, the Flight Path Observation. It is also natural to prescribe some boundary for a drone because drones cannot execute unlimited or infinite flight paths. Furthermore, prior elicitation studies on tablets and smartphones have inherent boundaries, either at the corners within the devices themselves, or only to the extent the human hand can move while holding the devices.

The participants were told they had to begin each gesture on or above an 'X' marker on the table. They were encouraged to provide verbal commentary of their thought process in designing the gesture for each state and to describe each gesture while performing it manually with the model drone. A researcher took written notes about the flight paths while participants provided commentary for each gesture.

3) Flight Path Observation: Participant flight paths were recorded in a Vicon motion-capture cage using a script that wrote a set of waypoints to a file and then had the drone fly from point-to-point in order to recreate a path that was "drawn" by the researcher walking the vehicle through the path. After the researcher used the previously described method to program the gestures indicated from the Flight Path Design, participants were next escorted to the Vicon motion-capture cage where they observed their gestures performed autonomously by an AscTec Hummingbird. As each of the seven gestures were being demonstrated, participants were encouraged to provide verbal feedback regarding the similarity of their expected gestures to the flight path of the drone. After the demonstration, participants were escorted back to the original room to thank them for their time and conclude the study. Of the 140 flight paths which were automatically replayed (20 participants for seven states), only four of them were described as incorrect representations of the gestures drawn on paper. All four inconsistencies were easily and quickly corrected to satisfy the participants' original drawings.

Some of the drawn gestures required representative substitutes because they were not physically repeatable by the sUAS. For example, Roll and Waggle maneuvers are not possible with the AscTec Hummingbird, so we chose Rock in Place as a reasonable substitute. Regardless, we achieved 140 out of

140 agreement between drawn gestures and replayed gestures.

D. Classification and Taxonomy for User-Designed Flight Paths

Having elicited 140 gestures, the final step was to create an objective classification and taxonomy to group them according to specific, common characteristics. Although related work provides a taxonomy for HCI, HRI, and human-to-sUAS interactions, no work has established a taxonomy for user-designed flight paths to communicate states. Also, some of the taxonomies from related work are for gesture sets for mobile phones and tablets, and not necessarily applicable to sUAS.

We classified each gesture along six categories: *Complexity, Space, Cyclicity, Command, Altitude, and Motion*. Within each category are multiple sub-categories shown in Table I.

1) Categories Selected from Related Work: We adopted the Complexity category from Ruiz, Li, and Lank to classify the movements of the drone. We also adopted the Space category from Chi et al. and Cyclicity from Duncan and Murphy to capture the spatial characteristics of the flight paths.

Complexity classifies the gesture either as simple or compound. Simple gestures are defined as a single movement along any direction and a compound gesture is a collection of simple gestures combined with spatial discontinuities. Discontinuities are inflection points, pauses in motion, or corners [16].

Space dimension describes a gesture's attention to its surroundings [3]. Indirect gestures have a multi-focused approach to a destination and can deviate from a straight line path, while direct gestures do not deviate from focus on the destination.

Taxonomy for User-Designed Flight Paths		
Categories from Related Work		
Complexity	*Simple*	Single movement.
	Compound	Collection of movements.
Space	*Direct*	Focused approach to a point.
	Indirect	Deviates from direct path.
Cyclicity	*Cyclic*	Repeated motion (same path).
	Random	Singular flight path.
Additional Categories		
Command	*Roll*	Left or right movement.
	Pitch	Forward or back movement.
	Yaw	Rotation.
	Throttle	Up or down movement.
Altitude	*Increasing*	Increase flight height.
	Decreasing	Decrease flight height.
	Variable	Increase and decrease.
	Stable	No height change.
Motion	*Rectilinear*	Only straight movement(s) and 90-degree turns.
	Curvilinear	Only curved movement(s).
	Rotational	Only rotates.
	Combinational	Combination of the above.

TABLE I: This taxonomy has six categories and is divided into two sections: Related Work Categories and Additional Categories.

