AN EXPLORATION OF UNMANNED AERIAL VEHICLE DIRECT MANIPULATION
THROUGH 3D SPATIAL INTERACTION

by

KEVIN PFEIL
B.S. University of Central Florida, 2010

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science, Computer Science
in the Department of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2013

ABSTRACT

We present an exploration that surveys the strengths and weaknesses of various 3D spatial interaction techniques, in the context of directly manipulating an Unmanned Aerial Vehicle (UAV). Particularly, a study of touch- and device- free interfaces in this domain is provided. 3D spatial interaction can be achieved using hand-held motion control devices such as the Nintendo Wiimote, but computer vision systems offer a different and perhaps more natural method. In general, 3D user interfaces (3DUI) enable a user to interact with a system on a more robust and potentially more meaningful scale. We discuss the design and development of various 3D interaction techniques using commercially available computer vision systems, and provide an exploration of the effects that these techniques have on an overall user experience in the UAV domain. Specific qualities of the user experience are targeted, including the perceived intuition, ease of use, comfort, and others. We present a complete user study for upper-body gestures, and preliminary reactions towards 3DUI using hand-and-finger gestures are also discussed. The results provide evidence that supports the use of 3DUI in this domain, as well as the use of certain styles of techniques over others.

Dedicated to my family, for their neverending support the entire way.

ACKNOWLEDGMENTS

Many thanks to my thesis advisory committee – Dr. Joseph LaViola, Dr. Gita Sukthankar, and Dr. Charlie Hughes. Special thanks to the members of UCF's Interactive Systems and User Experience Lab, especially Seng Lee Koh and Corey Pittman.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have an increasingly prominent role in society. The term "UAV" can be used to describe countless devices and aircraft, including planes, helicopters, missiles, blimps, etc., in various domains. There is no universally accepted definition of the term, except that regardless of device, there is no human on-board. In this thesis, we refer to UAVs as the robotic, unmanned, fixed-wing and rotor-wing aircraft that can be flown by a remote pilot. This form of UAV has been widely used in the military domain, providing a means for gathering intelligence of enemy activity or territory (such as the Desert Hawk[1] or Indago[2]). Various models have been developed and deployed over the recent years, and we now see UAVs such as the Predator[3] that carry explosive ordinance that can be accurately deployed towards specified targets. These aircraft are desirable due to the fact that no on-board pilot is needed - this helps decrease the likelihood of loss of life from the controlling party, and this also allows smaller aircraft to be used in a mission, which results in a stealthier operation, and even a reduced cost[4].

There is a clear benefit to using UAVs in the military, but there is also much opportunity for these devices to be used in other scenarios that can benefit the general public. With commercial UAVs costing less and less, a personal device can now be acquired for low cost. The Parrot AR Drone is one such device[5]. This quad-rotor aircraft can be flown for entertainment purposes fairly easily through a native touch-based interface, on a tablet or smart phone device. The on-board cameras allow pilots to inspect higher locations that cannot easily be reached; typical users fly the AR Drone around parks, fields, or other open spaces. The camera feeds have often been recorded and distributed, allowing viewers from around the world to see what various parts of the Earth

---

[1] http://www.lockheedmartin.com/us/products/desert-hawk.html

[2] http://www.lockheedmartin.com/us/products/procerus/quad-vtol.html

[3] http://www.ga-asi.com/products/aircraft/predator_b.php

[4] http://americansecurityproject.org/blog/2012/the-us-and-its-uavs-a-cost-benefit-analysis/

[5] http://ardrone2.parrot.com/usa/

look like from higher altitudes. This particular UAV is primarily used for entertainment, but there is much opportunity to expand its role into something more than just a recreational device.

Many companies, hobbyists, and municipal governments have identified potential applications for smaller forms of UAVs. In Germany, UAVs may be used to assist with crime deterrence; specifically, train stations may begin using these devices equipped with thermal sensors to detect graffiti misdemeanors in progress[6]. The field of agriculture can benefit from the use of an inexpensive unmanned aircraft to help detect problems with crops[7]. These examples are just two of many that have been discussed in [22], which describes even more applications outside of the military. New uses for these devices are constantly being discovered.

With a surge for the next wave of UAV applications upon us, it is beneficial to evaluate different methods of interaction, to help with providing future users with a well-rounded control experience. While R/C toys including cars and helicopters have been common for decades, some input devices that have ability to control them are comparatively new. The Nintendo Wii, introduced in 2006, was one of – if not the – first instances of a household device incorporating motion control. The Microsoft Kinect and Sony Move were introduced shortly thereafter, and all of these devices are becoming increasingly popular. While these were originally intended for video gaming, the Human-Computer Interaction (HCI) community has found alternative use for the technology [13][25][28]. We explore how commercially available motion control technology can be used to provide an interface with the Parrot AR Drone, including the Microsoft Kinect[8] and the Leap Motion[9] controller.

The problem we tackle is that of user interaction in the context of direct manipulation. Such scenarios find a system that is constantly receiving input from the user, resulting in continuous

---

[6]http://www.france24.com/en/20130527-german-railways-use-mini-drones-stop-graffiti

[7]http://www.uav-applications.org/gallery/doc/FieldTalk.pdf

[8]http://www.microsoft.com/en-us/kinectforwindows/

[9]https://www.leapmotion.com/

commands that specifically map to the translation, rotation, or other movements of an object – the teleoperation of various robotic platforms is a prime example. Direct manipulation does not heavily involve autonomy; rather, the object or system being manipulated by the user only performs tasks when explicitly requested.

Most military-grade UAVs are typically paired with some form of a ground control station (GCS), which is essentially a virtual or miniaturized cockpit that displays all of the sensor readings from the aircraft. These stations can primarily assist with navigation of the UAV by setting way-points for flight, making use of software such as the Systems Tool Kit[10]. In many GCS instances, a UAV can be directly controlled through the use of a keyboard, mouse, and/or joystick. The interaction that is used to generate way-points is not tackled in this thesis, but instead, the focus is primarily on finding new forms of input that can be used to directly navigate the UAV, by explicitly altering its navigation through translation and rotation. The use of computer vision systems is proposed, to augment the user experience in multiple ways. It does not seem intuitive to use a keyboard and mouse to navigate an aircraft that can not only move in 3D space but also rotate on 3 different axes. 3D spatial interaction, by contrast, has potential to allow this natural form of manipulation [3]. Since pairing motion control with UAVs is not yet well studied, we provide an exploration into the domain with a user study and discussion.

As mentioned, 3D spatial interaction has been used in many different domains, including video games, medicine[11], and robotics [13], with more implementations constantly being developed. Although these different domains benefit from the same technology, the interaction techniques used to control the various systems are often unique, or, at least, they do not translate well when applied elsewhere. For example, a vision system could allow a user to directly manipulate robot arms using a one-to-one mapping technique, with which the robot arms mirror the movements of the user's arms. However, this mapping may not be reliably reused in a scenario where

---

[10]http://www.agi.com/products/stk/modules/default.aspx/id/stk-free

[11]http://www.zdnet.com/blog/health/xbox-kinect-helps-surgeons-in-the-operating-room/277

the user would need to drive a virtual vehicle in a simulation, as an example. Likewise, two motion controlled video games – perhaps a first-person shooter and a real-time strategy game – would not use the same techniques due to the widely differing natures of the required playing styles. With this stated, it is important to acknowledge that some techniques have potential to map well to various domains, and proper consideration of these other techniques can still be given as part of an exploration.

Essentially, every domain and sub-domain can warrant its own types of interaction techniques. Some may be useful, while others may not be. User interface designers would be correct to study the interaction space in order to find what seems intuitive, easy, and comfortable for the end users in the specific context. We perform this exploration of the interaction space in the domain of direct manipulation of UAVs by utilizing inexpensive vision system applications.

Problem Declaration

There exists opportunity to create a better user interface for directly manipulating a UAV than what is found on a typical GCS, which mainly features the traditional keyboard, mouse, and/or joystick. We expect that by incorporating a form of 3D spatial interaction, a user will have a more natural experience with the entire system. Commands may become more intuitive, resulting in a more satisfactory experience, and perhaps a less boring one at that; recent research describes how UAV pilots are very prone to boredom, which may result in a hindered performance, as well as a negative outlook of the job [8].

Motion control has been tackled across many domains, but the gestures between them do not necessarily translate into usable substitutions for properly designed interactions. We therefore perfom an exploration into the domain UAV control using 3DUI.

4

Contributions

In order to efficiently perform this sort of exploration, interaction design must be paired with subsequent user studies. It is possible to develop interaction techniques that allow the user to completely command the UAV, but if user performance does not surpass or at least rival that of traditional input devices or pre-existing methods, then there may not be a substantial benefit to using the new interactions. Thus, we aim at evaluating the amount of time a user needs to complete a task with various techniques, as well as measuring the user disposition towards the interaction, in the form of qualitative metrics described in chapter 4. This thesis is based on our previously published paper in IUI [18].

The contributions of this work can be summarized by the following points:

- Development of five upper-body interaction techniques, with accompanying user study

- Development of three hand-and-finger interaction techniques, with preliminary design considerations

- Application of the developed gestural commands towards the flight of a quadrotor UAV

- Discussion of strengths and weaknesses of two computer vision system devices used for tracking - Microsoft Kinect and Leap Motion controller

Reader's Guide

Chapter 2 - A review of relevant literature in the fields of HCI, HRI, and VR are presented. Some work has been conducted to command UAVs with alternative modes of input, and this section describes this work, as well as parallel HRI and VR applications, including design and analysis of various interaction techniques.

5

Chapter 3 - Interaction techniques discussed throughout the thesis are described in this section. These techniques include upper-body and hand-and-finger gestures that are processed, to derive commands that are then communicated to the UAV.

Chapter 4 - The design and execution of a user study is described. This study targets quantitative and qualitative metrics, and this section introduces the trials and questionnaires used to gather this data.

Chapter 5 - The results of the user study described in Chapter 4 are discussed. The results are analyzed using statistical methods, and specific items of interest for each upper-body interaction technique are extrapolated upon.

Chapter 6 - The preliminary reactions toward hand-and-finger gestural commands for navigating the UAV are presented. Benefits and drawbacks of the developed techniques, as well as the qualities of a potentially optimal tracking system, are discussed.

Chapter 7 - Plans for future work are laid out.

Chapter 8 - Concluding remarks for the presented work are presented.

Appendix A - Example questionnaires used to measure participant demographics and qualitative metrics are displayed.

Appendix B - IRB Approval Letter

CHAPTER 2: LITERATURE REVIEW

Household UAVs are relatively new, and there has not been substantial work conducted, as of yet, to explore various interaction techniques with these devices. Because typical UAVs are at least semi-autonomous - particularly those in the military - the majority of interfaces are used to set flight patterns and way-points rather than directly manipulate the aircraft. However, there has been much work performed to enable users with more meaningful modes of input by diverging from traditional devices such as the mouse and keyboard. In order to maximize the user's understanding of the interaction, many techniques have been developed around metaphors, which provide the user with a cognitive example of how the system works. These works set the foundation for significant development of UAV 3D interaction.

Various forms of input and interaction techniques have been researched for use in virtual reality (VR). While VR at first glance does not seem to apply to the realm of HRI, there are indeed many techniques that can be applied with lessons to be learned. For example, navigation and manipulation have been explored significantly for virtual environments [17][20][24], and various robotic platforms have already been targeted for these new forms of input [13][28][30]. The Joyman interaction technique is one method for navigation, where the user leans in a direction for an avatar to respectively move [15]. The actual implementation finds a user standing on a type of swivel board that captures the direction of the leaning - essentially, the body becomes a human-sized joystick, a metaphor that is easy to understand. One of the developed interaction techniques in this thesis is in nature very similar to this technique, but has been further augmented to allow more commands besides 2D navigation.

The Handle Bar technique [24] is an example of manipulation of a 3D object in virtual reality. The underlying metaphor is that the object being manipulated is pierced by a skewer; the interaction finds the user moving both hands in a way to rotate the virtual skewer, thus rotating the object. The skewer can be placed within 2 objects, allowing a manipulation of a group. This

technique has been explored for the manipulation of many different 3D objects in various scenarios. In this thesis, one developed interaction technique shares in nature the gestures with the Handle Bar metaphor, but only one UAV is used, instead of multiple systems. Further, as the vision system we use is unable to accurately detect rotation in all axes, the developed technique includes additional gestures to allow all possible UAV commands.

Outside of VR, other interaction techniques have been developed for various applications. In Jeon et al. [11], three interaction techniques to select and manipulate objects on a large screen were attempted through use of a mobile phone. This technology has potential to become useful in business settings, as described by the authors. This work shows that through use of a non-traditional device, being the re-purposed phone, it is possible to interact with alternative methods. While the work does not directly relate to this thesis, it does provide evidence that the future of HCI is stepping towards non-traditional modes of input.

The control of various robotic platforms have been explored, including robot arms, humanoids, and others. Smith et al. [23] provides one such work as an application example of the Nintendo Wiimote to control a robot arm. The research exhibits the potential for operators to command an arm through simple pointing gestures. A basic user study was conducted, but it did not consider qualitative metrics to evaluate ergonomic advantages and disadvantages of using the Wiimote. Rather, results were used to analyze the accuracy of the input device when using filtering techniques.

Work conducted by Zuher et al. [30] shows that complete humanoid robots can benefit from 3D spatial interaction techniques. Through the use of a vision system, a robot copies the movements of the user's limbs as well as head. This work can provide the foundation for future telepresence applications, where humanoid "surrogates" are manipulated from a remote location, while a first person video feed is reported back to the operator.

Chen et al. [5] presents work that attempts to combine the use of hand-and-finger tracking with upper-body gestures, with the hopes of equipping users with a more intuitive method of

commanding a robotic arm. In their system, the hand is tracked, and two distinct gestures are recognized - open and close of the fist. When these commands are given, the robot arm's fingers respectively open and close. Another vision system detects and tracks the user's arm, enabling him or her to intuitively move the robot arm in a direct manner. When combined, the arm and hand can be used to directly manipulate the robot arm with a one-to-one mapping. This form of robotic telemanipulation is a prime example of how 3D spatial interaction can be used to naturally command robotic platforms.

The Digits device [12] has been presented as another medium to conduct hand-and-finger tracking. This vision system mounts on the wrist, allowing a constant view of the hand, which in turn provides potentially higher accuracy for tracking. Results show that this device can be efficiently used as an alternative to prior wearable devices such as the CyberGlove. The authors note that this device can be used for a multitude of domains, and HRI is just one of these. Digits was not commercially available at the time of this work.

As an example of UAV control with novel input, the Flying Head technique [9] offers an egocentric interaction with the aircraft without the need for a traditional controller. By placing a head-mounted display (HMD) on the user's head and capturing the physical navigation of the user, all movements can be directly mapped to a UAV's flight pattern in real-time. This technique can be used for telepresence, as noted by the authors, but has limitations. First, in order for the UAV to fly in a large area, the user must also have a large space to navigate in. Second, in order to freely move the UAV vertically, an additional controller must be used. One potential step toward solving both of these issues is the use of interaction techniques that allow users to remain stationary while still having ability to issue all possible commands. The developed techniques in this thesis provide examples of such interfaces, but do not make use of an HMD.

In Cummings et al. [7], a sketch-based interface is explored to allow users to interact with the UAV through a seemingly natural method. The system does not allow users to directly manipulate a UAV, but instead it is designed to enable the user to plan out requests for UAV support.

Through the interface, a user would sketch environmental factors, including known ally and hostile locations, which would affect the UAV's route. This method works very well in that it equips users with a faster mode of input to request support, compared to a simple mouse and keyboard combination that can be found on a typical ground station. However, the system does not allow for direct manipulation of the UAV, and the outcome may rely heavily on the autonomy.

A recent demonstration by Alapetite et al. [1] shows the potential for using eye gaze tracking as a mode of input for UAVs. In their demonstration, the authors show that an eye tracker can be used to give navigational commands in two dimensions, but ignoring altitude commands. While this hands-free approach shows potential for directly manipulating the pitch and roll of the UAV, there is no method to give the other two necessary commands, being altitude and yaw changes. Another problem with this technique is the fact that the user must be looking at the screen at all times, so that constant input is being given. An interesting on-screen change may cause eye gaze to drift away from the intended navigational target; this flaw can prove to render this input mode far less useful than others. Through the use of 3D spatial interaction, all commands are possible, and the user's viewpoint may also change without directly interfering with UAV flight patterns.

Lichtenstern et al. [14] has developed an environment that is able to demonstrate human-robot interaction with multiple UAVs simultaneously. In this prototype, the user performs explicit 3D gestures to select one UAV out of many. After the selection is achieved and confirmed, manipulation of the selected robot occurs with further 3D gestures. The command of multiple UAVs through 3D spatial interaction has not yet been explored, but this prototype serves as a vessel to support future work. This thesis explores the direct manipulation of one UAV and does not touch on multiple agents, but can be extended to do so.

