THESIS

COLLABORATING WITH ARTISTS TO DESIGN ADDITIONAL MULTIMODAL AND UNIMODAL INTERACTION TECHNIQUES FOR THREE-DIMENSIONAL DRAWING IN VIRTUAL REALITY

Submitted by

Brian T. Sullivan

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2023

Master's Committee:

Advisor: Francisco Ortega

Sudipto Ghosh
Cyane Tornatzky
Mayra Barrera Machuca
Anil Ufuk Batmaz

ABSTRACT

COLLABORATING WITH ARTISTS TO DESIGN ADDITIONAL MULTIMODAL AND UNIMODAL INTERACTION TECHNIQUES FOR THREE-DIMENSIONAL DRAWING IN VIRTUAL REALITY

Although drawing is an old and common mode of human creativity and expression, virtual reality (VR) has presented an opportunity for a novel form of drawing. Instead of representing three-dimensional objects with marks on a two-dimensional surface, VR permits people to create three-dimensional (3D) drawings in midair. It remains unknown, however, what would constitute an optimal interface for 3D drawing in VR. This thesis helps to answer this question by describing a co-design study conducted with artists to identify desired multimodal and unimodal interaction techniques to incorporate into user interfaces for 3D VR drawing. Numerous modalities and interaction techniques were proposed in this study, which can inform future research into interaction techniques for this developing medium.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Francisco Ortega, for his support and encouragement on this project, and for his openness to consulting people with a variety of skill sets and perspectives when conducting research. I would also like to thank the members of my committee for the time and commitment they provided to make this project possible: Dr. Sudipto Ghosh (Colorado State University), Dr. Mayra Barrera Machuca (Dalhousie University) and Dr. Anil Ufuk Batmaz (Concordia University) for also providing insights and suggestions when designing the study, and Cyane Tornatsky (Colorado State University) for also providing input helping us find art students with whom to work. I also thank all of the members of the NUI Lab, especially Richi Rodriguez, with whom I collaborated closely on this project.

I also want to thank my parents, Kathleen Maloney and Dr. Tom Sullivan, for brainstorming with me and for being loving and supportive through everything.

A final general thanks goes to everyone who contributed to this project, directly and indirectly.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Commercially available tablets and drawing tools have made digital drawing an accessible and popular medium of two-dimensional (2D) art that has proliferated over the past several decades (e.g., devices such as the Wacom Cintiq and iPad Pro and software such as Adobe Photoshop and ProCreate). Similarly, the increasing availability of Virtual Reality (VR) head-mounted displays will likely lead to similar growth for three-dimensional (3D) drawing as an artistic medium. Unlike 2D digital drawing, however, 3D drawing does not have as many analog (non-digital) counterparts to inform the design of the VR interface and user experience. Although there are 3D analog art forms, such as sculpture, that have been leveraged for some VR art programs focused on 3D modeling [1, 2], sculpting is distinct from drawing, and drawing in VR is likely to serve a role earlier in the creative process than actual 3D modeling [3–6]. Additionally, digital systems offer opportunities for user interactions that are less commonly possible in analog workflows, such as selecting and repositioning a part of the image that was already drawn or duplicating that selection. One such type of interaction would be multimodal interaction, where two or more modes of interaction are used in tandem to accomplish a task, such as using a pointing gesture to indicate an object in the environment and then speaking to specify an action to be done to that object. This is in contrast to unimodal interactions, where a single mode of interaction is used to accomplish the task. Although the ultimate objective would be to evaluate the usability of various multimodal interactions by how they perform for users, it would be beneficial to first identify promising opportunities in the VR drawing process for multimodal interaction.

A co-design study was conducted to identify multimodal and unimodal interactions that would improve a 3D VR drawing interface. Co-design studies involve the active participation of likely users to ensure that a system is designed in a way that is usable. Thus, this co-design study recruited the participation of artists. The multimodal and unimodal interactions that are identified through this study will then be able to inform subsequent studies, which can test the viability of such interactions in VR drawing. Additionally, the findings from such studies have the potential to be applied to other forms of extended reality (XR), such as Augmented Reality (AR) and various points on the mixed reality (MR) continuum [7].

The co-design approach for this study involved recruiting participants with art experience, which was defined as either being actively enrolled in a fine arts or graphic design degree program or having graduated from such a program. Although art skills can be learned in a variety of environments, this requirement helped to ensure that there was some commonality between the participants' backgrounds. Participation in this initial study was limited to artists to ensure that the responses would be more focused on achieving actual art and drawing goals. There is also precedent for working with art and design students to evaluate the usability of similar systems, since these applications are often intended to be used by such individuals in their work and practices [3, 5, 8, 9].

The objective of this study was to identify multimodal and unimodal interaction techniques in VR drawing that could improve the interface and drawing experience for users. Participants were presented with a unimodal VR drawing system and given the freedom to suggest any multimodal or unimodal interaction techniques they thought could be beneficial, regardless of the feasibility of the implementation. Thus, rather than asking participants to rate what kind of multimodal interaction techniques they would prefer for a given task, they were all taught about multimodal interaction and then asked to complete a series of drawing activities, during which they were free to propose any interaction techniques at any time. Two experimenters were present, with one monitoring technical support while the other asked participants at critical junctures if a multimodal interaction would help with what they were doing. These questions were always generic and phrased in such ways as, "Is there a multimodal interaction that would help you with this task?" This was to ensure that any proposed interaction techniques came from the participants and were not biased by a participant simply agreeing with a technique proposed by an experimenter. Examples of participants' drawings can be seen in Figure 1.1.

Figure 1.1: Examples of 3D drawings created by two participants in the study. Images labelled **A** are views of one participant's drawing, and those labelled **B** are views of the other participant's drawing. Both drawings were created in Open Brush using a VR headset.

Upon completion of the study, the collected data was processed to identify the different proposed multimodal interaction techniques—as well as interaction techniques that only identified individual modalities. Participants were also asked to rate their experiences using the system. From this, additional thematic analysis can be performed to identify specific multimodal and unimodal interaction techniques to be implemented for future studies that can evaluate the actual usability of those interactions. Future studies would also begin to work with non-artists, in addition to continuing to work with artists, in order to determine how to make systems usable across skill levels, and also potentially identify whether the needs and desires of artists differ significantly from those of non-artists when it comes to drawing in virtual reality.

1.1 Contributions

The desired outcome of this study is to offer insights into what interaction techniques artists may benefit from in the interface of a VR 3D drawing application. By conducting a co-design user study with artists, multimodal and unimodal interaction techniques were proposed, along with additional tools and features that were not present in the system used in the study. These proposed interaction techniques can be further explored, implemented, and tested, to ensure that they do enhance the usability of the system. This is also the case for the new features that were proposed. This opens numerous avenues for further investigation to advance this medium. These findings could also lead to the incorporation of 3D Drawing techniques into other VR activities, such as annotation for immersive analytics or staging and blocking for visual effects-heavy scenes in motion pictures.

Chapter 2

Literature Survey

Drawing[1] in three dimensions (3D) in virtual reality (VR) is a domain with clear applications in art and design, but can also have uses in other domains where quick visualization in spatial environments is necessary. As a different medium from 2D drawing, 3D drawing also offers greater immersion for those who are drawing and does not require the projection of 3D concepts onto a 2D plane [4]. This, coupled with the fact that the activity itself involves the complex interplay of many different technologies and user behaviors, means that there are many research questions that have been and continue to be explored in relation to drawing in virtual reality. The primary focus of this thesis is on identifying multimodal—and additional unimodal—interaction techniques that might be implemented in interfaces for VR 3D drawing programs, with an assumption that these insights will likely overlap in AR and other XR experiences.

2.1 Techniques to Aid in 3D Drawing

Although this thesis focuses primarily on multimodal inputs that users would use to interact with the system, there has been previous work exploring how different modes of feedback (output) from the system affects users' performance and experience. In the real world, people typically draw or paint onto physical surfaces, where they receive resistance and tactile feedback as they work [10], and this can even be seen in 2D digital art, where tools such as drawing tablets offer an analog to that experience. In 3D VR drawing, by contrast, users are typically drawing in midair with no such physical constraints or feedback. Elsayed et al. have identified how several different methods of generating haptic feedback in a pen controller can help to improve VR drawing [11].

[1] "Drawing" is a broad term, which encompasses processes ranging from making quick, personal sketches [10], used in design ideation, to the creation of detailed and refined finished works. Many of the previous works cited have focused primarily upon sketching and the ways in which it is used by designers for planning and ideation. Due to the nature of the study described in this thesis, "drawing" is being used, to account for the breadth of applications for such a system as both a sketching tool and a tool for creating a finished piece.

This sort of work may be applicable in the future, along with work such as An et al., which endeavored to enhance immersion in a VR automotive design process by having human "haptic helpers" provide physical props and haptic feedback that emulated designs [3].

An area in which users often struggle with XR drawing is the precise and accurate placement of strokes for drawing in three dimensions. While haptic feedback could aid in this, non-haptic solutions have also been studied. Barrera Machuca et al. observed that users' spatial abilities can affect their 3D drawings [12]. One way to possibly compensate for this could be generating smart visual guides that using either the controller or the world as a reference point to improve users' 3D drawings, regardless of their spatial ability [6]. Similarly, Kim and Bae presented a system in 2016 that could scan the poses of users' hands and create reference models in a 3D sketching program to aid in drawing concept designs for items with appropriate scale and ability to be handled relative to human hands [5]. It should be noted, however, that this study ran the 3D drawing program on a 2D drawing tablet, meaning that these reference hands' effects on drawing spatially may differ from what their effects would be in a VR environment.

Other research [13] has focused on ways to use 3D drawing to apply strokes to 3D virtual surfaces—similar to how one might paint a model or miniature. Arora and Singh studied midair drawing onto 3D surfaces and observed that users frequently emulated the curves of the 3D target object when drawing objects. This indicated that it was valuable for the system to mimic the stroke positions precisely when projecting them onto the 3D surfaces, rather than the more common method, which emulates the behavior of spray paint by projecting strokes onto the 3D surface based on the position and orientation of the controller. In this instance, rather than simply providing outputs and feedback to help guide users to draw more effectively in 3D, it was identified how the system could process inputs to behave according to users' needs and expectations.