Cyclicity was adopted from [6], where it was defined as "a judgment of whether the expression is cyclic or random in nature" and is based on assumptions in animal literature that unpredictable behaviors are used to display fitness, confuse predators, and startle observers. Here it is used as a measure of observability and likelihood for reception, because random flight paths are likely to be perceived in part by observers as their attention is gained.

2) Additional Categories: Because drones have more degrees of freedom than a tablet or mobile phone, we developed additional categories to account for expected differences in sUAS flight paths compared to categories developed in prior work involving human movement, tablets, and mobile phones. In defining these categories, there could be some overlap with related work.

The Command category maps to operator or autopilot input to the drone. Roll is movement in the left or right direction, pitch is movement in the forward or backward direction, yaw is rotational movement, and thrust is movement up or down in elevation. This category is related to Ruiz's "Dimension," which describes the number of axes involved in the movement of the gesture, but the description of commands allows an inherent representation of the number of axes of movement required to perform them, as opposed to simply counting the number of axes along which the drone must move to perform a particular command. There are four subcategories within the Command category, but we considered the possibility of collapsing these subcategories in combinations of two, three, or all four. For example, Command subcategories could also be "pitch and yaw," "yaw and thrust," or "pitch and yaw and roll." Our inter-rater reliability scores did not account for every possible subset of these four subcategories, only those which raters indicated were employed in the flight paths. The agreement scores, on the other hand, combined flight paths by collapsing categories which bystanders cannot easily perceive (such as pitch versus roll and intent behind behaviors), which is described further in Section IV.

Altitude is an additional category to describe one characteristic which was explicitly defined by each participant and is central to the understanding of the sUAS. This category is divided into increasing, decreasing, variable, and stable. That is, increasing or decreasing flight paths will end higher or lower from where they started, variable refers to paths that both increase and decrease, and stable refers to paths that are fixed in the Z plane.

Motion is another dimension that classifies the general shape of the gesture's movements. The gesture could consist of only rectilinear, curvilinear, rotational, or a combination of two or more of those sub-categories.

State	Most Common Flight Path	Count
Attract Attention	Up-down; Horizontal Circle; Descending	4
Sensor Lost	Yaw	4
Low Battery	Up-down	8
Signal Lost	Yaw	4
Area of Interest	Horizontal circle	8
Missed Goal/Target	Horizontal circle	6
Landing	Descending	13

TABLE II: The most popular flight path for each of the seven states, and the number of participants who designed those flight paths. Attract Attention had a three-way tie, with each flight path chosen by four participants.

E. User-Defined Gestures for sUAS Communications

Using the 140 gestures designed by our participants, we generated a combined gesture set for our seven states. We then grouped gestures with common features according to our taxonomy, and used these groupings to calculate our agreement scores to determine the level of consensus. Finally, we chose the most common gesture for each state as the representative gesture for that state as shown in Fig. 2. The most commonly designed flight paths are listed for each state in Table II.

Five flight paths were prevalent in the gesture set: descending, horizontal circle, up-down, left-right, and yaw. Thirteen participants used descending to indicate Landing and seven participants used descending to indicate Low Battery. Nine participants used a horizontal circle for Area of Interest. Eight participants generated an up-down flight path to indicate Low Battery. Four participants used yaw for Signal Lost and four used yaw for Sensor Lost. For Attract Attention, three groups stood out: four participants used horizontal circle, four used up-down, and three used left-right. Three groups also stood out for Sensor Lost: four participants used yaw, three used left-right, and three used descend.

For each of the categories, participants tended towards a single-entry subcategory from our taxonomy, and the most common subcategories are shown for each state in Table III. Due to strong agreement among several of the taxonomy subcategories, we can make some inferences about how to communicate states more intuitively. For example, participants preferred simple and stable altitude gestures for Attract Attention, but preferred compound and decreasing or variable altitude gestures for Sensor Lost. Rectilinear gestures were preferred for most states, with the exception of curvilinear gestures for Area of Interest and Missed Goal/Target.