Ng et al. [16] describes an experimental interaction technique to control UAVs with a falconeering metaphor; the concept behind this style of interaction is that a UAV assumes the role of a bird, and the human becomes its "trainer". The research is aimed at exploring gestural commands that would be processed by the UAV, and not at the direct manipulation. This particular study in-

volved a wizard-of-oz experiment, to explore human reaction towards the UAV's movements. With the current market expanding, household UAVs may soon be common; the interactions with these devices must not be intimidating. This research may assist in determining what kinds of gestural commands would make a larger range of audiences more confident with these new robots; this thesis does not enter this domain, and rather assesses ergonomic factors with UAV control with 3D spatial interactions.

Work has been performed by Song [25] that demonstrates recognition capabilities of the entire human body while also extracting the hand data, including position and pose. While the work does show accurate recognition of hand and body gestures, the hand poses that were studied have been classified as "canonical" and are very distinct. It remains to be seen if, at a larger distance, smaller hand movements, including rotation on all 3 axes, can be accurately detected by a computer vision system. In this thesis, a preliminary exploration of hand and finger gestures makes use of a device that is used in close proximity to the user's hand (see chapter 6).

CHAPTER 3: UPPER BODY INTERACTION TECHNIQUES

Interaction technique design is somewhat dependent on the input device used to form the gestural commands. We studied both upper-body and hand-and-finger based input to form these commands. For our upper-body tracker, the Microsoft Kinect sensor is used. The following section describes the benefits and drawbacks of this device.

Input Device

We used the Microsoft Kinect (see figure 3.1) for upper-body interaction technique development. This device is considerably inexpensive and can be programmed to provide custom interaction techniques. The benefit to using this device over other motion controllers is the fact that the user does not need to interface with a physical device, such as the Nintendo Wiimote or the Sony Move, and that it is very easy for developers to make use of the provided SDK. This enables the user to freely maneuver the entire body as well as fingers, when appropriate, to minimize physical constraints. Although similar in the regards that each device provides motion control to a user, each are inherently different in regards to capabilities.

The Microsoft Kinect is a computer vision-based system that utilizes an array of sensors to detect the human body. By combining input from an array of cameras, the Kinect can estimate object depth in relative 3D space and the joint positions of the human body. This depth and skeletal data can then be used to detect human posture or gestures, which can be formed into input commands. The Kinect also allows for speech recognition, but we did not use this feature.

Figure 3.1: The Microsoft Kinect

Besides its low cost, one benefit of the Kinect is the fact that no physical devices need to be used in order to interact with a system. A user has the opportunity to interface with this system using movements of the body by itself, which ultimately provides opportunity for the most intuitive interaction. One drawback of this device is the possibility for occlusion, where body parts closer to the device veil other parts of the body and prevent detection. Another drawback is the fact that finger tracking is not a completely solved problem, and it can not be accurately achieved at a longer range due to hardware limitations. In typical scenarios, the Kinect needs to be far enough from the user in order to see the entire body; this distance combined with the resolution of the Kinect cameras ultimately results in a lack of the finger detection potential.

The Kinect is supported with a software development kit, enabling programmers or user interface designers to form their own interaction techniques. Utilizing this development kits, we created custom techniques to support flight of a UAV subject.

The UAV

We used a single type of UAV, the Parrot AR Drone (see figure 3.2). This quad-rotor is equipped with two HD cameras one faces forwards and one faces down and it also utilizes sensors that assist

with the platform's stability. The price, programmability, and capabilities of this UAV makes it a desirable device. Using a wireless connection, a computer or smart phone can send commands to the UAV in order to navigate it through 3D space as well as change its heading. The possible commands it can process include moving forwards/backwards, moving left/right, turning left/right, and moving up/down. These commands can also be aggregated to form complex commands, such as moving forward while turning left. Traditional input devices can be used to send commands to this UAV, but 3D spatial interactions can potentially be used to positively augment user capabilities in the context of direct manipulation.

Figure 3.2: The Parrot AR Drone

We used 3D spatial interaction techniques to command the UAV, by developing upper-body gestures. There are perhaps an infinite number of potential interaction techniques that can be developed, but not all of these make sense in the context of direct manipulation, and not all of these make sense in the context of UAVs. During the developing of interaction techniques described in the following sections, we considered multiple factors that ultimately defined our design process. These factors are later explored as qualitative metrics.

Design Process

Per the previous definition of direct manipulation, we made use of constant tracking for the user's upper body or hands. Because of this design constraint, it was ensured that the gestures were created in a way so that they exhibited features that would be favorable to the user, as often as possible.

In order to develop a successful set of gestural commands, we used the paradigm that the interaction techniques must have a comfortable "do nothing" state, or the "null" command. In the case of many techniques, the arms are constantly tracked – allowing the user to rest the arms at the sides is not a feasible option, as that particular pose may also translate into a command. One goal was the design of entire techniques while also allowing for the most comfortable null command, when possible. Along the same lines, the rest of the entire technique must also be favorable and intuitive.

In order to formulate these command sets, we first observed the nature of the UAV. It is rather plain to see that there are not many components of the Parrot AR Drone that can be manipulated; in fact, the only moving parts are the rotors that allow for navigation and rotation. Compared to other robotic platforms such as a humanoid, the UAV has fewer dimensions that can be directly manipulated. This ultimately makes the gesture interface design less difficult; however, even though various robotic platforms may share the same base commands, such as a UAV's "move forward" and a humanoid robot's "move forward", possible gestures that would suit one platform may not be as natural when applied to the other. For instance, using a walk-in-place gesture to command a humanoid robot does not seem to be an applicable gesture for a UAV. Thus, the movements of the device were observed, and potential metaphors believed to support the design of gestural command sets were identified. These gestural commands allow a user to perform all of the tasks identified above.

The following sections describe our developed interaction techniques for upper body ges-

tures. We used heuristic values of relative 3D coordinates to determine if a command is being performed by the user. In many cases, gestures as performed if one joint of the body has a specific 3D coordinate that is greater than that of another joint; for instance, if the left hand is higher than the right hand, that may trigger a command. However, this would be triggered even if the hand is slightly higher, perhaps by centimeters - this may sometimes be undesirable. To counteract this effect, many gestural commands require a larger difference of 3D coordinate values between joints, in the form of a buffer. As an example, the left hand being no less than 5 inches higher than the right hand would trigger a command.

The values for each interaction space are device specific and may not be reusable across different platforms, but the logic tests for these heuristics are more universal. For example, moving one hand above the head would result in that hand's Y value to be larger than the head's, in 3D space. Using heuristics allowed us to omit the development and implementation of a gesture recognition algorithm. Our definition of direct manipulation agrees with the use of heuristics over recognizers; heuristics do not require statistical models that machine learning techniques may use, and further do not require additional time that would be used to train a classifier. Heuristics values devised around intuitive metaphors allows for a fast and efficient method of detecting gestural commands. For each technique in the following section, the associated tables describe the parts of the body used to form commands; note that many times, more than one heuristic check must be met to properly form a command. Where denoted by an asterisk (*), a buffer was used on top of the base heuristic check, to ensure slight differences of 3D coordinate space values between the joints did not accidentally trigger a command.

Upper Body-based Interaction Techniques

Using a vision system to track the upper body, five interaction techniques were developed, based off of metaphors believed to be intuitive, in regards to the UAV. Names were assigned to the interaction

techniques, and are referred to as such:

- First Person

- Game Controller

- Standing Proxy

- Seated Proxy

- The Throne

After observing the nature of the UAV, we identified particular metaphors as having potential to convey increased user understanding of the commands. The first and perhaps most natural technique is the First Person technique.

First Person Technique

The First Person technique is built on the metaphor of the user pretending to mimic the movements of an airplane, as can be seen by youth at play. By spreading the arms to take shape of airplane wings, the user would lean the body in ways that compliment the pose. For instance, by leaning forwards, the UAV would receive the forwards command. Likewise, by leaning backwards, the UAV would move backwards. The same is true for leaning left or right; these commands would have the UAV strafe in the respective left or right directions. In order to achieve turning, the user would simply need to rotate the torso in the desired direction, so that the shoulders were placed in an explicit complimentary position. The UAV can be instructed to ascend or descend when the user lifts or lowers the arms, respectively. Vertical climb is sometimes inverted in flight simulators, but a direct mapping was instead incorporated, so that moving arms up yields ascension of the UAV, and moving arms down yields descent. There may be a case for user preference in regards for this inversion, but this was not a target as a point of study. Figure 3.3 shows the gestural commands.

After developing and analyzing the First Person interaction technique, our first reaction was that it would probably be the most intuitive metaphor, as it was the first one that came to mind during development. Because of this, we believed that users should be able to have a more natural and therefore better interaction.

An obvious benefit to using this technique is that while the user is engaged in control of the UAV, it is possible to form complex commands; for instance, the user can lean forwards and tilt the body to a side, denoting a forward strafe motion is desired. A user could also turn the body while leaning in order to indicate translation and rotation to be achieved simultaneously. However, a completely complex command is difficult to give; for instance, the user would find difficulty in turning the body, leaning forwards, tilting to a side, and moving the arms far enough up or down for vertical translation. Attempting such a command would also provide support for an argument against this interaction technique, which is that it is rather uncomfortable. Maintaining the default pose for a prolonged period of time puts much stress on the arms, and performing any translation command can put additional stress on the back. While this drawback is acknowledged, this technique was still expected to perform favorably due to the perceived intuition and ease of use.

Figure 3.3: The First Person gestural commands. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Table 3.1: Joint Requirements per Command - First Person. All conditions in a command's row must be met. An asterisk denotes the usage of a buffer in addition to the difference between the joints.

	Left Wrist	Right Wrist	Left Shoulder	Right Shoulder	Head	Pelvis
Forward	-	-	-	-	Closer than Pelvis to tracker*	Further than Head from tracker*
Backward	-	-	-	-	Further than Pelvis to tracker*	Closer than Head to tracker*
Strafe Left	Lower than Pelvis	Higher than Head	-	-	Lower than Right Wrist	Higher than Left Wrist
Strafe Right	Higher than Head	Lower than Pelvis	-	-	Lower than Left Wrist	Higher than Right Wrist
Turn Left	-	-	Further than Right Shoulder from tracker*	Closer than Left Shoulder to tracker*	-	-
Turn Right	-	-	Closer than Right Shoulder to tracker*	Further than Left Shoulder from tracker*	-	-
Move Up	Higher than Head	Higher than Head	-	-	Lower than both Wrists	-
Move Down	Lower than Pelvis	Lower than Pelvis	-	-	-	Higher than both Wrists

Game Controller Technique

The next developed technique is called the Game Controller, based off of the metaphor of the user's arms assuming the role of traditional joysticks found on controllers for recent video game consoles. In most games that have a user's avatar translating and rotating in 3D space especially in first person shooters - common, default control schemes have translation performed through use of the left joystick, and the rotation achieved through use of the right joystick. This scheme was adapted and applied to the interaction technique.

Originally, the base pose indicating no command was keeping the arms up in a manner where each elbow forms a 90 degree angle, so to keep the hands aligned to the head. To move the UAV forwards, the left arm would be moved forwards. The UAV could be brought backwards by moving the left arm back. UAV strafing could be achieved by moving the left arm to the left or to the right. For rotation, the right arm would need to be moved to the appropriate side. Vertical climb is achieved by moving the right arm forwards, and descent is achieved by moving the right arm backwards. Figure 3.4 shows the gestural commands.

Although the interaction technique was based off of a common layout for control, it was not found as desirable for two reasons. First, control of the vertical movements was perceived to be very unnatural. Rather than moving the arm forward or backward, it would be more appropriate to move the arm up or down. Secondly, keeping the arms in the air was very uncomfortable. Thus, we altered the entire interaction by rotating both arms 90 degrees forward and downward, so that the base pose was to keep the hands around the waist, with the elbows vertically aligned with the hands. In doing so, the interactions needed to form the commands were also changed. In order to move the UAV forwards, the left arm rotates so that the hand points downwards, and to move backwards, the left arm would rotate so that the hand points upwards. Strafing and rotation is still achieved by moving the corresponding arm left or right. Vertical translation has been changed to now achieved by moving the arm up or down, respectively; inversion here is also not used.

Contrary to the First Person interaction technique, the metaphor of the Game Controller seemed to be a little weaker, as it is based off the nature of a more traditional input device. However, compared to the First Person technique, the arms are in a more comfortable pose, and the most complex of commands can also be given using this technique, a very positive trait. Since the arms are independent of each other, and can be moved to the side and vertically, all four major commands can be aggregated to form a complex command. This technique was expected to allow the user to have the most potential when operating the UAV, since a wider variety of commands can be given.
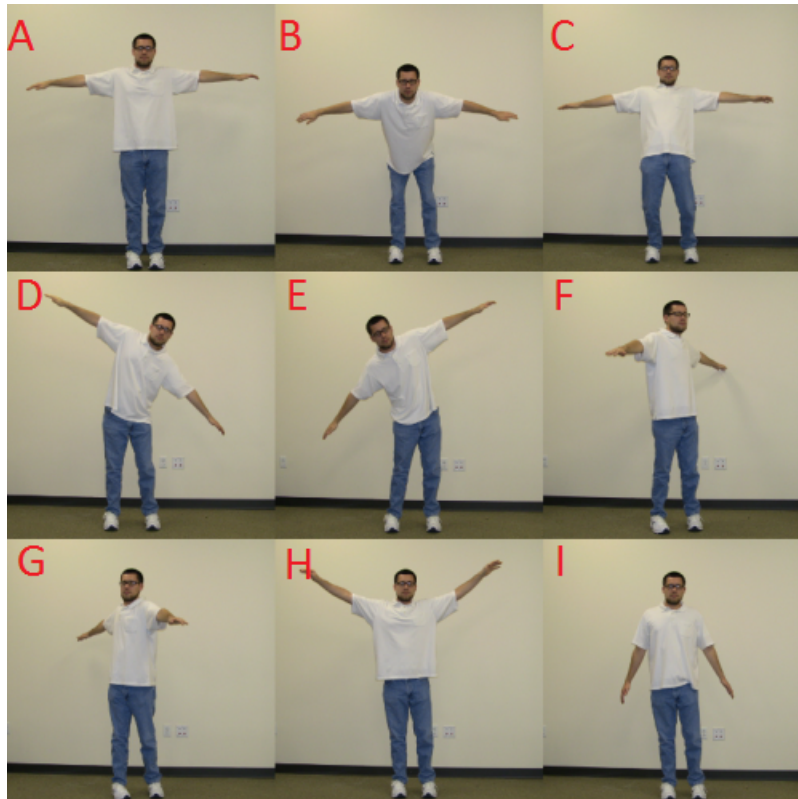

Figure 3.4: The Game Controller gestural commands. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Table 3.2: Joint Requirements per Command - Game Controller. All conditions in a command's row must be met. An asterisk denotes the usage of a buffer in addition to the difference between the joints

	Left Wrist	Right Wrist	Left Elbow	Right Elbow
Forward	Lower than Left Elbow*	-	Higher than Left Wrist*	-
Backward	Higher than Left Elbow*	-	Lower than Left Wrist*	-
Strafe Left	Further left than Left Elbow*	-	Further right than Left Wrist*	-
Strafe Right	Further right than Left Elbow*	-	Further left than Left Wrist*	-
Turn Left	-	Further left than Right Elbow*	-	Further right than Right Wrist*
Turn Right	-	Further right than Right Elbow*	-	Further left than Right Wrist*
Move Up	-	Higher than Right Elbow*	-	Lower than Right Wrist*
Move Down	-	Lower than Right Elbow*	-	Higher than Right Wrist*

Standing Proxy Technique

The third developed interaction technique is the Standing Proxy technique. This is built on the metaphor of the user imagining holding the UAV, where the movements of the imaginary device would be mirrored by the real one. By keeping the arms resting in front of the user, the "null" command would be given. This position becomes the base location to begin any other command. To make the UAV move forwards, the user would perform a pushing gesture; the arms would both be moved forwards. Likewise, in order to fly the UAV backwards, the user would perform a pull gesture; both arms would be retracted towards the chest. To have the UAV strafe to the left or right, the user would perform a "tilt" gesture. This is achieved by moving one hand over the other, as if to lift one wing of the aircraft up, and push the other wing down. The UAV strafes to the left if the right hand is above the other, and it strafes to the right if the left hand is above the other. In order to have the UAV rotate, the user would simply move one hand forwards and one hand back, as if pushing one wing forward and one wing backwards. If the left hand is pulled in, the UAV turns counter-clockwise. If the right hand is pulled in, the UAV turns clockwise. Lastly, vertical climb or descent can be achieved by moving both hands upwards, or downwards, respectively, as if to lift or lower the UAV. Figure 3.5 shows the gestural commands. Because some of these commands were very dependent on the individual's interaction space, and due to the device's limitations in sensing accurate depth, some calibration was needed to ensure commands were being successfully met. The user's maximum reach, retracted arm position, and a "down" position (used for the down command) were recorded before usage. These calibrations were used for heuristic checks, as seen in table 3.3.