2.2 Multimodal Inputs for 3D Drawing

At this time, VR drawing systems primarily make use of two controllers, one in each hand, meaning the interactions are performed bimanually while ultimately remaining unimodal in nature. For instance, in the Open Brush program [14], the controller in the user's dominant hand is used for drawing strokes and selecting menu items, while the controller in the user's non-dominant hand often serves as a menu system for selecting brush styles, colors, and performing other operations. Because of the limited, and sometimes cumbersome nature of such a system, research into additional modes of input is being explored.

One approach to a kind of multimodal interaction was the VRSketchIn system developed by Drey et al., which created a hybrid between 3D and 2D sketching [15]. This system allowed participants to draw in midair in 3D but also draw 2D elements to be projected into the 3D space on a physical tablet, which had a visual representation provided in the VR environment [15]. This system made use of a pen controller and provided a multimodal interaction technique in the sense that users would shift from the midair techniques of 3D drawing, to a 2D process that offered haptic feedback. This system did not, however, explore opportunities to use gestures, speech, or gaze to enhance the drawing experience or process.

Xu et al. offered a gestural and bimanual form of interaction to aid in VR sketching [16]. While users were actively drawing with their dominant hands, they could use their non-dominant hands to construct reference scaffolding that could act as guides for drawing in 3D. These scaffolds were constructed from combinations of five pre-defined primitive shapes (a surface, a curved surface, a sphere, a cylinder, or a cone), which were generated using a series of gestures. Similarly, other researchers have developed systems that allowed 3D objects to be sculpted using gestures [1, 17]. In each of these instances, however, the modality selected was a gesture, and the motivation was rooted more in similarities to sculpting than to drawing. Thus, there remains a question about whether other multimodal interaction techniques could be helpful with the focus on drawing, as well as what those modalities may be.

2.3 Co-Design Studies

Much of the previous work described in 3D drawing has tested the usability of certain techniques, either on their own or in comparison to other techniques [1, 5, 6, 11–13, 15, 16]. Although this approach provides many insights, other studies [3, 4, 8, 9, 18] have employed a co-design approach wherein experimenters worked more collaboratively with participants who have experience in the domain for which this technology is being designed. For instance, Israel et al. conducted a pair of studies reported in a single paper [4]. The first of their studies recorded focus group sessions with furniture designers who proposed uses and needs for a 3D sketching program, and the second had different furniture designers participate in a user study where they completed sketching tasks in either 2D or 3D [4]. Being focused on sketching, the responses from the focus group emphasized the need for the process to be quick and intuitive, with limited technical distractions. The participants in the user study portion indicated that drawing in 3D helped them to think and work spatially and that they could use the process in the early stages of product design. A different study, by Weise et al., explored how participants improved at drawing basic shapes in 3D over time [18]. This involved evaluations of the shape drawings after each session based on shape quality and the time taken to draw, as well as self-reports by the participants rating their subjective workload, or how much effort they felt went into performing the tasks. The results from Weise et al.'s study indicated that the quality of the shapes drawn improved, although the time taken to draw the shapes did not change. Inspired by previous work such as these two studies [4, 18], this thesis sought to identify what sorts of multimodal and unimodal interaction techniques users would like to have available in the VR drawing interface. For this reason, a co-design approach was implemented for the user study.

Using insights gleaned from surveys of co-design studies for developing technology in healthcare domains, it was observed that it is important to ensure that the user group for the study is well-defined and to remember that even within a well-defined user group, there will be heterogeneity among participants [19]. Additionally, because co-design studies are qualitative in nature, and participants are welcome to offer insights and suggestions in a freer form than surveys and questionnaires account for, researchers must consider more than just the direct topic of interest [20]. These insights are relevant to consider when describing the data collected from this user study. While the participant pool was very selective and defined, artists will likely differ in their preferences and priorities—although there may also be considerable overlap. Similarly, while the main focus of this thesis and user study is interaction techniques, other comments on the interface and experience should also be noted. That information could not only be relevant for improving the user experience, but could also open additional avenues for exploration and research.

Chapter 3

Methodology

The following describes the methodology of the study. This includes the specifics of the apparatus used for the study, a description of the open-source drawing software used, and the procedure of the study itself.

3.1 Apparatus

This project required a VR headset with two handheld controllers to use 3D virtual reality sketching software in order to evaluate the opportunities for multimodal input in such a system. Thus, an HTC Vive Pro Eye headset was used, along with its included controllers, in order to use Open Brush as the drawing program of choice.

An HTC Vive Pro Eye VR headset was used for this study (Figure 3.1), in part because of its gaze tracking capabilities [21]. Gaze tracking was not strictly necessary for the first study, since the user's input was intentionally unimodal.

While a wireless adapter module is available for the HTC Vive, it proved to be unreliable. Therefore, the the VR headset was tethered (i.e., connected directly) to the computer. The computer was an Alienware Aurora Ryzen Edition R14 running Microsoft Windows 11 Home with an AMD Ryzen 9 5900 12-Core CPU (32 GB RAM), an NVIDIA GeForce RTX 3080 GPU (26 GB RAM).

Participants watched software tutorial and multimodal interaction briefing videos on an iPad. Video and audio from the co-design sessions were recorded using a GoPro camera.

Open Brush version 2.2.0 was used for this study and was downloaded from the Open Brush GitHub repository [14] and run from the Unity Hub. Open Brush was selected for being a stable VR drawing program with a wide array of features. The fact it was open-source left the opportunity open to modify it directly for future research where new interaction techniques may be implemented.

Figure 3.1: An example of an individual wearing an HTC Vive Pro Eye headset and holding the headset controllers. As illustrated in this image, participants did not have the headphones near their ears, so that they could hear the interview questions from the experimenters.

3.2 Three-Dimensional Drawing with Open Brush

Open Brush is an open-sourced branch of Google's Tilt Brush software, which offers numerous features for three-dimensional drawing in virtual reality, including different brush styles, color selection, and the ability to select and change brush strokes [14]. Users of Open Brush interact with the program by using two hand controllers (Figure 3.2). The user's dominant hand is used to draw, select, delete, rotate, and reposition strokes in the drawing environment, as well as changing brush size and calling up and using a mini pop-up menu. Meanwhile, a user's non-dominant hand acts as a palette of sorts, containing an array of menu options, including the brush styles and color selection, as well as undo and redo operations. A few features also require both hands working in conjunction, such as resizing a selected object or resizing the drawing environment as a whole.

Because Open Brush uses bimanual unimodal input in the form of controllers, it was well-suited for this initial co-design study. The lack of existing multimodal inputs ensured that participants would not be biased toward options that were already present. Open Brush's unimodality also kept the actual performance of a hypothetical multimodal interaction technique from affecting how participants perceived the various multimodal input options. Additionally, the fact that Open Brush is an existing system with stable implementations of the relevant features was another factor in selecting it for this initial study, as this once again ensured that participants would be less distracted by the performance of the existing features and could instead speak to additional features and interaction techniques that they wanted in such a system.

The ultimate goal of this research into 3D drawing in virtual and augmented reality is to determine what multimodal and unimodal interaction techniques would improve the interface design for those who need or wish to draw in three dimensions. Eventually, this needs to account for users of any skill level, so that it is neither overwhelming for novices nor limiting for expert artists. Future stages of this research will address this by recruiting participants of varying skill levels and

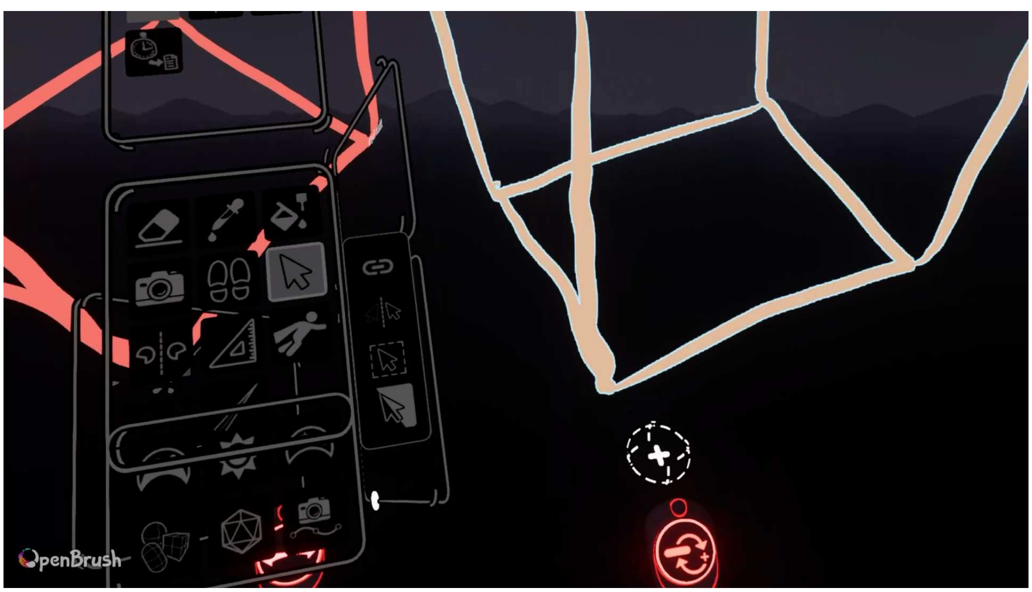

Figure 3.2: An example of the Open Brush interface, as seen by the user wearing the headset. In this image, the left-hand controller contains the menus for choosing brushes, changing colors, and accessing modification controls (pictured). The right-hand controller performs actions that affect the world, such as making strokes and selecting strokes (pictured). Controller roles can be swapped to accommodate users who are left-handed.

quantitatively evaluating participants' performance when using the system with the various input modalities. Before such experiments can be conducted, however, it must be identified what multimodal inputs would be most beneficial to implement and test in the future. As such, this initial user study specifically recruited individuals with some amount of formal, college-level art training to participate in a co-design study and propose possible multimodal inputs that they would potentially find useful. Based upon these results, a VR drawing prototype program can be implemented that would allow for the experimental testing of the various proposed input modalities.

For this initial study, the primary interest was multimodal and unimodal interaction techniques, with the main focus being on the modalities of gesture, gaze, speech, and pen. Participants were asked to propose how one or more modalities would assist in completing various parts of drawing tasks.

3.3 Co-Design Study

For this co-design study, individuals were recruited who had at least some formal art training at the post-secondary level. Sessions ran for approximately 1 hour, and participants were compensated for their time, either with extra credit in the art class of a cooperating professor or with $20 in Amazon Gift Cards. Each participant ran the study individually with one or two experimenters present. When two experimenters were present, one would interact with the participant while the other provided technical support to monitor recordings and hold the cable tethering the headset to the computer. If only one experimenter was present, that experimenter would perform both jobs.

