

Drone & Me: An Exploration Into Natural Human-Drone Interaction

Jessica R. Cauchard

Jane L. E

Kevin Y. Zhai

James A. Landay

Stanford University, Department of Computer Science
353 Serra Mall, Stanford CA 94305-9035
{cauchard, ejane, kzhai, landay} @stanford.edu

ABSTRACT

Personal drones are becoming popular. It is challenging to design how to interact with these flying robots. We present a Wizard-of-Oz (WoZ) elicitation study that informs how to naturally interact with drones. Results show strong agreement between participants for many interaction techniques, as when gesturing for the drone to stop. We discovered that people interact with drones as with a person or a pet, using interpersonal gestures, such as beckoning the drone closer. We detail the interaction metaphors observed and offer design insights for human-drone interactions.

Author Keywords

Drone; UAV; quadcopter; Wizard-of-Oz; elicitation study.

ACM Classification Keywords

H.5.2. Information interfaces and presentation: User Interfaces: User-centered design.

INTRODUCTION

Personal drones are becoming increasingly present in our everyday environments. They are primarily being used for outdoor activities, such as film capture, agriculture, Search and Rescue, entertainment, and delivery. In the future, we expect that drones will become partly, if not fully, autonomous, and they will be able to support people in their everyday lives. Even with full autonomy, however, people still need to communicate with personal drones and have the control needed to make requests and express intention.

Imagine a personal trainer drone, that could accompany a user on a run [12], giving real-time feedback. As a tour guide, it could adjust language and show points of interest. In either scenario, it is unwieldy for the user to control the drone with a remote while paying attention to the setting. In addition, as drones become autonomous, remotes become redundant since the drone can compute its optimal path without needing any user input. Also, in collocated

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp '15, September 7-11, 2015, Osaka, Japan.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3574-4/15/09...\$15.00.

<http://dx.doi.org/10.1145/2750858.2805823>

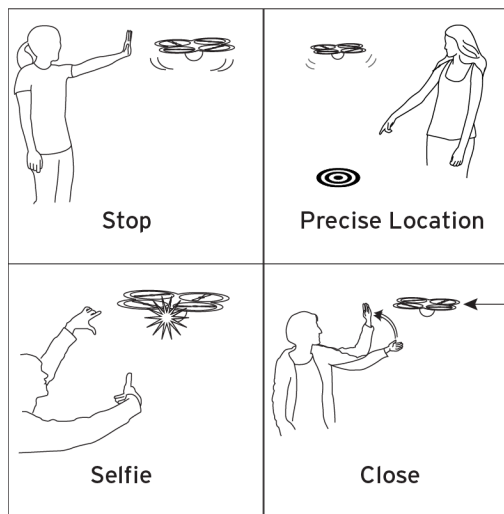


Figure 1. Example of user-defined gestures (high agreement)

scenarios it is unnatural to use a remote to interact with an agent that people treat like an intelligent being [14].

The need for natural interaction is supported by the Human-Robot Interaction (HRI) literature. As drones have different characteristics than ground robots, such as not allowing touch interaction, it is unclear whether existing techniques can be adapted to flying robots. Our user-centric design strategy seeks to understand how users naturally interact with drones. We present a 19-participant WoZ elicitation study that shows that users felt extremely comfortable interacting with a drone. Participants used metaphors drawn from interacting with a person or a pet, called for the drone by name, encouraged it, and trusted it enough to bring it to an almost unsafe distance. We see that the preferences in the use of voice and gestures vary across tasks, leading us to conclude that no single modality would provide suitable natural interaction. We consider multimodal interaction a major challenge in the future and conclude by presenting design insights for Human-Drone Interaction (HDI).

RELATED WORK

A variety of work investigates the nature of HRI. Projects explored multimodal [1], gestural [4], and even prop-based interactions techniques [8]. Guo et al. [5] show people feel comfortable using a variety of interaction techniques. Specifically, with a robot dog, people felt it was more natural to use a Wii controller than a keypad.

HDI applications vary from running with drones [11, 12], filming [7], creating flying displays [15, 18], and looking at dynamically rechargeable flying objects [9]. Researchers have also explored communicating feedback, such as intent and directionality [19, 20]. Yet, there are major differences between traditional robots and drones, which can fly freely, and cannot safely be touched, requiring new interaction techniques that are well suited for drones.

