NAIST-IS-DD1461211

Doctoral Dissertation

Creating Immersive 3D User Interfaces for Professional 3D Design Work

Max Krichenbauer

June 12, 2017

Department of Information Science
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Max Krichenbauer

Thesis Committee:

Professor Hirokazu Kato	(Supervisor)
Professor Kiyoshi Kiyokawa	(Co-Supervisor)
Associate Professor Christian Sandor	(Co-Supervisor)
Assistant Professor Takafumi Taketomi	(Co-Supervisor)
Professor Gudrun Klinker	(Technical University of Munich)

Creating Immersive 3D User Interfaces for Professional 3D Design Work*

Max Krichenbauer

Abstract

Before the rise of the personal computer as a universal workstation in the 1990s, animations and visual effects for movies and television were generally created with physical media such as clay or miniatures and puppets which were hand-crafted. With the success of early computer generated visual effects such as used in the film Jurassic Park, and computer generated 3D animation movies such as Toy Story the PC gradually began to replace various physical media for production. The User Interfaces (UI) have not changed much since, with the great majority of artists using a mouse, keyboard, and 2D computer screen to create and control virtual 3D worlds. With the advent of Virtual Reality (VR) and Augmented Reality (AR) technology in recent years, a new venue has been opened to provide more intuitive and efficient work environments for creating 3D designs and animations: immersive 3D User Interfaces (UIs) allow the user to enter into a three-dimensional work environment that engulfs him, instead of just looking at a flat PC workstation screen in front of him, which allows for a more natural and efficient way of interaction. However, most research in this area until now has not been focused on professional application and has failed to be adopted by professional artists. In order to bridge the gap between research prototypes and real-world adoption, more application-oriented research is required. In this work I present my ongoing research efforts focusing on the various factors and problems of immersive 3D user interfaces for 3D design as well as reflect on experiences from creating a prototype 3D user interface aimed at professional artists.

I have performed an analysis of the current work situation of 3D artists through both a survey and individual observation and analysis of the workflow. Through the

---

*Doctoral Dissertation, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1461211, June 12, 2017.

survey, which 54 media professionals from around the world participated in, I found out details about their current workflow and situation. The individual observation of two artists who I recorded at work and then analyzed the video footage provides insight into what percentage of time artists spend on which subtask. I developed at prototype 3D UI aimed at professional design and performed a formative user study with eleven artists, which gives insights into the possibilities and limitations of current VR and AR technology and allowed me to improve the prototype. I have further performed two summative user studies through which I tried to find quantitative evidence for effects on work performance from 3D UI related human factors. One of this was the possible effect of positional head-tracking on task completion time on a 3D selection and transformation task. Previous research on this topic has been inconclusive and was performed with outdated technology. In my user study, I was not able to find any statistically significant effect from positional head-tracking on task completion time. The other study was the first direct comparison between AR and VR on 3D work performance in 3D interaction tasks. Surprisingly, this study showed significant advantages of AR work environments over VR work environments even in cases where prior known factors did not apply. Finally, I released my research prototype 3D UI to the public where it has been successfully adopted by 3D artists around the world. The lessons learned through this process are provided in the final chapter of this work.

Keywords:

Augmented Reality, Virtual Reality, Computer Aided Design, User Study, Usability

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Bridging the gap between research in the lab and application in the industry is a continuing challenge in all areas of computer science. Specifically though, applying Augmented Reality (AR) and Virtual Reality (VR) 3D User Interfaces (UIs) in professional work environments has proved more challenging than consumer adoption, in spite of extensive research being performed. Fite-Georgel reviewed over 50 publications on AR systems and found that only two of those eventually made it to adoption in the industry [28]. In my research work, I focus on a particular application of AR and VR UIs for 3D computer aided design (CAD) with a special focus on the professional entertainment industry, such as 3D animation movies, visual effects, and video games. This thesis is the summary of my work in this particular field.

1.1. Background

A great part of todays media consists at least in part of Computer Generated (CG) graphics. Not only games and animations, but many movies and television productions rely heavily on visual effects created on the computer. What used to be done with models and miniature figures is now almost exclusively created with 3D Computer Aided Design (CAD) software.

This was not an easy transition for artists, who had to switch from what was a spatial interaction task of moving real objects to a very abstract task of moving a mouse over a table to control virtual 3D objects represented on a flat screen. Famously, Jurassic Park—the first major movie that made extensive use of computer graphics—required engineers to build physical puppets which were used as input devices [46]. These figures resembled the dinosaurs which they were used to animate and were filled with sensors to translate every motion into a computer graphics software. This effortfull and expensive approach was necessary because the experienced stop-motion animators could not easily get used the abstract work of controlling a purely virtual Dinosaur with a mouse and keyboard. Over the years however, the sheer power of computer graphics algorithms to create stunning and convincingly realistic imagery increased to the point where it almost totally superseded traditional means of content creation. Given the advantages of spatial interaction, experimentation with physical figures as input devices never stopped [7]. However, the complexity of modern CG productions makes the use of physical rigs almost impossible. The complexity of natural lifeforms make lifelike joint configurations hard to realize, muscle, hair and cloth simulation must be omitted, and animation cannot be played back and corrected on physical figures, forcing the animator to work "straight forward" without the ability to go back and adjust previous poses. Furthermore, any changes is the character design would require the creation of a new physical input device, making productions very inflexible and expensive. Thus, today 3D design is commonly performed with mouse and keyboard [47].

With the rise of VR and AR technology, the possibility has opened up to use 3D user interfaces for 3D CG content creation. In fact, 3D design was one of the first applications to be successfully demonstrated as AR application [21] on Ivan Sutherland's famous first head-mounted display [68]. Since then, with great regularity, new prototype systems for 3D design work have been proposed, implemented, and tested in research laboratories around the world. However, real-world application by 3D designers is still waiting. Despite ever better hard- and software and ever more thought-out user interfaces, few if any production companies and artists have adopted 3D UIs for their daily work [47].

One of the reasons for this may be that the hardware is still not evolved enough to allow professional application. However, it can be argued that the early computer systems that were gradually adopted in the 1990s also were a far cry from the well evolved computers used today. Thus, technical maturity alone cannot be a valid explanation for a lack of adoption, as it is adoption that drives technical maturity. Another reason may be that the improvements in performance and comfort possible in 3D user interfaces are so small that they do not justify adoption. This obviously depends heavily on the specific type of work and the UI used to perform it. Finally, difficulties in communicating the progress achieved in science and development to the potential users might be a hindering factor. In this case, a better understanding of the work situation of professionals by the researchers and developers of 3D UIs may be required.

Whatever the reasons may be, there is a need for research exploring the 3D design work world in order to bridge the gap between research laboratories and real-life application. As long as the factors which are important to adoption are not clear, every prototype system developed in the laboratory can only generate limited knowledge.

1.2. Motivation

3D design work is by definition a form of spatial interaction. Whatever user interface is used, the goal is always the creation or manipulation of objects in a 3-dimensional virtual space. Therefore, this type of work is can benefit greatly from adoption 3D user interfaces in one of their many forms. Firstly, a 2D computer screen representing a 3D virtual space is inherently ambiguous and make it harder to understand 3D relations between objects. Secondly, using a 2D input device such as a mouse or stylus to manipulate objects in the scene limits the interaction to one plane. While many UI metaphors have been developed to make this task as intuitive and efficient as possible, they can only approximate the natural spatial interaction that we are used to perform daily in the real world for all our life.

Thus, 3D UIs can offer a more efficient way of working. Since many artists during studies or in their leisure time experiment with physical media such as clay sculpting or stop-motion animation, the fundamental concept of spatial interaction and it's advantages should be obvious to most professionals. However, it is in the specific metaphors and details of the UI that make it either viable or unfit for use. There is therefore both a need and great potential to conduct research towards the realization of usable 3D UIs in the context of professional 3D design.

1.3. Challenges in Developing User Interfaces for 3D Design

Developing and deploying UIs for professional 3D artists pose a variety of problems that are different from the common lab-based design approach.

Firstly, researcher may often do not have a detail understanding of the way how professionals in various fields work. This makes it hard find solutions to problems or may lead to trying to solve problems that do not exist in real life. Research on the workflow and work environment are therefore necessary, as well as understanding the complex software products currently used.

Furthermore, professional users differ in the ways in which they adopt new technology, since they do so in relation to their ability to earn a livelihood, not for personal enjoyment or leisure. This means that the focus switches from enjoyment and intuitiveness to productivity and sustainability. These differences in thinking must be reflected in the research.

Another challenge is the limited access to members of the target audience. Much UI research is conducted with graduate students, since they are easily and cheaply acquired as experimental subjects. Professional 3D artists are often busy and will not readily spend their leisure time in research laboratories. This makes it harder to gather data and important feedback.

1.4. Proposed Solution

My approach to addressing these challenges was to begin by doing research on the current work environment and how state-of-the-art UIs are used. I was aided by personal experience of working in the games and visual effects industry and having some personal contacts. Next, I tried to deduce a set of universal requirements and then design a prototype UI to meet the needs of the target audience. On top of a formal evaluation of the prototype, I also performed experimental quantitative research on detail UI factors and how they affect work efficiency. Finally, I attempted to validate my theoretical findings by offering the prototype to professionals for real-life adoption. This spans the complete cycle from fundamental research, conception, experimentation, evaluation, and adoption.

<table>
<tr><th>Step</th><th>Task</th><th colspan="2">Contribution</th><th>Chapter</th></tr>
<tr><td rowspan="5">1.
Learn / Understand what is “useful”</td><td rowspan="2">Analyze the current situation</td><td colspan="2">Survey</td><td rowspan="2">Ch 3.</td></tr>
<tr><td colspan="2">Observation</td></tr>
<tr><td rowspan="3">Research factors of the novel technology</td><td>Input devices</td><td>Thumb-based gloves</td><td>Ch 5.</td></tr>
<tr><td rowspan="2">Output devices</td><td>Head-Tracking</td><td>Ch 5.</td></tr>
<tr><td>AR vs VR</td><td>Ch 6.</td></tr>
<tr><td rowspan="3">2.
Build it</td><td colspan="3">Requirements Analysis</td><td rowspan="3">Ch 4.</td></tr>
<tr><td colspan="3">UI Design</td></tr>
<tr><td>Iterative Refinement</td><td colspan="2">Formative User Study</td></tr>
<tr><td rowspan="2">3.
Deliver it</td><td>Get people to use it</td><td colspan="2">Release it to public</td><td rowspan="2">Ch 7.</td></tr>
<tr><td>Learn from feedback</td><td colspan="2">Collect Feedback</td></tr>
</table>

Figure 1.1. The outline of my work presented in this thesis.

1.5. Hypothesis

My research hypothesis was: “It is possible to develop a 3D immersive UIs for professional 3D design that allow such an advantage over current 2D systems that professionals will adopt them even at a non-negligible price. The factors that are required can be deduced by study of the current work environment, experimental studies, and iterative formative development. Furthermore, if such a UI would be implemented, it will be readily adopted by professional users.”

1.6. Approach

My proposed solution was realized within five projects. Figure 1.1 gives an overview of steps along the way. Please note that the chapters do not follow the logical order or the steps. This is because in order to perform user studies on detail factors, I fist needed to develop a working prototype.

1.7. Research Contributions

Part of the contributions discussed in this thesis have already been presented in my Master’s thesis. These include the following:

- Through an online survey I provided up up-to-date information on the work situation and attitudes of professional 3D designers.
- Based on this survey I deduced a set of requirements related to professional 3D design work which can be used by UI designers as a reference.
- Through a pilot user study with three participants I provide some insight in how artists evaluate UI technologies.

In addition to these, I have made the following contributions during my doctoral course :

I provide

- an additional evaluation of the prototype UI based on a larger formative user study with 3D artists which can be helpful to consider for future 3D UI design attempts.
- quantitative data on the effect of positional head-tracking on task performance in the 3D design context from a user study performed with 3D artists.
- quantitative evidence for the superiority of AR systems over VR systems for spatial interaction, both when using 3D and 2D input devices.
- information from my practical experience of bringing an experimental immersive 3D UI to market for adoption by professional 3D designers.

In this work, I present both the previously published and novel research contributions in context. Following the question “what makes 3D user interfaces fit for adoption by professional users in a 3D design context”, I have investigated both the current state of the art 2D user interfaces and conducted user studies aimed to create both quantitative and qualitative knowledge on the topic. I present my research towards finding the factors that are important for the adoption of 3D UIs to 3D design work and my efforts to design and deliver such user interfaces.