Our initial reaction to this interaction technique is that it is favorable; the metaphor that the interaction is built on seems to be very understandable, since it is almost a direct mapping to the physical movements a person would use in order to actually guide the UAV. However, since all of the commands are built by using both arms together, some complex commands could be lost.

For instance, turning while moving forward is difficult, since the forward command requires both hands to move forwards, but a turn command requires one hand moving backward. Despite this drawback, this technique still offered opportunity to be very intuitive.
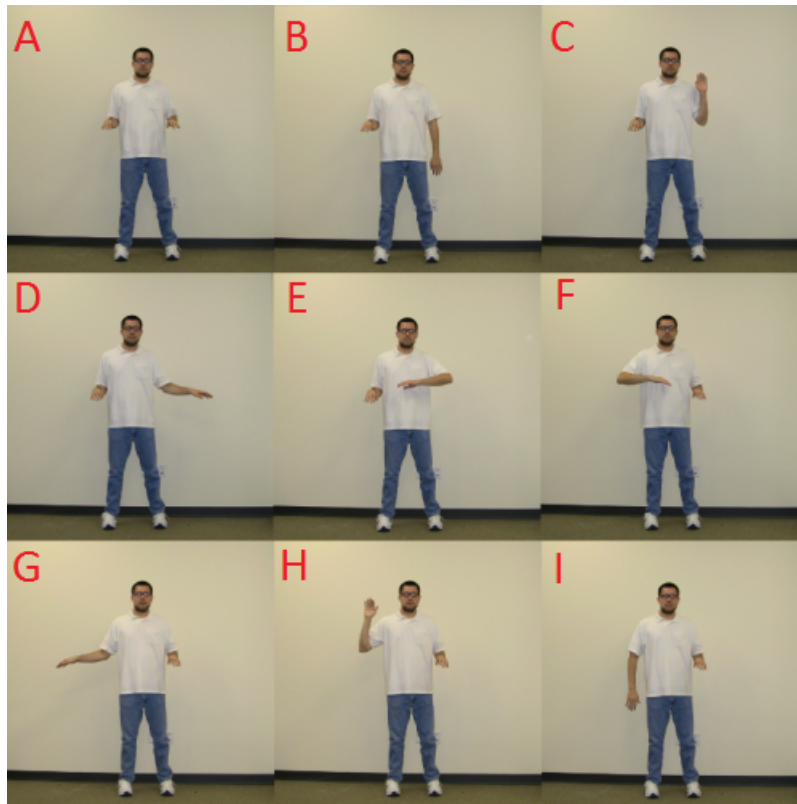

Figure 3.5: The Standing Proxy gestural commands. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Table 3.3: Joint Requirements per Command - Standing Proxy. All conditions in a command's row must be met. An asterisk denotes the usage of a buffer in addition to the difference between the joints. The calibrations are relative 3D coordinate points defined by the user's individual interaction space.

	Left Wrist	Right Wrist	Head	Extend Calibration	Retract Calibration	Down Calibration
Forward	Meets the Extend Calibration	Meets the Extend Calibration	-	Meets both Wrists	-	-
Backward	Meets the Retract Calibration	Meets the Retract Calibration	-	-	Meets both Wrists	-
Strafe Left	Lower than Right Wrist*	Higher than Left Wrist*	-	-	-	-
Strafe Right	Higher than Right Wrist*	Lower than Left Wrist*	-	-	-	-
Turn Left	Meets Retract Calibration	Meets Extend Calibrate	-	Meets Right Wrist	Meets Left Wrist	-
Turn Right	Meets Extend Calibration	Meets Retract Calibration	-	Meets Left Wrist	Meets Right Wrist	-
Move Up	Higher than Head	Higher than Head	Lower than both Wrists	-	-	-
Move Down	Meets Down Calibration	Meets Down Calibrate	-	-	-	Meets both Wrists

Seated Proxy Technique

These previous techniques, though developed on unique metaphors, are all similar in that the user would stand in order to perform the commands (although it is possible to achieve these in other postures). A fourth interaction, the Seated Proxy, was developed and based off of the Standing Proxy. The primary difference between the two techniques is the change in posture standing vs. sitting; of course, the seated trait suggests that the user would have a more comfortable and perhaps less distracting interaction, due to relieving the strain on the legs. This was thus created in order to maximize the comfort level of the interaction. While this is the main difference between the two techniques, an alternative method to control the strafing of the UAV was also identified. Rather than imagining tilting the UAV in the corresponding left or right direction, the user would imagine dragging the UAV to the left or the right. Both hands simultaneously would move to the left or right of the center of the torso, for the respective strafe left or right command. This follows the physical gestures of forwards and backwards, and was believed to assist in keeping the interaction easy to remember. Figure 3.6 shows the gestural commands. As is the case with the Standing Proxy, some calibration of the user's interaction space was required in order to form some commands. Table 3.4 assists in describing these cases.
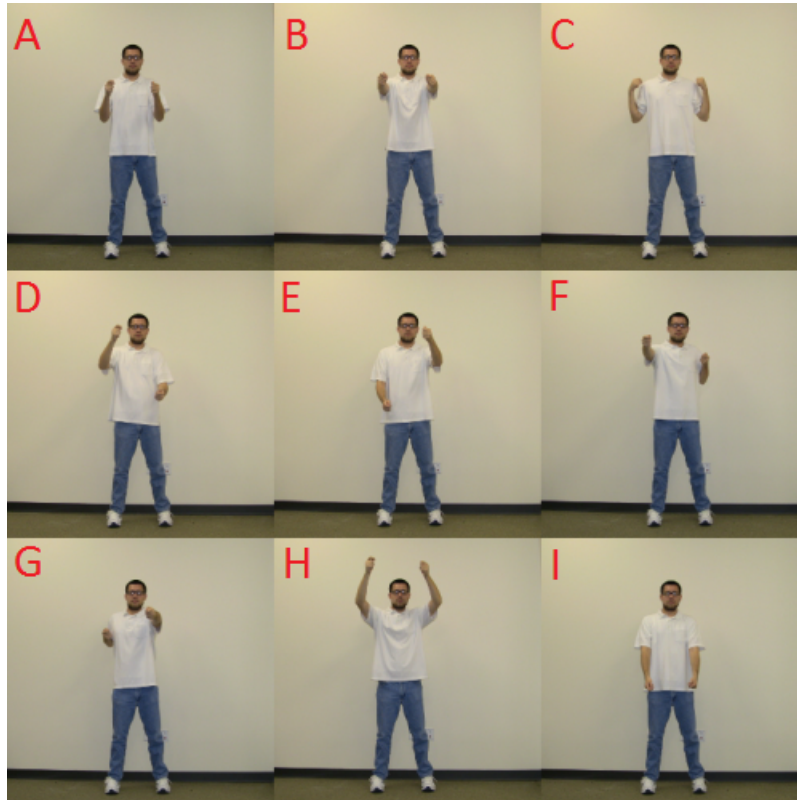
Figure 3.6: The Seated Proxy gestural commands. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Table 3.4: Joint Requirements per Command - Seated Proxy. All conditions in a command's row must be met. An asterisk denotes the usage of a buffer in addition to the difference between the joints. The calibrations are relative 3D coordinate points defined by the user's individual interaction space.

	Left Wrist	Right Wrist	Head	Extend Calibration	Retract Calibration	Down Calibration
Forward	Meets the Extend Calibration	Meets the Extend Calibration	-	Meets both Wrists	-	-
Backward	Meets the Retract Calibration	Meets the Retract Calibration	-	-	Meets both Wrists	-
Strafe Left	Further Left than Head	Further Left than Head	Further Right than both Wrists	-	-	-
Strafe Right	Further Right than Head	Further Right than Head	Further Left than both Wrists	-	-	-
Turn Left	Meets Retract Calibration	Meets Extend Calibrate	-	Meets Right Wrist	Meets Left Wrist	-
Turn Right	Meets Extend Calibration	Meets Retract Calibration	-	Meets Left Wrist	Meets Right Wrist	-
Move Up	Higher than Head	Higher than Head	Lower than both Wrists	-	-	-
Move Down	Meets Down Calibration	Meets Down Calibrate	-	-	-	Meets both Wrists

The Throne Technique

The final technique we developed is The Throne. This was built off of a metaphor in which the user assumes the identity of a monarch. Typical depictions of monarchs show commands to subordinates given in the form of pointing, as if to say "go there" or "do this". This was the basis for the metaphor and gestural command set. In order to have the UAV move forward, the user would point one hand in front. The UAV comes backwards when the user would bring the hand back, to align with the shoulder. Strafing from either side can be achieved by moving the hand into the appropriate direction. The hand does not need to be outstretched to its fullest potential, but just far enough so that the intent for strafing is easily identified. Commands for vertical climb and descent are given by pointing either up or down. Using the opposite hand, turning can be achieved by pointing left or right. Figure 3.7 shows the gestural commands. Similar to the Proxy techniques, the Throne required some calibration, as described in table 3.5.

The main benefit to this technique is the fact that the user only needs to move one hand in order to send the majority of commands; it is very easy to achieve a desired translation through a simple pointing gesture. All types of complex commands can also be given using this technique, which is found to be highly desirable. All translation commands can be aggregated by pointing forwards, to a side, and then either up or down, and the opposite hand could simultaneously be used to provide rotation, resulting in a completely complex command for the UAV.

With interaction techniques defined and developed, a user study was designed and executed to analyze quantitative and qualitative metrics.
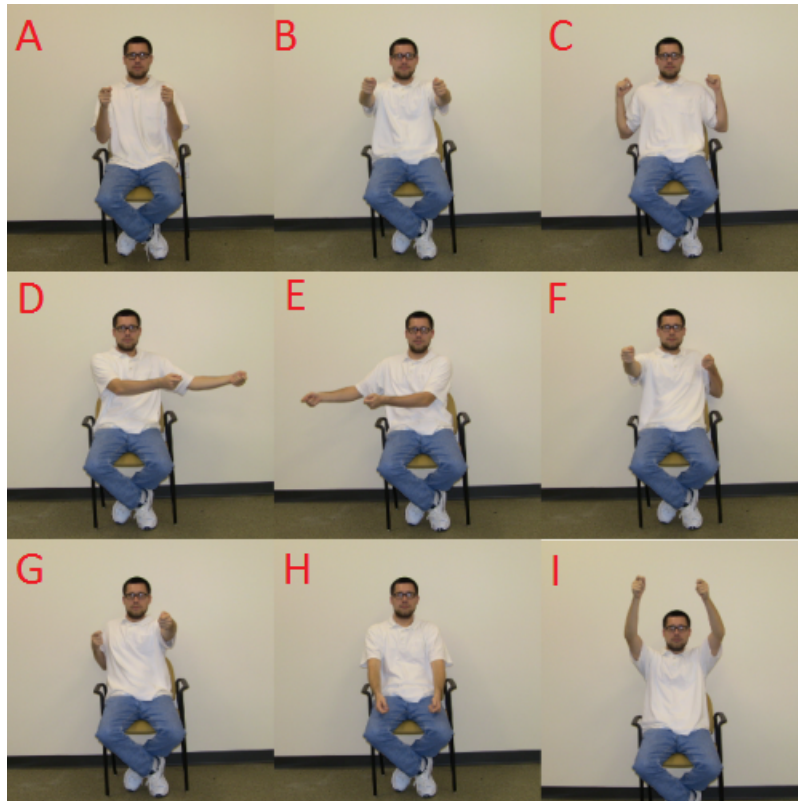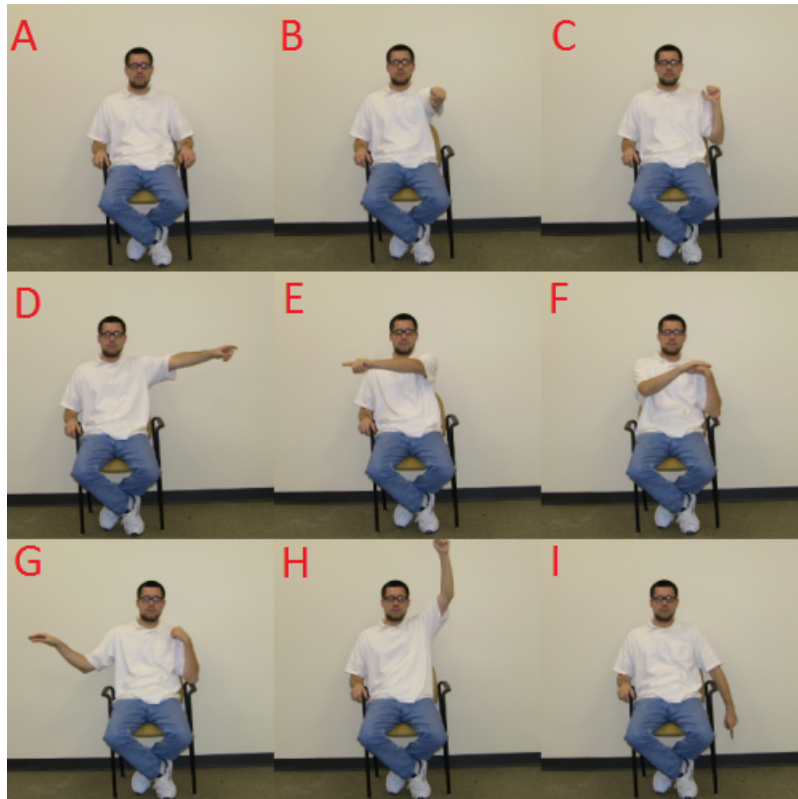
Figure 3.7: The Throne's gestural commands. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Table 3.5: Joint Requirements per Command - The Throne. All conditions in a command's row must be met. An asterisk denotes the usage of a buffer in addition to the difference between the joints. The calibrations are relative 3D coordinate points defined by the user's individual interaction space.

	Left Wrist	Right Wrist	Head	Rest Calibration
Forward	Higher than Rest Calibration*, Closer than Rest Calibration to tracker*	-	-	Lower than Left Wrist*, Further than Left Wrist to tracker*
Backward	Higher than Rest Calibration*, Further than Rest Calibration from tracker*	-	-	Lower than Left Wrist*, Closer than Left Wrist to tracker*
Strafe Left	Higher than Rest Calibration*, Further Left than Rest Calibration*	-	-	Lower than Left Wrist*, Further Right than Left Wrist*
Strafe Right	Higher than Rest Calibration*, Further Right than Rest Calibration*	-	-	Lower than Left Wrist*, Further Left than Left Wrist*
Turn Left	-	Higher than Rest Calibration*, Further Left than Rest Calibration*	-	Lower than Right Wrist*, Further Right than Right Wrist*
Turn Right	-	Higher than Rest Calibration*, Further Right than Rest Calibration*	-	Lower than Right Wrist*, Further Left than Right Wrist*
Move Up	Higher than Head	-	Lower than Left Wrist	-
Move Down	Lower than Rest Calibration	-	-	Higher than Left Wrist

CHAPTER 4: USER STUDY - UPPER BODY INTERACTION TECHNIQUES

As this work is an exploration into UAV direct manipulation using 3D gestural interfaces, there are specific points to address, in order to answer initial, key questions, and to open the door for future work. Specifically, the aim was set at determining how individual interaction techniques are more effective than others in this domain, and how changes in the nature of interaction affect the user's performance when accomplishing tasks. A second goal was to analyze qualitative metrics that would assist in determining how these changes affect the user's perception of the technique. As most of today's UAV controllers are typically traditional input devices, a third goal was to see if the 3D spatial interaction techniques would allow for a more effective mode of interaction compared to a traditional mode of input, while also determining if there are certain metaphors that harbor the most potential for understanding the command set.

Participants

14 students (10 male and 4 female) from the University of Central Florida were recruited from the ISUE Lab. 11 are graduate students. The ages ranged from 20 to 37 and the median age is 28. Only 2 have ever interacted with the AR Drone prior to the user study, but half reported prior experience using remote controlled vehicles. 10 of the participants have used a Microsoft Kinect before.