Table III also suggests several other insights about novice users' preferences. Most of the gestures could be classified as simple, indirect, random, some combination of roll and throttle and pitch, stable altitude, and rectilinear. These subcategories could also represent types of gestures which bystanders could more easily perceive and interpret. Curvilinear motion predominated for only two states, Area of Interest and Missed Goal/Target, both of which are mission-centric. Similarly, compound motion predominated for only two states, Low Battery and Signal Lost, both of which are vehicle-centric. Lastly, participants generally preferred random cyclicity motions for all seven states, and showed a preference for indirect motions except for Landing. Based on these observations, we can conclude that for all seven sUAS states, participants generally preferred: simple (Complexity); indirect (Space); random (Cyclicity); some combination of roll, throttle, and pitch (Command); stable (Altitude); and rectilinear (Motion).

F. Inter-rater Reliability for Taxonomy

In order to assess the usefulness of the taxonomy categories and to classify the individual states according to common subcategories, two raters were obtained to independently assign each of the 140 user-generated flight paths to a single subcategory within each taxonomy category. After their independent assessments, their results were compared in order to calculate Cohen's Kappa and assess their agreement according to [11].

Complexity (0.881), Motion (0.907), Command (0.92), and Altitude (0.914) were considered "Almost Perfect" agreement, while Space (0.79) and Cyclicity (0.641) were considered "Substantial Agreement." Of note is the wide distribution of participant agreement within categories (as shown in the parentheses of Table III), with 17 participants designing "Simple" paths for Area of Interest compared to a relatively even split between "Simple" and "Compound" for Sensor Lost. Similarly, though not reflected in the table, Signal Lost had a three-way tie in the Command category between "Yaw," "Throttle," and "Roll and Pitch" with 4 participants' paths in each compared to Low Battery which had 13 participants using only "Throttle." These distributions, particularly those with high agreement among the participant paths, suggest common perceptions about *how* to convey these states even if designers would like to stray from the very straightforward paths with high agreement.

IV. Results

Our agreement scores for each state, after analyzing all 140 gestures and categorizing them according to our new taxonomy, are shown in Table IV along with whether the state is "mission-centric" or "vehicle-centric." A state is mission-centric if that state needs to be communicated to an operator or bystander to indicate the sUAS has encountered difficulty completing its mission. A state is vehicle-centric if it simply reflects an error state that is inherent to the sUAS and irrelevant to its mission.

As an intermediary step, we combined gestures into common flight paths with respect to drone orientation or bystander perspective. That is, because the drone will not be aware of the location of a bystander and a bystander will not be aware of the drone's orientation, flight paths such as left-right and front-back were combined into a single gesture for the purposes of calculating agreement scores. Such combinations were *not* used in the inter-rater reliability scoring for our taxonomy.

Ruiz implicitly suggests that the minimum threshold for consensus is 0.1 [16], and participants achieved that level of consensus for all seven states. Our participants achieved a stronger agreement score of 0.2 or higher for three states: Low Battery, Area of Interest, and Landing. Of note, Landing had the strongest agreement score by far, and Landing is the sUAS state which is arguably the most important for a bystander to understand and act accordingly.

As part of the survey between construction and observation of the flight paths, we recorded participants' opinions on whether they were confident their gestures would effectively communicate the seven states to other bystanders. Their confidence was measured on a 1-5 Likert scale, with average confidence shown in Table IV. Participants were most confident in their gestures for Attract Attention (4.4), Landing (4.2), and Area of Interest (4.05).

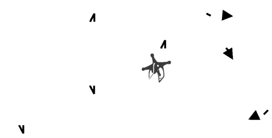
After participants viewed their gestures replayed on the sUAS, we asked them to indicate which states they were

TABLE III: Summary of the taxonomic subcategories with most participant agreement (as classified by the raters) for each state. Bold text indicates strong agreement amongst participants.