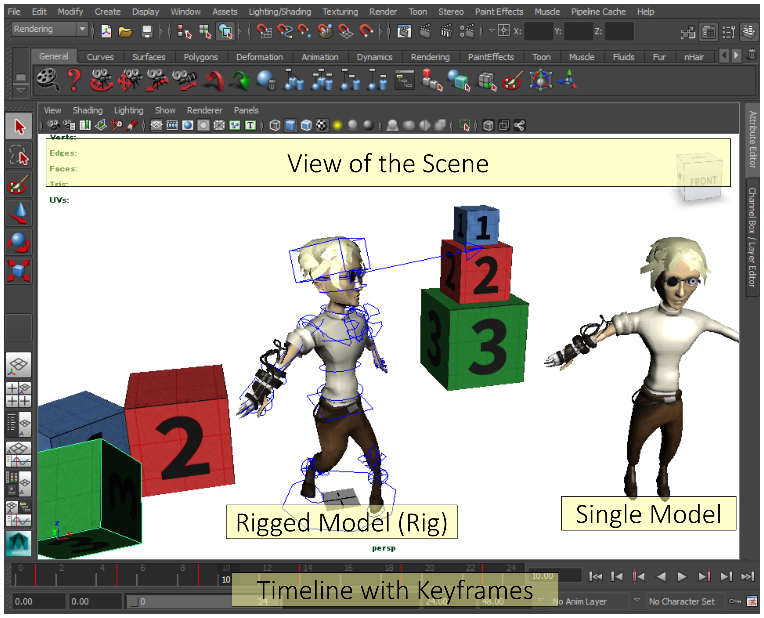


Figure 1.2. The default Maya User Interface, with several of the common terms indicated.

1.8. Terms and Definitions

In this section I will clarify terms commonly used in both media production [25] as well as AR and VR research, in order to avoid confusion. Please see Figure 1.2 for an example of a 3D design workplace.

1.8.1 Terms used in 3D design and animation production

A “3D Artists” or “3D Designer” is a term used to refer to creative people who are using computer software to generate 3-dimensional virtual objects, characters, and worlds, either to be used in video games or to generate visible 2D images (both still images and animations).

In the visual effects and animation industry, technicians and engineers working towards enabling artists to perform their work are usually called Technical Directors (TDs). The TDs work is usually focused on computer management and programming. However, a thorough understanding of the creative process and the tools used is required.

A single virtual 3D object is commonly referred to as a “model”. A “scene” is a union of one or several models, together with auxiliary data and effects to form a unified segment, for example the data required to render one shot for a film.

The creation of virtual 3D objects by an artist is called “Modeling”. It is not necessary that a real-world reference for the model to be created exists or is available. While some modeling techniques focus on creating a virtual representation of a real object, many artists work solely from their imagination.

“Animation” is the process by which the illusion of motion (and, in extension, life and personality or the observed figure) is created artificially by an artists. The most traditional form of animation is to create a series of drawings which are then viewed in rapid succession. The human mind will be tricked into interpreting the changes as motion and subsequently the illusion that a drawn cartoon figure is “alive” can be achieved. In computer animation, the computer takes on the work of generating a number of images from a virtual 3D world. Again, changes in this world will create the illusion of motion. The artist will have to define the state of the virtual object or world at several points on virtual timeline. These stated—often called “poses” in the case of virtual characters—are called “key-frames”. The animation software can then interpolate between the key-frames to generate a continuous motion. The process of recording the motion of real objects and actors to use in CG productions is usually not called Animation but is instead referred to as “Motion Capture” (if the recorded motion data is directly used by the animation system) or “Rotoscoping” (if a display of the recording is used by an artist as a reference which is to be closely re-created as original animation). In professional production, the term “Animation” is further restricted to describe only those motion patters that are used to bring some form of active agent to life. The motions of non-living things such as fluids, cloth, or rigid objects are commonly called “Simulation”.

In order to allow virtual models to perform complex motions and in order to make the task of Animation as convenient to the artist as possible, elaborate virtual control structures are created, which are called “Rigs”. A Rig usually includes virtual handles to control the model, a virtual skeleton system, and mathematical formulas on how the computer is supposed to interpret the interaction between all parts. The work to create such a Rig is called “Rigging”.

“Rendering” is the process by which a computer system generates a visible 2D image from an internal 3D representation of objects. In 3D design the term is usually not used for the interactive display on the screen that allows the user to interact with the scene, but only for the separate process of generating images for viewing alone, for example when generating the final result. This kind of rendering is usually not performed in real-time and often involves a complex management system and a separate computer system dedicated to image generation.

The interconnected steps required to create CG images or content for games and their integration into a work-flow is referred to as “Pipeline”. The Pipeline usually consist of a number of networked computers and programs used to organize the work preformed by artists working together on the same project. Specialized technicians working on creating and maintaining these systems are usually called Pipeline TD. Even in case of one artist performing every part of the work by himself, the logical dependencies of some steps on other steps dictate the cascading work structure of a Pipeline. While the dependencies may me linear, the execution of work is usually iterative, where the output of later stages is used to correct or improve prior stages.

1.8.2 Terms used in 3D user interface research

User Interfaces are usually categorized in “2D” (also called “traditional” UIs) and “3D UIs”. 2D UIs are performed on a flat computer screen and using a 2 degree-of-freedom (DOF) input device such as a mouse. A stylus tablet may have an additional DOF regarding the angle of at which the stylus is held, but since it cannot be lifted off the tablet without losing tracking, this third DOF cannot be used intuitively for 3D interaction. In “3D UIs” both the display technology and the input devices allow 3-dimensional perception, usually through the means of stereo vision and tracking an input device in 3 DOF. 3D input devices actually often offer 6 DOF of interaction: three DOF translation (forward, sideways, and upwards) as well as 3 DOF of rotation (yaw, pitch, and roll). This reflects the natural interaction with hands in the real world where simultaneous translation and rotation are possible and common. Some input devices offer more than 6 DOF which then usually require bimanual operation [30].

The action of changing the user point of view on the scene without changing the scene itself has different terms. In design software it is often called “camera control”, since it changes position parameters of the virtual camera used to render the scene. In VR technology, the terms “navigation” and “locomotion” are used to describe the same concept, while giving additional focus on how the user experiences this perceived motion. Common metaphors used to achieve navigation include "grabbing-the-air" and "world-in-miniature" [15].

A "head-mounted display" is a system of computer screens and lenses that is worn on the head that allows the user to observe the images rendered on the screens. These can be used both for virtual reality (where the HMD occludes the whole field of view and does not allow the user to observe the environment) and augmented reality (where the HMD only adds certain virtual images to the real environment). There are two types of HMDs for AR. "Video See-Through" (VST) are similar to VR HMDs in that they block the complete field of view of the user, but instead use video cameras attached to the front of the HMD which are then displayed in the HMD, allowing the user to observe the real environment, thus combining reality and virtual content [69]. "Optical See-Through" (OST) HMDs use an optical combiner. The user is able to observe the real environment through the combiner, while at the same time virtual imagery is projected onto the combiner. In the eyes of the user, the two sources of light will be combined into an augmented reality experience [68]. See Figure 1.3 for an illustration of the different types of HMDs.

An alternative to HMDs are "Cave Automatic Virtual Environment" or "CAVE" systems in which projectors are used to display the user interface onto the walls surrounding the user [23]. Often, special glasses or other headgear are worn when using the "CAVE" in order to achieve 3D vision.

"Tracking" describes the calculations performed to find the position and orientation of an object in 3D space, often the users head (point of view) or hands. Two types of tracking are commonly employed. "Inside-Out Tracking" uses data gathered from sensors inside or attached to the object to be tracked. This can be achieved with video cameras that take pictures of the environment. The technique to calculate the position and motion of a camera solely from the image that it makes is called "visual odometry". This often requires the computer system to keep a virtual representation of the real world which is created and extended upon ad-hoc at runtime. These systems are called "Simultaneous Localization and Mapping (SLAM)". "Outside-In Tracking" on the contrary relies on external sensors in the environment that provide data with which to measure the target objects' position. Often these are color or infra-red video cameras that track specific details on the surface of the object. Visual markers attached to an

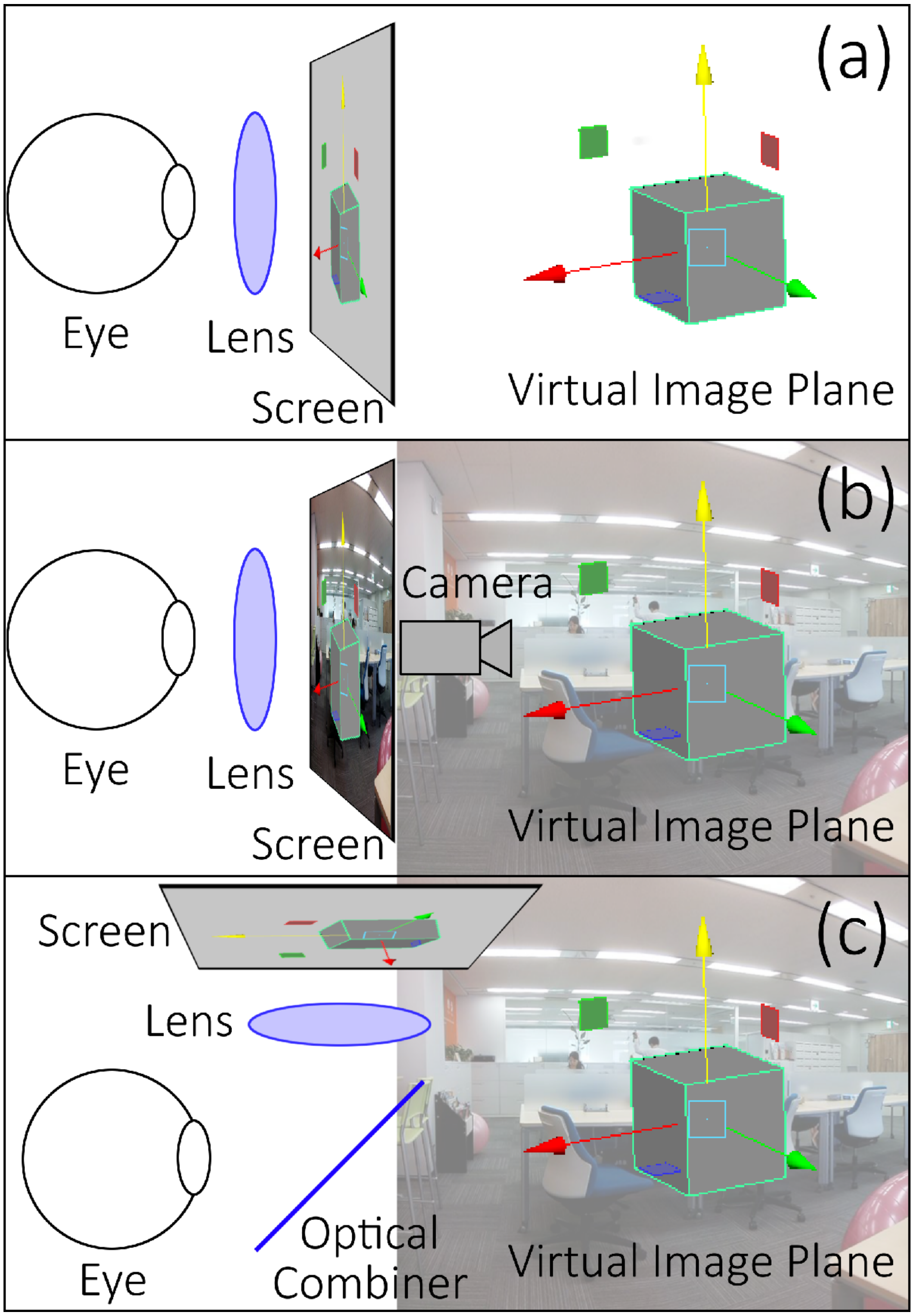


Figure 1.3. Illustration of the technical set-ups of different types of HMDs. (a) VR HMD. (b) Video See-Through (VST) HMD. (c) Optical See-Through (OST) HMD.

object or in the environment for visual tracking are called “fiducial markers”. A very elaborate version or outside-in tracking are Motion Capture (MOCAP) systems, which can track the entire range of motion of the human body. Finally, inertial measurement units (IMUs) that measure acceleration forces are available. These are often not referred to as either inside-out or outside-in tracking, though it can be argued that they track relative to natural physical phenomena such as the earth’s gravitation or magnetic field.