Prior work investigated controlling a drone using face poses and hand gestures [13] as well as a multimodal falconry metaphor [14]. This last project leads us to believe that drones can be “socially” adapted and accepted. Others studied upper-body gestural interaction in a controlled lab where users were given specific interaction metaphors [16]. In contrast, we do not specify the type of interaction to obtain user-defined gestures.

There is a history of user-defined interaction techniques and gesture elicitation studies for new technology [2, 3, 10, 22], such as tabletops [23], mobile devices [17], and TVs [21].

USER STUDY

To explore interactions and better understand the metaphors and relationships that occur when users interact with drones, we ran a user-defined interaction elicitation study.

Methodology

We simulated the drone autonomous behavior and reactions to user input. We chose an outdoor space to operate the drone safely and with added flexibility compared to indoors or to pre-programming the drone’s movements. The experimenter using the remote stayed behind the user but could not be fully hidden to keep direct sight of the drone and participant for safety reasons. We found that even with the WoZ, users felt in control of the drone. Each task was described on a card to avoid verbally biasing the users’ actions and modality choices (Table 1). Users were asked to perform any action to get from start to end.

Name	Following
Start	The drone is flying around
End	The drone is following you

Table 1. Example of a task as written on a card.

To let users interpret the task freely, we did not show the effect of the actions (*referents* in [23]). For example, when getting the drone to follow, its comfortable relative position with respect to the participant was different for each person and seeing the referent could have biased the interaction.

Participants

19 volunteers (12 m), 19 to 38 y.o. ($\mu=25$) were recruited from our institution and nearby companies. Their training was in engineering (7), CS (6), other sciences (1), and non-scientific fields (5). They were rewarded \$15 for their time.

Apparatus, Setting, and Tasks

We used a DJI Phantom 2 (29x29x18cm) with prop guards around the propellers. The study was run outdoors, partially

protected from the wind by trees and a building. 18 tasks (Table 3) with different levels of complexity were presented in a random order to avoid interaction effects and to minimize the impact of learning and fatigue.

Procedure

Each session lasted 1 to 1.5h. One experimenter filmed and interviewed participants while another controlled the drone. Participants were informed of the WoZ and we emphasized that they should ignore the experimenters’ presence when interacting with the drone. We asked participants to not worry about technical capabilities and to interact in the manner that felt most natural for each task.

The instructions and the task cards (face down) were positioned on a stand. The participant picked up the top card and read it out-loud to confirm that they had understood the task. They would then interact with the drone. After each task, the participant was prompted to recall and explain their actions using a post-task think-aloud technique. Participants also rated their interaction in terms of suitability and simplicity. After completing the 18 tasks (Part 1), the participant was given a sheet with suggestions for interaction techniques. They were then asked to complete a second time 4 representative tasks (Part 2) that covered a range of category types and complexity (Table 3, blue italic). In Part 2, we looked at whether participants changed their interaction strategy after given suggestions.

RESULTS

The data collected includes transcripts, videos, and post-task and post-experiment interviews during which we collected qualitative feedback on the users’ experience.

User-Defined Gesture Sets per Task Type

Out of the 418 interaction tasks, 4 were misunderstood by the users, and removed from further analysis. We found 216 unique interactions: 96 gestures (body gestures not restricted to hands and arms), 59 sounds, 53 combinations of gesture and sound, and 8 with a prop. Given the low usage of props, we do not count them in further analysis. We use gesture and sound only as they encompass the vast majority of interactions, as shown in Table 2¹.

Tasks Performed	Gesture	Sound	Both
All	86%	38%	26%
Representative tasks (Part 1)	88%	37%	28%
Representative tasks (Part 2)	70%	57%	33%

Table 2. Percentage of use of interaction modalities.

Many participants initially expressed discomfort in talking with the drone. Gestures are quick and allowed for precise adjustments and continuous control throughout the interaction. Over the course of the study, participants felt more confident in the drone’s ability. As illustrated by the

¹ Note that rows do not sum to 100% because interactions that used both modalities are counted in all columns.

increase from 37% to 57% in using sound for the representative tasks in Part 2, participants started giving voice commands. This suggests that rapport can be established, allowing humans to accept collocated drones.

We determine an *agreement score* per referent, A_r , per modality type, based on [22] where P_r is the set of proposed interactions for referent r , and P_i is the subset of identical interactions for that referent.