3.3.1 Participants

There were 13 participants in total (Average Age = 22.77, SD = 2.42). Two participants had graduated with degrees and the remaining participants were currently enrolled in art-related degree programs. The majors/degrees represented were Art (5), Visual Art (1), Integrated Visual Studies (3), Electronic Art (2), and Graphic Design (2). Among the participants, two reported never having used a virtual or augmented reality headset before. One other reported having used one "once for a very short period of time" and another specified having used a headset twice. All other participants simply responded "yes" to having used a virtual or augmented reality headset.

3.3.2 Tutorial & Training

Upon agreeing to participate in the study, participants completed a preliminary questionnaire that included questions about participants' past experience with virtual or augmented reality headsets, as well as asking them to estimate the average amount of time they spend playing certain video and computer games and using certain kinds of console controllers each week. These questions could be used to contextualize participants' backgrounds and determine whether familiarity with VR or particular gaming systems may have influenced the interaction techniques they suggested. Participants then watched an approximately 4-minute tutorial video [22] that was produced for this study and demonstrated the basic drawing features of Open Brush. The tutorial video was custom-made, to ensure that all participants were aware of the basic features of Open Brush and that the tasks could be replicated in the training. This video consisted of a recording of the actions from the view of the VR headset along with a voice-over narration that described how the action was performed. The script for this tutorial video can be found in Appendix A.1.1.

After the tutorial video ended, the participants were shown the controllers and headset. The participants then donned the headset and carried out the same actions demonstrated in the tutorial video, with the instructions read aloud by the experimenter. During this tutorial, participants were welcome to ask how to accomplish the specified actions and experimenters could provide verbal assistance. When participants completed the tutorial, they were offered the chance to take a 2-minute break before proceeding to the next stage.

3.3.3 Briefing on Multimodal Interaction

Once participants completed the tutorial stage of the study, they were briefed on multimodal interaction and the goal of the experiment. To be briefed on the basic concepts of multimodal interaction, participants watched an approximately 5-minute video that was produced for this study [23].

A custom video was made to ensure that it contained all of the information relevant to the study, so that participants would have a common baseline of knowledge.

The video consisted of visuals that included written information and definitions, along with simply animated illustrations of certain actions or concepts. There was also voice-over narration that provided information and definitions, as well as describing the situations illustrated by the animations. The script for this voiceover can be found in Appendix A.1.2. The video described the difference between unimodal and multimodal interaction and also gave definitions and examples of speech, gesture, gaze, pen, and bimanual interaction[2] as modalities to consider.

After the video briefing ended, an experimenter would read from a script (Appendix A.1.3), which explained the goal of this initial study and told the participants that they would be asked to complete certain drawing tasks in Open Brush and were free to use any tools available in the system. (Participants were not limited to only using the tools demonstrated in the tutorial stage.) The participants were also told that experimenters would be asking questions about possible multimodal input opportunities they drew. At the end of the briefing, participants were given the opportunity to ask questions and those questions were answered.

Because this was a co-design study, the objective of the experiment was not kept secret from the participants. This could ensure that participants knew what was of primary concern and would focus more on multimodal inputs than they would on other features—although the participants were welcome to comment on anything relating to the system.

3.3.4 Drawing Activities

Following the tutorial and briefing stages of the study, participants donned the headset and complete two drawing activities. The first activity was to draw a dog. Participants were told that they could draw the dog in any artistic style they wished, so long as it was 3-dimensional. Participants had 10 minutes to complete this activity. This time limit ensured that participants

[2]Bimanual interaction can be considered a form of multimodal interaction in itself, due to the presence of two hands. However, since it could be coordinated with another modality, such as speech, it was treated as a unique modality option in this study.

had time to complete the activity but did not overwork the drawing. A dog was selected as the object to draw, as it would involve drawing many organic shapes and would be complex enough to take time for artists to draw while not being overly intricate. Participants could draw in any style, ensuring that they could assess and comment on the system for how they would most likely use it. Experimenters watched a video feed of what the participants saw in the headset and would ask the participants to suggest multimodal inputs that would assist them with the various parts of drawing the dog. Participants were also invited to volunteer comments about the system or interface if or when they so chose. Additionally, experimenters could provide some verbal assistance with accomplishing a task *if* a participant asked for it. When this occurred, it typically led to a question about possible multimodal inputs that could have assisted the participant.

When the first drawing activity was completed (either because time was up or the participant clearly stated being done), participants were offered the opportunity to take a 2-minute break before proceeding to the next drawing activity.

The second of two drawing activities was to draw an outdoor environment for the dog that participants drew in the first activity. For this activity, participants had 15 minutes to draw and were asked to, at minimum, draw the ground, a path, and a tree. Once again, they were welcome to draw in any style they wished, so long as the drawing was 3-dimensional. The participants were also welcome to draw additional items in the environment if they wanted to and had time remaining. In all other ways, the interview process for the second drawing activity was the same as that of the first.

During the drawing activities, audio and video of the participants and experimenters were recorded using a GoPro camera and its built-in microphone. The view of Open Brush from within the VR headset was also recorded.

3.3.5 System Usability Scale & Post-Interview

After completing the drawing activities, participants were asked to complete a System Usability Scale (SUS) to rate their experience using the system [24]. An experimenter then verbally asked follow-up questions pertaining to the participants' experiences and possible opportunities for multimodal interaction (Appendix A.2.3). Participants were then asked if they had any remaining questions and offered their choice of compensation.

The GoPro recording started during the drawing activities continued through this stage up until the moment before asking participants about compensation. Participants' choices in compensation were omitted from the recordings for the privacy of the participants.

3.3.6 Data Analysis

The video/audio recordings of each participant's drawing activities were synchronized to the corresponding recordings of the headset views for simultaneous viewing and analysis. The audio recordings were also isolated from the video and transcribed using the transcription feature of the Microsoft Word Web App [25]. These transcriptions were then processed into a spreadsheet that separated each sentence for analysis. Analysis was then performed on a sentence-by-sentence basis, noting the actions observed in the video recordings and making note of interaction techniques proposed by the participants, as well as other features and comments mentioned by participants.

Following this, unique interaction techniques were noted and counted according to the modality (or modalities) proposed, as well as whether they were unimodal or multimodal in nature. Then these counts were compared using the Kruskal-Wallis rank sum test, to determine whether the differences in frequency of reporting were significant.

Chapter 4

Results

4.1 Preliminary Questionnaire

On the preliminary questionnaire, participants reported playing various genres of video and computer games for an average of 14.93 hours (SD = 12.82) each week. Participants' playing habits varied considerably, with the most hours reported by a participant being 40 hours a week and the least (reported by two participants) being 0 hours a week. The most commonly reported amount of time playing games each week was 10 hours. The most played genre of game was fantasy/adventure with an average weekly play time across participants of 4.42 hours (SD = 5.16) (Figure 4.1). Regarding console controllers, the average reported time spent using any controller per week was 15.33 hours (SD = 12.82), with a range from 0 to 40 hours per week, and the most frequently occurring amount of time being 6 hours. The most popular controller type was "keyboard and mouse" with an average weekly time rating of 6.00 hours (SD = 11.03). Other controllers reported to be used for one or more hours per week were "touchscreen" (Average: 4.31 hrs/week, SD = 9.46), "Playstation controller" (Average: 2.00 hrs/week, SD = 2.58), "Xbox, Xbox 360, and Xbox One controller" (Average: 1.38 hrs/week, SD = 2.96), and "Nintendo DS and Nintendo 3DS" (Average: 1.23 hrs/week, SD = 2.86). Some participants also wrote in additional controllers under "Other." One participant reported using Oculus Quest 2 controllers 3 hours a week, which would amount to an average of 0.23 hrs/week (SD = 1.31) across all 13 participants, and three participants reported using Nintendo Switch controllers, with their weekly time using these controllers leading to an average usage of 1.31 hrs/week (SD = 2.63) across all 13 participants (Figure 4.2).

One participant elected to sit for the drawing activities and would use the controllers to change the position and size of the space while drawing, while everyone else chose to stand and walk around their drawings in addition to moving them with the controllers. Examples of two of the participants' drawings can be seen in Figure 1.1.

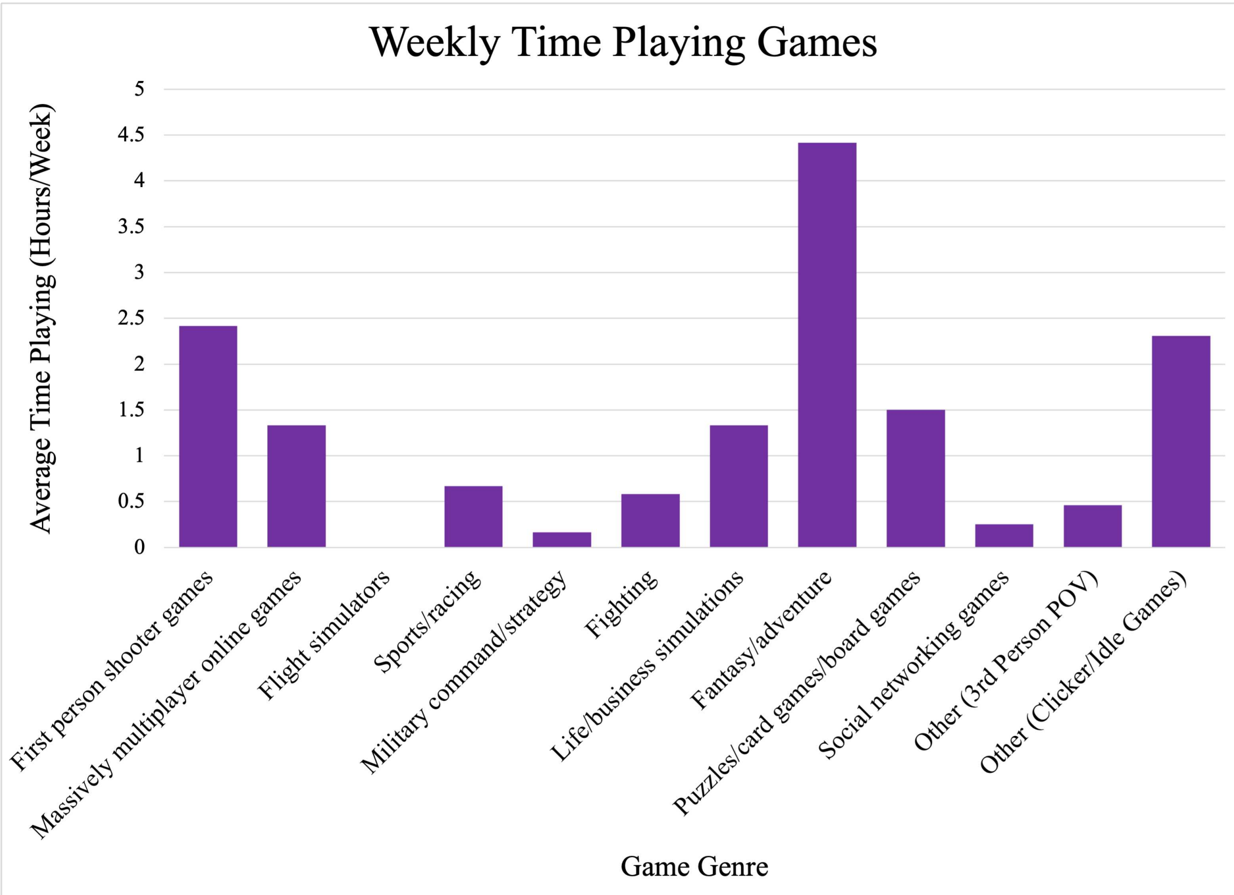

Figure 4.1: The average time playing the indicated game genres each week. The total average time playing games was 14.93 hours/week (SD = 12.82).