CHAPTER 2

Factors for Applying 3D User Interfaces to Professional 3D Design

2.1. Introduction

Different from theoretical or explorative research, application-focused research has to follow a different set of criteria. Namely, it is of greatest importance to identify those factors that matter most to practical application and find ways to address these issues. In this case, UI designers need employ user-centric research methods to investigate the background of their users such as their work style, habits, way of thinking and nomenclature used in their line of work, as well as typical or possible use cases of the technology being evaluated.

For this, it is important to first consider the various research methods and organize them into a system where they can build another. In this chapter, I explain about the various factors that come into play in regard to 3D UIs and 3D design and how they affect the research process. This organization proved useful in segmenting the larger goal and selecting those factors for closer analysis that could be reasonably approached during my doctoral course.

2.2. Understanding the User

An understanding of the work environment for which one is trying to apply one's research is of utmost importance. Such an understanding comes in three forms: personal understanding of the subject matter, a top-down analysis of the circumstances, and in-depth detail observations.

The first type, understanding of the subject matter, gives the basic understanding of the workflow and work environment which is required to add meaning to observations and form hypotheses for experiments and well as suggesting possible explanations for the results of such experiments.

The second type of top-down analyses adds scientific validity to what would otherwise be only anecdotal evidence. By collection and statistical analysis of key metrics over a larger group of individuals of the target user group, we are able to determine which experiences and demands are characteristic for the whole group of target users.

However, this data is very abstract and does not lend itself easily to deduce concrete steps to take to improve the situation. Therefore, it is necessary to add a third type of understanding which comes from observing users in their real-life work and analyzing their behavior. While this type does not provide the same level of generality as the second type—some behaviors may be individual and not representative of the larger group of users—they allow us to find the concrete sources of the results found in the second type of analysis. It is again the first type, understanding of the subject matter, that allows us to make educated assumptions which behaviors are likely to be universal and important enough to justify further analyses in an experimental setting.

2.2.1 Understanding of the Subject Matter

The most basic understanding required for successfully conducting application-oriented research comes from personal education and experience and can be described as the ability to do the work that the target user group performs. In my case, undergraduate studies in Media Technology and Design [1], internships and prior employment in media production companies have provided me with a general broad understanding of the topics and tasks usually encountered by 3D artists, including an extensive knowledge

---

[1] https://www.fh-ooe.at/en/hagenberg-campus/studiengaenge/bachelor/media-technology-and-design/

of the use of 3D modeling and animation software. Through this, I found that a disproportional amount of work resulted just from working around the limitations of the UI such as the difficulty to understand the spatial relations of the 3D content correctly. I furthermore found through personal contacts and anecdotes how and where artists were dissatisfied with their current UI, and that many had developed health problems in their arms and hands due to overuse. I also found that many owned a 3D input device of some kind but were not using it consistently. The main reason appeared to be that they could not perform the complete workflow with it and switching between the 3D input device and the mouse was tiresome.

For this research, I further relied on reviewing video tutorials such as those provided by The Gnomon Workshop and [2] and Pluralsight [3], which usually show a seasoned artist performing the work and thus provide good insights into the workflows of professionals.

2.2.2 Top-down analysis of the circumstances

In order to add validity to assumptions based on the first kind of understanding, statistical analysis of a larger representative group of the target population is required. Pitkow and Recker spearheaded the approach of using the web to learn more about users of certain technologies [61]. In Chapter 3 Section 3.2 I report the results of an online survey that I performed in 2014 in which 54 media production professionals from 17 countries described their daily work experience in both quantitative and qualitative terms.

2.2.3 Detail observation

Since the survey can give only very generalized insights into the work situation of artists, it is difficult to identify the causes of specific results and give meaning to the answers provided by the participants. To gather more concrete insights into the particulars of the artist's workflow, I invited two artists in order to observe their normal everyday work flow. Following the recommendations by Carvalho [19], I recorded their work performance on their normal project work with a video camera while leav-

---

[2] https://www.thegnomonworkshop.com

[3] https://www.digitaltutors.com/software/Maya-tutorials

ing the room in order to avoid the Hawthorne effect, and later analyzed the resulting video. The details and analysis of these recordings are described in Chapter 3 Section 3.3.

2.3. User Interface Considerations

In exploratory research, the goal is usually to find new ways to achieve certain tasks that rank higher on any scale of intuitiveness, enjoyability, or any performance metric, regardless of the cost of acquisition or technical complexity. Application oriented research targeted at businesses is different from this approach in that it must mainly consider financial concerns compared to productivity gains. For a normal user, the price and time investment of adopting a new technology is not of great concern for research because the evaluation whether it is worth the cost is one of individual preference, desire, and wealth. Businesses, on the other hand, may be unwilling to commit to even minor expenses if they don't expect an increase in productivity or, on the other hand, may readily expend even more money than that can currently amount (by taking up credit or outside investments) if they can calculate a future increase in profits that justifies the expense. These financial concerns are associated with concerns regarding the quality of the work or product that the business is trying to sell. Again, individuals who perform the work for their own enjoyment or altruistic contribution have a different than those who attempt to achieve a profit. Finally, the health and well being of the user is evaluated differently in a business context due to the repeated and extended use and due to the fact that the system is not used for enjoyment but in order to secure one's livelihood in the long term.

2.3.1 Financial Concerns

Since adopting a novel technology requires an initial investment—both monetary and in human labor—the optimal decision whether commit to such a step can be calculated using the net present value [57]:

$\sum_{T}^{t=1} \frac{C_t}{(1+r)^t} - C_0$

where $t \in [1T]$ are the time periods under consideration, C_t is the net cash inflow during each time period t r is the discount rate, and C_0 is the initial investment costs, in this case, the cost expended to acquire the VR or AR hardware and software.

Using this formula, a studio manager or freelance artist can compare the cost of switching to a 3D UI system to the expected increase in profits provided through the increase in productivity.

Thus, the focus of application-oriented research shifts from exploration and trying to find more intuitive interaction metaphors towards trying to estimate the effect size of potential benefits in order to allow professional end users to make informed choices. For instance, Mapes and Moshell developed a glove-based 3D UI and evaluated it's efficiency to the mouse in a user study [53]. While most of the time, the mouse was superior, the authors managed to train one participant on the system until he was six times faster using the 3D UI as compared to the 2D UI. This, however, took several hours of training. While this may be disappointing for researchers trying to find more intuitive UI metaphors, qualitative results like these are extremely valuable to professionals and managers because the training period can be calculated as an investment into future performance improvements.

In Chapter 7 I report on artists and studios that have made the financial decision to adopt 3D UIs in their workflow.

2.3.2 Quality Concerns

Productivity is commonly thought of as throughput, meaning the time it takes to perform a certain task. The lower the task completion time the more units can be produced in any given time, thus raising total output.

Often, however, completion time can be lowered at the expense of quality. A reduction in task completion time can only then be seen as an increase in performance if the quality of the output stays the same. On the other hand, performance improvements can also mean increasing the quality of the output while keeping production time constant.

Quality is an intrinsic cue of the product and cannot be changed without altering the nature of the product itself, while price is considered an extrinsic cue that is not part of the product itself [75]. There is no simple formula connecting the two variables, since the requirements may be different for each individual case. In some cases, even huge improvements in quality may not be seen as beneficial, if such a level of quality is not recognized and honored by the consumer.

In the case of design, the calculation becomes even more complex because the output is an artistic creation. The evaluation of quality thus becomes highly subjective. Some viewers may even value a reduction in quality as having a “rough charm”, or may dismiss increases in quality as qualitatively inferior if the new style does not meet their expectation.

In the real world, artists and content producers often carve out a specific niche of style, quality, and cost that they serve. They acquire customers who are willing to pay a reasonable price for the work the producer is able and willing to create. Often, changing the recipe may result in losing customers, even if the quality increases in the eyes of the producer. Thus, changing the style and quality of output is more of a marketing decision than a technical one.

Since it is hard to objectively measure artistic quality and since even increasing quality may not be desired by many users, this leaves only one formula to scientifically measure performance improvements: if a novel method is able to produce the same type and quality of output in a lower time frame, the resulting percentage of saved time can be seen as equal to the total productivity increase.

In Chapters 5 and 6 I present two user studies that were conducted by fixating a certain outcome as baseline quality and then measured time expenditure to reach this outcome as a measure of performance. Additionally, in Chapter 4 Section 4.5 I present the results of a questionnaire filled out by artists after using an experimental 3D UI in which they attempt to estimate the potential performance benefits from using 3D UIs.

2.3.3 Health Concerns

Professional artists—unlike hobbyists—can be expected to use their main system for long hours at a time every day. Furthermore, professional users do not expect their work to be fun, but instead to allow them to sustain their livelihood in the long term. This raises major concerns regarding the health and well-being of the artist. Chapter 3 Section 3.2 provides some insight into the health situation in current work environments using 2D UIs.

One of the main concerns when switching away from the mouse is the fear of arm strain. Similar concerns have researched previously regarding the use of a mouse an input device for CAD work [18]. The risk of developing Computer Related Upper Limb Musculoskeletal (ComRULM) disorders are well researched for traditional 2D

UIs [55], but not yet well understood for novel 3D UIs. However, since many traditional tasks such as stop-motion animation require manual labor, it can be assumed that a method of implementing 3D UIs in a way that does not cause long-term stress or chronic injuries must exist. It is therefore important to consider this as a human factor in interface design. Chapter 4 section 4.3 discusses concepts to make the use of 3D UIs more comfortable.

However, the requirement for in-depth medical understanding and the inability to perform long-term user experiments have limited my ability to address these concerns in more depth in my research. Furthermore, since the experience of strain or pain are subjective, simple Likert scale questionnaires may give a distorted image of the condition [10]. I have therefore refrained from claims that personal experience or reported comfort of test subjects are evidence for the medical safety of UI concepts researched during my doctoral course.

2.4. Technical Factors

I have limited the scope of my research to those 3D UIs that are realized using HMDs. CAVE systems that work with back-projection differ in that the users head is not inhibited as much [23]. However, the complexity of the set-up and space requirements make them more difficult to adapt in a studio environment. Fish-tank VR systems are a more simple version of CAVE where only one 3D monitor is used to display the virtual world [3, 11]. This reduces the set-up complexity but also limits the system because the field of view and possible view direction is now very limited. Furthermore, both CAVE and Fish-tank systems make it difficult to create AR environments, since they introduce a physical barrier on which the content is displayed making it impossible to display virtual content in front of physical objects. Current-generation head-mounted displays are sufficiently comfortable to allow analyzing general factors and requirements regarding 3D user interfaces as long as the use-time does not extend over several hours. Therefore, with the exception of long-term effects such as eye strain or neck strain, HMDs are more suitable to explore possibilities for bringing 3D UIs to creators. I have omitted any analysis of HMD hardware technology because such research is highly complex and requires a great amount of understanding and special equipment to perform. Instead, I focused on the user interface and input devices.

As of the state of the art at the time of this research, the possibilities for creating 3D UIs using HMDs present themselves as a continuum of increasing performance and complexity, which can be broadly categorized into three parts.

In the first category, an off-the-shelf consumer smartphone is placed into a harness that can be worn on the head. It includes lenses that allow the eyes to focus on the smartphone screen. Tracking is achieved by the smartphone's internal Inertial Measurement Unit (IMU, also called gyroscope). Thus, only rotational head-tracking is possible, since current-generation accelerometers are not sufficient to realiably track translational motion. Thus, even when the user moves his head from side to side or walks around, he will not be able to alter his point of view. On the other hand, no set-up or initialization procedure is required which makes these systems very portable. In the following, I will refer to this concept as "mobile VR". This is different from what is commonly called "mobile AR", in which the user holds the device in front of him and the image of the internal camera is displayed on the screen, augmented with virtual content by the use of computer vision algorithms. Mobile AR systems are usually not attached to the user's head because smartphones lack the required stereo cameras required for stereo vision, and even for monoscopic AR the smartphone camera is usually not placed correctly (on the corner of the device rather than in the center) and does not meet the strict latency requirements for AR systems. Therefore, while it is technically possible to create hands-free head-mounted mobile AR systems, the application is quite rare in the real world, and I have therefore neglected it in this work.