Devices and Software

Using the Microsoft Kinect, the five developed interaction techniques were implemented in C-sharp on top of an existing program designed by Hobley et al. that allows a computer to fly the UAV [1]. A computer exhibiting a 64-bit version of the Windows operating system was used,

---

[1]https://github.com/shtejv/ARDrone-Control-.NET

including 8GB RAM, a 2.3 GHz Intel I7 processor, a 1920 x 1080 resolution on a 17" screen, and a 2GB NVIDIA GeForce GTX 660M graphics card. In order to display the video feed from the UAV, we used FFPlay [2]. The smart phone application designed by Parrot was run on an HTC Evo 4G, but it can be run on any IOS or Android device containing an accelerometer.

Test Space

Task goals were designed for users to accomplish in a large, indoor environment; while the UAV is intended for primarily outdoor activities, incontrollable factors including inclement weather patterns would affect the data. Thus, a 15 meter long, 6 meter wide rectangular space was partitioned to fly the UAV, within an enclosed area. The edges of the defined space were not borders, but they rather served as guidelines. There was additional space beyond the boundaries that were cluttered with obstacles, and bumping into these objects would typically result in the UAV crashing. The height of the room was roughly 5 meters, allowing plenty of space for the UAV to navigate vertically.

In the corners of the rectangular space, square way-points of side length 1.2 meters were marked. These served as the indicators for how the UAV would achieve its primary objective - simple navigation. A small image was also mounted on the wall of a side boundary, roughly 3 meters high. This served to have the UAV achieve its secondary objective - observation. Lastly, near the middle of the test area, the start/end position for the UAV was marked, providing the final objective precision navigation. These three tasks fully represent the sort of objectives all typical UAVs must undergo. Specialized UAV activities, such as targeting ground units for tactical strikes, were not represented.

---

[2]http://ffmpeg.org/index.html

Trial Design

For every individual being run through the user study, a pre-questionnaire was given before performing the actual tasks. The pre-questionnaire was used to briefly capture the demographics of the participant. Specifically, information was requested about prior experience with a UAV or any remote controlled vehicles; likewise, it was important to identify users with previous Kinect experience. The complete pre-questionnaire can be found in the Appendix.

Each participant used every interaction technique. Before beginning with any technique, a pseudo-random order of use was determined. Due to the small number of participants recruited, the orders were selected to help reduce the amount of positional bias that an individual technique could take on. For instance, any technique could benefit from being used mostly last, as by the time the participant reached that particular interaction, a maximum understanding of how the UAV works would have been achieved; the other techniques used before would prompt this knowledge. Likewise, techniques used primarily first could suffer. There exists a certain learning curve for understanding how the UAV works, and it could potentially result in the first technique used having the worst task completion time. The order selection was thus pseudo-directed to ensure no technique suffered or benefited from these scenarios.

User Setup

Upon determining the order of the techniques, the user was set-up to begin. For every individual technique, the metaphor behind the design of the gestural controls was explained. Then, the participant was explicitly instructed on how to perform each command; the proctor demonstrated each command and the participant mimicked the gestures. By instructing the participant of the metaphor and the gestural commands in this way, the users were equipped with the potential to create a mental mapping that connects the 3D interaction technique with a more understandable scenario. The users were then allowed to become acquainted with the technique through practice.

The user was positioned roughly 4.5 meters in front of the vision system, and was faced in a direction that allowed for a complete view of the test area. From the user's view, the test area was longer than it was wider. This was determined to be easier than if the opposite was provided, as the user would need to look further to either side during flight. If this occurred, the user could absentmindedly turn the entire body to compensate, which would result in the sensor failing to provide accurate data.

Before tasked with maneuvering through the actual course, the user was given 5 minutes of preparation time, to become familiar with the gestural control scheme. During this time, the user was allowed to perform any commands as desired while flying the UAV. Typically, the participants utilized this time to help with remembering the commands. It was not a requirement to use all 5 minutes, and in many cases, the full time allotment was not used. Rarely, a participant declined the training time this typically occurred during preparation for the second of the two Proxy techniques, as they are nearly identical.

UAV Setup

Before the trial run began, the UAV was positioned near the center of the rectangle, but slightly closer to the user than not. It was preferred that the user have a better view of the UAV during take-off and land, to defeat potential parallax error. In a 3D environment, a view of an object may be misleading due to this natural effect, and a lack of contextual clues may prevent an accurate determination of position; for example, in this study, the user may have a difficult time flying the UAV to a checkpoint, as there is no clue that allows the user to know it is within the boundaries of the checkpoint. The UAV was initially aligned with the user's heading, so that all of its directions matched those of the user. This allowed the user to begin the task with an intuitive mindset.

36

Course Completion

In order to complete the entire course, the user was required to fly the UAV from the take-off point, through the four way-points and back to the center in a figure 8 manner, and then view the image that is mounted on the wall. After performing these tasks, the UAV was required to land back at the take-off spot. In the event that the UAV crashed at any of the run, it was placed back at the starting position, and the user was required to start from the beginning. A total of 3 crashes were allowed before the participant moved to the next technique. If 3 crashes occurred, the maximum allowed trial time for that technique was recorded 5 minutes were allotted, but this limit was never surpassed by any participant's run.

The trials began by completing a figure 8 with the checkpoints. The figure 8 technically can be completed in a variety of ways, and the user was allowed to do so as seen fit i.e. with any combination or pattern of commands but with a few conditions. The user must fly the UAV to the furthest two checkpoints before moving to the closer checkpoints, although he or she was allowed to select the first checkpoint that was maneuvered to. After moving through the first two points, the participant was required to turn the UAV around, so that it faced him or her. This forced an understanding of flying the UAV while its heading was inverted, from the participant's point of view. This requirement was only needed during flight from the second point to the third point; after this, the participant was allowed to take any course of action to continue the trial. The remaining two checkpoints were then navigated to, and after this, the UAV flew back to the center of origin, to complete the figure 8.

Because the checkpoints were very far from the participant, it was very difficult to determine if the UAV made it through one successfully. Thus, a line judge, one of the proctors, was used to provide information to the participant about the remaining distance to the checkpoint. Once the UAV reached its destination, the line judge would communicate a message of success to the participant.

The second leg of the course requires the user to fly the UAV in a way that enables him or her to clearly view an image that is mounted on the wall. The participant was asked to fly to a recommended area that was near the image; if the UAV was too far away being able to view a larger space that enveloped the image that attempt did not count. Sweeping motions were also not permitted the UAV was required to view the image without moving for a rather extended period of time, which was at least enough to verify success of the task. The image was allowed to be displayed in the corners of the screen, as long as the entirety was displayed. This was considered a successful attempt. When the line judge determined that the UAV was in an appropriate location that allowed a clear view of the image, or when the image was confirmed to be clearly viewed from the computer screen, the participant was allowed to fly the UAV in any manner back to the starting point. Because the image was on a side wall, this further enforced the participant to master UAV flight from alternative angles, and this also required the user to locate an object using the on-board camera, a utility that is one of the primary features of almost any UAV.

To complete the course, the user would navigate the UAV back to the landing pad. The nature of the course design finds the landing pad behind the UAV after it views the image. This seemingly trivial task could be rather difficult, as the quickest route to completion is to have the UAV move horizontally from the participant's perspective, which is only achieved by moving backwards from the UAV's perspective. To complete the trial, the UAV needed to line up with the landing pad and land on it. The whole UAV did not need to fall within the landing pad boundaries, but rather needed to at least be tangent to it by touching it. The landing procedure of the AR Drone is automated, and tends to stray off of a linear descent. Thus, attempts where the UAV was given the land command while it was over the landing pad were counted. If the landing was very obviously unsuccessful, however, the timing of the trial was not stopped until the UAV was corrected.

An illustration of the course design is shown in figure 4.1.

Figure 4.1: Course design and task objective layout. The user was required to complete a figure 8, view an image mounted on the wall, and successfully land at the origin.

Metrics

After the trial was concluded, the time was marked from rotor start-up to rotor shut-down. This quantitative data was recorded, and in-between trial survey was issued. This questionnaire was aimed at capturing the qualitative measurements of the participant's user experience. The survey specifically targeted eight metrics through the following statements:

- The interface to fly the Drone was comfortable to use.

- The interface to fly the Drone was confusing to me.

- I liked using the interface.

- The gestures to fly the Drone felt natural to me.

- It was fun to use the interface.

- I felt frustrated using the interface.

39

- It was easy to use the interface.

- The Drone always moved in an expected way.

Each of these metrics used a 7-point Likert scale to record data, where 1 denoted "Strongly Disagree" and 7 denoted "Strongly Agree" For the two inherently negative questions (confusion and frustration), the recorded values were inverted to align in polarity with the other questions. If the participant had any additional comments, he or she was encouraged to write in an allocated space for each statement. After the survey was completed, the proctors verbally verified all responses with the participant, to ensure the intended answer was marked.

This in-between questionnaire was given to the participant immediately after each interaction technique was attempted and completed or failed. After all of the techniques were tried and all surveys were completed, one final questionnaire that asked the participant about the summary of their experiences was issued. In this final survey, the participant was asked to rank each of the interaction techniques on the previous statements, as well as rank them overall, with all factors considered. No ties in rank were allowed, forcing the participant to distinguish the favorite out of a group of highly regarded techniques. The in-between and post questionnaires are visible in the Appendix.

CHAPTER 5: UPPER BODY INTERACTION TECHNIQUES USER STUDY RESULTS

After all data was collected, analyses on both quantitative and qualitative metrics were performed to determine if statistical differences exist between the interaction techniques. The quantitative measurement was analyzed first. Mean completion time for each technique is shown in 5.1. Throne had the absolute worst performance out of all techniques, and there was no statistical difference between the others. The Proxy techniques performed similarly the more traditional input, the touch-based smart phone application. While no statistical difference was found between these techniques, one positive result is that the 3D spatial interactions at least rival the more traditional forms of input. With more participants, however, it is believed that statistical difference would have been found between at least one of the interaction techniques, when compared to the smart phone application.

Analysis of Quantitative Metrics

A 6-way repeated measures ANOVA analysis was used to test for significant differences in the mean completion times between all interaction techniques and smart phone application in the study. If there were any significant differences found between the groups, a matched-pair t-tests was used to look for interesting differences between any 2 sets of interaction techniques in the post-hoc analysis. For instance, when using the smart-phone application as the control, a t-test comparison was performed with each of the other techniques at α=.05, resulting in a total of five comparisons. Type I errors inside the t-tests are controlled by using Holm's Sequential Bonferroni adjustment [10].

Significant differences were found between the interaction techniques and smart-phone application in their trial completion times ($F_{5,13} = 4.201, p < 0.002$). However, using pairwise t-tests with the smart-phone application as the control, no significant differences were found with

the interaction techniques. Completion times from 'The Throne' technique was the cause of the significant differences between the groups, due to less than half of the participants not being able to complete the trial. Pairwise t-tests were the conducted with the 'The Throne' technique as the control instead. Significant differences between 'The Throne' and the other interaction techniques were found, except for the smart-phone application (Seated Proxy: $t_{13} = 3.164, p < .007$; Standing Proxy: $t_{13} = 3.037, p < .01$; First Person: $t_{13} = 2.796, p < .015$; Game Controller: $t_{13} = 2.607, p < .022$). Figure 5.1 implies that the mean completion time for the smart-phone application is comparable with The Throne technique.

Although users had ample time to become familiar with 'The Throne', the gestural commands were not well-tuned. Additionally, a longer reaction time was perceived for the users executing gestures from the neutral position. This occurred because users were not able to gain control of the UAV when it acted unexpectedly. It as hypothesized initially that users would prefer a comfortable pose to execute their most frequent interaction task, being the "do nothing" command, but the commands requiring more movements ultimately ruined the experience.

Analysis of Qualitative Metrics

A non-parametric Friedman test was used on the post questionnaire qualitative metrics to check for any significant differences in their medians. If any significant differences were found between the groups, Wilcoxon signed rank tests were used to look for interesting differences between the interaction techniques with the smart-phone. Type I errors were controlled by using Holm's Sequential Bonferroni adjustment [10]. Quite similar to the quantitative time completion analysis, no significant differences were found in any the qualitative metrics, except for "Fun" ($\chi^2 = 14.776, p < 0.01$) and "Likability" ($\chi^2 = 12.342, p < 0.03$). The interaction technique that was significantly different than the others was the Standing Proxy, which benefited from a greater appreciation from the participants.
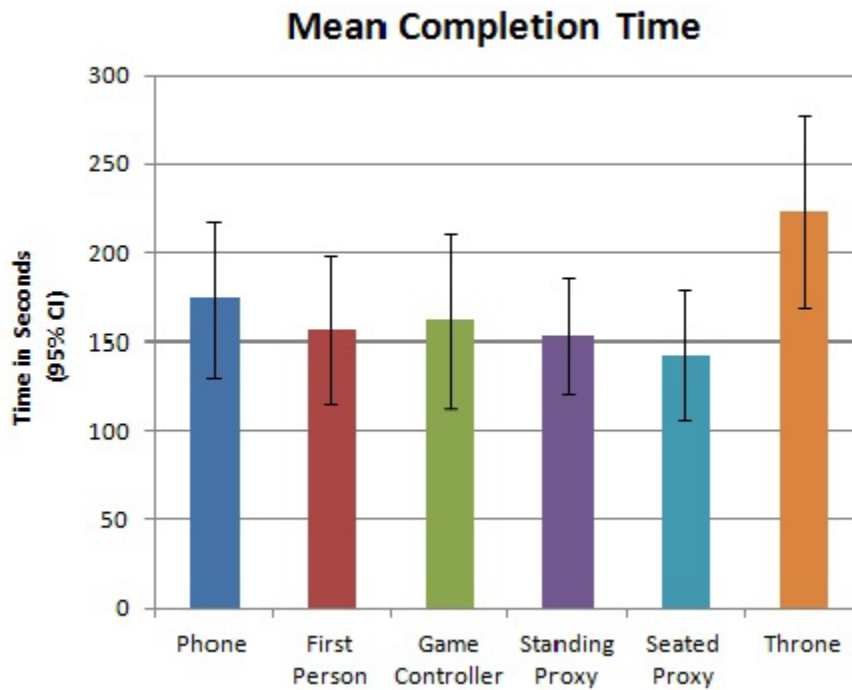
Figure 5.1: Mean completion times for each of the techniques.

For each of the following rankings metrics, a higher number indicates a better result. The maximum value is 6, which is of course the number of techniques being ranked.

Overall Rankings

Figure 5.2 depicts the overall participant perception of the techniques. The Standing Proxy received the most maximum-value votes, and The Throne received an overwhelming number of last-place votes. The mean and median rankings for each technique are as follows:

- Standing Proxy 4.71 (4.5 median)

- Seated Proxy 4.2 (5 median)

- First Person 3.57 (3.5 median)

- Game Controller 3.57 (3.5 median)

- Smart Phone 3.28 (2.5 median)

- The Throne 2.5 (2 median)
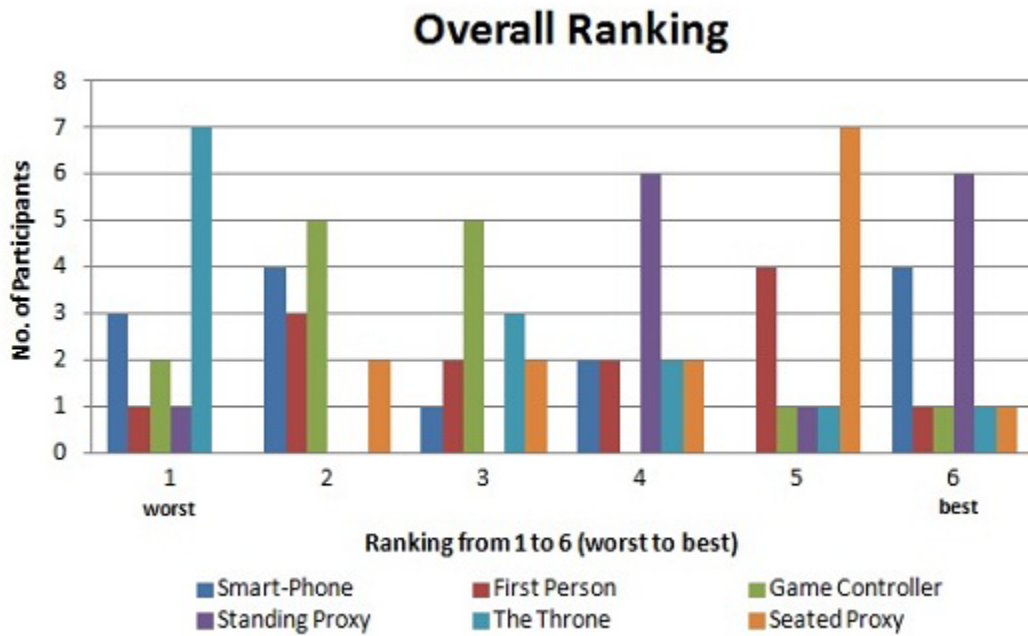
Overall Ranking

Figure 5.2: Overall participant rankings of each technique

Out of all developed metrics, the Proxy techniques were the ones that were regarded as better to use than the others. The Standing version of the technique was also regarded as the better of the two; this is interesting and counters initial reactions. It was believed that since users would sit during the entire interaction, they would regard the gestures as more comfortable to use. This was not the case, but no conclusions can be drawn from this, as data was only recorded for interactions with a maximum length of five minutes.