4.2 Proposed Interaction Techniques

Across the 13 participants, 125 unique kinds of interaction techniques were proposed that either made use of multiple modalities or specified a desire to use a single modality (full list found in Appendix B). Of these proposed interaction techniques, 84.00% of them only mentioned a single modality, with the remaining 16.00% explicitly stating a desire to use multiple modalities in tandem. The difference between frequency of proposing unimodal and multimodal interaction

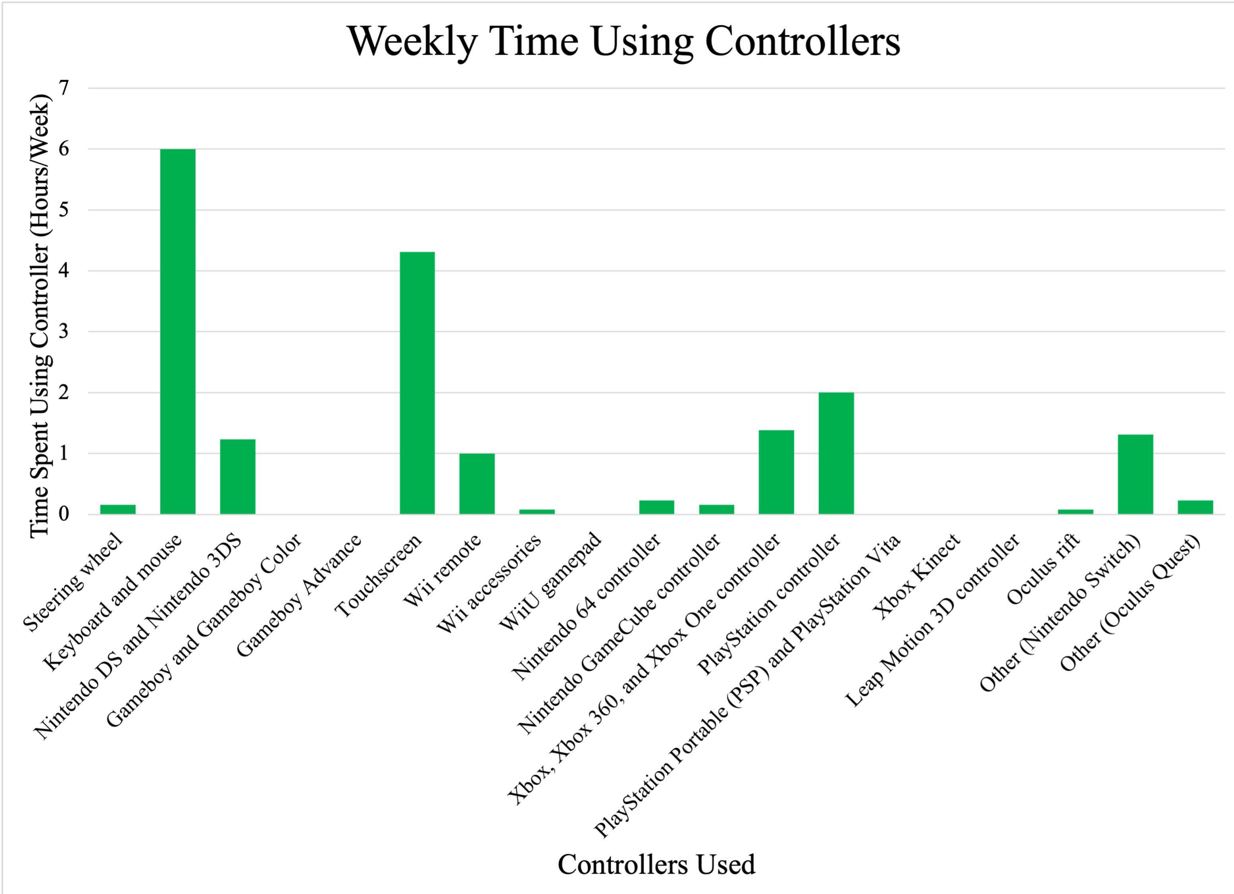

Figure 4.2: The average time using the indicated gaming controllers each week. the total average time using the controllers was 15.33 hours/week (SD = 2.58).

techniques was found to not be significant (Chi square = 1, df = 1, p = 0.32). The frequency of mentions of the different modalities, with distinctions between unimodal and multimodal interactions, can be observed in Figure 4.3. Some features and interaction techniques overlapped between two or more individual modalities, however, and participants would sometimes express a desire for multiple unimodal options (e.g., asking for speech *or* gesture to select all strokes, but not asking to use the two in conjunction, as would be the case for a multimodal interaction).

4.2.1 Proposed Modalities

There was found to be no significant difference (Chi square = 5, df = 5, p = 0.42) in the frequency with which various modalities were mentioned for the proposed interaction techniques. Because of this, no single modality can be identified as being the most popular, which could have

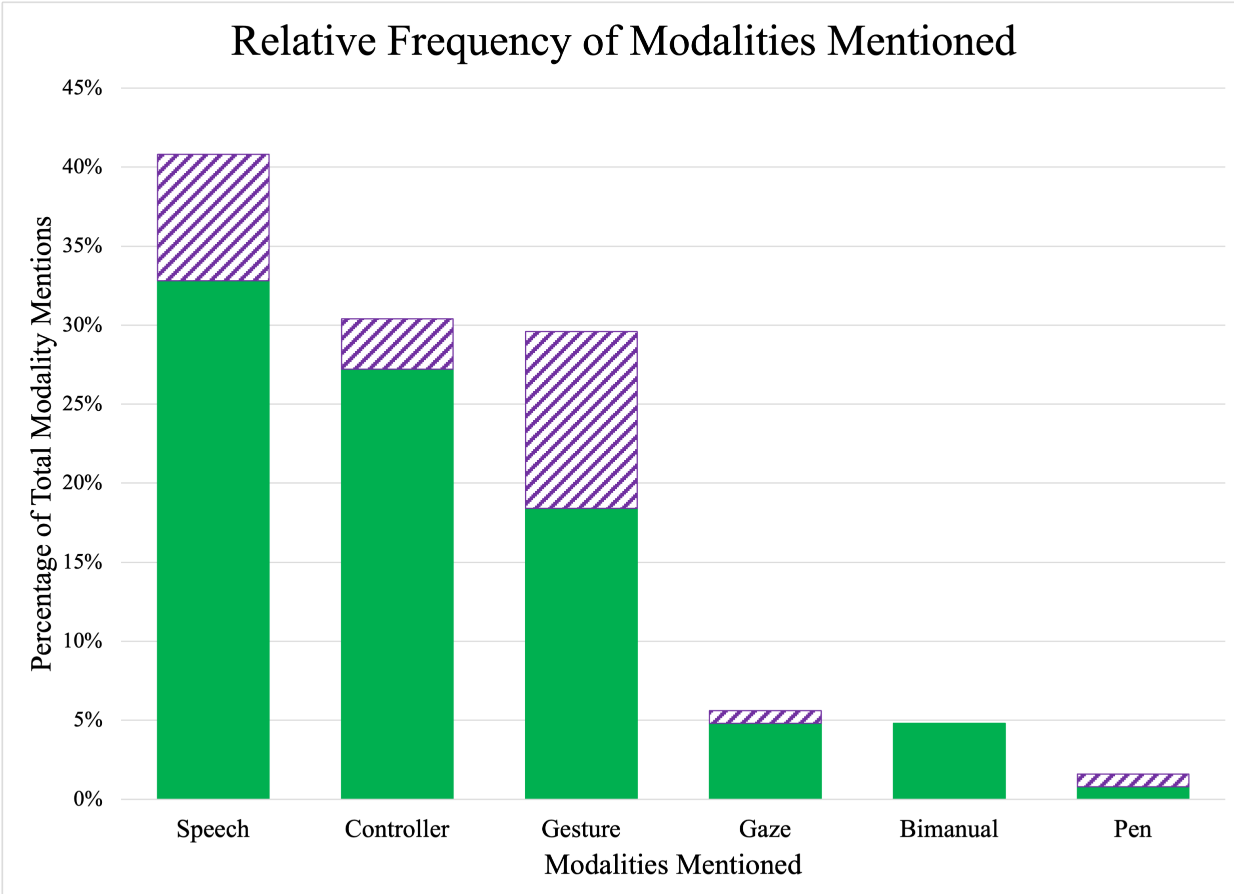

Figure 4.3: Proposed modality-related interactions, sorted by the frequency of mentioning. For each bar, the solid portion indicates times when only that specified modality was mentioned, while the portion with diagonal lines indicates times when the bar's modality was mentioned to be used in tandem with one or more other modalities.

impacted how future interaction techniques were implemented. Conversely, it also means that none of the modalities of interaction mentioned should be entirely excluded. Among the proposed interaction techniques, **speech** consisted of 41.60% mentions, with unimodal speech constituting 78.85% of those proposed interaction techniques, and multimodal interaction techniques involving speech constituting 19.23% of them. The multimodal interaction technique involving speech proposed to combine speech with gesture or speech with both gesture and the controller. Many of the unimodal speech interaction techniques involved different types of cloning or duplication, as well as methods of selecting individual strokes, multiple strokes, all strokes, or a subset of strokes that are grouped into an object. Some participants also mentioned wanting to use speech to generate shapes or change brushes. It is also worth noting that, while speech was a frequently proposed modality of interaction, some participants mentioned concern that speech would be distracting in a drawing environment. This may be supported by the observation that many of the participants would stop drawing when they spoke or answered questions, suggesting that the act of speaking may conflict with the act of drawing to some extent.