The second group of currently available systems for 3D UIs are the more elaborate "full" VR systems which consist of a dedicated HMD and external tracking units. The Oculus Rift and HTC Vive are currently the main options in this field. Such systems are usually more expensive and require a more elaborate setup in the form of installing the tracking system and calibrating the space. However, they also offer positional head-tracking.

The most advanced group of systems are HMD based AR systems, either VST or OST. In the OST-HMD category, the Microsoft Hololens is the most prominent example. VST systems usually consist of choosing one of the VR systems in group two and attaching special cameras to them such as the ovrvision stereo camera system

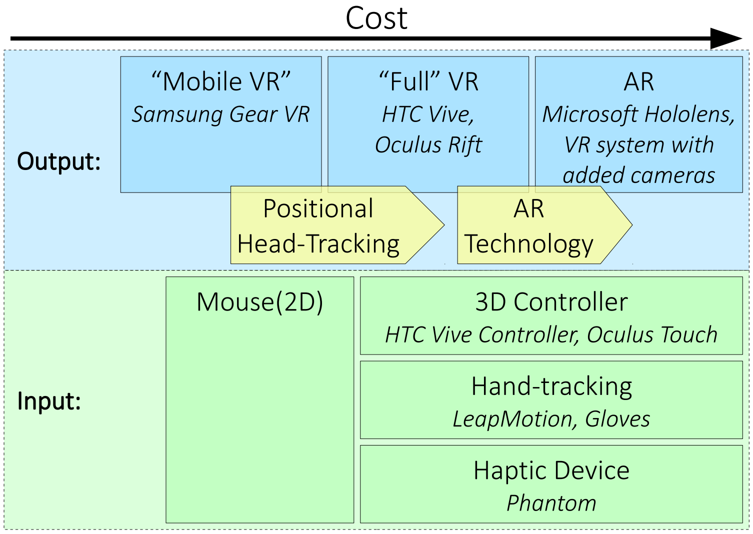


Figure 2.1. Current-day space of options available when considering 3D UIs.

[4]. This final category allows the user to observe the real world and virtual content at the same time. They are therefore often used for tasks that require interaction with real objects such as in manufacturing.

Thus, when considering adoption using the present-day available systems, the dividing lines are Positional Head-Tracking and AR technology. Each of these increases the cost and complexity of the system, but may bring some advantage to productivity. Another dimension that affects productivity is the input device used. Here, the traditional 2D mouse can be replaced with a variety of novel 3D input devices at a premium, but the differences between these devices is less clear cut. See Figure 2.1 for an illustration. I will discuss the factors in detail in the following sections.

2.4.1 Positional head-tracking

While rotational tracking can be achieved through an inertial measurement unit at a low cost, positional tracking is more complicated to realize. Bhatnagar reviewed different technologies can be employed to achieve this kind of tracking, including magnetic, acoustic, optical and mechanical [8]. The Oculus Rift HMD requires the user to acquire and set up external sensors for an "outside-in" tracking scheme. The most

---

[4] http://ovrvision.com

prominent current examples for “inside-out” tracking systems that require external devices is Valve’s “Lighthouse” system that is used in the HTC Vive VR HMD [5]. An alternative way is to use a camera attached to or included in the HMD and determine its position by analyzing the camera images with computer vision algorithms. This is technically also a case of “inside-out” tracking, but does not require an external room-mounted tracking system. One way to achieve this kind of tracking is to use fiducial markers which can efficiently be detected and then calculated in relation to the HMD. The other way is to use natural features in the surroundings to simultaneously generate a virtual map of the area and track the camera’s motion in it at the same time, an approach which is commonly called “Simultaneous Localization and Mapping” (SLAM). This method can be performed both with monocular as well as stereo cameras, and can be improved upon if depth-cameras use used.

This second approach is particularly well suited for VST HMDs, since the camera image processing is already in place. However, even OST HMDs like the Microsoft Hololens use computer vision instead of an external tracking system. It can even be applied to “mobile VR”, where the device camera is used to perform tracking without showing the image to the user.

All of these methods however significantly raise the cost of the system as compared to neglecting positional tracking. External trackers need to be acquired and installed, and computer vision algorithms are computationally expensive, requiring more expensive processors.

In Chapter 5 I present the results of a user study that aimed to find potential benefits of positional head-tracking that would justify the expense.

2.4.2 Augmented Reality

AR systems can be basically divided by the method by which they combine the real and virtual content. Video See-through (VST) systems work by editing a live video stream, inserting virtual content. They are thus very closely related to VR systems with the additional requirement of video cameras and computational cost of video processing such as image undistortion and rectification. Optical see-through (OST) systems work by overlaying placing an optical combiner before the eyes of the user,

---

[5] http://gizmodo.com/1705356768

overlaying the natural view with virtual objects. These systems still require somewhat more challenging to build and calibrate than VST systems.

It must be stressed that all AR systems require positional tracking of the device in one of the ways described in the last section. This is because the virtual content must be registered in 3D with the real world. An exception to this are 2D "heads-up displays" (HUDs) such as the Google Glass which are also sometimes called AR. However, these are not AR systems in the strict sense of Azuma's definition [4], which requires 3D registration of the virtual content with the real world. For the purpose of this work, however, 2D HUD systems are unsuited for 3D design work. I have therefore omitted them in my research. Thus, it can always be assumed that AR systems are more expensive than comparable VR systems, as they share the same requirements and only add to the complexity.

In Chapter 6 I present the results of a user study that aimed to find potential benefits of the ability to see the real environment that would justify the expense.

2.4.3 Input methods

While it is less common, it is possible to use a traditional 2D mouse for interaction by displaying the cursor in the AR or VR environment. Advantages are the high maturity of the device, familiarity of the users, and the low cost. Novel 3D input devices can provide higher DOF interaction, but the distinction between them is not as clear-cut as in the case of viewing devices.

3D user input can be achieved either by natural hand motions and gestures or by using some kind of device or controller. Whether one approach is superior to the other is not easy to determine and generally application specific. While direct hand manipulation is generally seen as more natural, it must be noted that most physical real-world tasks require the use of a tool anyway. Thus, a physical controller may be more natural in some settings because it gives the same tactile feedback as holding a real tool would.

Current-generation hand tracking systems such as LeapMotion generally lack behind comparable 3D input devices in terms of precision and robustness. Because of the complex shape of the hand and self-occlusion during certain gestures, it is much more challenging to implement and computationally expensive to perform than a button on a controller that simply closes a circuit. A middle-way that closely mimics direct hand interaction while simplifying the process are glove systems. Not only does the glove make tracking easier because it can be color-coded or fitted with trackable markers, but also sensors inside the glove can detect touch and pinch gestures reliably even under the most challenging conditions. A great number of glove-based input devices have been developed and evaluated over the years [24].

Another option is the use of haptic feedback devices. These are stationary and thus limit the range of motion, but provide tactile feedback to the interaction.

The user studies in Chapters 5, 6, and Chapter 4 Section 4.3 were conducted using either a pinch glove or a printed 3D input device, as well as a traditional 2D mouse.

CHAPTER 3

Research on the Current Use of 2D User Interfaces

3.1. Introduction

Over the recent years, a great number of 3D UIs were developed and evaluated in the lab for all kinds of purposes. However, most of these experiments concerned themselves more with the technical challenges and possibilities. Few if any were based on a solid analysis and understanding of the current work flow that they are trying to supersede or improve upon. This limits their possibilities for later real-world adoption. In order to keep my own research more applicable, I started by gathering concrete information on current 2D user interfaces for 3D design. I did so by performing a survey with media professionals, asking them about their work environment and situation. I furthermore recorded two artists at their work and analyzed the videos, in order to gain a better understanding of the workflow in detail. The details and results of this analysis are described in this chapter.

3.2. Survey on 3D Design Work Environments

In 2014, I performed an online survey with media production professionals. The results of this survey were published in [47] and my subsequent Master's thesis.

I used the on-line survey platform surveygizmo.com in order to reach professionals from around the globe, and distributed it via various channels, including my own personal contact network, 3D art forums and freelancer recruiting websites. The participants came from various walks of 3D design including visual effects, games, advertising, and architecture. I did not limit the survey to one particular type of work and did not attempt to reproduce the actual real-world distribution of 3D artists based on nationality, profession, size of company, or other demographics. Detail results may therefore differ based on what sub-group or country is considered, but the survey nonetheless provides some insight into the work situation of 3D artists.

The questionnaire consisted of several parts and had some degree of built-in automation. In the first part, participants were asked for a self-assessment of their professionality and skills. I used this self-assessment as a first threshold to limit the sample to experienced media production people. As the log files showed, a number of people tried to join the survey, but felt not sufficiently proficient in media design work to describe themselves as professionals. In this case, the survey form automatically excluded the participant from the survey results.

On the last page of the questionnaire (after all answers had been completed) I asked the name and company or portfolio website of all participants. I did so in order to protect the survey from fraudulent responses, and ensured the participant that their names would not be published. Since I asked for the identity of the participant only after all questions were completed, the risk of social desirability bias was minimized, and it can be argued that contrarily the validity of the survey results improved, since people who were not serious about their answers were hesitant to give their real name. I removed answers of participants who could not be identified from the sample. I did not contact each artist individually to check if they were in fact the right person, but the timing of responses after informing different groups of people about the survey gave confidence that the stated people did in fact take the survey themselves. The final sample consisted of 54 participants from 17 countries (see Figure 3.1 and Table 3.1).

The complete survey questionnaire including all questions, illustrations, and comments is in Appendix A.

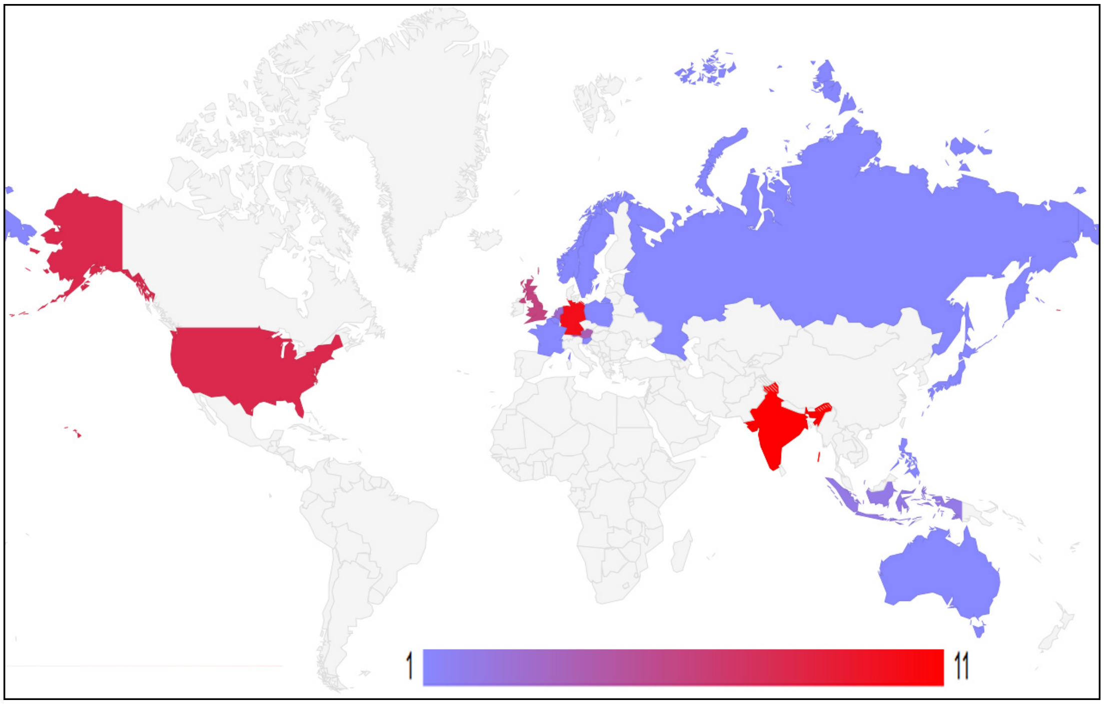


Figure 3.1. Distribution of participants.

The survey consisted of 46 questions relating to various aspects of 3D design work: self-assessment of professionalism, use of 3D software functionality, the role of collaboration in 3D design, used user interfaces, and potential novel user interfaces. Some of these questions allowed the participants to leave additional comments, which turned out to be very constructive and insightful.