The difference in gestures for strafing is believed to be the cause for the drop in rank. Whereas the Standing Proxy technique exhibits a gesture similar to turning a wheel, the Seated Proxy technique requires the user to more purposefully bring both arms to a further location. This exaggerated movement resulted in a poorer score for the Seated version.

Ranking Natural

Intuition was perhaps the most highly sought metric to study, as one of the research goals is to identify a more natural user interaction. Figure 5.3 depicts the participant rankings of "natural". The mean and median values are as follows:

- Standing Proxy 4.57 (5 median)

- First Person 4.5 (4 median)

- Seated Proxy 4.14 (4.5 median)

- Smart Phone 3.28 (3 median)

- Game Controller 2.29 (2 median)

- The Throne 2.21 (3 median)

According to user feedback, the technique initially believed to be very natural the First Person technique received the highest score. The participants expressed their appreciation for a very simple metaphor to understand, which assisted them in performing the gestures required to move the UAV as necessary. One observation is that during instructing how to give the proper commands, most users were able to perform them without a verbal explanation; they felt as if the gestures were second nature, which is a positive outcome for this interaction technique.

The Proxy techniques were also ranked highly. Users expressed that the underlying metaphor to fly the UAV imagining holding the UAV in their hands greatly assisted them in understanding, remembering, and performing the gestures. The other techniques did not receive high ranks, especially The Throne, which was the least intuitive for the participants.
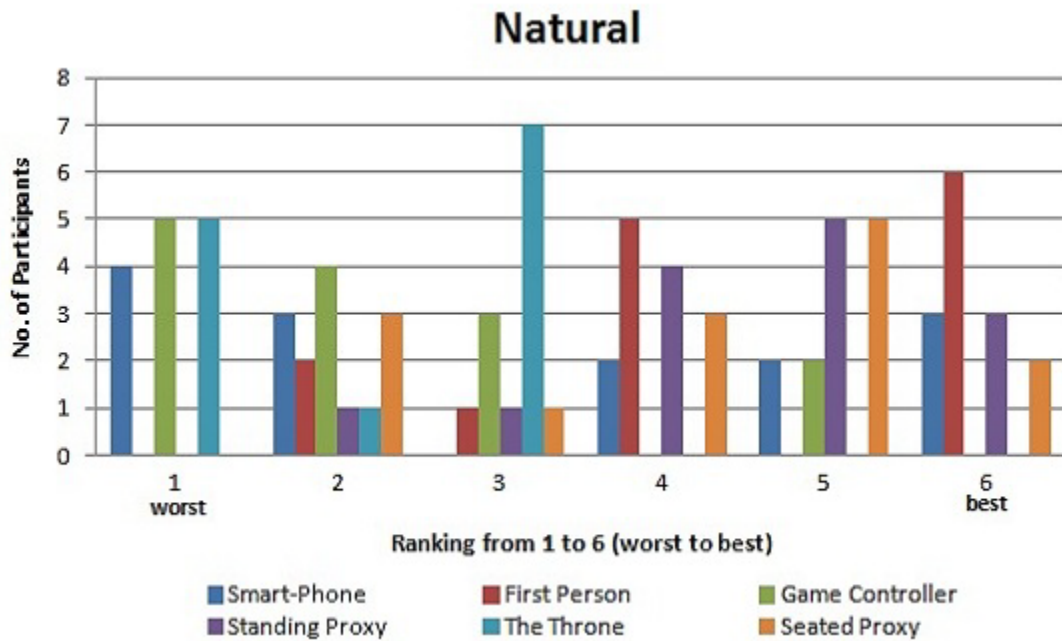
Figure 5.3: Participant ranking of perceived technique Naturalness

Ranking Confusion

Figure 5.4 illustrates the rankings of how confusing the techniques were perceived to be. This metric could be considered the antithesis of the "natural" metric, but was aimed at capturing data about the actual gestural control. The data between the two metrics are almost identical; the means and medians for the rankings of confusion are as follows:

- Standing Proxy 4.57 (5 median)

46

- First Person 4.36 (4.5 median)

- Seated Proxy 4.21 (4 median)

- Smart Phone 3.35 (3 median)

- Game Controller 2.5 (2 median)

- The Throne 2 (1 median)

The Throne was clearly the most confusing technique to use, per the participant feedback. Many crashes occurred during its use, resulting in the worst task completion times, and this poor performance is attributed to the fact that navigation could be achieved through the use of a single hand. Likewise, the Game Controller's performance suffered from the same effect. The main navigation could be performed through a one-handed gesture. The smart phone application was somewhat confusing to use, but observations of the participants, who typically over-steered the phone, reveals the cause for mistakes by the internal sensor; the UAV would either stop moving or would move in the completely opposite direction.

Figure 5.4: Participant ranking of perceived Confusion

Ranking Comfort

Figure 5.5 illustrates the participants' rankings of the interaction techniques in terms of how comfortable they seem to be. This metric was used to determine which, if any, were more optimal than others, from an ergonomic standpoint. A technique is far less appealing if it requires users to contort into poses that are undesirable, which warrants the inclusion of this metric. The ranking scores are as follows:

- Standing Proxy 4.5 (4.5 median)

- Seated Proxy 4.36 (4.5 median)

- Smart Phone 4 (4 median)

48

- Game Controller 2.92 (3 median)

- First Person 2.79 (2.5 median)

- The Throne 2.42 (1.5 median)

The data reveals that the Proxy techniques were much more comfortable to use than the other developed interaction techniques. The smart phone application also received a high ranking score, which was expected due to the minimal movements required to manipulate the UAV. The in-between questionnaire results indicate high scores for all three interaction techniques, but comments reveal that sometimes, the UAV moved unexpectedly when using the smart phone, which caused much apprehension to give further commands; this apprehension caused a mental discomfort, rather than the physical interaction being uncomfortable. The same principals apply to the Game Controller and The Throne techniques, which also require minimal movements to command the UAV. Instead of invoking mental discomfort, the First Person technique was simply more physically uncomfortable due to the strain applied on the arms and the back of the user, resulting in a low ranking score.
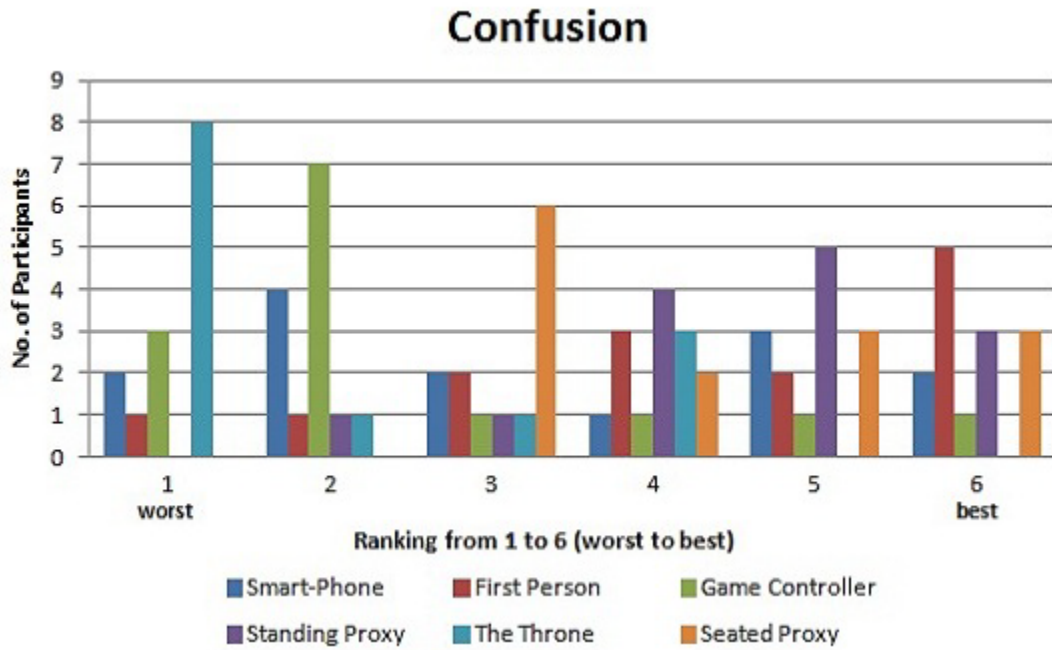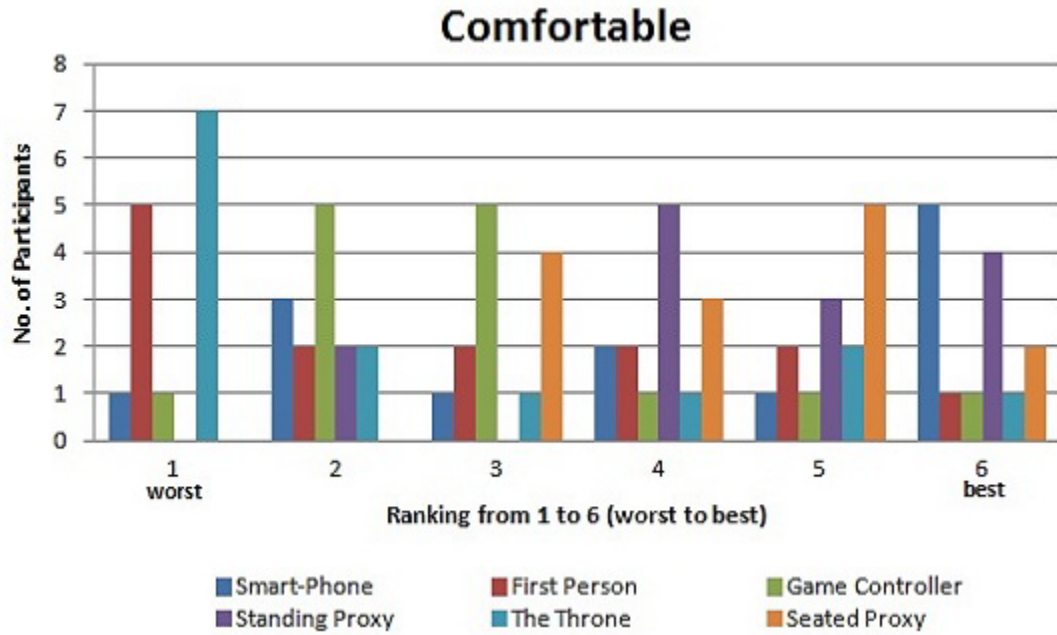
Figure 5.5: Participant ranking of perceived Comfort

Ranking Fun

Figure 5.6 depicts the participant rankings of the interaction techniques on their perception of fun. This metric was included in an effort to identify which form of interaction attracted the most user interest. There have been recent news articles and studies [8] reporting that UAV pilots in the military are very inattentive to their jobs, which has been identified as a problem in effectiveness. 3D interaction techniques, being more novel than traditional forms of input, may help alleviate lackadaisical attitudes these pilots may exhibit on the job. The rankings are as follows:

- Standing Proxy 4.58 (5 median)

- First Person 4.14 (4 median)

- Seated Proxy 4 (4.5 median)

- Game Controller 3 (3 median)

- The Throne 2.93 (3 median)

- Smart Phone 2.36 (2 median)

Results indicate a strong disinterest in using the smart phone application in the presence of the 3D interaction techniques, and even though it performed well compared to The Throne, it was still not as fun for the participants. This may be because the use of 3D gestures to control systems like the UAV are relatively new, whereas the traditional forms of input have been around for decades. User comments did not seem to provide much evidence for this either way, and although the majority of the participants have used the Kinect before, they may still prefer the gestural interfaces.
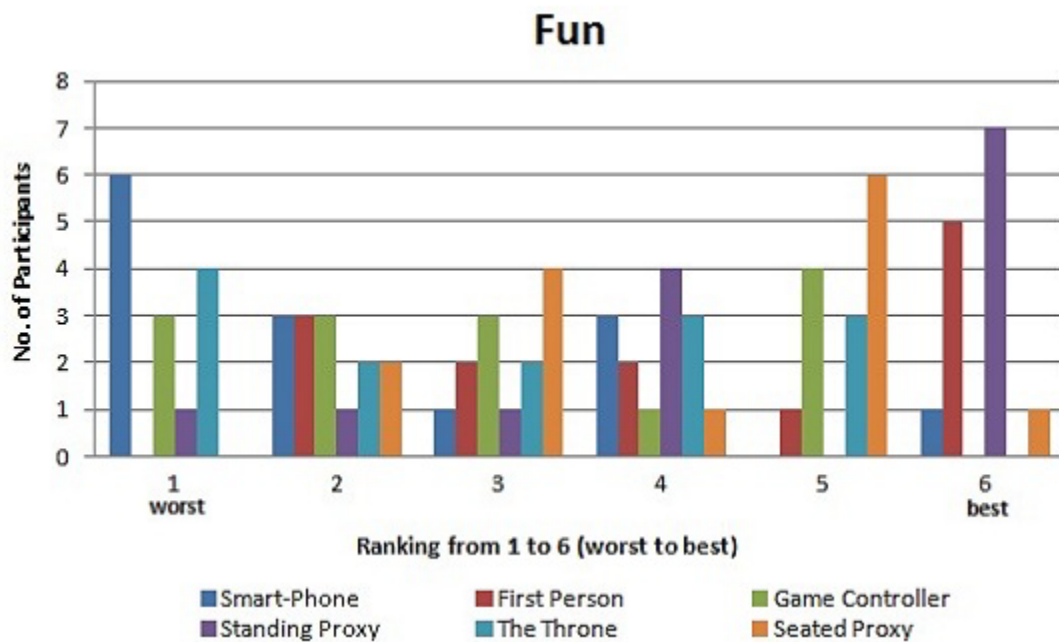

Figure 5.6: Participant ranking of perceived Fun

Ranking Likability

Figure 5.7 depicts the user rankings of the Likability metric. This metric was included in the hopes that additional data for user perception would be revealed. To like an interaction is very vague, as it could be considered a culmination of many different factors. It was not intended to be a secondary "overall" ranking, but instead it was intended to be more of a straight-forward manner to determine whether or not the participant would use this form of interaction again. The scores for means and medians of each technique are as follows:

- Standing Proxy 4.5 (4 median)

- Seated Proxy 4.29 (5 median)

- First Person 3.71 (4 median)

- Game Controller 3 (3 median)

- The Throne 2.93 (2.5 median)

- Smart Phone 2.64 (1.5 median)

The Proxy techniques outperformed the others for this metric. Interestingly, the smart phone application was again ranked the lowest, on average, and half of the participants ranked it last. It is unclear why using the smart phone was considered the least liked, as the demographic data and the participant comments do not reveal any insightful information. However, these results differ from the overall rankings, and they are almost aligned to the Fun rankings. These results may also be due to the novelty of the 3D interaction approach.