Controllers were mentioned in 30.40% of the proposed interaction techniques. As with speech, the majority of interaction techniques using the controllers did not include another modality. 89.47% of controller mentions discussed using only the controller. Several of these mentions identified the controller as acceptable for the tasks being performed or the result of an inability to devise an alternative modality for the task. For instance, when asked whether any multimodal interactions could help with drawing the tree (trunk, branches, and leaves) in the second drawing activity, several participants tried to think of an alternative to the controller, but ultimately said they could not think of any alternatives to the controller. The 10.53% of controller-related proposals that involved multimodal interaction proposed combining use of the controller with gesture, use of the controller with gesture and speech, or use of the controller with gesture and pen.

Gesture constituted 29.60% of proposed interaction techniques. 62.16% of gesture-related interaction techniques mentioned only gesture as the modality for interaction. While gesture was sometimes mentioned vaguely, without specifying any particular kind of gesture to accomplish the task in question, there were several instances of participants requesting a specific gesture for the tasks. This included scrubbing motions to add textures to an already-drawn stroke, chopping to split a stroke into two separate strokes, or pointing to select distant objects. There appeared to be a desire among many participants for gestures to extend their reach in various ways in the virtual environment. This sometimes meant selecting distant strokes or objects, but it also involved repositioning and rescaling the world, either to change their perspective on the drawing or to give them access to a part of the drawing that would otherwise be difficult to reach. As with speech, there were also proposals to use a gesture, such as a wrist flick, to switch between brushes. The remaining 37.84% of gesture-related interaction techniques proposed the use of gesture multimodally with either speech, the controller, gaze, a combination of speech and the controller, or a combination of the controller and a pen.

Gaze constituted 5.60% of the unique interaction techniques proposed. 85.71% of those mentioned interaction techniques were unimodal, while 14.29% were multimodal with gesture. For unimodal interactions, gaze was proposed as a method for selecting strokes and objects, moving objects, cloning objects, rotating drawings, recalling the color palette, or interacting with objects that are far away or out of reach. For multimodal interaction techniques, it was paired with gesture for interaction, manipulation, and duplication of objects.

Bimanual interaction constituted 4.80% of the unique interaction techniques proposed, and 100% of those proposals were unimodal interaction techniques. These techniques were often proposed for the creation or modification of strokes. This included drawing with both hands, splitting strokes, and reshaping strokes. There was also a bimanual interaction technique proposed by two participants that asked for the ability to use one hand to hold an object while drawing or "painting" on it with the other hand. This appears to be analogous to how one might paint a model or figurine.

The other modality mentioned was the use of a **pen**. 1.60% of proposed interaction techniques pertained to the pen, having been mentioned twice in total. One mention of the pen suggested using it unimodally as a way to draw tree branches, and the other mention proposed that it be used in coordination with the controller and gesture.

4.2.2 Applications of Multimodal Interaction Techniques

The combinations of speech and gesture usually entailed pointing at an area and using speech to have the system perform a function on the indicated area, such as selecting strokes, cloning strokes, erasing strokes, filling the area with color, adding shapes, or moving objects. Often, this was proposed as a way to simplify the process of selecting strokes, or to allow the selection of strokes that are not within reach of the participant. Speech was also proposed to be used in conjunction with gesture and the controller to generate pre-made shapes.

Gesture was proposed for multimodal use with the controller to manipulate strokes and select all of the strokes in a "layer" grouping. With gaze, gesture was proposed to be used for interacting with, manipulating, and duplicating objects. It was also proposed to be used in conjunction with both the controller and the pen to make a selection that would then be filled with a texture.

4.3 System Usability Scale

The System Usability Scale (SUS) was presented as 10 statements for participants to rate using a 5-point Likert-type scale. For each statement, 1 indicated that the participant strongly disagreed with the statement, while 5 indicated that they strongly agreed. The average rating of each SUS statement across the participants can be seen in Figure 4.4.

Overall, participants rated the system quite positively. The average rating for the statement, "I think that I would like to use this system frequently," was 4.13 (SD = 0.89), and the statement "I thought the system was easy to use" received an average rating of 4.25 (SD = 0.77), with its most frequent rating being 5. Similarly, the functions in the system were rated as "well integrated" with an average of 4.19 (SD = 0.66), and participants generally agreed that they "felt very confident using the system" (average rating: 3.94, SD = 0.93) and that they "imagined that most people would learn to use the system quickly" (average rating: 3.94, SD = 0.77).

The positive rating of the system was further supported by the participants' disagreement with the more negative statements on the SUS. An average rating of 1.56 (SD = 0.63) suggests that participants did not find "the system unnecessarily complex," nor did they find it very cumbersome to use with an average rating of 1.75 (SD = 0.68). They also did not think that there was "too much inconsistency in the system" (average rating: 1.38, SD = 0.5). On average, participants also did not think they "would need the support of a technical person to use this system" (average rating: 1.88, SD = 1.02) or that they "needed to learn a lot of things before [they] could get going with this system" (average rating: 2.28, SD = 1.26). However, there were a few participants who rated these statements with 4's or 5's, indicating agreement. Nonetheless, the most frequent response to "I think that I would need the support of a technical person to be able to use this system," was 2 (disagree), and the most frequent response to "I needed to learn a lot of things before I could get going with this system," was 1 (strongly disagree).

4.4 Requests for Additional Features

Amid the various proposals for multimodal and unimodal interaction techniques, participants would occasionally request additional features be implemented. The additional features described here were presented by at least two participants. A variety of interaction techniques would often be proposed by different participants for any one of these features.

A desired feature mentioned by several participants was an ability to group strokes into objects or "layers" (in a grouping sense, rather than a dimensional sense of stacking items on top of each other). For instance, if drawing the head of a dog, these participants would want every stroke constituting the head to be grouped so that they could say, "Select the head," and have the dog's head selected for subsequent moving and modification. One participant compared this to the layers used in Adobe Photoshop. This proposal was often presented by participants while they were trying to select each stroke in a part of their drawings without accidentally selecting unwanted strokes. A desire to treat a grouping of multiple strokes as a single object was coupled with gesture, speech, gesture *and* speech, controller, gaze, and bimanual interaction, so no one interaction technique was clearly preferred.

Participants also mentioned with some regularity a desire to generate shapes—both 3D and 2D. Some participants wanted to be able to generate specific complex objects, such as a tree branch with leaves on it, but the requests were usually for basic shapes, such as spheres, cubes, or circles. The desire for basic shapes tended to be rooted in two intended uses: sculpting and reference. For sculpting, participants wanted to generate basic shapes and then distort and modify them as they would in a 3D modeling program, such as Blender. While this did not necessarily pertain to 3D drawing, when participants discussed using generated shapes for *reference* this was meant as an aid for drawing in three dimensions. Several of the participants appeared to struggle to draw three-dimensionally—especially for the dog activity—and they believed that this could be aided with a three dimensional object around which they could draw. This may be similar to how, in 2D drawing, people sometimes create underdrawings of geometric shapes, such as circles and rectangles, as aids for drawing accurate or desired proportions. Speech was the primary modality associated with generating shapes, although there were several times where generating shapes was proposed without specifying any particular kind of interaction technique. There was also one participant who wanted to use speech in conjunction with pointing (gesture), in order to generate a shape at an indicated location.

Other features that participants proposed were an ability to fill a space or area with color or texture and some kind of brush organization. For filling with color or texture, this sometimes meant filling the area within closed strokes with a surface that would have the desired color or texture, but other participants indicated a desire for a solid object to fill an area outlined by strokes. Speech, gesture, speech *and* gesture, the controller, and controller with gesture and pen were all proposed interaction techniques for this, alongside instances where the feature was suggested without specifying an interaction technique. For brush organization, it appeared to mostly be a desire to have similar brushes grouped under certain labels, in order to either view those brushes using the controller or be able to summon a particular group of brushes using speech.