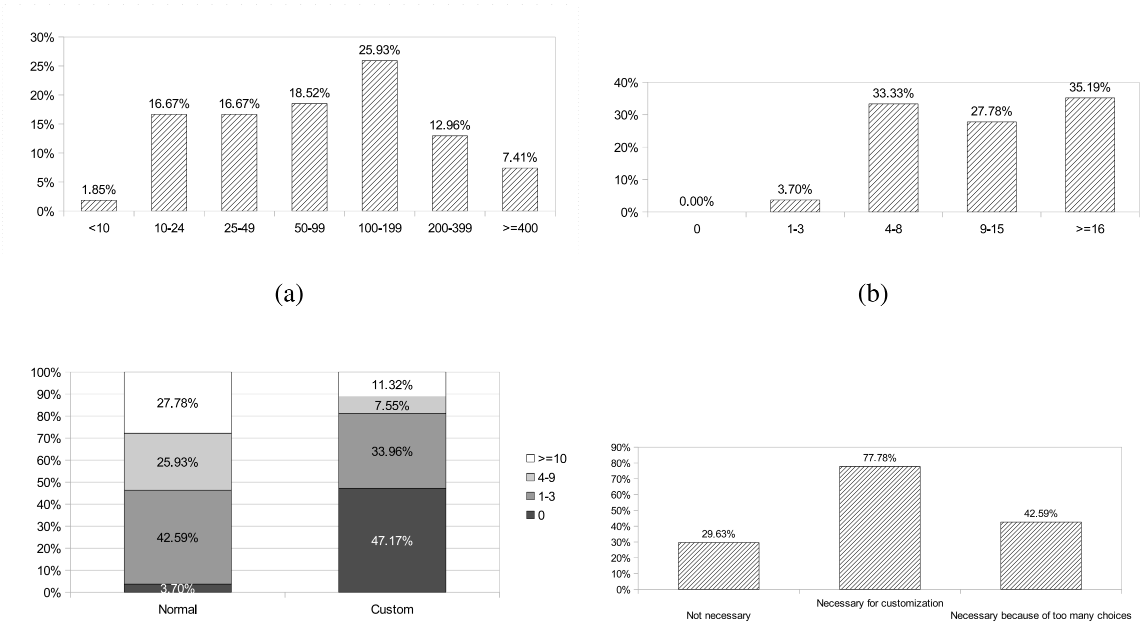
Use of 3D software: One set of questions related to the way that artists used their professional tools daily. I asked them how many of the functions that their 3D software offers they actually use regularly. I discovered that the average 3D artist utilizes up to a hundred functions (Figure 3.2(a)). Furthermore, I asked how many keyboard hot-keys they habitually used to increase their efficiency and found that using 16 or more hot-keys was not a rare exception while using less than or 3 hot-keys is uncommon (Figure 3.2(b)). When asked whether they use scripts or plug-ins for their software that were not part of the original set of functionality, only 3.70% stated not to require any additional functionality and many artists use more than just a few additional scripts or plug-ins (Figure 3.2(c)). Similarly, 52.83% of the participants in this survey stated

India	11
Germany	10
United States	8
United Kingdom	6
Austria	4
Netherlands	3
Indonesia	2
Japan	1
Poland	1
Sweden	1
Slovenia	1
Philippines	1
Belgium	1
France	1
Australia	1
Russian Federation	1
Norway	1
Total:	54

Table 3.1. Number of participants by country.

to use custom scripts or plug-ins, that were developed at their own company or were ordered to be created exclusively for their own requirements. 77.78% stated that they believe custom scripts and plug-ins will always be a requirement in order to adjust the software to production requirements (Figure 3.2(d)).

Navigation and scaling: Furthermore, I asked about the use of translational navigation and zooming/scaling during work on a 5-bin Likert scale, ranging from “constantly” to “never” (Figure 3.3). Navigation in this context means that the viewpoint in the virtual 3D scene is repositioned in order to show the content from a different angle or to show a different part of the content. Zooming or scaling then refers to changing the size of the viewpoint to either focus on some details (zooming in) or include more content for an overview (zooming out). Both operations do not alter the content since they only affect the view of the user. Most artists concurred that they use these



(a)

(b)

(c)

(d)

Figure 3.2. (a) “Estimate how many different tools and functions of 3D CG software you regularly use.” (b) “Estimate how many keyboard hot-keys for your 3D software you use regularly.” (c) Estimation and classification of additional scripts or plug-ins used with 3D software; *(normal)* means publicly available, *(custom)* means developed in-house or ordered exclusively. (d) “How do you feel about employing plug-ins and scripts in your work-flow?”

navigational devices regularly or even constantly during their work.

Collaboration: To get an impression of collaboration in the studio environment, I divided the various types of collaboration into four categories. *Non-interactive review* was defined to be the process of presenting the content without being able to edit it at the same time. *Interactive review* similarly was described as receiving feedback while simultaneously being able to edit the content. *Divided collaboration* was defined as the process of multiple artists working on the same content by means of dividing it either by separation or taking turns. *Full collaboration* finally defined working together on the same content at the same time. I asked the participants both how regularly they performed each type of collaboration and how important they consider it—whether

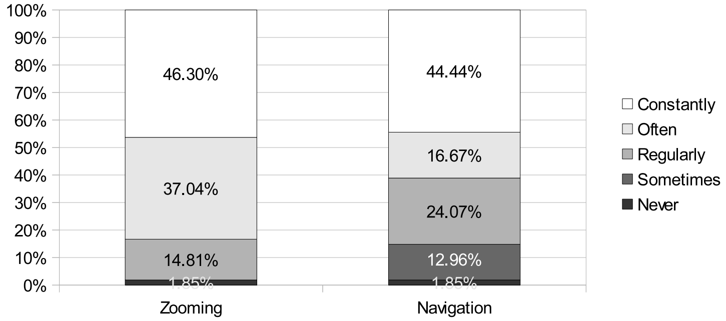
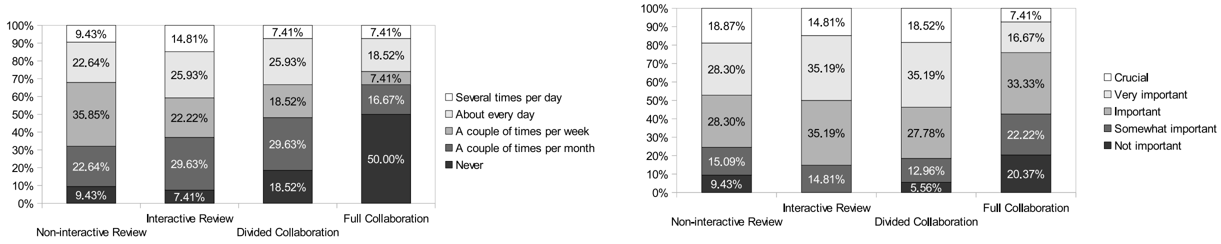


Figure 3.3. Navigation and zooming.



(a) (b)

Figure 3.4. Frequency (a) and perceived importance (b) of various types of collaboration.

based on experience or expectation—for their professional work. The *Interactive review* was considered the most common and important form of collaboration, but even full collaboration, while relatively rare, was considered important by the majority of artists. Figure 3.4(a) and 3.4(b) show the details.

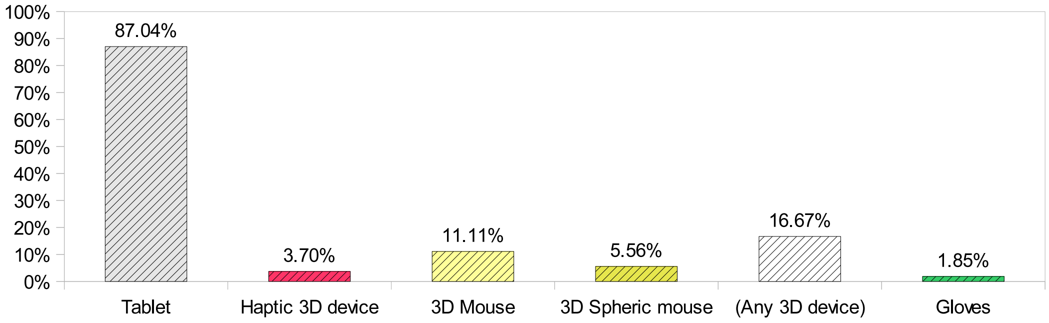
Work conditions: I asked a number of questions regarding the artists work environment and daily work context. When asked for their average daily work hours, the average time stated was 6.87 hours (2.27 hours standard deviation), where the common duration of one work session until taking a break was on average 3.02 hours (2.25 hours standard deviation).

When asked for their current main input device, all artists expressed working with a 2D device, 75.47% with a mouse and 24.53% with a pen tablet. Not a single participant stated to use an alternative or 3D user input device for work. More than 16.67% of artists reported that they previously tried 3D user input devices, either haptic 3D input device, 3D mouse, or a 3D spheric mouse (haptic 3D input device: 3.70%; 3D mouse 11.11%; 3D spheric mouse 5.56%). 1.85% stated to have used a glove-based input device (Figure 3.5(a)). None stated to have tested a device I did not anticipate.

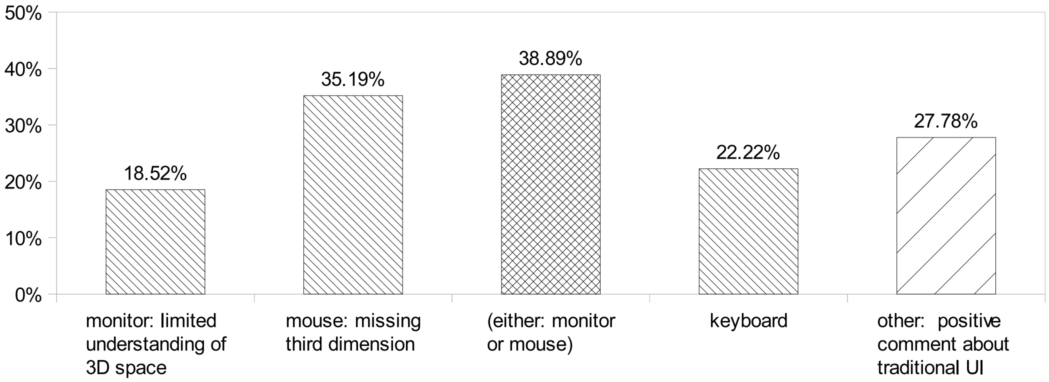
I further asked for the participants' opinions regarding the use of 2D UI devices for 3D work and found that 38.89% of the participants are missing the third dimension (monitor: 18.52%, mouse: 35.19%). However, 27.78% also expressed explicitly positive opinions about 2D UIs (Figure 3.5(b) and Table 3.2).

When asked how their arms, hands, and wrists feel after several days of working with 3D software, 46.67% stated "normal", 46.67% expressed that their arms felt "tired" and 6.67% even stated that their hands "hurt". However, when asked whether they perform some kind of countermeasures to deal with work related problems in their arms, wrists or hands, a total of 31.48% stated that they employ some kind of counter measure (special hand-gear: 1.85%; special exercises: 29.63%; take medication: 1.85%; seeing a physician or health expert: 3.70%). (Figure 3.6(b) and Table 3.3).