Like
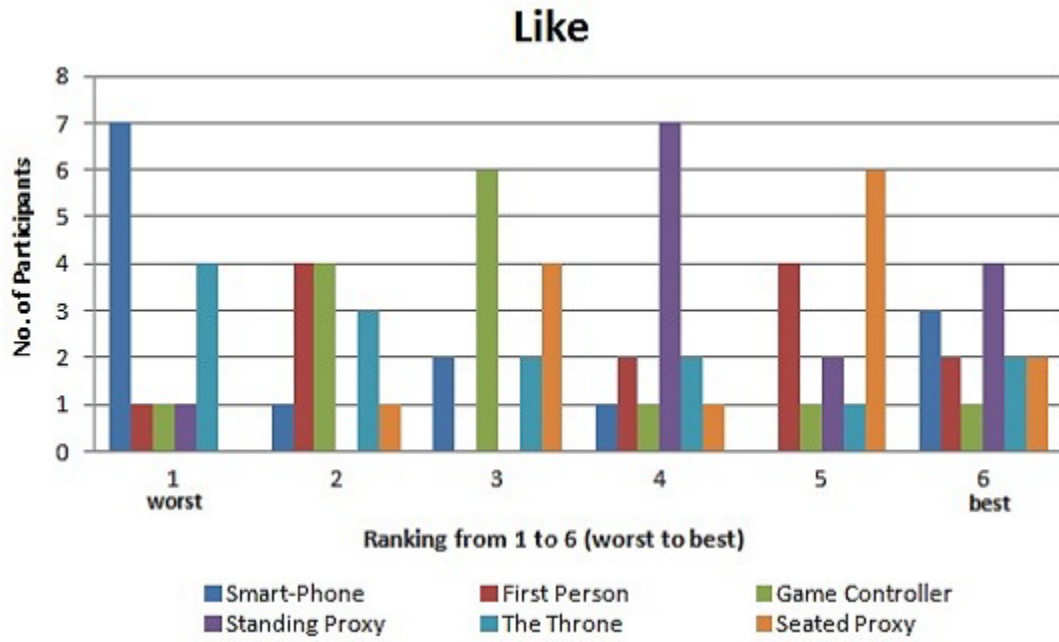
Figure 5.7: Participant ranking of perceived Likability

Ranking Easiness

Figure 5.8 depicts the user rankings of how easy the interaction was found to be when flying the UAV. It is important to evaluate how difficult an interaction is to perform or even perform correctly, even though it may be very easy to understand. If it is a challenge to provide input, or difficult to remember how to provide that input, then the overall interaction becomes less valuable. The rankings are as follows:

- Standing Proxy 4.36 (4 median)

- Seated Proxy 4.29 (5 median)

- First Person 3.86 (4 median)

- Smart Phone 3.79 (3.5 median)

- Game Controller 2.43 (2 median)

- The Throne 2.29 (1.5 median)

Again, the Proxy techniques outperform the others. The participants found it hard to use The Throne and the Game Controller techniques, which is partially attributed to the fact that incorrect commands are very easy to give observations of the participants reveal that much of the time, errant commands were applied. Because the "do nothing" command for the Game Controller is to keep the hands level, there is opportunity to let gravity bring the hands towards the forwards and move down commands. Participants were observed performing these commands and not realizing this, until after the UAV crashed or it moved very far from the next checkpoint. By contrast, the Proxy techniques' commands require both hands to move in a coordinated fashion, with little opportunity for an inadvertent command to be given.
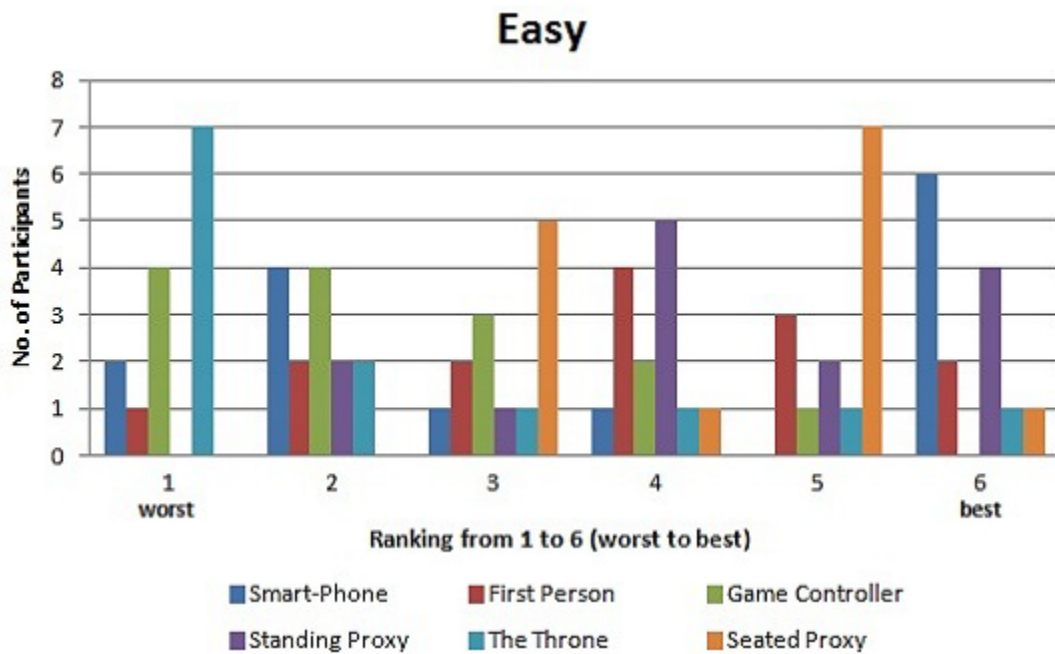
Figure 5.8: Participant ranking of perceived Easiness

Ranking Frustration

Figure 5.9 shows the results for the rankings of participant frustration. This metric was selected to assist in determining if negative performances eventually caused the user to feel fed-up, which would indicate a lack of desire to use the interface. The scores are as follows:

- Standing Proxy 4.07 (4 median)

- Seated Proxy 3.86 (3.5 median)

- Smart Phone 3.71 (4.5 median)

- First Person 3.71 (3.5 median)

- Game Controller 3.07 (3 median)

- The Throne 2.57 (2 median)

This particular metric, while finding the Proxy techniques on top of the rankings, did not find much difference in the techniques, with the exception of The Throne, which was widely considered the most frustrating technique. The rating data supports this result, revealing that only The Throne was frustrating. Due to the short amount of time spent on any technique, it may be preliminary to draw any conclusions from this metric. With a longer time of interaction, it is expected for users to become more and more frustrated when encountering errors, but, on the other hand, it is also expected that the user becomes more and more familiar with the technique, resulting in fewer errors. It is clear, however, that The Throne was frustrating, as half of the participants were not able to complete the course with it.
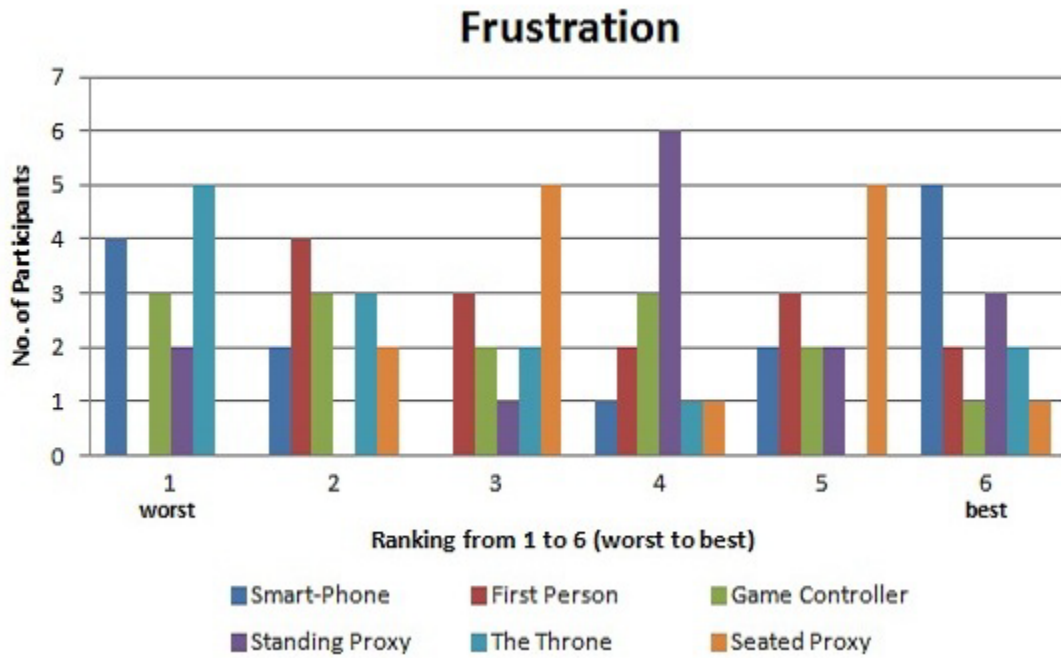
Frustration

Figure 5.9: Participant ranking of perceived Frustration

Ranking Expectation

Lastly, Figure 5.10 illustrates how the participants' expectations were met when commanding the UAV with each interaction technique. This metric is valuable as it assists in determining if the user is able to effectively recall the correct gestural commands needed to fly properly. The rankings for this metric are as follows:

- Standing Proxy 4.5 (4 median)

- Seated Proxy -4.29 (5 median)

- Smart Phone 3.71 (3.5 median)

57

- First Person 3.57 (3.5 median)

- Game Controller 2.64 (2.5 median)

- The Throne 2.29 (2 median)

The Proxy techniques once again top the rankings, and The Throne is shown to be the technique that commands the UAV in an unexpected manner. The rating data indicates that all of the techniques with the exception of The Throne commanded the UAV in an expected way. No proper conclusions can be drawn from this metric, with the exception that The Throne was considered the worst, and the Proxy techniques were ranked higher than the others, even though difference in performance was almost negligible.

Figure 5.10: Participant ranking of perceived Expectation

Lessons Learned

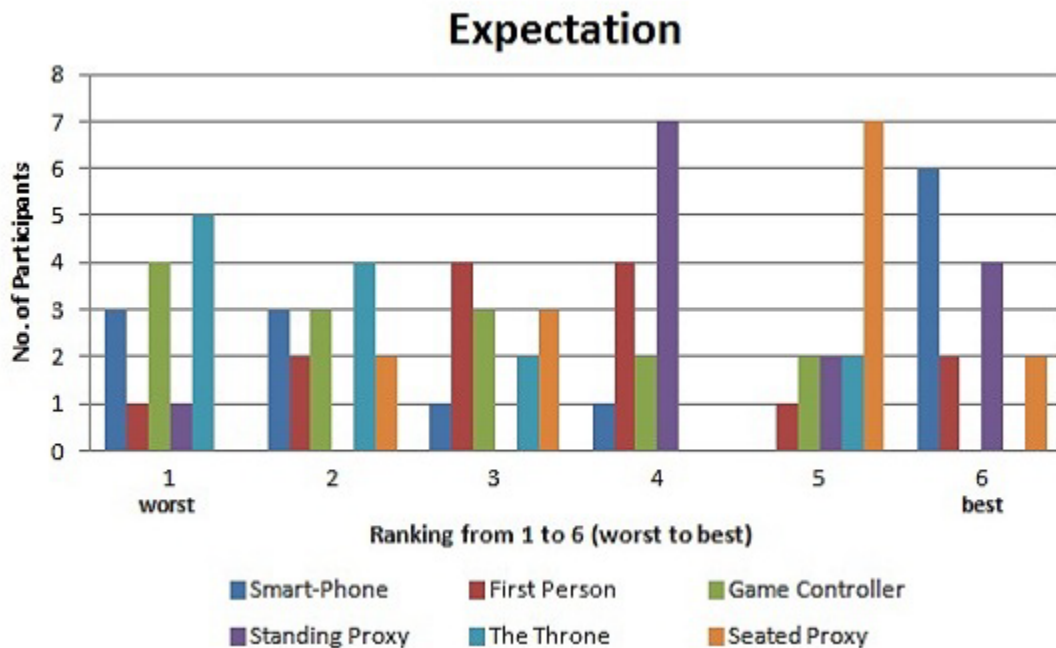While it may not be possible to directly generalize the results of this user study, much can be learned and applied to future implementations of 3D spatial interaction to other forms of robots, including other models of UAVs. It would seem that by using metaphors as a basis to develop an interaction technique, users may have an easier time learning or understanding a new command or gesture. Opposed to just instructing users on the "what to do", metaphors offer clues on the "how to do". However, it is not enough to simply provide a command set based on metaphor and expect positive results. Each interaction technique must be evaluated to determine the strengths and weaknesses, and to ensure that while it is understandable, it is also ergonomic. The following sections offer an evaluation of the developed techniques.

First Person Technique

The First Person technique was originally regarded as highly favorable, as the metaphor was the first one to be identified after observing the UAV. The gestural commands in the interaction are very simple and easy to understand, and this technique was also considered the most intuitive to use by the participants. The use of this technique favorable, but not the best. Participant feedback indicates this was a demanding technique to use, even though the total interaction time was very short. The need to hold the arms outwardly during the entire interaction ultimately detracts from the value, as it brings about much discomfort. This, coupled with the leaning required to achieve UAV navigation, strains the user and renders this technique as a lesser interaction. To make this technique more usable, it would be ideal to decrease the exaggeration needed for the commands for example, moving backwards is achievable by leaning very far back; a slight lean would be more desirable. However, in doing so, the vision system would need to be able to recognize a smaller change in posture, which could lead to false positives as well as false negatives. This technique is not recommended for use in an environment where the UAV needs to be controlled

for an extended period; however, this could be suited for shorter interaction times, such as during recreational purposes, as it is considered fun to use.

Game Controller Technique

The Game Controller interaction technique was originally believed to be an excellent avenue into commanding the UAV in a more complex way, but with smaller amounts of effort needed. The metaphor behind the gestural commands was based off of a traditional form of input - a basic remote controller found with video game consoles. Because these types of controllers have been around for decades, this was thought to be a very understandable metaphor. The gestural commands can be conducted with either arm, and, when used together, all four major commands can be given at one time. This was regarded as a very positive attribute of the interaction. However, the user feedback indicates this was not a good technique to use. The nature of the interaction was not found to be very intuitive and was instead found to be very confusing to use, even though the mean task completion time was not too dissimilar from the others. The trouble with this technique is the fact inadvertent commands can be given. Many participants accidentally let an arm drift into a position that would trigger an accidental command, resulting in user confusion. Interestingly, out of all individual trials, this interaction technique tied with one other for the shortest completion time, at 1:13. We attribute this success as minor as it may be to the fact that complex commands are very easy to be given. While performance and perception of this technique are very poor among the participants, it is expected that with much more training, users would be able to successfully use this technique.

The Throne Technique

The Throne was initially perceived to be a success in that very simple gestures were able to fly the UAV. All that is required to perform almost every command is a simple wave of the hand. The metaphor behind it was thought to be a rather vivid one, but in reality it is an extended pointing

metaphor. The resulting task completion times find this technique to be the very worst - half of the participants could not complete the course with The Throne. As with the Game Controller technique, the problem with providing a command set that is very easy to perform is that inadvertent and incorrect commands are also very easy to perform. When users become confused of the UAV orientation, the simplicity of the interaction actually becomes a hindrance; we observed many participants try to correct the UAV's flight patterns during a mistake, but more often than not, recovery was not possible. During pilot testing, expert users were actually able to complete the course swiftly with this technique. This form of technique is not suggested for use, as it seems much more training would be required for a user to have any benefit.

Standing and Seated Proxy Techniques

The two Proxy techniques were great successes, as they not only received the fastest average task completion times, but they were perceived by the participants as the best, overall. With the exception of two metrics "natural" and "confusing" they were both best and second best. For the two other rankings, the First Person slightly overtook the seated version. According to user feedback, the strafing was the difference between these two techniques. While the metaphor of holding the UAV in the hands remain the same, the two strafes were different the command for the standing version, putting one hand over the other as if to turn a steering wheel, was found to be more intuitive to perform, opposed to dragging both hands across the body. This shows that even the slightest change in a gestural command set can make a difference for the user.

In contrast to the Game Controller and The Throne techniques, the Proxy techniques require the users to move both arms in coordinated tandem to perform a command. This is a natural prevention of inadvertent commands. Because the gestural commands are technically more exaggerated than the other command sets, the user needs to explicitly perform a command there is little opportunity to "accidentally" give a command. Because of the performance in both quantitative and qualitative metrics, out of the developed techniques reported here, either of these techniques

would be recommended for UAV control.

CHAPTER 6: HAND-AND-FINGER INTERACTION TECHNIQUES - INITIAL REACTIONS

We developed interaction techniques using the Leap Motion device, for hand-and-finger gestures. A complete user study was not conducted for these, however, as at the time of this work the Leap Motion SDK was not complete[1]; it was not feasible to draw major conclusions. However, some points of interest were gathered after exploring the device in the context of these techniques.

Input Device

Similar to the Kinect, the Leap Motion device (see figure 6.1) is also an implementation of computer vision, but instead of tracking the entire body, it is used to detect and estimate the human hand and finger positions in a relative 3D space; it does not have the capability to detect the entire user body. This device can be used for a variety of applications, but it is only useful in scenarios that can benefit from accurate hand and finger gestures. One desirable feature of this device is the fact that the complete interaction space is very small, especially compared to the Kinect, where the interaction space can be very large. When constrained to a small, bounded area, natural hand and finger gestures would trump whole-body gestures on this scale.

Because of the accuracy in determining a hand and its extended fingers in relative 3D space, this device ushers in a new layer of input previously not reliably achievable. This will assist in studying future forms of interaction techniques. However, this device also has drawbacks. Occlusion is very possible, but the nature of fingers typically will not interfere with tracking; if any occlusion problems are encountered, it seems the thumb is typically involved, or the hand is rotated so that the palm is perpendicular to the sensor. The main detriment associated with this device is the fact that it seems to be designed as a stationary device - and soon, embedded in laptops - which

---

[1]At time of this work, the Leap Motion was at development version 0.7.6.

introduces various problems. In an interface that require a less proximate distance between the Leap Motion and the user, it becomes less physically feasible to achieve interaction. However, it may still be used effectively in interactions where the user needs to remain in one location.