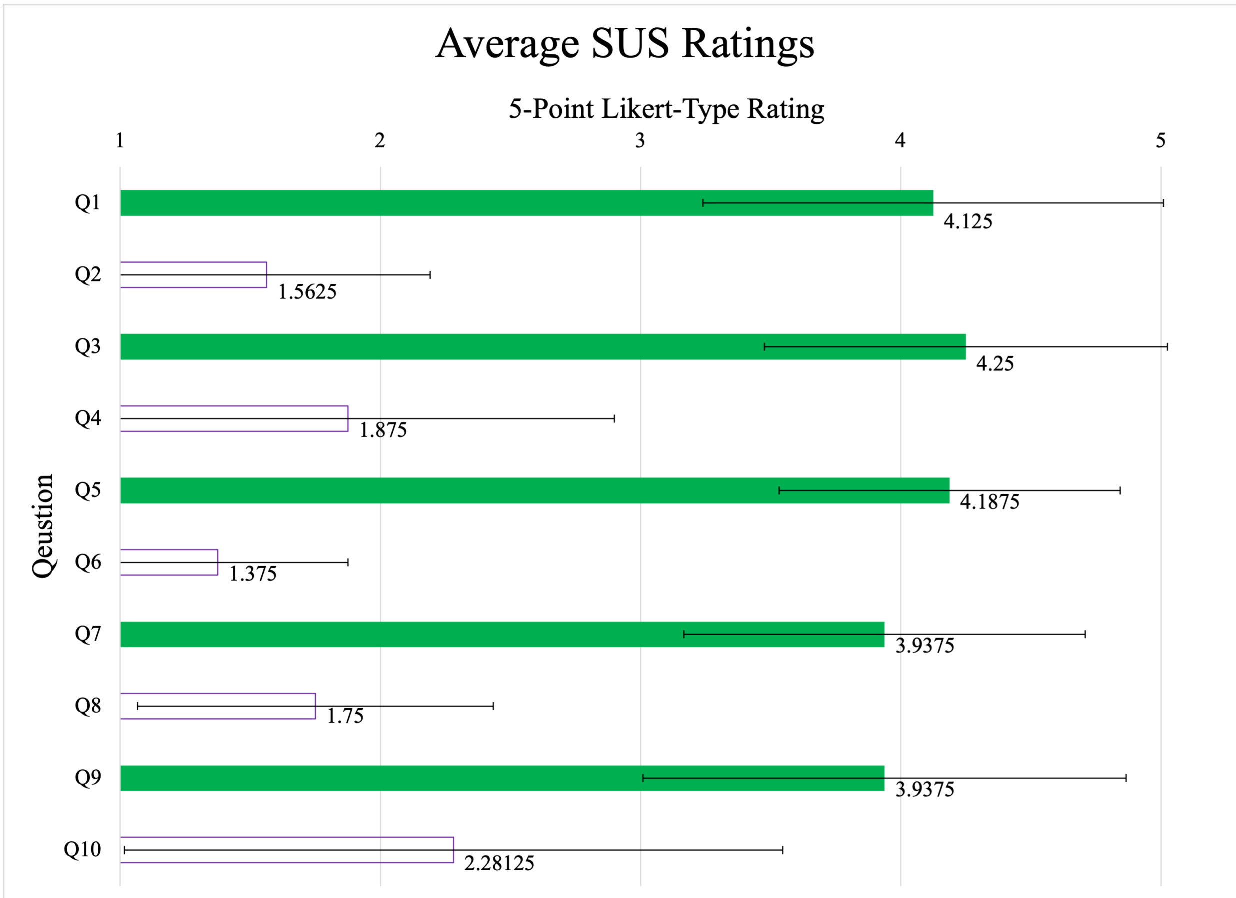

Figure 4.4: Averaged ratings to System Usability Scale (SUS) statements on a 5-point Likert-type scale. Solid bars indicate the SUS statement was **positive**, whereas outlined bars indicate the SUS statement was **negative**. A rating of **1** indicated the participants **strongly disagreed**, and a rating of **5** indicated that the participants **strongly agreed**.

Chapter 5

Discussion & Future Work

5.1 Discussion

There were not significant statistical differences in the frequency with which modalities were mentioned. If a specific modality had emerged as a extremely popular, that could have emphasized the importance of that particular modality. Nonetheless, the objective was to identify possible interaction techniques for a variety of modalities, to be used both unimodally and multimodally. In this, this study was quite informative.

The artists who participated were generally interested in having additional interaction techniques, including unimodal and multimodal interactions, for 3D drawing. Aside from the controller, which is already present in the system, speech and gesture appeared to be particularly promising modalities to explore. For both of these modalities, selection of strokes or objects was an action that commonly received suggestions for alternatives to the controllers, which suggests that the current method of selection may be cumbersome for some people. This is further supported by the fact that participants wanted speech and gesture to select multiple strokes at once, as well as selecting all strokes, and these selection behaviors were presented for other modalities as well. Another possible reason for using these additional modalities for selection could be to decrease the time it takes to switch between the drawing and selection modes on the controller. This is similar to the desire to have speech or gesture to switch between brushes. Presumably, these would all be interaction techniques that would allow users' focus to remain on the activity of drawing, rather than sifting through menus to find certain features, which may be used with regularity.

One curiosity about the proposed interaction techniques from participants were the caveats mentioned about speech as a modality for interaction in the system. Some participants commented that speech would be distracting to use while drawing, yet some of those participants also proposed speech as a modality they wanted. Although participants had been informed of multiple different modalities to consider for interaction techniques, it is possible that the preferences for proposing speech were due to familiarity with that modality as a form of interaction. Digital voice assistants, such as Apple's Siri or Amazon's Alexa, are increasingly ubiquitous, and voice-operated computer systems are frequently depicted in popular fiction, such as the Marvel comics and films and *Star Trek*. Familiarity may have accounted for the mention of the controller as well, because that was the mode of interaction actually available to the participants for the drawing activities and was thus a default option. Of course, the utility of speech for 3D drawing in VR would be best determined through active testing of such a system. The fact speech was the most commonly proposed modality for new interaction techniques suggests that such investigation would be warranted.

Overall, this study has offered initial insights into what multimodal and unimodal interaction techniques might be for 3D drawing in VR. Further analysis of the data will likely reveal more specific information that can be synthesized to inform how to implement and test multimodality for such a system moving forward. Drawing in virtual reality remains an open area of exploration. While systems such as Open Brush are available for general use, there are still many opportunities to refine the interfaces of these tools to allow users to focus on expressing themselves in this novel medium, rather than being concerned with the limitations of the interfaces and systems.

5.2 Limitations of the Study

This study intentionally focused on gathering suggestions from artists for multimodal and unimodal interaction techniques in 3D VR drawing. This was done considering that artists may be frequent users of such systems and would also be better able to identify and explain their desired interactions. Nonetheless, focusing on artists means that the desires of naïve users remain unknown. Ideally, 3D drawing programs would be usable at all skill levels, such that it is neither overwhemling for a non-artist nor restrictive for an experienced artist. Thus, it will be important to eventually include naïve users in the design process.

Another limitation of this study is that participants were welcome to propose any interaction techniques they wanted. This helped to ensure that participants felt free to suggest a technique without concern for how realistic that technique would be to use or implement. At the same time, this means that some of the proposed interaction techniques may not actually work. It may not be feasible to implement a proposed interaction technique, or even if it is, it may not actually make the interface more usable. Of course, these issues can be explored in future work.

Akin to that limitation, it would have been beneficial to gain further insight and context into the participants and their proposed interaction techniques. Since this study focused on art, it would have been beneficial to know not only what kinds of games participants played but what kind of art they make and the tools they use to make it. Questions to rate the time spent working in particular media would have provided additional insights, as would insights into how often participants used various traditional and digital art tools.

Similarly, it may have been beneficial to have consistently asked participants to explain why they proposed certain interaction techniques. This could have decreased the amount of *post hoc* speculation involved in motivating the proposed interaction techniques. Conversely, participants may not have always had clear reasons for desiring a particular interaction technique.

5.3 Future Work

This thesis has provided an overview of the various modalities that participants proposed, as well as insights into the participants' general feelings about the usability of the system. At this stage, however, there remain too many proposed interaction techniques to identify what to implement. Proposals from the participants would need to be further refined to determine how to implement the modalities for particular interaction techniques. Thus, work in the immediate future will be to perform a full thematic analysis upon the data collected [26]. In this process, the proposals identified here will serve as codes that can then be organized into themes, wherein further patterns can be identified. Additionally, a thematic analysis may identify themes pertaining to other aspects of 3D VR drawing that do not necessarily involve multimodality, which opens additional opportunities for investigation.

Following the full thematic analysis and determination of which interaction techniques are most promising, these identified interaction techniques will be need to be implemented and tested for actual usability. Since Open Brush is an open-source program, it may be possible to modify a branch of that program to implement the desired interaction techniques into that system. However, the breadth of options available in Open Brush could be distracting or prohibitively complicated for future studies where a more limited number of features are being evaluated. Thus, an in-house VR sketching program is currently under development using the Unity game engine. This system may be used for future stages of research, when the new interaction techniques are implemented and a more constrained feature set is preferable.

Once the new interaction techniques are implemented in a system, it will be possible to study the actual usability of these techniques. While it is valuable at this stage to know what artists think could help them use a 3D VR drawing program, the idea of such interaction techniques and features may seem more usable or useful to them than the techniques would be in practice. Thus, future research will necessitate evaluating the system in ways other than suggestions from participants. They will need to evaluate various metrics, such as the time it takes to complete a task using certain modalities, the accuracy of the drawing performance, and how satisfied the participants are with the end results and the experience as a whole.

Another possible area of future work would be to collaborate with artists who already use 3D VR drawing in their practice. Such individuals would have a considerable amount of experience. As a result, these hyper-experienced artists may have insights or desires for interaction techniques and features that would not occur to other artists, such as those who participated in this study. After all, this study's participants *were* experienced artists, but they were relative novices to Open Brush and 3D VR drawing, meaning their insights may have differed from those of artists who have used tools such as these more extensively.

Chapter 6

Conclusion

This thesis has described a co-design user study that recruited artists to propose multimodal and unimodal interaction techniques for drawing in three dimensions (3D) in virtual reality (VR). For the study, participants wore the HTC Vive Pro Eye VR headset [21] to draw in the open-source 3D VR drawing program, Open Brush [14]. Before drawing, participants were taught the basics of using Open Brush and briefed on the concept of multimodal interaction through pre-recorded videos.

Once briefed, participants were asked to complete two short 3D drawing activities (10 and 15 minutes in length). While the participants drew, they were asked to suggest interaction techniques that would help them complete the various tasks involved in creating the drawings. Most of the interaction techniques that participants suggested involved unimodal interactions, with speech, controllers, and gesture being the modalities proposed most often. Multimodal interaction techniques were less frequently proposed, but were present, especially when using gesture and speech in tandem.

From this, it can be concluded that the notion of additional modalities is appealing to artists for both multimodal and unimodal interaction techniques in the interfaces of 3D VR drawing programs. Future work will involve identifying which of the proposed interaction techniques are feasible to implement and use. This can then help to guide the further design and development of these interfaces.

Bibliography

[1] Sung-A Jang, Hyung-il Kim, Woontack Woo, and Graham Wakefield. Airsculpt: A wearable augmented reality 3d sculpting system. In *Distributed, Ambient, and Pervasive Interactions: Second International Conference, DAPI 2014, Held as Part of HCI Interational 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings 2*, pages 130–141. Springer, 2014.

[2] Anirudh Sharma, Sriganesh Madhvanath, Ankit Shekhawat, and Mark Billinghurst. Mozart: a multimodal interface for conceptual 3d modeling. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 307–310, 2011.

[3] Sang-Gyun An, Yongkwan Kim, Joon Hyub Lee, and Seok-Hyung Bae. Collaborative experience prototyping of automotive interior in vr with 3d sketching and haptic helpers. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*, pages 183–192, 2017.

[4] Johann Habakuk Israel, Eva Wiese, Magdalena Mateescu, Christian Zöllner, and Rainer Stark. Investigating three-dimensional sketching for early conceptual design—results from expert discussions and user studies. *Computers & Graphics*, 33(4):462–473, 2009.

[5] Yongkwan Kim and Seok-Hyung Bae. Sketchingwithhands: 3d sketching handheld products with first-person hand posture. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 797–808, 2016.