Novel UI devices: Finally, I presented artists with a series concepts of novel user input devices, as they were previously proposed by scientists and asked them to give their opinion on these in the context of professional 3D content creation. I asked them not to worry about the precise details of the implementation but to focus on the general concept instead. Each of the novel input devices was displayed in the form of an illustration and a short descriptive text. One was the use of a haptic input device (force feedback 3D pen) with an AR overlay, another was a tracked 3D mouse or expladdle as suggested by Fiala [27] and Shaw and Green [65]. The next presented system was the use of gloves instead of a tangible object. I furthermore presented the possibility of employing a mannequin or physical rig that is tracked, similar to the system proposed by Barakonyi [7]. Finally, I proposed to the participant the use of voice commands to control the software, as utilized by Girbacia [31]. Since all of these concepts are based on AR research prototypes that are not publicly available, it is highly unlikely that participants had previously experienced such systems. However, the comments that the participants provided for each concept show that they not only understood the idea but also reflected on various issues surrounding it such as the temperature when using

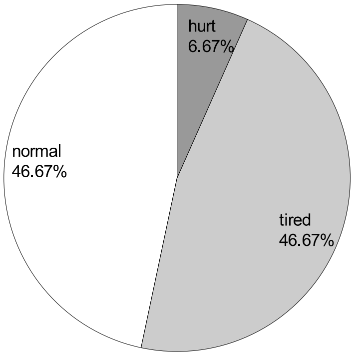


(a)

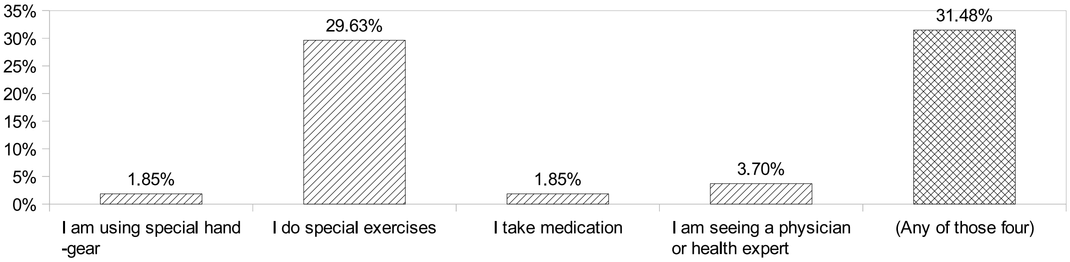


(b)

Figure 3.5. **(a)** Q24: "Which alternative / special input devices have you used for your work before (including testing them)?" **(b)** Q25: "How do you feel about the currently common 2D user interfaces (mouse, keyboard & 2D screen) for 3D CG work?" (1) "Working with a 2D monitor is making it difficult to work efficiently because of the limited understanding of 3D space." (2) "The keyboard is not suited for this type of work (because of layout, size of keys, missing icons or other reasons)." (3) "Working with a mouse is inefficient because of the limitation to 2D interaction (missing third dimension)."



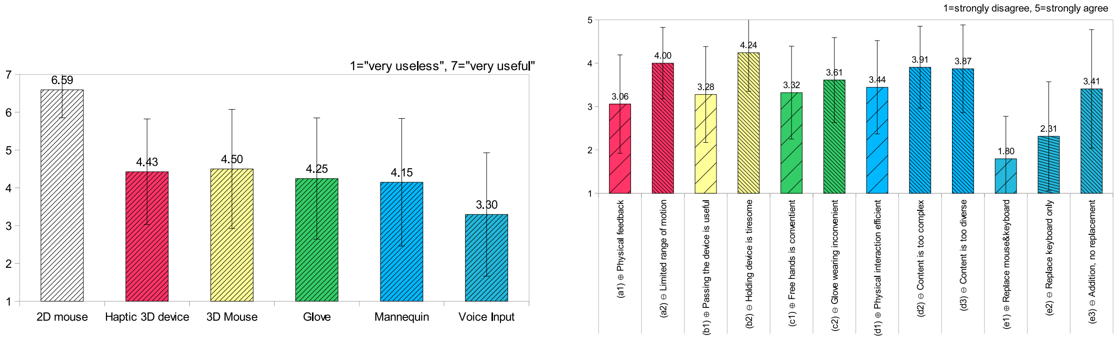
(a)



(b)

Figure 3.6. (a) Q28: "How do your arms (including wrists and hands) feel after several days of working with 3D CG software?" (b) Q29: "Are you currently taking any active measures to deal with work-related strain or pain in the arms, wrists or hands?"

glove-based systems in how weather. I asked the participants to rate their agreement to common arguments regarding the advantages and disadvantages of these systems in form of a Likert Scale (Figure 3.7(b)). It became apparent that media professionals agree considerably more with the shortcomings of novel user input devices than with their expected merits. I also asked them to rate the usefulness of the presented devices on a Likert scale from 1 ("very useless") to 7 ("very useful"), comparing them to each other as well as to the traditional mouse as a user input device for 3D content creation (Figure 3.7(a)). Strikingly, all 3D input devices scored significantly lower (between 4.50 and 4.25) than the traditional 2D mouse (6.59). Voice input scored even lower than the 3D UI devices (3.30).



(a)

(b)

Figure 3.7. (a) “Please rate the presented devices in respect to their usefulness, comparing them to each other as well as to the traditional mouse as user interface device.” (b) Opinions (positive and negative) on presented user input devices.

3.3. Direct Observation and Analysis of 2D UI Use

Next, I began working with artists directly to understand their work environment. In August and September 2015, I performed a pilot study with one professional artist and one amateur, in which I examined their habitual 2D UI workflow. I instructed them to continue working on their own projects as they normally do, while I recorded their actions with a video camera pointed at their computer. I left the room while recording, to avoid influencing the participants’ behavior. The participants worked on their own computers which featured a mouse and keyboard as the only input devices. Both of the participants engaged in the editing of polygonal models.

In order to gain quantitative data from the recordings I performed a time analysis for which I categorized the type of task they were engaged in at any moment in time. To facilitate this analysis, I developed a small software tool that allowed defining chunks of time as belonging to a certain category in a drag-and-drop fashion. The tool was developed using Processing [1] and can be downloaded from github [2].

Analyzing the artist’s 2D workflow, I found several interesting actions in which both the amateur and the professional substantially engaged. One was the frequent

---

[1] https://www.processing.org/

[2] https://github.com/max-krichenbauer/VideoAnalyzer

and rapid change of viewpoint (camera position) on a very small scale. This camera motion did not allow the artist to see a different object or a previously hidden side of the object, and often ended very close to the original position. This behavior—where the artist returns to the original viewpoint and continues working on the same part of the object or scene as before—makes up about 42% of all viewpoint changes (3D scene navigation) and about 8% of total work time on average. The intended purpose of these actions seems to be to gain a better spatial understanding of the virtual 3D object. Since the monitor is monoscopic and no parallax effect from head motion is available, the visible image is ambiguous in its depth and 3D shape. A slight "wiggle" of the virtual camera produces a parallax motion that gives the artist a better understanding of the virtual scene. This finding reflected previous findings from the survey in the previous section, where the participants reported that they used the camera controls of their 3D software extensively or even constantly.

In comparison, the artists spent between 4% (Professional) and 15% (Amateur) of the work time looking at reference material. The largest share of time was spent on 3D object transformations—between 47% (Professional) and 49% (Amateur).

3.4. Conclusions

In this chapter, I presented the results of my research efforts to provide a better understanding of current-day professional 3D design work. I performed both a wide-range survey and a detailed analysis.

As I originally expected, there is room for improvements in the current UIs. Q25 (Figure 3.5(b)b) revealed that more than a third of the artist perceived their current UI devices as insufficient in some way, and in Q28 (Figure 3.6(a)b) showed that more than half of artists experience tried or even painful hands from their wrist. As was shown in Q24 (Figure 3.5(a)a) only a minority had previously experienced 3D user interface devices, so a lack of exposure may be a contributing factor.

At any rate, the information gathered from this survey allowed me to derive the requirements for a 3D UI that would meet the artists' demands. I present this work in the following chapter.

While I focused on media production in general, both a look at different fields such as CAD for engineering and a more in-depth analysis of sub-fields such as video games may produce different, equally interesting results. This, however, is beyond the scope of this work.

Different software has different commands, which often makes it unfriendly to switch between software.
Why only negative options? I don't mind the 2D UI.
It's fast, responsive, and precise. That's what is most important.
Using specialized navigation tools offers little improvement over navigating like a FPS or with camera targets, but adds a separate tool which means I would have to move one of my hands off of the keyboard or mouse.
They do the job just fine.
Appropriate but could be enhanced.
I have no problems using a 2D interface for 3D work.
Mouse & keyboard combo works really well with an interface that knows and uses their strengths.
No problems working in 3D space with current interfaces.
Work as expected.
I use a 3D monitor and mouse and feel that it is sufficient for creating 3D models.
Not sure.
They work fine, really.
Currently common user interfaces do a good job about 80 to 90% of the time. At specific use cases additional interfaces could close a gap.
No problems here.
It's fine.
Perfect.

Table 3.2. Comments on Q25.

Just a mouse pad with cushion.
No, but i should.
I exercise with two 40 pound weights.
I’m using a standing desk at home to encourage better posture.
No special measures taken.
Quick break.
Being careful.
I work out.
Stretches.
I go rock-climbing.
Transitioning to using pen tablet as main mouse.
I am still young and fit.
I set my mouse to really sensitive so I don’t need to move my hands as much.
Improved desk and chair height and positioning.

Table 3.3. Comments on Q29.

CHAPTER 4

Conception and Formative Evaluation of a 3D User Interface

4.1. Introduction

Based on the insights into the current work methods and environment I gained from the work described in the previous chapter, I derived the concept for a 3D UI that could meet the demands for professional adoption.

I started by identifying the key requirements that come from the work and then tried to find concepts that could be used to address each of these requirements.

Throughout the process, I stayed in touch with 3D media professionals and eventually invited three of them to test the prototype user interface in a pilot study.

After improving and formalizing my approach I performed a larger scale formative user study designed to test my assumptions and find problematic issues with the current approach.

These steps are described and discussed in detail in this chapter. The requirements analysis (Section 4.2), UI design (Section 4.3), and pilot user study (Section 4.4) were previously published in my Master's thesis. The formative user study (Section 4.5) was conducted later during my PhD research.

4.2. Requirements Engineering

This section describes the specific requirements that I deduced from my analysis of the work and work environment of 3D design professionals. They were published as part of a paper at the International Symposium on Mixed and Augmented Reality (ISMAR) in 2014 [47] and subsequently in my Master's thesis.

In this chapter, I present information on specific requirements for the professional 3D content creation workflow. These are: an ergonomic design that can be used comfortably for a long period of time, support for collaboration among artists and supervisors, a high amount of features as they are commonly used in 3D design, support for navigating the content, support for 2D and alphanumeric operations, and an increase in productivity over the traditional 2D UI. The sections are in no particular order, as all points are independent requirements that cannot be weighed easily.

4.2.1 Ergonomic design

As the survey has shown, 3D artists tend to work for several hours before taking a break and spend a lot of time working with their software every day. Even with the quite comfortable and sophisticated traditional UIs, this becomes an issue for the artists' health and well-being. It is therefore important to provide UIs that are comfortable to use over an extended time without strain or tiredness. Previous publications only provided quick user tests and thus never discussed problems encountered during extended use, which are vital for professional application. For example, some researchers suggested systems where the users interact while standing [38, 58]. While this may be comfortable or even healthy for a short time period, it is unlikely to be adopted for extended use. Furthermore, the device used for viewing the virtual or augmented environment must be considered. Using a tablet or other mobile device requires the artists to move it around whenever the viewing position must be changed. As the survey has shown, this is basically constantly the case. Hence a non-intrusive viewing device like a monitor or HMD is more appropriate.

4.2.2 Collaboration support

Since professional production is not an individual effort of a single artist, some kind of collaboration is obviously necessary. According to the survey and my own experience, many types of collaboration are commonly performed and necessary for producing high-quality results.

In the most basic way, this consists of showing the content to some another person while the user is simultaneously being able to perform changes. The traditional use case being that the supervisor or a colleague visits the artist at his work desk, and comments on the current state of the work. Traditional 2D UIs support this naturally because the monitor screen is visible to everyone, and via pointing and talking, feedback can be communicated. Novel UIs, however, must be designed with this requirement in mind. This requires some way of bidirectional information transfer: other people than the artists must be able to perceive the virtual content, and must be able to communicate with the artists about the content. AR has an inherent advantage over Virtual Reality (VR) in this respect because all users in a shared augmented environment can see each other without the need for virtual representation. Hence pointing at specific parts of the content while talking with each other directly is possible. For VR and remote collaboration, similar communication methods must be provided.

However, reviews are not the only type of collaboration used or wished for by artists. While traditional PC UIs are inherently single-user (even when several mouses and keyboards are connected, the single focus makes real collaboration impossible), some types of simultaneous work through quick alteration, file versioning or separation of tasks are also common or at least estimated to be useful by most artists. Novel systems can break with the traditional single-user paradigm and offer true real-time multi-user collaboration either locally or remote via network.

4.2.3 High amount of features

While in industrial applications precision and correctness is usually the greatest challenge [66], media production challenges us with an intense amount of functionality. Because there is little requirement of technical correctness and a huge pressure to produce more and more stunning expressive imagery, the accumulated feature set by far exceeds any existing AR or VR application.