Figure 6.1: The Leap Motion Controller (from https://www.leapmotion.com/product)

Interaction Techniques

We developed interaction techniques that only required the hands to perform. Similar to the development of techniques for the upper body, metaphors were targeted, to allow for an easy understanding of the gestural commands. Since the vision system provides an interaction space similar to that of upper body, except on a much smaller resolution, some of the prior interaction techniques were replicated by using just the hands. These techniques translated to the full body ones rather well, but others that were unique in comparison were attempted. In total, three interaction techniques deemed worthy of preliminary study were developed, and are presented as such:

- The Throne

- Scaled Proxy

- First Person

The Throne Technique

One developed technique was reminiscent of The Throne, which mainly incorporated pointing gestures. In this technique, through the use of just one hand, the UAV can receive all possible commands. To send the forwards or backwards command, the hand is moved in front or back of the vision system. Likewise, strafing is achieved by moving the hand to either the left or the right of the device. Vertical climb and descent is achieved by moving the hand far above the device or by moving the hand to a position slightly above it. Turning can be achieved through a rotating gesture of the hand; by constantly circling one finger in a clockwise manner, the turn right command is sent. Similarly, circling the finger in a counter-clockwise manner triggers the turn left command. Figure 6.2 shows the gestural commands.

This technique was expected to be somewhat favorable in the sense that only one hand needs to be used, and because the gestures are reasonably easy to understand. Complex commands can be given through this technique, but turning while navigating somewhat counter-intuitive. The circling gesture takes much attention away from actual navigation, which was considered a major detriment, but is otherwise a decent proposal for this command.
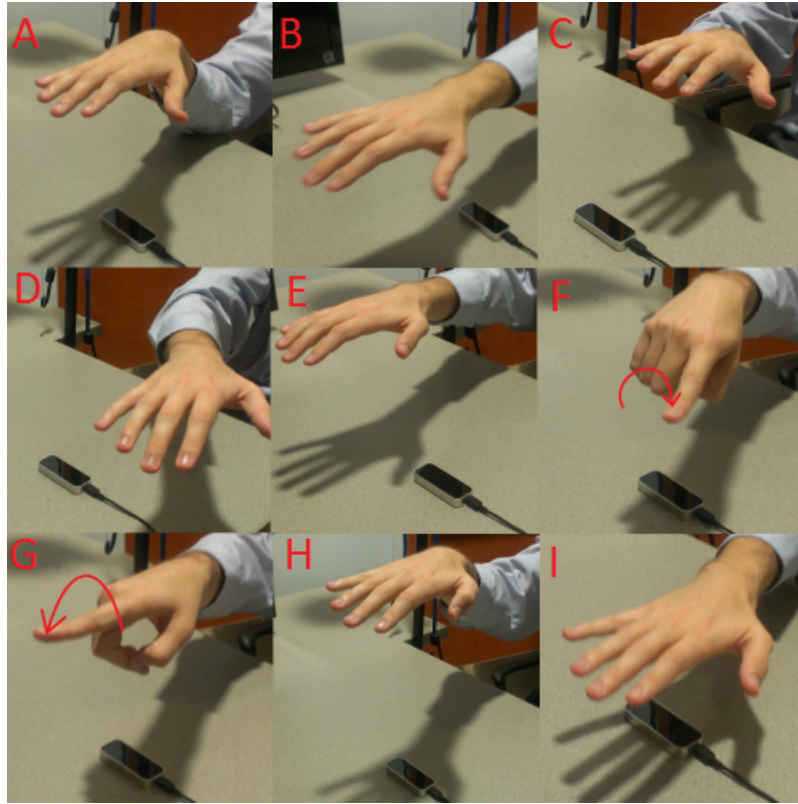
Figure 6.2: The Throne's gestural commands - hand and fingers. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

First Person Technique

The second developed technique is a different interpretation of the first person metaphor that was explored using upper-body gestures. In this version, the mindset needs to change in order to support the metaphor; in the full body technique, the user was able to move in a way that made the whole interaction more egocentric, where the user's explicit and perhaps more natural motions were directly mapped to UAV movements, as if the user's body was the UAV. In this technique, one hand can be moved in a way that would seem to directly map to the UAV's nature. The entire gestural command set would seem to be egocentric, but from the hand's point of view. Visualizing

the fingertips as the forward heading of the UAV, the hand gestures invoke an almost identical result from the UAV.

The base position for the hand is to hover flatly over the vision system so that the palm faces the sensor. In order to turn the UAV, the hand would rotate its yaw to either side so that the fingertips are pointing left or right. Strafing is achieved by rotating the hand's roll, so that the palm faces more towards the left or right than downward. The command for moving forwards and backwards is given by tilting the hand either forward or backward, so that the fingertips are, respectively, either below or above the wrist, which also results in the palm facing towards the user or away from the user. Lastly, vertical climb and descent can be achieved by simply moving the hand up or down; moving close to the sensor results in a down command, and moving far above the sensor results in an up command. Figure 6.3 shows the gestural commands.

Similar to The Throne technique, this one also only requires the use of a single hand in order to perform all possible commands. Comparing the two techniques, this one does not require as much movement as the Throne. All commands, except the vertical commands, can be given without translating the hand in 3D space. The rotation of the hand alone can be used in order to efficiently determine how the UAV should move. This technique is seemingly easy to understand, as the first person metaphor still applies, though the mindset requires a change. It is expected that this results in very positive task completion times and qualitative responses.
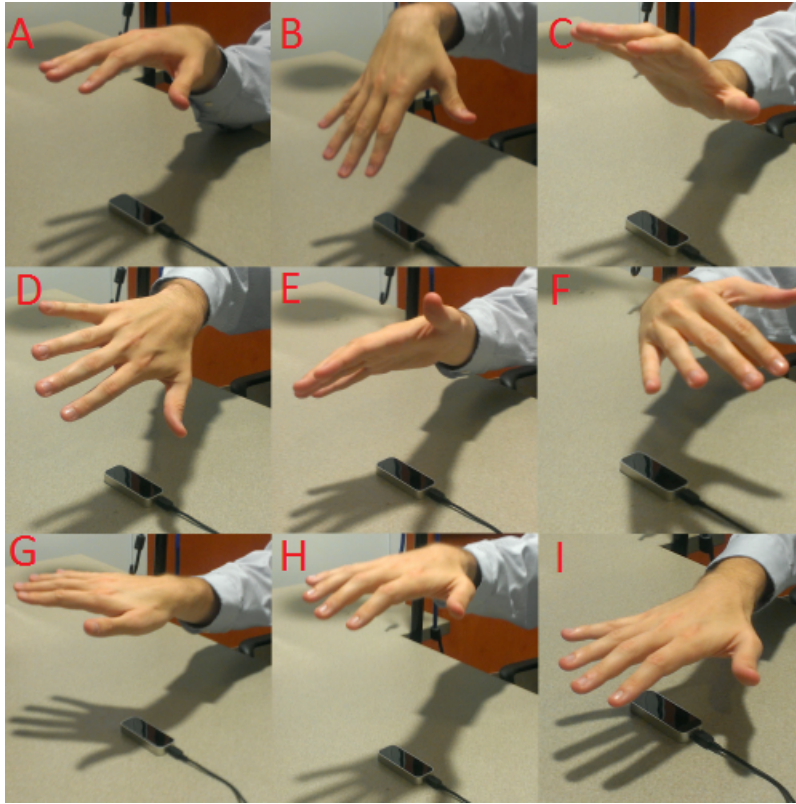
Figure 6.3: The First Person gestural commands - hand and fingers. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

Scaled Proxy Technique

The Scaled Proxy technique was developed in an effort to create a scaled version of the upper body Proxy technique. A reasonable set of commands was created to meets this goal, where the use of the two hands are involved only. Where the full body technique was achieved by moving both arms in an exaggerated manner, this version is essentially a reduced implementation; moving arms in a slight manner still achieves the same result. To command the UAV forwards or back, both wrists must be moved in front or behind the center of the device. Vertical climb and descent is achieved by moving both hands far above the sensor or close to the sensor. Strafing is performed by moving

one hand over another, as if to create a steering motion. Likewise, turning is achieved by moving one hand forward and the other backward. Figure 6.4 shows the gestural commands.

This technique is perceived to be a success, in terms of mimicking the upper body gestures. The entire gesture set has been replicated but scaled down to allow a user's hands by themselves to have complete control over the UAV. Similar to the original gesture set, this is also believed to be a rather easy-to-understand method of achieving task goals.
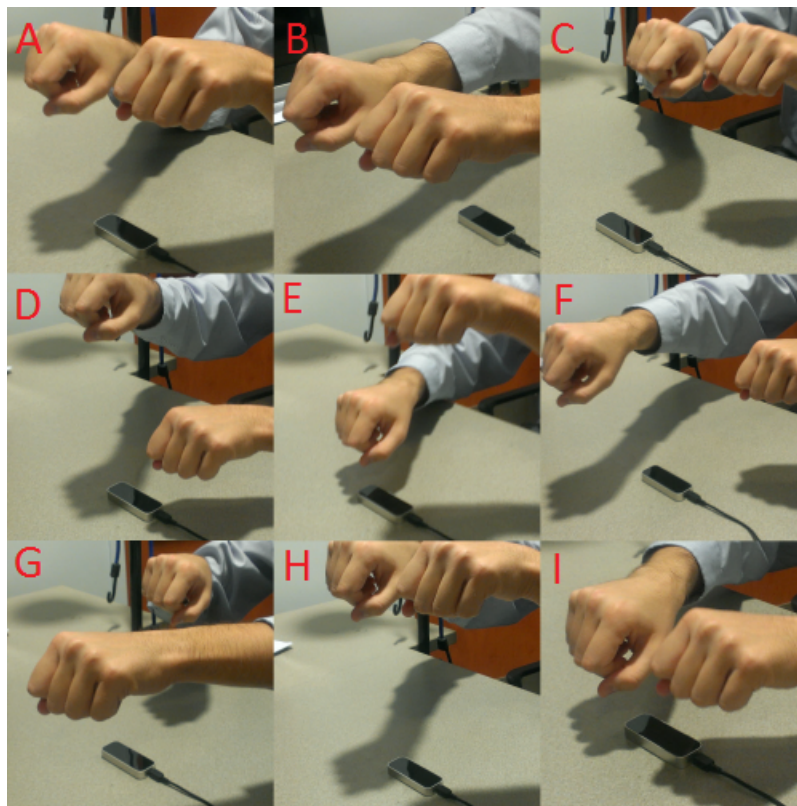

Figure 6.4: The Scaled Proxy gestural commands - hand and fingers. A: Base pose. B: Move Forward. C: Move Backward. D: Strafe Left. E: Strafe Right. F: Turn Left. G: Turn Right. H: Move Up. I: Move Down.

First Person

The hand-based first person interaction technique involves rotation on all three axes in order to send commands to the UAV. Imagining the hand as an aircraft and moving it in ways that directly map to the UAV is one very easy way to interact with the device. However, there are some downsides to using this technique, where rotation is key.

Firstly, to interact with our selected vision system, which is designed to lay flat on a desk or embedded in a laptop, the user would need to position the hand directly above it at all times. When sitting at a desk and just using the hand, this would seem to be very comfortable to use - however, after a prolonged period of time rotating the hand, it becomes tiresome and eventually difficult to keep the hand in the air. The counter this, resting the elbow on a surface to keep the hand prolonged in air would become the go-to pose - this is where problems with the technique are revealed.

Since rotation is required to give commands, it would seem crucial that all angles of rotation could be achieved - this potential constraint can not exactly be met by human nature. Rotating the hand to reveal the palm upwards is only achievable when rotating outwardly (clockwise for the right hand, counterclockwise for the left hand). It is not possible to rotate the hand in the opposite direction. This interaction technique warrants rotation of the hand from a starting position - palm faced downwards - to perhaps 90 degrees in either direction. This simply cannot be comfortably met.

Likewise, the turn commands have a bias in favor of the direction associated with rotating the hand inwardly (turn right, for the left hand, and turn left, for the right hand). The nature of the wrist limits outward rotation. Fortunately, it seems that by keeping the hand parallel with the sensor, there is no difference between the ranges of the move forward and move backward commands.

Use of this command can be not only tiresome but very difficult for users if not tuned

properly. Due to natural human limitations, it seems that although this is an easy metaphor to understand, the actual interaction is quite difficult and thus not recommended for use.

The Throne

The Throne was the worst 3D interaction technique for the upper body gestures, as it was built around an extended pointing metaphor; however, it seems that in the realm of hand-based interaction, where pointing is a very natural gesture, it has a chance to become useful. This technique finds the user moving the hand about the 3D space defined by the sensor. Navigation is achieved by just moving the hand to certain coordinates away from the center. This requires absolutely no rotation, so human physical constraints do not interfere with the interaction. Moving the hand opposed to rotating it feels a little more exaggerated than it needs to be, but the benefit is that no uncomfortable boundaries limit the movements.

The turning of the UAV through this technique is the biggest hindrance. It is achieved by rotating one finger either clockwise or counterclockwise, a method that seems to feel very distracting. The gesture must be performed continuously, until it is both detected, processed, and sent to the UAV to a desired stopping point; in reality, this is less of a "direct manipulation" style of interaction. While it can be performed in parallel with navigation, performing both at the same time may not feel as natural as when performing each separately. However, when successfully performing the parallel gestures, the circling motion may not be accurately recognized, especially when aiming the palm in a manner that occludes any piece of the hand.

This technique would benefit from an alternative manner to conduct the turning, so that all of the commands feel natural and more direct. Overall, this interaction technique seems much more appropriate and comfortable than the First Person technique, but there is much room for improvement in terms of providing the best method.

Scaled Proxy

The Proxy technique was the most successful interaction technique for the upper body gestures. In an attempt to recreate that success, but on a smaller scale, this technique was designed to allow users to interact with two hands simultaneously. This technique finds the user moving the hands in a way that almost exactly maps to the original technique. All navigation can be achieved by moving both hands together to a specific 3D coordinate, with the exception of the strafing, which can be performed by moving one hand over the other.

The rotation is very similar in nature to the strafing, except the hand movements are performed on a different axis. The scaled interaction, as a whole, seems to be quite successful, and would be expected to perform very well in a formal user study. We observe that because there is a smaller interaction space with a hand tracker, there is much more opportunity for a single hand to trigger an unwanted command. When using both hands at the same time, this chance is significantly reduced. This makes the Scaled Proxy technique very valuable.

Future Hand-and-Finger Tracking

Initial reactions towards these three interaction techniques find only the Scaled Proxy to feel successful, due to the nature of the interactions. It is simply more comfortable to use due to the minimal strain on the wrists and arms. However, since pointing is one of the most basic hand gestures, more work into alternative interaction techniques should be conducted to accommodate it.

Hand tracking in general is still not a well-explored technology, although work for devices such as the Leap Motion seem to be making much progress. One example includes the Digits device, which attaches to the wrist to provide constant tracking of the hand; this was not available for purchase at the time of this work. This device seems to be more of a finger tracker, but could be equipped with a gyroscope to measure rotation - translation may not necessarily be tracked. Of

course, the Leap Motion allows the alternative, tracking both rotation and translation, but from a detached perspective. Another future technology is the Myo [2], which has potential to detect hand rotation, translation, and finger muscle movements without the use of a vision system. If all data can be captured, gestures would not be constrained to a smaller interaction area, unlike the relative space for the Leap Motion.

In the future, optimal hand and finger trackers should be able to overcome problems not solved today. Occlusion is currently being tackled indirectly through data fusion, for instance combining four Kinect sensors [29]. Using similar technology, multiple Leap Motion devices (or other sensors) can be aggregated to form a better representation and more accurate recognition. This would allow the development of interaction techniques that are not feasible with current technology. A more natural interface would allow the user more freedom, opposed to constraining the hand to a specific interaction space, while still detecting all possible degrees of freedom.

---

[2]https://www.thalmic.com/myo/

CHAPTER 7: FUTURE WORK

In the previous sections, various interaction techniques have been developed and analyzed in hopes of finding a more intuitive user interface to command a UAV. While a decent variety of techniques and styles have been attempted, there are many that are not reported here - perhaps there are more natural techniques that have not been identified as of yet. This thesis focuses primarily on the 3D spatial interaction through upper body gestures, with some discussion on hand-and-finger gestures. Proper user studies for these preliminary interaction techniques are warranted.

Exploring 3DUI with other modalities that allow direct manipulation of a UAV should also be explored. HMDs can be used to provide a different avenue of 3D spatial interaction, with heavy focus on head tracking. [9] provides one potential application that uses HMDs for UAVs, but there may be a better style of interaction. As implied by the authors, vertical navigation may still require work with their current style. We plan on exploring the design implications of a more first-person approach to UAV command, especially associated human factors.