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

Table 3 summarizes the agreement scores per task and modality. For calculating agreement, we consider gesture and sound separately [10] to avoid overlap in counting individual interactions and using both modalities simultaneously. In a few cases, some participants built a sequence of gestures/sounds into their interaction. Each interaction is counted separately, resulting in some over-counting, causing some agreement scores to be greater than 1 (impacted scores highlighted in grey). Based on prior work, scores above 0.5 show strong agreement (44% of interactions, in bold).

Category	Task name	Gesture	Sound	Both	
Navigation	Within body frame	Fly closer	0.80	0.61	0.28
		<i>Fly higher (to user's height)</i>	0.24	0.72	0.38
		Fly lower (from user's height)	0.44	0.69	0.38
		Fly sideways (small delta)	0.40	0.38	0.50
		Stop by me	0.87	0.72	0.68
	Outside body frame	Fly further away (far)	0.37	0.48	0.36
		Fly sideways (large delta)	0.31	1.00	1.00
		<i>Fly to a precise location</i>	0.65	0.33	0.33
		Fly higher	0.30	0.69	0.33
		Fly lower	0.23	0.67	---
Action	General motion	Stop motion (when flying)	1.00	0.80	0.68
		Land	0.28	1.16*	0.33
		Take off	0.32	0.28	0.28
	Relative to user	<i>Follow</i>	0.56	0.42	0.43
		Stop following	0.51	0.28	0.22
	Photo	Get attention	0.26	0.72	0.22
		<i>Take a 'selfie'</i>	0.37	0.63	0.19
	Take a picture of a tree	0.94*	0.85	0.59	

Table 3. Tasks and agreement scores per modality.

Navigation Strategies

For navigation, most people used a repeated waving or continuous sweep, mapping the drone's movement directly to their arm (as a pointer with a line extending out from the hand). Within the body range, people were more likely to use smaller motions with additional interactions using body parts as reference frames for the drone's target flying level. For specific locations, users initially hesitated pointing at the target or describing it verbally because of the lack of precision. As trust increased, they were more willing to depend on the drone's ability to identify its spatial context.

Pictures

Taking a photo is one of the most complex tasks. For selfies, many realized that on top of framing and focusing, a counter would be useful. This caused people to use sound to avoid interfering with posing (8 participants in Part 1 and 12 in Part 2). Yet, agreement scores were high across modalities. Almost all used the word "picture" for both selfies and photos. People often were less confident in their choice of gestures here, but in fact there were two that almost everyone used: 1. using two hands as a frame, 2. holding an invisible camera and clicking the shutter button. People also suggested having a screen on the drone or the use of props to frame or adjust camera settings.

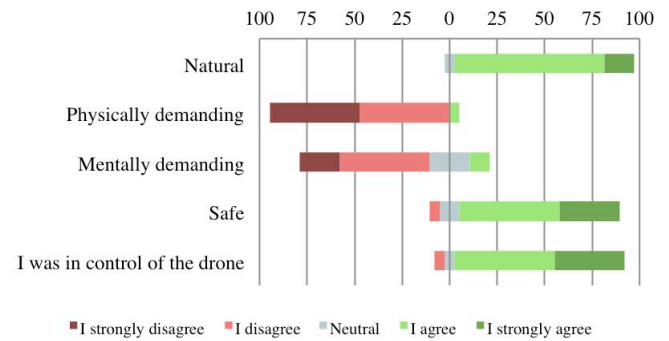


Figure 2. Subjective ratings in percentage of participants for interacting with the drone using a 5-point Likert scale.

Qualitative Survey Data

Figure 2 shows subjective ratings for how natural, physically and mentally demanding the interaction was, how safe it felt and whether the participant felt in control.

90% of the participants stated they felt they were in control and 95% felt it was natural to interact with the drone. None of the participants reported being tired or feeling fatigue at any point in the study. Most participants were keen on continuing to interact with the drone. People commented on their own faults in properly directing the drone. "It landed rather roughly and I didn't mean for it to, but that's also because I didn't really know what I was doing with my hand gesture" [P9] or on the drone's abilities "I don't know if the drone is looking at me" [P1], [I ran] "just to see if the drone would catch up to me" [P3].