These functions include tools like translation or rotation, actions like extruding or collapsing a polygon, operational modes like interacting with single vertices or complex group hierarchies of objects, mathematical operations like the boolean difference of two bodies, computer graphics features like different types of light sources, materials or cameras, or support for industry standards like automation APIs, scripting or file types. As the survey illustrates, these features are actually actively used by the artists, and that fast input methods like hot-keys are a necessity. Previous prototypes for AR UIs commonly only implement some basic tools like transforming primitive shapes. Even more problematic: since most artists extend their feature set with additional scripts and plug-ins (sometimes even exclusively tailored to the artists' requirements) it is impossible to even list the required features.

While the technical challenge seems daunting, it is actually easily solved: instead of trying to implement all required functionality, one can easily build on existing software either by integrating open-source code or by offering the novel UIs in the form of a plug-in to existing software.

The real challenge lies in the UI design: no previous system known to me attempted to provide access to a similar amount of functionality while at the same time staying flexible and intuitive. Novel approaches are required to organize complexity where menu lists and keyboard hotkeys become unfeasible.

4.2.4 Fast and intuitive navigation

Another important factor that stems from the complexity of the content processed in professional production is the need to navigate through the content. In this context I use the term *navigation* in its traditional meaning of changing the relative position of the user's viewpoint and the virtual environment or content without altering the content (editing). In immersive VR environments, the term and reasoning are intuitive to understand because the user has a sense of "moving" through the environment. AR applications typically don't incorporate any navigation other than physical (by moving the camera or moving a fiducial marker) because since the user can still see the real environment, there would be no sense of personal motion. Transitional UIs that mix both VR and AR such as the MagicBook [9] usually provide some means of transitioning between VR and AR, but no means of navigating AR other than physical.

However, in digital media production, the content is too big and complex for physical navigation to be feasible. Some scenes may incorporate whole cities, where the artist is working on figures in individual streets while trying to keep an overview of the complete scene. The survey confirmed this: no one stated that they can perform their work without any form of zooming or navigation – on the contrary, most artists navigate their content often or constantly during work.

This means that applications have to provide an intuitive and quick way of allowing user navigation. Especially "zooming" operation where the relative scale of the content to display is changed has previously been neglected in AR UI design.

4.2.5 Support for 2D and alphanumeric operations

3D design consists of a greater number of sub-tasks, not all of which are 3-dimensional in themselves. Some tasks are inherently 2-dimensional in nature, even though they relate to 3D CG. An example for this is texture mapping, where the "UV" coordinates of vertices of polygonal models must be laid out on a 2D map, trying to achieve a balance between local distortions in the texture, seams in texture mapping, limited space on the 2D map and ease-of-use for texturing artists.

Another problem that is difficult to realize in 3D UIs is the need for alphanumerical input. This is often required in 3D design in the form of file names or inputting exact parameter numbers. Moreover, in rigging, it is often required to perform some minor scripting on the 3D content, such as defining a relationship between several objects to implement some mechanics.

One way to offer alphanumeric input is via voice input over a microphone, as proposed by Girbacia et al. [31]. However, a talking-out-loud approach is not very suited for a shared workplace environment. Using throat microphones might alleviate the problem at the expense of increased cost, decreased comfort and quite likely a loss in reliability. The survey reflects that: 21.2% of the participating artists disagree to any usefulness of voice input, and only 36.4% see it as a possible replacement for traditional mouse-and-keyboard user input.

Therefore, in these cases, providing a mouse-and-keyboard option may be advantageous especially for AR applications where the user can still see the real environment.

4.2.6 Increase in productivity

The final point may be surprising: one of the greatest challenges is to design a UI that actually matches or even outperforms traditional 2D UIs in terms of effectiveness such as speed of operations. It is easy to believe that when employing 3D UIs, the spatial understanding will in itself increase ease of use and thus user performance. As the survey shows, artists are open to alternative 3D input devices, but critical in their adoption.

In my user tests, however, I had to discover that this is not the case: spatial understanding is limited, even when stereo vision HMDs are used, and the limited advantage of some depth perception is hardly enough to encourage switching the system.

The artist have developed the skills to work with 2D UIs with high speed and precision. Even though 3D software often requires the artist to click at objects that are only a few pixels in size, experienced artists are able to perform their work very quick, as I could see in the training videos. Since time is an important factor, the pressure for fast and efficient UIs can easily be understood. The survey has also shown the limited hope of artists that alternative UI concepts can live up the speed and precision of working with a traditional 2D mouse.

Some of this gap can be attributed to the maturity of input devices, as the first mice also suffered from rather slow and unreliable tracking until modern LASER-mice made the interaction much faster and more precise. However, even the first mice had to provide a sufficient advantage over purely keyboard based interaction, else they would not have been developed further. In the same manner, AR / VR UIs must provide an advantage now in spite of their limited maturity, in order to justify and promote further development.

4.3. User Interface Design

Based on this the requirements identified in the previous section I developed a concept for a 3D UI to meet the demands of professional 3D design work. An illustration of the concept system can be seen in Figure 4.1. In this system, the artist is comfortably seated at his work desk, wearing a stereo video-see-through head-mounted display (HMD), which displays the virtual content in the work area. The artist's hands are tracked to allow direct manipulation within arms reach. In my prototype system, I

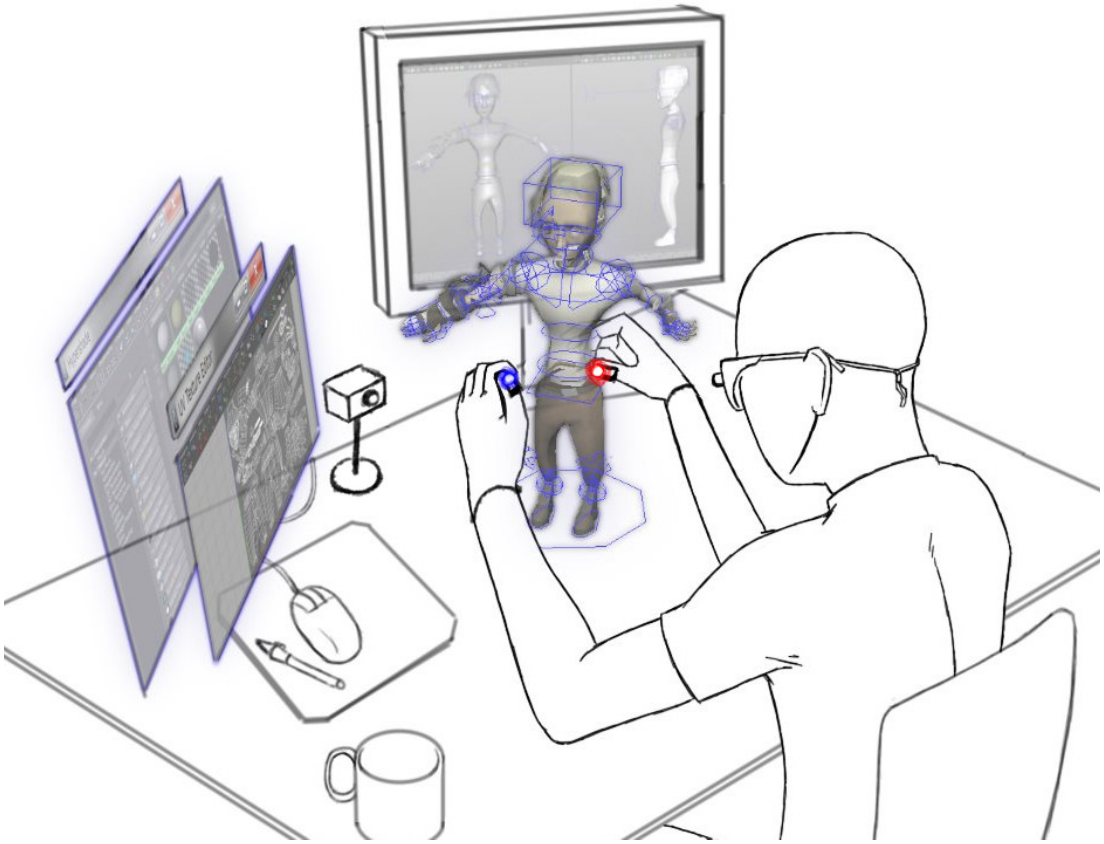


Figure 4.1. Illustration of the proposed system. The user is wearing an HMD with front-facing cameras. These are used both for AR (VST) and to track the position of the thumbs via attached color LEDs. Thumb rotation is tracked via IMUs attached to the back of the thumb. The HMD position is tracked via an external tracking system (placed on the table). The user is wearing pinch gloves to facilitate detecting hand gestures. These are connected wirelessly to the computer via Bluetooth connection.

used pinch gloves to which I attached color LEDs to facilitate positional tracking, as well as IMUs for rotational tracking. Data was transmitted wirelessly via Bluetooth. The HMD was an Oculus Rift DK2 and therefore tracked via a single external tracking camera.

The 3D UI should be implemented as a plug-in or extension of the respective software package that the artist is used to, similar to ARpm [27]. This allows seamless transition of work to the new 3D UI and back in case this is required while providing all the features that the artist is used to and requires. It further allows production to continue with the current production pipeline, including the use of proprietary file types or custom scripts.

4.3.1 Ergonomic direct hand interaction concept

I decided to use gloves as input devices in order to allow bimanual direct hand interaction within arms reach as the main input vector. This is the most natural and versatile way of interaction and closest to the traditional style of working with real models for stop-motion animation. While it is often criticized that 3D interaction is more tiring and less precise than traditional 2D mouse interaction, the higher DOF can make it more efficient for 3D interaction [53]. Furthermore, it is possible that the requirement to move one's arms may result in health benefits in the long term due to utilizing the muscles more naturally, especially since the current 2D UI is known to be a common cause of health issues itself [55]. Therefore, I choose hand interaction for my research prototype UI.

The thumbs of both hands are the frame of reference to a pinch-based interaction concept as proposed by Piekarski [58], because all natural pinch gestures are performed with the thumb and it is more stable during the pinch gesture than the finger. This turns the thumbs into the cursors of the UI, allowing a natural "grabbing" interaction metaphor, similar to grabbing real objects. Multi-finger interaction is not supported, as even in the case of stop-motion-like animation only one joint is manipulated at any time. This also makes it unnecessary to track all fingers individually or the whole hand shape, resulting in easier implementation. This is similar to the Tinmith-hand system [59, 60] but optimized for within-arms-reach use. The actual cursor is visualized by the UI as an arrow on top of the thumb for precision (just using the visual tip of the thumb is not precise enough to select single vertices of a 3D model) and easier understanding.

It is implemented to lie a little ahead of the tip of the thumb, so that the immediate area around the cursor is not occluded by the thumb itself.

The key interaction functionality (editing 3D content, using 2D elements and navigation of the scene) is implemented on both hands in the same way. This not only allows both right-handed and left-handed users to use the system with one hand, but also allow simultaneous bimanual interaction. Shaw et al. [65] showed the feasibility of such an approach.