[6] Mayra D Barrera Machuca, Wolfgang Stuerzlinger, and Paul Asente. Smart3dguides: Making unconstrained immersive 3d drawing more accurate. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–13, 2019.

[7] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[8] Daniel F Keefe, Daniel Acevedo, Jadrian Miles, Fritz Drury, Sharon M Swartz, and David H Laidlaw. Scientific sketching for collaborative vr visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):835–847, 2008.

[9] Sander Van Goethem, Jouke Verlinden, Regan Watts, and Stijn Verwulgen. User experience study on ideating wearables in vr. *Proceedings of the Design Society*, 1:3339–3348, 2021.

[10] Rahul Arora, Mayra Donaji Barrera Machuca, Philipp Wacker, Daniel Keefe, and Johann Habakuk Israel. Introduction to 3d sketching. In *Interactive Sketch-based Interfaces and Modelling for Design*, pages 151–177. River Publishers, 2023.

[11] Hesham Elsayed, Mayra Donaji Barrera Machuca, Christian Schaarschmidt, Karola Marky, Florian Müller, Jan Riemann, Andrii Matviienko, Martin Schmitz, Martin Weigel, and Max Mühlhäuser. Vrsketchpen: unconstrained haptic assistance for sketching in virtual 3d environments. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2020.

[12] Mayra Donaji Barrera Machuca, Wolfgang Stuerzlinger, and Paul Asente. The effect of spatial ability on immersive 3d drawing. In *Proceedings of the 2019 on Creativity and Cognition*, pages 173–186, 2019.

[13] Rahul Arora and Karan Singh. Mid-air drawing of curves on 3d surfaces in virtual reality. *ACM Transactions on Graphics (TOG)*, 40(3):1–17, 2021.

[14] Open Brush. Open brush. https://openbrush.app, 2022.

[15] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[16] Xinchi Xu, Yang Zhou, Bingchan Shao, Guihuan Feng, and Chun Yu. Gesturesurface: Vr sketching through assembling scaffold surface with non-dominant hand. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[17] Tim McGraw, Esteban Garcia, and Drew Sumner. Interactive swept surface modeling in virtual reality with motion-tracked controllers. In *Proceedings of the Symposium on Sketch-Based Interfaces and Modeling*, pages 1–9, 2017.

[18] E. Wiese, J. H. Israel, A. Meyer, and S. Bongartz. Investigating the learnability of immersive free-hand sketching. In *Proceedings of the Seventh Sketch-Based Interfaces and Modeling Symposium*, SBIM '10, page 135–142, Goslar, DEU, 2010. Eurographics Association.

[19] Rhys Bevan Jones, Paul Stallard, Sharifah Shameem Agha, Simon Rice, Aliza Werner-Seidler, Karolina Stasiak, Jason Kahn, Sharon A Simpson, Mario Alvarez-Jimenez, Frances Rice, et al. Practitioner review: Co-design of digital mental health technologies with children and young people. *Journal of child psychology and psychiatry*, 61(8):928–940, 2020.

[20] Chrysanthi Papoutsi, Joseph Wherton, Sara Shaw, Clare Morrison, and Trisha Greenhalgh. Putting the social back into sociotechnical: Case studies of co-design in digital health. *Journal of the American Medical Informatics Association*, 28(2):284–293, 2021.

[21] HTC Corporation. Htc vive pro eye. https://www.vive.com/us/product/vive-pro-eye/specs/, 2019.

[22] Brian T. Sullivan and Richi Rodriguez. Open Brush Tutorial — NUILab YouTube. https://youtu.be/T9CqyMVyFrI, 2023. [Accessed 14-Jun-2023].

[23] Brian T. Sullivan and Richi Rodriguez. What is Multimodal Interaction in VR/AR? — NUILab YouTube. https://youtu.be/jw_zmJfdbxU, 2023. [Accessed 14-Jun-2023].

[24] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[25] Microsoft. Transcribe your recordings: Onenote for microsoft 365, word web app. https://support.microsoft.com/en-us/office/transcribe-your-recordings-7fc2efec-245e-45f0-b053-2a97531ecf57, 2023.

[26] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

Appendix A

Scripts & Questions

A.1 Scripts

What follows are the scripts to the videos used in the tutorial and briefing stages of each session of the study, along with a script that was read in-person by an experimenter following the briefing video.

A.1.1 Open Brush Tutorial Video Script

This script was recorded and played over video captures of the headset view in Open Brush demonstrating each task.

1. To swap the functionality of each controller, touch their bottom ends together as shown. This works in both directions.

2. On the menu controller, slide your thumb horizontally on the trackpad to get to the brush selection menu. Using your drawing controller as a pointer, select the brush in the upper-right corner of the third page that is labelled "Wire."

3. Slide your thumb on the trackpad of your drawing controller to change the size of the brush stroke. The size is indicated by the diameter of the circle at the end of your drawing controller.

4. Navigate to the Color Picker menu and change your brush color to green.

5. Now, draw a cube. Make sure that it is actually three-dimensional. Do not draw a 2D representation of a cube, as you might on a sheet of paper.

6. Open the Quick Menu by pressing the button above the trackpad on your drawing controller. Then, choose the "Selection" button. Now, select the strokes of the cube by hovering over them and pulling the trigger.

7. Move both controllers to be in front of you and press the grip buttons on the sides of both controllers to enter Resize Mode. You can make the cube larger by moving your hands apart and make it smaller by moving your hands together.

8. Using only the selection controller, press the grip buttons on the sides and move your hand to rotate the cube to have a corner pointing toward the ground.

9. To change the color of the cube, select a color. Then, open the Quick Menu and choose the "Repaint" option. You can now change the color of any stroke by touching it with the controller and pressing the trigger.

10. First, select the strokes of the object you wish to clone. Then, with the selection controller near the object, press and hold the trackpad until the object is duplicated. You can now move this copy around, if you wish, by pressing the grip buttons of the controller, like when you rotated it.

11. If you wish to undo an action, press the left side of the menu controller's trackpad. To redo, you can press the right side of that trackpad.

12. To erase or delete a stroke, use the Quick Menu on your drawing controller and choose the "Eraser" option. Then, you can delete any stroke that the controller comes into contact with while holding the trigger.

A.1.2 Multimodal Interaction Video Script

This script was recorded and then accompanied by simple stick-figure animations illustrating the actions and concepts being described. Note that text that is **BOLD AND CAPITALIZED** *indicates a change in topic or section and is not read in the narration recording.*

INTRODUCTION

Although interaction can mean both input and output when working with computers and other digital electronics, our focus today is on the input component of the interaction–although comments about output are welcome as well.

MODALITIES

A modality is a type of input. The modalities we encourage you to consider today are: Gestures, Controllers, Speech, Gaze, and Pen.

Gestures can be broad motions, which generally involve moving your hands and arms, but can also include head motions. Controllers are input devices that have inputs such as buttons, trackpads, and triggers, and can also record motion as input, for both large gestures and smaller gestures, known as microgestures. Speech is spoken language that tells the device how to behave. Gaze is the use of eye movements, staring, and blinking to interact. And finally, pens are a type of controller that emulates the form factor of a pen. It should be remembered, however, that a pen in a 3D environment will behave differently from a pen in a 2D environment.

UNIMODAL VS. MULTIMODAL

Before we define Multimodal Interaction, let's define Unimodal Interaction:

Unimodal Interaction is any type of interaction that only uses a single type of input or output. For something like 3D sketching in virtual or augmented reality, this could be a controller, gestures, gaze, speech, or a pen. The idea is that you only have one way of controlling or interacting with the system. An example of unimodal inputs would be if you had a unicorn in your three-dimensional space. If you wanted to move it to the right, you could use a controller to select the unicorn and move it, or you could use speech to say, "Move the unicorn to the right."

With Multimodal Interaction, you could combine multiple input methods to accomplish a task. For instance, you could use a controller to point at the unicorn and then simply say "move it to the right."

A Multimodal Interaction is any type of interaction that combines two or more inputs. This could be a combination of gesture and speech, as just described, or it could combine gesture and gaze, gaze and speech, using both hands in tandem to accomplish tasks (also called “bimanual” interaction)…The list goes on.

CATEGORIES

The major categories of Multimodal Interaction are Synchronous vs. Asynchronous, Symmetrical vs. Asymmetrical, and Dependent vs. Independent.

With Synchronous interactions, you can do the inputs at the same time. For instance, you could be selecting objects with the controller while simultaneously saying to change the color of these objects.

For an Asynchronous interaction, the objects might be selected with the controller, and then a spoken instruction is made to change the color.

Symmetrical interactions are usually bimanual in nature, meaning that they make use of both hands. In this situation, it is where each hand is mirroring what the other is doing, such as for painting mirror images, perhaps for butterfly wings. Asymmetrical interactions, on the other hand, have the hands performing different tasks.

A related–but different–category of bimanual input would be Dependent vs. Independent interaction. A Dependent interaction would be something where the actions of both hands are different but rely upon each other to accomplish a task, such as using one hand’s controller as a color palette and the other as a brush. Meanwhile, an Independent interaction would be one where both controllers can act as brushes and summon menus and color palettes, regardless of what the other controller is doing at any given moment.

REVIEW

To review: A Unimodal interaction is one where all actions or tasks are completed using a single type of input. Conversely, a Multimodal interaction will make use of two or more input types to accomplish tasks.

The common categories for Multimodal Interaction are Synchronous vs. Asynchronous, Symmetrical vs. Asymmetrical, and Dependent vs. Independent. The modalities that you can consider are Gesture, Controllers, Speech, Gaze, and Pen.

We encourage you to consider all of these options when suggesting possible multimodal interactions today.

A.1.3 In-Person Multimodal Interaction Script

[This script was read to participants live by an experimenter, following the Multimodal Interaction Briefing Video.]

Thank you for choosing to participate in this study. Today, you will be participating in a co-design study to identify possible opportunities for multimodal interaction in 3D sketching. To do this, you are asked to complete a set of 3D drawing tasks in Open Brush. As you perform the tasks, we will ask questions about what you are doing, to better understand how you are using the system and identify possible opportunities for multimodal interaction.

Please feel free to make comments on the system yourself as well, and you are welcome to use any tool in Open Brush that will help you complete the tasks. You are not limited only to the tools you learned in the tutorial. Do you have any questions?

A.2 Survey Questions

What follows are the Questions asked in the various surveys. The format of the questioning (e.g., verbal back-and-forth vs. written responses) is noted before each set of questions.

A.2.1 Preliminary Questionnaire

This written questionnaire was presented on paper and collected demographic data and was completed by participants at the start of the study.