This illustrates that the WoZ was well received and did not affect the results of the study. Some mentioned being careful about not losing control [P8], worried about the drone's height as "it could fall and break" [P16], losing sight of the drone [P3, P14], or vice versa: "It should know and be smart to not get lost, like a dog" [P13].

DESIGN INSIGHTS

Throughout the study, we found trends in terms of the interaction metaphors participants used as well as the feedback they gave on aspects of the interaction that they desired in the future. This section presents design insights based on these trends.

Interaction Metaphors

Several interaction metaphors emerged during the study. We observed that users treated the drone as if it were an animate being: a person, a group of people, or even a pet.

Interacting with a Person

The most popular metaphor was interacting with the drone as a person. All but one mentioned this analogy. We observed this phenomenon in the words used when talking to the drone: “ok, we’re good”, “let’s go”, “come this way” [P2], “please” and “thank you” [P5], along with the word “respect” when explaining their interaction in the post-task interviews and the fear of being impolite [P6]. One participant asked if they “should have been more gentle [with the drone]” [P7]. Some users cited task-specific metaphors, such as navigation tasks that felt like helping someone park [P2, P3, P8]; getting the drone’s attention was like getting attention in a classroom [P1, P11] or a stadium [P16]; having the drone follow them would be similar to leading a tour group [P16]; and getting the drone to stop moving or following was like in the army [P8].

Interacting with a Pet

Another popular metaphor was interacting with the drone as if it were a pet, mentioned by 16 out of 19 participants. Most participants compared it to a dog: “*I’m almost starting to command it like I would a dog. Like, ‘stay, go over there, go fetch.’*” [P18]. Some participants also referred to the drone as mosquitos [P18] or bumblebees [P7] due to the noise it makes, while others thought of it as a bird [P1, P12]. We also saw this interaction strategy when participants called the drone by whistling at it like they would a dog [P5], talking about its “under-belly” [P10], and saying “all right boy” [P13] and “good job”, “good drone” [P11] when the drone did what was expected.

Naming the Drone

The pet interaction metaphor continued when [P8] said he would call the drone Nick, after his own dog. [P2] gave the drone an ID number, seven other participants decided to call it “Drone”, while [P5] felt that “Ferdie” would be more appropriate. Similarly [P11] said she would call the drone to get its attention the same way she would call a friend.

Safety and Proxemics

When designing the study, one main concern was to ensure the participants’ safety. We thought users might be afraid of the drone and be uncomfortable interacting with it. Instead, their reactions could not be further from our expectations. 16 participants reported feeling safe interacting with the drone (Figure 2). Some appeared more concerned about the drone’s safety [P13, P15, P16]. “*I’m not really worried to get hurt, but I don’t want to also hurt the drone*” [P17]. Similarly, as users became comfortable with the drone, they got closer to it than we expected. In our preliminary look at proxemics, 7 participants brought the drone within their *intimate space* (1.5ft), 9 in their *personal space* (4ft), only 3 preferred to have the drone in their *social space* (10ft) at closest, and none in the *public space* (> 10ft) [6].

We found several factors created discomfort, such as propeller noise or the wind they generated, and resulted in three users preferring the drone to be located in their *social space*. When asked what aspect of the drone made users uncomfortable, these participants again worried more about the drone’s safety than their own. Participants also built trust with the drone, ensuring that it would stop when they would ask it to do so, “*The more that I learn to trust it, the more I would feel comfortable not saying as much*” [P12].

Feedback

We did not implement a feedback system for this study. We found in a pilot study that participants wanted feedback for the “Take a photo” task. We added a nod where the drone tilts forward to signal that a photo was taken. Some users commented on the drone not being as still as they expected when stopped (hovering) and that an additional confirmation would help their interaction. Participants were specific about the type of feedback they wanted, using light or sound confirmations, such as a shutter sound for pictures [P5, P14], or responding “ok” or “I’m leaving” [P6]. [P13] suggested adding eyes to see where the drone is looking. Others asked for a display of what the drone is seeing (i.e., camera feed) on the drone itself [P9, P10], on a tablet/phone [P11], or on a head-mounted display [P17].

Emergency Landing

Several participants mentioned they would like an interface for emergency landing in case anything was going wrong or they had to immediately stop their activity.

FUTURE WORK

The next step is the implementation of the best technical solution to support multimodal Input/Output for HDI, taking into consideration the need for multiple modalities based on the context of use. The scenarios of use we described can then be implemented. We will also look at proxemics in the 3D space for HDI. As several users mentioned feeling attached to the drone, it would be interesting to further study human emotion towards drones and how they differ from interacting with ground robots.