State	Complexity	Space	Cyclicity	Command	Altitude	Motion
Attract Attention	Simple (12)	Indirect (9)	Random (10)	**Roll and Throttle (7)**	Stable (8)	Rectilinear (12)
Sensor Lost	Simple (9)	Indirect (11)	Random (10)	**Roll and Throttle (6)**	Stable (10)	Rectilinear (10)
Low Battery	**Compound (12)**	Indirect (12)	Random (11)	**Throttle (13)**	Tie: **Decreasing (8)** and **Variable (8)**	Rectilinear (15)
Signal Lost	**Compound (11)**	Indirect (13)	Random (15)	No Majority	Stable (13)	Rectilinear (9)
Area of Interest	Simple (17)	Indirect (13)	Random (14)	**Roll and Pitch (9)**	Stable (13)	**Curvilinear (10)**
Missed Goal/Target	Simple (14)	Indirect (13)	Random (16)	**Roll and Pitch (7)**	Stable (14)	**Curvilinear (9)**
Landing	Simple (11)	**Direct (10)**	Random (12)	**Throttle (12)**	**Decreasing (16)**	Rectilinear (14)

(a) Horizontal circle flight path. Nine participants used this path to indicate Area of Interest and four used it to indicate Attract Attention.

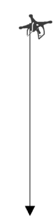
(b) Yaw flight path. Four participants used this path to indicate Sensor Lost and four used it to indicate Signal Lost.



(c) Spiral flight path. One participant used this path to indicate Signal Lost.

(d) Up-and-down flight path. Eight participants used this path to indicate Low Battery and four used it to indicate Attract Attention.

(e) Left-and-right flight path. Three participants used this path to indicate Attract Attention and three used it to indicate Sensor Lost.



(f) Descending flight path. Thirteen participants designed this path to indicate Landing and seven used this path to indicate Low Battery.

Fig. 2: User defined gestures to convey requested states.

most and least confident about in their gestures' ability to communicate those states. Attract Attention was the state for which participants expressed the most confidence (12 of 20 participants), and Signal Lost had the least confidence (eight of 20 participants). These ratings are also consistent with the Likert ratings assigned after creating the gestures.

State	Agreement Score	Type	Confidence
Attract Attention	0.155	mission	4.4
Sensor Lost	0.125	vehicle	3.2
Low Battery	0.245	vehicle	3.5
Signal Lost	0.125	vehicle	3.2
Area of Interest	0.265	mission	4.05
Missed Goal/Target	0.145	mission	3.5
Landing	0.455	vehicle	4.2

TABLE IV: Agreement scores across all participants and average individual confidence for gestures elicited from participants for each of the seven states. Individual confidence reflects the mean Likert value (1-5, 5 being "strongly agree") from participants that they were "confident in their gesture for [state]" after describing it to the experimenter.

V. Discussion

Our elicitation study with 20 participants resulted in the development of the first user-designed gesture set for sUAS to communicate states. Three states had relatively strong agreement scores: Landing (0.455), Area of Interest (0.265), and Low Battery (0.245). The agreement scores were also reflected in the participant confidence in the ability of their gestures to communicate the state for Area of Interest, Landing, and Attract Attention. These results are encouraging, because those states are arguably the most important for bystanders to understand and act accordingly, especially as sUAS become more prevalent in everyday consumer services. Further research into sUAS gestures can help reduce public concerns about increasing drone interactions, and engineers should be able to implement these gestures on pre-existing sUAS without special hardware.

In order to generalize the gestures designed by the participants, the common and uncommon subcategories from the taxonomy give us guidance about their underlying thought process when considering those states. For example, the use of curvilinear motion for Area of Interest and Missed Goal/Target shows an agreement from participants that these states are likely circling around a space rather than transitioning to another space as in the other states. Another interesting note is that most of these states would require action on the part of the operator or bystanders and the participants defaulted to Indirect motions, which [18] recommended to increase valance or arousal, in contrast to the Landing state which was Direct.