Additional work will be conducted on commanding various robotic platforms in a first-person context. Specifically, we plan on working with humanoid robots towards the goal of completely immersive teleoperation. Using 3D spatial interaction as a foundation for control, users will be able to command robots while seeing what they see. Alternatively, other modes of interaction can be studied, in an attempt to find even more styles of control. Brain Computer Interfaces (BCI) may be one such mode these greatly differ from 3DUI as they no longer involve physical input from the user, but rather they depend on the brainwave activity of a user for operation. This technology is already being explored for various usage, and may one day be harnessed for UAVs in the future [6][19]. It would be interesting to use BCI in conjunction with natural gestures, as a combinatorial method of providing context for UAV commands. Another mode of input that will be explored is that of speech recognition; while speech input requires more processing and thus has more latency between commands, adding this layer of input on top of 3DUI would enable a

user to provide greater context within the commands [4]. A study combining all forms of input would be conducted, to see if adding all of these modes of input increases the accuracy, decreases the task completion time, or augments the user experience in a positive manner.

Moving away from direct manipulation, further work will be conducted to optimize the user experience when commanding an autonomous UAV. Specifically, 3D spatial interaction can be used to assist pilots in updating flight patterns or sending way-points to autonomous aircraft. As [7] explores a sketch-based approach for requesting UAV support, 3DUI may be another potential avenue, as well as BCI and speech input, or a combination of these modes.

Additional work will be conducted on commanding multiple UAVs as part of a sequential or parallel interaction. For example, a pilot may be in command of two or more UAVs at any given time, and would perhaps need to give flight commands to each, either at the same time or one after another. Perhaps for both scenarios, user studies that measure task completion time when commanding multiple UAVs with alternative modes of interaction will be studied.

Manipulation of other robotic platforms utilizing 3DUI will be studied. Besides UAVs, unmanned systems including ground and underwater vehicles share the same principles and can be used to assist society with a variety of tasks. Studies will be conducted in an effort to identify interaction techniques most suitable for either platform. Additionally, non-vehicular robots will be targeted for direct manipulation using 3DUI. The humanoid robot as well as the robotic arm can be used for future teleoperative tasks, and work to identify proper interaction styles will be conducted.

CHAPTER 8: CONCLUSIONS

For this research, two devices implementing computer vision algorithms were used to develop interaction techniques to manipulate a quad-rotor micro UAV. While this aircraft shares many principals with military-grade UAVs, it takes a very different shape when compared to those currently used in combat. First, the UAVs such as the Predator, Reaper, and Desert Hawk are fixed-wing craft and fly like a standard airplane, whereas the UAV in this study has the ability to hover and turn without affecting its location. Second, most UAV pilots make use of a GCS as well as the on-board camera views, but do not have much opportunity to directly view the body of the aircraft, as it could theoretically be hundreds of kilometers away. In this study, the UAV was in the participants' viewpoint at all times, but the on-board camera feed was always available. Additionally, many GCS's implement tools such as the STK to assist the pilot with visualizing the aircraft. This work did not incorporate nor warrant the use of such a tool, but that is an effective application that can be considered.

Despite these differences, the input modality between remote and proximate UAVs are still similar. We expect the replacement of traditional input devices with 3D spatial interaction techniques to be seamless when considering an entire system.

Additional work is warranted to confirm the results of the user study presented in this thesis, for distant UAVs as well as other aircraft or robotic platforms. However, reactions by participants show that 3D spatial interaction may be a suitable replacement for traditional input devices in this domain. As the 3DUI were all considered as more fun than the smart phone application, we expect that pilots would exhibit decreased levels of boredom on the job while maintaining task completion efficiency, as evidenced per the user study results.

Apart from military scenarios, the experience with recreational or household UAVs can benefit from these gestural commands as well. Future everyday applications of UAVs may include assisting with tedious work, like checking drainage gutters on rooftops, surveying large yards or

farmland [22], or even retrieval of small objects [27]. In order to make these tasks less cumbersome and easier to perform, it may eventually become the norm to include a form of 3D spatial interaction.

This research, while reviewing only 5 upper body and 3 preliminary hand and finger interaction techniques, provides evidence that this form of input may be just as efficient as traditional input devices, while also serving as a more intuitive, fun, and liked channel through which to operate a UAV.

APPENDIX A: SAMPLE QUESTIONNAIRES

AR Drone / Kinect Gestural Interfaces Study – Kevin Pfeil & Seng Lee Koh

Participant Number_____

Please answer all questions below. Read the directions before answering any question. If you have any questions, feel free to ask.

Please fill in the blanks or circle the correct answer corresponding to that question.

Age _____

Gender   Male   Female

Occupation _____

Degree _____

I have used a Microsoft Kinect before.   True   False

I have flown an AR Drone before.   True   False

I have flown remote controlled vehicles before.   True   False

I play video games regularly.   True   False

I know a programming language.   True   False

AR Drone / Kinect Gestural Interfaces Study – Kevin Pfeil & Seng Lee Koh

Participant Number_____

Interface Name_____

Please answer all questions below. Read the directions before answering any question. If you have any questions, feel free to ask.

The following questions use a scale from 1 – 7, with 1 meaning "Strongly Disagree" and 7 meaning "Strongly Disagree". Circle the number that most closely corresponds to your answer.

Please rate the following factors using the described scale:

    1.) The interface to fly the Drone was comfortable to use.

1   2   3   4   5   6   7

Additional Comments:

    2.) The interface to fly the Drone was confusing to me.

1   2   3   4   5   6   7

Additional Comments:

    3.) I liked using the inteface.

1   2   3   4   5   6   7

Additional Comments

4.) The gestures to fly the Drone felt natural to me.

1   2   3   4   5   6   7

Additional Comments:

5.) It was fun to use the interface.

1   2   3   4   5   6   7

Additional Comments:

6.) I felt frustrated using the interface.

1   2   3   4   5   6   7

Additional Comments:

7.) I quickly understood how to use the interface.

1   2   3   4   5   6   7

Additional Comments:

If you have further comments, please record them here in the appropriate spaces. If you do not have remarks for a question, you may skip it.

Things I **liked** most about using this interface:

Things I **did not like** most about using this interface:

Suggestions for improving the interface:

Other Comments:

AR Drone / Kinect Gestural Interfaces Study – Kevin Pfeil & Seng Lee Koh

Participant Number_____

Please answer all questions below. Read the directions before answering any question. If you have any questions, feel free to ask.

The following questions use a ranking scale from 1 – 6. For an individual question, do not repeat a number. If you do not remember which interface is which, please ask. The proctor will remind you.

Please rank the gestures from 1 – 6 using the described scale.

1.) Rank the interfaces by their level of comfort to use, where 1 means "least comfortable" and 6 means "most comfortable."

Android Phone ___
First-Person Metaphor ___
Game Controller ___
Proxy Manipulation ___
The Throne ___
Seated Proxy ___

2.) Rank the interfaces by how confusing they are to use, where 1 means "least confusing" and 6 means "most confusing."

Android Phone ___
First-Person Metaphor ___
Game Controller ___
Proxy Manipulation ___
The Throne ___
Seated Proxy ___

3.) Rank the interfaces by how likable they were, where 1 means "liked least" and 6 means "liked most."

Android Phone _____
First-Person Metaphor _____
Game Controller _____
Proxy Manipulation _____
The Throne _____
Seated Proxy _____

4.) Rank the interfaces by how natural the interface felt to use, where 1 means "least natural" and 6 means "most natural."

Android Phone _____
First-Person Metaphor _____
Game Controller _____
Proxy Manipulation _____
The Throne _____
Seated Proxy _____

5.) Rank the interfaces by how fun the interface was to use, where 1 means "least fun" and 6 means "most fun."

Android Phone _____
First-Person Metaphor _____
Game Controller _____
Proxy Manipulation _____
The Throne _____
Seated Proxy _____

6.) Rank the interfaces by how frustrating they were to use, where 1 means "least frustrating" and 6 means "most frustrating."

Android Phone _____
First-Person Metaphor _____
Game Controller _____
Proxy Manipulation _____
The Throne _____
Seated Proxy _____

7.) Rank the interfaces by how easy they were to use, where 1 means "hardest" and 6 means "easiest."

> Android Phone   ___
> First-Person Metaphor   ___
> Game Controller   ___
> Proxy Manipulation   ___
> The Throne   ___
> Seated Proxy   ___

8.) Rank the interfaces by how using them met your expectations of the Drone's behavior, where 1 means "met expectations least" and 6 means "met expectations most."

> Android Phone   ___
> First-Person Metaphor   ___
> Game Controller   ___
> Proxy Manipulation   ___
> The Throne   ___
> Seated Proxy   ___

9.) With all factors considered, rank the interfaces by your disposition towards them, where 1 means "worst interface overall" and 6 means "best interface overall".

> Android Phone   ___
> First-Person Metaphor   ___
> Game Controller   ___
> Proxy Manipulation   ___
> The Throne   ___
> Seated Proxy   ___

Thank you for participating in the study. Your input is much appreciated and we are sincerely grateful for your time. If you have any additional comments, please provide them below.

APPENDIX B: IRB APPROVAL LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

From : **UCF Institutional Review Board
FWA00000351, IRB00001138**

To : **Kevin Pfeil cc: Jospeh La Viola (Faculty Supervisor)**

Date : **July 05, 2013**

IRB Number: **SBE-13-09458**

Study Title: **3D gesture metaphors for UAVs**

Dear Researcher,

Thank you for submitting the information regarding your Doctoral dissertation, as
requested by the IRB office. As you know, the IRB cannot approve your research
study because it was already completed prior to IRB review.

However, IRB Designated Reviewer Patria Davis has reviewed the study materials and
determined that if this proposal had been submitted to the IRB prior to conducting the
research; the study would have met the criteria for Expedited review and likely would have
been approved as being minimal risk to human subjects.

You may use the data collected for your dissertation, but remember in the future,
you must obtain IRB approval prior to conducting research that involves
human participants or their identifiable data.

If you have questions, please phone the IRB office at 407-823-2901.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Patria Davis

IRB Coordinator

LIST OF REFERENCES

[1] Alapetite, A., Hansen, J. P., and MacKenzie, I. S. Demo of gaze controlled flying. In Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, NordiCHI 12, ACM (New York, NY, USA, 2012), 773774

[2] Bowman, D. A., McMahan, R. P., and Ragan, E. D. Questioning naturalism in 3d user interfaces. Commun. ACM 55, 9 (Sept. 2012), 7888

[3] Bowman, D., Kruijff, E., LaViola, J., & Poupyrev, I. (2005). 3D User Interfaces: Theory and Practice. Addison-Wesley.

[4] B. Burger, I. Ferran, F. Lerasle, and G. Infantes. 2012. Two-handed gesture recognition and fusion with speech to command a robot. Auton. Robots 32, 2 (February 2012), 129-147. DOI=10.1007/s10514-011-9263-y http://dx.doi.org/10.1007/s10514-011-9263-y

[5] Chen, N., Tee, K. P., Chew, C.-M., and Yan, R. Intuitive interaction for robotic grasping. In Proceedings of the Workshop at SIGGRAPH Asia, WASA '12, ACM (New York, NY, USA, 2012), 105111.

[6] Chun Sing Louis Tsui and John Q. Gan. 2007. Asynchronous BCI control of a robot simulator with supervised online training. In Proceedings of the 8th international conference on Intelligent data engineering and automated learning (IDEAL'07), Hujun Yin, Peter Tino, Will Byrne, Xin Yao, and Emilio Corchado (Eds.). Springer-Verlag, Berlin, Heidelberg, 125-134.

[7] Cummmings, D., Fymat, S., and Hammond, T. Reddog: a smart sketch interface for autonomous aerial systems. In Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling, SBIM 12, Eurographics Association (Aire-la-Ville, Switzerland, Switzerland, 2012), 2128.

[8] Cummings, M.L., Mastracchio, C., Thornburg, K.M., Mkrtchyan, A.,. Boredom and Distraction in Multiple Unmanned Vehicle Supervisory Control, Interacting with Computers, Vol. 25, No. 1, pp. 34-47, 2013.

[9] Higuchi, K., and Rekimoto, J. Flying Head: head-synchronized unmanned aerial vehicle control for flying telepresence. In SIGGRAPH Asia 2012 Emerging Technologies, SA '12, ACM (New York, NY, USA, 2012). 12:1-12:2.

[10] Holm, S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6 (1979), 6570.

[11] Jeon, S., Hwang, J., Kim, G. J., and Billinghurst, M. Interaction techniques in large display environments using hand-held devices. In Proceedings of the ACM symposium on Virtual reality software and technology, VRST 06, ACM (New York, NY, USA, 2006), 100103.

[12] Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In Proceedings of the 25th annual ACM symposium on User interface software and technology, UIST '12, ACM (New York, NY, USA, 2012), 167176.

[13] Lambrecht, J., Kleinsorge, M., and Kruger, J. Markerless gesture-based motion control and programming of industrial robots. In Emerging Technologies Factory Automation (ETFA), 2011 IEEE 16th Conference on (2011), 14.

[14] Lichtenstern, M., Frassl, M., Perun, B., and Angermann, M. A prototyping environment for interaction between a human and a robotic multi-agent system. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, HRI 12, ACM (New York, NY, USA, 2012), 185186

[15] Marchal, M.; Pettre, J.; Lecuyer, A *Joyman: A human-scale joystick for navigating in virtual worlds* 3D User Interfaces (3DUI), 2011 IEEE Symposium on , vol., no., pp.19,26 19-20 March 2011.

[16] Ng, W. S., and Sharlin, E. Collocated interaction with flying robots. In RO-MAN, 2011 IEEE (2011), 143149.

[17] Ouramdane, N., Otmane, S., Davesne, F., and Mallem, M. Follow-me: a new 3d interaction technique based on virtual guides and granularity of interaction. In Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications, VRCIA 06, ACM (New York, NY, USA, 2006), 137144

[18] Pfeil, K., Koh, S. L., and LaViola, J. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In Proceedings of the 2013 international conference on Intelligent user interfaces, IUI 13, ACM (New York, NY, USA, 2013), 257266.

[19] Riccardo Poli, Caterina Cinel, Ana Matran-Fernandez, Francisco Sepulveda, and Adrian Stoica. 2013. Towards cooperative brain-computer interfaces for space navigation. In Proceedings of the 2013 international conference on Intelligent user interfaces (IUI '13). ACM, New York, NY, USA, 149-160. DOI=10.1145/2449396.2449417 http://doi.acm.org/10.1145/2449396.2449417

[20] Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In Proceedings of the 9th annual ACM symposium on User interface software and technology, UIST 96, ACM (New York, NY, USA, 1996), 7980

[21] Poupyrev, I. and T. Ichikawa, Manipulating objects in virtual worlds: categorization and empirical evaluation of interaction techniques. in Journal of Visual Languages and Computing, 1999. 10(1): p. 19-35.

[22] Quaritsch, M., Stojanovski, E., Bettstetter, C., Friedrich, G., Hellwagner, H., Rinner, B., Hofbaur, M., and Shah, M. Collaborative microdrones: applications and research challenges. In Proceedings of the 2nd International Conference on Autonomic Computing and Communication Systems, Autonomics 08, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (ICST, Brussels, Belgium, Belgium, 2008), 38:1–38:7.

[23] Smith, C., and Christensen, H. Wiimote robot control using human motion models. In Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on (2009), 5509–5515.

[24] Song, P., Goh, W. B., Hutama, W., Fu, C.-W., and Liu, X. A handle bar metaphor for virtual object manipulation with mid-air interaction. In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI 12, ACM (New York, NY, USA, 2012), 1297–1306

[25] Song, Y., Demirdjian, D., and Davis, R. Continuous body and hand gesture recognition for natural human-computer interaction. ACM Trans. Interact. Intell. Syst. 2, 1 (Mar. 2012), 5:1–5:28.

[26] Sturman, David J. Whole-hand Input. Ph.D dissertation, Massachusetts Institute of Technology, 1992

[27] Thomas, J., Polin, J., Sreenath, K., and Kumar, V. Avian-Inspired Grasping for Quadrotor Micro UAVs. In to appear IDETC/CIE, ASME (2013).

[28] Urban, M., and Bajcsy, P. Fusion of Voice, Gesture, and Human-Computer Interface Controls for Remotely Operated Robot. 2005.

[29] Williamson, B., LaViola, J., Roberts, T., and Garrity, P. Multi-Kinect Tracking for Dismounted Soldier Training, Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012, 1727-1735, December 2012.

[30] Zuher, F., and Romero, R. Recognition of human motions for imitation and control of a humanoid robot. In Robotics Symposium and Latin American Robotics Symposium (SBR-LARS), 2012 Brazilian (2012), 190195.