1. Age: ______

2. Sex: ______

3. What is your current academic standing? ____________

4. What is your major? ____________

5. Have you used an Augmented Reality headset? ______

6. Are you currently wearing corrective lenses (contacts or glasses)? ______

7. Do you currently have or have you ever had a heart condition, seizures, or epilepsy? ______

8. If you have answered "Yes" for any of the above conditions, please explain: _________

9. Have you consumed any of the following in the past 24 hours?

    Caffeine ______

    Alcohol ______

    Nicotine ______

10. Indicate the average number of hours per week you spend using computers (personal or work combined) ______

11. Indicate the average number of hours per week you spend playing the following types of games:

    First person shooter games ______

    Massively multiplayer online games ______

    Flight simulators ______

    Sports/racing ______

    Military command/strategy ______

    Fighting ______

    Life/business simulations ______

    Fantasy/adventure ______

Puzzles/card games/board games ______

Social networking games ______

Other ______

Please specify ____________________

12. Indicate the number of hours per week you spend playing games using the following controllers:

Steering wheel ______

Keyboard and mouse ______

Nintendo DS and Nintendo 3DS ______

Gameboy and Gameboy Color ______

Gameboy Advance ______

Touchscreen ______

Wii remote ______

Wii accessories ______

WiiU gamepad ______

Nintendo 64 controller ______

Nintendo GameCube controller ______

Xbox, Xbox 360, and Xbox One controller ______

PlayStation controller ______

PlayStation Portable (PSP) and PlayStation Vita ______

Xbox Kinect ______

Leap Motion 3D controller ______

Oculus rift ______

Other ______

Please Specify ______

13. How much overall experience, if any, do you have with the JTAC trainign or operational display systems or control devices (in years)? ______

A.2.2 System Usability Scale (SUS)

This written System Usability Scale (SUS) was presented on paper, and participants were instructed to mark the box they most agreed with for each statement, corresponding with numbers on a 1-5 Likert-type scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree." An example of the rating boxes is provided below:

[1] [2] [3] [4] [5]

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system

A.2.3 Follow-Up Questions

Experimenters would verbally ask these questions of participants following completion of the SUS. Participants would then respond, with clarifying questions sometimes being asked in an ad hoc manner, based on participants' responses.

1. What part of the drawing process would you have most liked to have had an option for multimodal input/output?

2. Did you at any point feel like your focus on drawing/creating was distracted or encumbered by the user interface?

3. In what situations/contexts might you find a tool such as this useful?

4. Were there any other multimodal interaction opportunities that you would recommend?

Appendix B

List of Proposed Interaction Techniques

What follows is a list of the unique interaction techniques proposed by participants in the study. If two or more participants suggested the same interaction technique with the same modality, it is treated as a single item. It has been noted if the interaction techniques are unimodal or multimodal in nature, as well as what modality/modalities were proposed.

- gesture to change brush size — Unimodal Gesture
- Scrubbing (gesture) to add texture to surface — Unimodal Gesture
- chopping (gesture) to split a selected stroke — Unimodal Gesture
- gesture for undo, gesture for redo — Unimodal Gesture
- gesture to create empty object and fill it with brush — Unimodal Gesture
- gesture to select recently used brushes — Unimodal Gesture
- gesture to turn non-drawing controller panel — Unimodal Gesture
- gesture, shake to turn the panel — Unimodal Gesture
- gesture, similar to sculpting — Unimodal Gesture
- hugging, gesturing to select an object — Unimodal Gesture
- push down with hand to move scene downwards. — Unimodal Gesture
- gestures to work with large scale objects — Unimodal Gesture
- extend gestures beyond user's actual reach — Unimodal Gesture
- gesture to resize world — Unimodal Gesture

- move non-drawing hand up and down (gesture) to resize world — Unimodal Gesture
- gesture to erase lines — Unimodal Gesture
- head movements (gesture) or gaze to rotate drawing — Unimodal Gesture
- wrist flick (gesture) to switch to previous brush — Unimodal Gesture
- say "select all" (speech) or spread arms out (gesture) to select all strokes in drawing — Unimodal Gesture
- gesture to make line straight — Unimodal Gesture
- modify generated shape with hands (gesture) — Unimodal Gesture
- pointing (gesture) to select distant object — Unimodal Gesture
- standing on tiptoes, gesture/body movement to move user perspective upwards — Unimodal Gesture
- speech to split a selected stroke — Unimodal Speech
- Speech to fill in space between major strokes — Unimodal Speech
- speech to change brush size — Unimodal Speech
- speech to add texture to brush — Unimodal Speech
- speech to add leaves to tree — Unimodal Speech
- speech to indicate style of generated shape/object — Unimodal Speech
- Speech to select an object/strokes — Unimodal Speech
- speech for brush categories — Unimodal Speech
- speech for multiple stroke selection — Unimodal Speech

- speech for switching between brushes — Unimodal Speech
- speech for undo / second theme — Unimodal Speech
- speech to clone dog ears — Unimodal Speech
- speech to clone objects or groups of strokes — Unimodal Speech
- speech to clone/duplicate — Unimodal Speech
- speech to create 3D objects — Unimodal Speech
- speech to duplicate strokes/objects — Unimodal Speech
- speech to fill-in space between strokes (surface) — Unimodal Speech
- speech to select group of strokes on a named layer — Unimodal Speech
- speech to switch a brush — Unimodal Speech
- speech using shortcuts to select favorite colors/brushes — Unimodal Speech
- speech, brush selection, switch between brushes — Unimodal Speech
- speech, fill feature (volume) — Unimodal Speech
- speech, auto generate content — Unimodal Speech
- speech, minor adjustments only — Unimodal Speech
- speech to specify number of clones to make — Unimodal Speech
- speech to generate a shape — Unimodal Speech
- speech to resize world — Unimodal Speech
- speech to select specific object in the drawing — Unimodal Speech
- speech to select all — Unimodal Speech

- speech for commands — Unimodal Speech
- speech to set brush color — Unimodal Speech
- say "select all" (speech) or spread arms out (gesture) to select all strokes in drawing — Unimodal Speech
- speech to select object, AI assistance — Unimodal Speech
- speech to select brushes — Unimodal Speech
- speech to search for a brush type — Unimodal Speech
- speech to select last drawn object — Unimodal Speech
- speech to create intersecting lines for a bubble. — Unimodal Speech
- speech for placing shape on X, Y, or Z plane — Unimodal Speech
- speech to move selected objects — Unimodal Speech
- speech to move object to top of space — Unimodal Speech
- speech or holding a button to beautify/straighten a stroke — Unimodal Speech
- controller for duplicating objects — Unimodal Controller
- select and duplicate with controllers — Unimodal Controller
- controller for copy/paste — Unimodal Controller
- controller for drawing branches — Unimodal Controller
- controller for duplication strokes — Unimodal Controller
- controller to erase part of a stroke — Unimodal Controller
- controller for moving strokes to another location — Unimodal Controller

- controller for selecting color — Unimodal Controller
- controller for selecting everything — Unimodal Controller
- controller for selecting the brush — Unimodal Controller
- controller for sketching the tree, better control for the tree strokes — Unimodal Controller
- controller for straight line — Unimodal Controller
- controller to switch between previously used colors — Unimodal Controller
- controller, double-click to switch between tools — Unimodal Controller
- controller for fill feature — Unimodal Controller
- rotate controller to access menu items — Unimodal Controller
- controller to adjust stroke size using the trackpad circularly — Unimodal Controller
- controller for changing colors — Unimodal Controller
- controller for beautification of liked-shapes into actual shapes — Unimodal Controller
- controller dedicated button for selection — Unimodal Controller
- change brush size on menu controller — Unimodal Controller
- controllers fine for drawing path — Unimodal Controller
- additional selection options from one button — Unimodal Controller
- snapping for menu sections. — Unimodal Controller
- just use controller brush to make ground — Unimodal Controller
- tapping controllers to switch between selected colors — Unimodal Controller
- controllers as-is for drawing roots and branches on tree — Unimodal Controller

- button to select all strokes in drawing — Unimodal Controller
- controller controls for undo/redo are good — Unimodal Controller
- button to straighten line — Unimodal Controller
- more trackpad control for tool selection. — Unimodal Controller
- swipe up on trackpad to open a menu — Unimodal Controller
- controller to draw leaves on tree — Unimodal Controller
- speech or holding a button to beautify/straighten a stroke — Unimodal Controller
- pen to draw the tree branches — Unimodal Pen
- gaze to clone objects — Unimodal Gaze
- gaze to move selected strokes, elapsed animations — Unimodal Gaze
- gaze to recall color palette — Unimodal Gaze
- gaze to select objects/strokes by looking at them — Unimodal Gaze
- Gaze to place strokes in far-away locations — Unimodal Gaze
- gaze to rotate drawing — Unimodal Gaze
- bimanual drawing with both hands — Unimodal Bimanual
- bimanual selection of objects/strokes — Unimodal Bimanual
- bimanual, hold object with one hand while drawing with another — Unimodal Bimanual
- bimanual stretching of a straight object — Unimodal Bimanual
- pull apart (gesture) to split a selected stroke — Unimodal Bimanual
- spreading (gesture) to create broad strokes — Unimodal Bimanual

- pointing (gesture) & speech for erasing strokes — Multimodal Gesture-Speech
- bimanual pointing (gesture) & speech to make basic shapes — Multimodal Gesture-Speech
- speech and pointing (gesture) to fill an area with color. — Multimodal Gesture-Speech
- speech or pointing (gesture) to fill a wireframe — Multimodal Gesture-Speech
- speech & bimanual pointing (gesture) to make basic shapes — Multimodal Gesture-Speech
- bimanual pointing (gesture) & speech to fill in area — Multimodal Gesture-Speech
- pointing (gesture) & speech to generate specific shape — Multimodal Gesture-Speech
- gesture+speech for cloning strokes — Multimodal Gesture-Speech
- pointing+speech to move object — Multimodal Gesture-Speech
- gesture/controller for distorting or manipulating strokes. — Multimodal Gesture-Controller
- controller/gesture, using layers to select everything — Multimodal Gesture-Controller
- controller, pen, and gesture to make selection to then be filled by a texture — Multimodal Controller-Gesture-Pen
- gesture, speech, and controller to create premade shapes — Multimodal Controller-Gesture-Speech
- gesture and gaze for interaction, manipulation of object, duplication — Multimodal Gesture-Gaze