CONCLUSION

Given the current increase in popularity of personal drones, we need to create natural interaction techniques to best support users. We executed a WoZ elicitation study with nineteen participants performing a range of tasks with different levels of complexity, and found high agreement scores between participants on how they naturally interact with the drone. We found strong agreement on nearly half (44%) of the gesture, voice, and multimodal interactions that felt intuitive to participants. This was due to most participants interacting with the drone in a similar way to how they would with a person or a pet. We contribute a set of design insights to develop Human-Drone Interaction. We expect drones to become smaller and quieter so that they will resemble humming birds, flying by the user and coming into play when needed. Giving people natural, easy control will enable incorporating drones into our daily lives.

REFERENCES

1. Alvarez-Santos, V., Iglesias, R., Pardo, X.M., Regueiro, C.V. and Canedo-Rodriguez, A. 2014. Gesture-based interaction with voice feedback for a tour-guide robot. *J. Vis. Comun. Image Represent.* 25, 499-509.
2. Connell, S., Kuo, P.-Y., Liu, L. and Piper, A.M. 2013. A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*, 277-280.
3. Cooke, N.J. 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies.* 41, 801-849.
4. Ende, T., Haddadin, S., Parusel, S., Wusthoff, T., Hassenzahl, M. and Albu-Schaffer, A. 2011. A human-centered approach to robot gesture based communication within collaborative working processes. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'11)*, 3367-3374.
5. Guo, C. and Sharlin, E. 2008. Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 121-130.
6. Hall, E.T. 1969. *The hidden dimension*. Anchor Books New York.
7. Higuchi, K., Ishiguro, Y. and Rekimoto, J. 2011. Flying eyes: free-space content creation using autonomous aerial vehicles. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, 561-570.
8. Ishii, K., Zhao, S., Inami, M., Igarashi, T. and Imai, M. 2009. Designing Laser Gesture Interface for Robot Control. In *Proceedings of INTERACT '09*, 479-492.
9. Kyono, Y., Yonezawa, T., Nozaki, H., Ogawa, M., Ito, T., Nakazawa, J., Takashio, K. and Tokuda, H. 2013. EverCopter: continuous and adaptive over-the-air sensing with detachable wired flying objects. In *Proc. of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp '13 Adjunct)*, 299-302.
10. Morris, M.R. 2012. Web on the wall: insights from a multimodal interaction elicitation study. In *Proc. of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*, 95-104.
11. Mueller, F., Graether, E. and Toprak, C. 2013. Jogobot: jogging with a flying robot. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*, 2845-2846.
12. Mueller, F.F. and Muirhead, M. 2015. Jogging with a Quadcopter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 2023-2032.
13. Nagi, J., Giusti, A., Caro, G.A.D. and Gambardella, L.M. 2014. Human Control of UAVs using Face Pose Estimates and Hand Gestures. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI'14)*, 252-253.
14. Ng, W.S. and Sharlin, E. 2011. Collocated interaction with flying robots. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '11)*, 143-149.
15. Nozaki, H. 2014. Flying display: a movable display pairing projector and screen in the air. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*, 909-914.
16. Pfeil, K., Seng Lee, K. and LaViola, J. 2013. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*, 257-266.
17. Ruiz, J., Li, Y. and Lank, E. 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 197-206.
18. Scheible, J., Hoth, A., Saal, J. and Su, H. 2013. Displaydrone: a flying robot based interactive display. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays (PerDis '13)*, 49-54.
19. Szafir, D., Mutlu, B. and Fong, T. Communication of intent in assistive free flyers. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI '14)*, 358-365.
20. Szafir, D., Mutlu, B. and Fong, T. 2015. Communicating Directionality in Flying Robots. In *Proceedings of ACM Human-Robot Interaction (HRI '15)*, 19-26.
21. Vatavu, R.-D. and Zaiti, I.-A. Leap gestures for TV: insights from an elicitation study. 2014. In *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '14)*, 131-138.
22. Wobbrock, J.O., Aung, H.H., Rothrock, B. and Myers, B.A. 2005. Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*, 1869-1872.
23. Wobbrock, J.O., Morris, M.R. and Wilson, A.D. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, 1083-1092.