Participants expressed some difficulty when creating communicative flight paths, which implies creating a generalized communication system of gestures for sUAS is challenging. One recommendation would be to show participants video vignettes about the states to provide context and make them easier to understand. For example, one vignette could ask the participants to imagine they are walking a dog and they encounter a drone attempting to deliver a package to a neighbor's house, but the drone experiences an error and must land immediately. Then the participants would be asked to describe a gesture the drone should take to indicate Landing.

Another insight is that our participants had likely never seen some of these failure states in sUAS. We assumed they would understand all seven states based on previous interactions with smart phones, tablets, or hobbyist drones. It might be better to explain the failure states, possibly also through video vignettes, to make sure participants had a common understanding of the states. However, this creates the risks of anchoring participants to certain gestures, or influencing them to create gestures which were similar to what they saw in the videos.

There were a handful of outlier gestures which were more creative, such as the horizontal spiral path shown in Fig. 2c which one participant used to indicate Signal Lost. Other outliers included making a figure 'X' in vertical space for Signal Lost, and making two vertical squares (one parallel, and one perpendicular to the operator) for Area of Interest.

VI. Threats to Validity and Limitations

The primary threat to the validity of this study is that we attempted to create an objective taxonomy to classify sUAS gestures from members of the general public, and we grouped them according to our own taxonomy. Other researchers could easily develop their own different taxonomies and produce different agreement scores, but we nevertheless believe our taxonomy and classifications are a fair representation of consensus for several flight paths and states. In particular, we successfully recorded and replayed 136 out of 140 of the handwritten gestures, and easily fixed the other four quickly to the satisfaction of the participants.

A minor threat to validity is that all users were presented the states in the same order and with the same amount of time. It is possible some participants felt pressure to create their gestures within the specified time limit, which could have resulted in fewer distinct gestures. However, all participants except one completed their handwritten drawings within the time limit and most had several minutes leftover.

This study is inherently limited by the problem of communicating states and sUAS possible flight paths to members of the general public. This was evident when a few participants suggested upward motions to indicate Low Battery, reflecting understandable ignorance about sUAS energy consumption. Even participants with prior drone experience were not necessarily experts with sUAS. Some participants struggled to create gestures for the states likely due to this lack of familiarity with sUAS. If all participants were experts with sUAS, there might have been less consensus because they would have understood the full set of possible maneuvers the sUAS could perform and might have created more varied gesture sets. However, the goal of this study was to create a gesture set which general bystanders could understand.

VII. Future Work

A natural next set of follow-up experiments would be to conduct "reverse" elicitation studies to determine whether the gestures could be understood to communicate states, either through video or in-person replays of gestures. One such study was performed in [5] on an initial set of gestures, but could be refined based on this work. Other research could assess whether these states can be more effectively communicated or achieve higher consensus with specialized hardware to include light and sound.

Based on the relative preferences of the taxonomy subcategories shown in Table III, it would be worth investigating a dichotomy of gesture sets for sUAS states depending upon whether the states are mission-centric or vehicle-centric. The intuition is that bystanders would expect different classes of sUAS gestures to indicate a mission-centric error versus a vehicle-centric error. Vehicle-centric errors are more likely to require bystanders to avoid specific areas as opposed mission-centric errors which might require operator interaction. If that is the case, vehicle-centric gestures arguably should take less time perform than mission-centric gestures in order to be effective. Furthermore, Table III suggests certain types of motions might be more natural or intuitive for bystanders to interpret. Follow-up studies could help confirm this by investigating whether gestures from minority subcategories are effective at communicating sUAS states.

Further research, such as the previously mentioned "reverse" elicitation study, could be performed to determine whether the more creative outlier gestures can efficiently communicate states, because some fit neatly within the preferred subcategories from our taxonomy. That is, the spiral flight path chosen for Signal Lost in Fig. 2c is simple, indirect, random, roll and pitch, stable altitude, and curvilinear. These subcategories were also preferred for Area of Interest and Missed Goal/Target.

VIII. Conclusion

This paper presented an elicitation study to elicit gestures from participants recruited from the general public to communicate seven key sUAS states to operators and especially bystanders. The agreement scores showed promise that a common gesture set can be created and implemented for current sUAS. Future work could refine the gesture set and confirm it conforms to how bystanders would expect sUAS to behave, hopefully increasing the general public's acceptance of increased commercial drone usage in everyday life.

The concept of elicitation studies can be broadly applied within the social robotics community to understand the naive assumptions that are common amongst users, classify them, and then make recommendations for candidates to test through wide-scale user studies.

References

[1] Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. Drone and me: An exploration into natural human-drone interction. 2015.

[2] J.R. Cauchard, K.Y. Zhai, M. Spadafora, and J.A. Landay. Emotion encoding in human-drone interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2016.

[3] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. 2000.

[4] A. Dragan and S. Srinivasa. Integrating human observer inferences into robot motion planning. *Autonomous Robots*, 2014.

[5] Brittany A Duncan, Evan Beachly, Alisha Bevins, Sebasitan Elbaum, and Carrick Detweiler. Investigation of communicative flight paths for small unmanned aerial systems. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 602–609. IEEE, 2018.

[6] Brittany A Duncan and Robin R Murphy. Effects of speed, cyclicity, and dimensionality on distancing, time, and preference in human-aerial vehicle interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):13, 2017.

[7] Brian Gleeson, Karon MacLean, Amir Haddadi, Elizabeth Croft, and Javier Alcazar. Gestures for industry: intuitive human-robot communication from human observation. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 349–356. IEEE Press, 2013.

[8] C. Huang and B. Mutlu. Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems (RSS)*, 2013.

[9] Mishel Johns, Brian Mok, David Michael Sirkin, Nikhil Manjunath Gowda, Catherine Allison Smith, Walter J Talamonti Jr, and Wendy Ju. Exploring shared control in automated driving. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 2016.

[10] Robert M Krauss, Palmer Morrel-Samuels, and Christina Colasante. Do conversational hand gestures communicate? *Journal of personality and social psychology*, 61(5):743 1939–1315, 1991.

[11] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[12] Alan Levin. Alphabet and chipotle are bringing burrito delivery drones to campus, 2016.

[13] V. Ng-Thow-Hing, P. Luo, and S. Okita. Synchronized gesture and speech production for humanoid robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[14] G. Prati and L. Pietrantoni. The semantic specificity of gestures when verbal communication is not possible: the case of emergency evacuation. *International Journal of Psychology*, 2013.

[15] L. D. Riek, T. C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson. Cooperative gestures: Effective signaling for humanoid robots. In *In Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 61–68, 2010.

[16] Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. 2011.

[17] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin. Generation and evaluation of communicative robot gesture. *International Journal of Social Robots*, 2012.

[18] M. Sharma, D. Hildebrandt, G. Newman, J.E. Young, and R. Eskicioglu. Communicating affect via flight path: exploring use of the laban effort system for designing affective locomotion paths. In *Human-Robot Interaction (HRI), ACM/IEEE International Conference on*, 2013.

[19] Curtis Silver. Amazon prime air delivery drones fly into our hearts with a package of 5 pounds or less, 2016.

[20] D. Szafir, B. Mutlu, and T. Fong. Communication of intent in assistive free flyers. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2014.

[21] D. Szafir, B. Mutlu, and T. Fong. Designing mechanisms for communicating directionality in flying robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2015.

[22] Gilbert Tang, George Charalambous, Philip Webb, and Sarah R Fletcher. Users' understanding of industrial robot gesture motions and effects on trust. *Contemporary Ergonomics and Human Factors*, pages 116–123, 2014.

[23] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. Maximizing the guessability of symbolic input. 2005.

[24] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. 2009.