As shown in the previous section, 3D design is a complex task requiring a large number of features to be readily available. In order to create an efficient UI, I implemented eight buttons on the glove with two complementary contact areas to trigger them: the tip of the thumb and the palm of the hand. Glove based systems have previously been shown to support a great number of buttons comfortably [24], so I considered this an additional advantage of gloves over hand-held pointing devices. However, it is important that the UI design does not require the artist to perform any complex hand gestures that may become tiresome or lead to false detection of gestures (“false positives”). Therefore, the eight buttons all worked through a simple touch contact with the tip of the thumb or palm of the hand. Double-clicking was tested but finally not used because it turned out to be too difficult to perform the gesture. The thumb can most easily and comfortable touch any area on the inner side of the fingers, and the ring and little finger can comfortably touch the palm when the hand is closed. Thus, this system allows simultaneous triggering of several combinations of buttons comfortably. To utilize this potential, I implemented the tip of the little finger to assume a function similar to an *Alt* key on a keyboard. While the *Alt* key is suppressed, the functions of the other keys is exchanged with an alternative function. The *Alt* button of each hand affect not only this hand, but the other hand as well, similar to a keyboard where the left and right *Alt* key have the same global effect on the keyboard. Thus, the number of functions provided through the remaining buttons rises to 14 per hand. However, the most common three functions of 3D interaction, 2D window operation, and navigation on both hands are mirrored on both hands to allow bimanual operation within each of these three functions. The other combinations were mapped to various menus or interaction methods. Figure 4.2 shows which functions and menus are mapped to which areas on the hand, in the case where the right hand is the primary (dominant) hand and the left hand is the secondary (non-dominant) hand. Handedness can be switched in

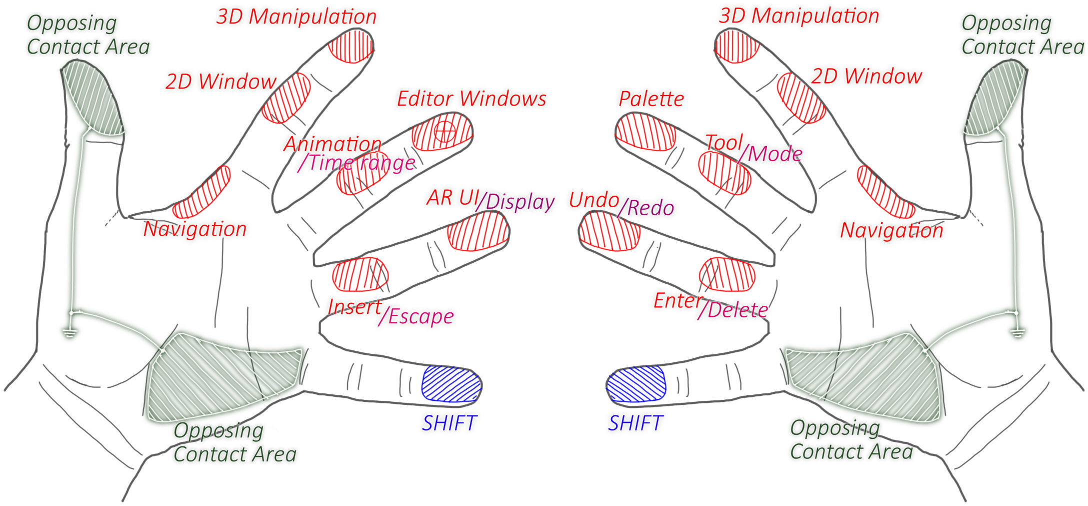


Figure 4.2. Mapping of functions to the hand. Due to the *Alt* key, the number of functions amounts to 14 per hand, which is similar to the number of hotkeys that artists commonly memorize on a keyboard (see Chapter 3).

the prototype at any time.

4.3.2 Collaboration

Novel 3D UIs can allow collaboration on two levels: interactive reviews and simultaneous interaction. In the more simple case of interactive reviews, a second user is able to view the virtual content and comment on it, without being able to edit it himself. In my system, I implemented cloning the video stream of the artists HMD (the dominant eye's view only) to the desktop in order to allow additional users to observe the virtual content.

Simultaneous interaction is much harder to achieve both technically and from UI standpoint. Shen et al. [66] and Raymaekers et al. [62] explored principles for remote collaboration such as concepts for floor control. A local collaboration approach promises even more efficiency as the participants can see each other as they collaborate, which increases the natural communication especially in face-to-face opposing workspaces [43]. This however conflicts with the requirement for simple navigation. If several artists share the same global coordinate system, they would interfere with each others' work when they change it. Kiyokawa's proposed separation between personal [43] and public spaces can be applied here, where separated coordinate systems can be offered to every user as well as a global coordinate system for exchange. For my own system, I omitted simultaneous interaction in order to keep the design, development, and experimentation work manageable.

4.3.3 Menus for managing a high amount of tools and operations.

Professional modeling software like Maya provides a great number of functions and tools. 2D UIs are at a significant advantage here because of the WIMPS metaphor and simple mouse interaction. However, in order to provide the advantages of 3D UIs such as spatial interaction and high DOF manipulation seamlessly It is necessary to provide these tools in a menu structure that can comfortably hold a large number of entries. Upon analyzing the natural range of motion of the hand, I decided on a palette-shaped marking menu (see Figure 4.3). When the user holds down the "palette" key, the menu appears at his hand and he can drag the pointer over the icon of the tool he wants to execute. Marking Menus[50] have a long history and have proved their usefulness in many settings. However, for this hand-based 3D UI I found that the user can reach the entries more easily in a palette than in a radial layout, especially while resting the elbows on the table. I also experimented with 3D arrangements but found it too confusing and finally decided to use a flat arrangement, adding two icons to the menu to switch through multiple levels of tools. The tools and icons were created by automatically parsing Maya's "Shelf" menu which holds most of the tools that Maya offers, as well as custom features that users imported in the form of scripts and plug-ins.

4.3.4 2D Editor Windows

Some of the functionality provided by 3D modeling softwares lies in additional editor windows that allow non-3D tasks such as texturing or defining software settings. In order to offer these to the user in a 3D UI, I implemented the functionality that all additional editor windows also automatically appear as virtual floating 2D menus in the virtual environment as soon as they are opened.

Using 2D windows in a 3D UI is notoriously clumsy, mostly because of the spatial

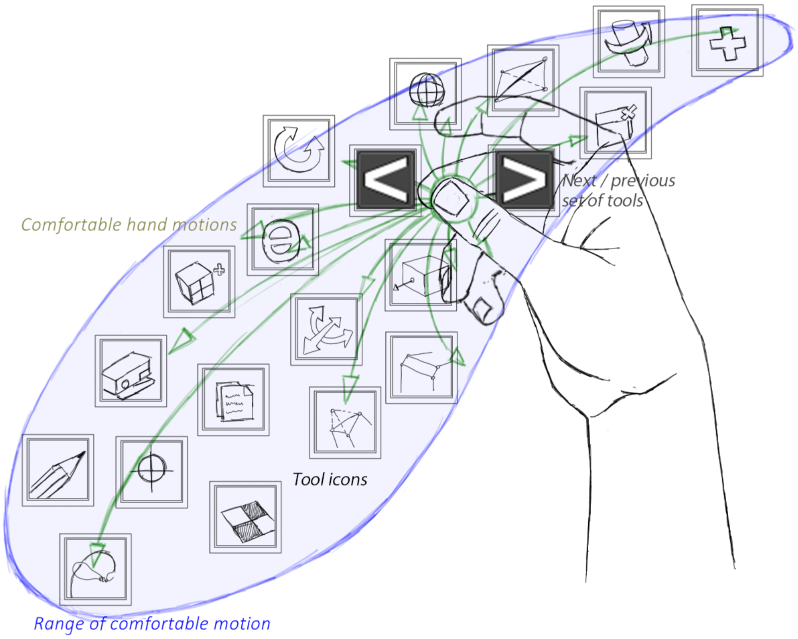


Figure 4.3. Tool menu concept.

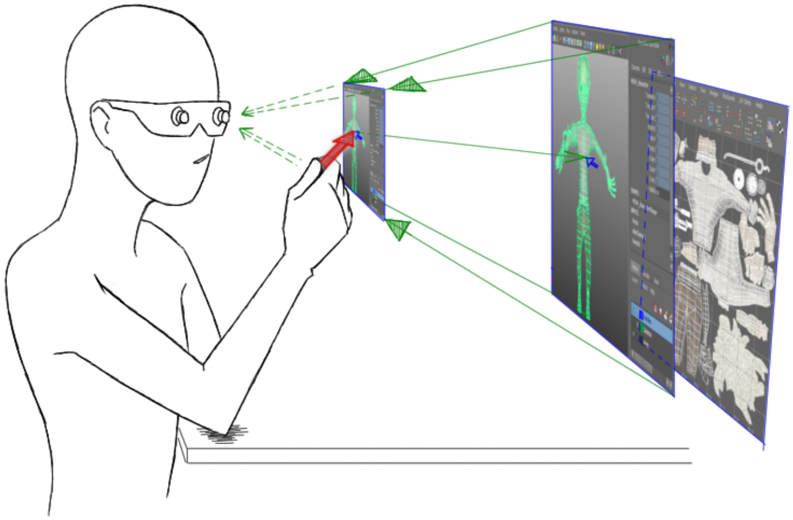


Figure 4.4. 2D window interaction. Inactive windows face the user at a background distance beyond arms reach. Ray casting from the dominant eye over the cursor is used for interaction. When a window becomes active, it is moved forward, changing its size accordingly so that the optical impression will not change.

discrepancy between the cursor, the window surface, and the eye positions in stereo viewing. In order to make the interaction as convenient as possible, I followed the approach outlined below. The cursor is at a 3D position, usually in front of the floating window frame, and its corresponding 2D position within the window is determined by ray casting. In stereo vision, the user sees with two eyes, which makes the resulting 2D position ambiguous. Even when the dominant eye of the user is known, it remains somewhat confusing, and prior research has shown that ocular dominance is not as strongly determined as commonly believed [6, 42]. Therefore, the window distance is interactively updated to match the position of the hand in order to eliminate the discrepancy. Thus the user does not have to try and match the position of the floating window, the floating window instead matches the position of the hand, similar to the old adage “If the mountain won’t come to Mohammed, Mohammed must go to the mountain”. This is akin to the “ Scaled-world grab” described by Mine et al. [54]. See Figure 4.4 for an illustration of the process. When inactive, the window is displayed on a background plane beyond arms reach. When it is interacted with, it is moved towards the user’s eye to match the distance of the user’s hand. Since a “jump forward” would be startling the user, the scale factor of the window is updated accordingly to keep the same visual size. Thus, the dominant eye will not notice any motion, while the binocular disparity from the other eye is removed. This is automatically updated throughout the interaction to compensate of the user’s inability to restrict hand motions to one single plane in space. When the interaction with the window ends, it is sent to the background, again updating the scale factor to keep its apparent size constant.

4.3.5 Navigation

Previous prototypes mostly relied on physical motion of the head and body to allow navigating virtual content. But in the case of artists working on whole buildings or even models of whole cities, this approach is insufficient. Therefore, I implemented a “navigation” button on the top of the hand that offers a *“grabbing-the-air”* navigation metaphor [15], similar to the one used by Mapes and Moshell [53], but with added rotational control in three dimensions. When one hand is used, the world follows the rotation of the hand. When both hands were used, rotation followed the position of both hands, as if the user was holding the table surface. See Figure 4.5 for an illustration of the navigation metaphor. This technique to be intuitive not only in use cases environment where the user is engulfed in virtual content—where it gives the impression of pulling oneself through a zero-gravity environment—but also when working on single object—where it appears to the user as if he would hold the object in their hands. Both transformations are mathematically identical. Another advantage of this approach is that it naturally realizes a "frame-of-reference" concept for hand interaction [15], where the user can hold an object—or rather the virtual worlds origin—in the non-dominant hand while still being able to edit the content with the dominant hand. Zooming in on certain parts of the 3D model also becomes intuitive using bimanual interaction. By mapping the distance between the hands to a scale factor, the users can "stretch out" any area of interest until it fills the entire field of view. The most intuitive way to implement this is to calculate the scale factor such that the hands both stay at the exact same place in the virtual scene. Thus, when moving the hands apart to double the distance between them, the virtual world must also be displayed at twice the size so that the hands again reside at the same virtual spot in space. In the case of one-handed interaction, the *Alt* button can be pressed, which switches the single-hand navigation to a zooming operation instead.

4.3.6 Improved manipulation efficiency by utilizing the hands DOF

In order to bring about a measurable boost in user performance, I focused on the one of the main advantages of 3DUIs, which is the increased degrees of freedom (DOF) offered intuitively by the input devices. Traditionally the mouse can only offer 2 DOF. A 3D input device usually offers 6 DOF—3 DOF of translation (forward, sideways, and upwards) and 3 DOF of rotation (yaw, pitch, and roll). Even when the used input device is designed to provide a high number of degrees of freedom, the actual interaction is limited to quasi-simultaneous operations where the user switches between he offered degrees of freedom such as in the systems proposed by Fröhlich [29] and Simon [67]. For the complete set of 3D transformations, the 3 DOF of scale (width, height, and depth) are missing. By using both hands for direct in-reach interaction, it is possible to offer object manipulation with truly simultaneous bimanual 9DOF manipulation including 3DOF translation (x, y, z), 3DOF rotation (yaw, pitch, roll), and 3DOF scaling (width, height, depth).

Starting from the more simple use case, such a tool offers 6DOF (translation and rotation) by using one hand and simply dragging the selected objects or components

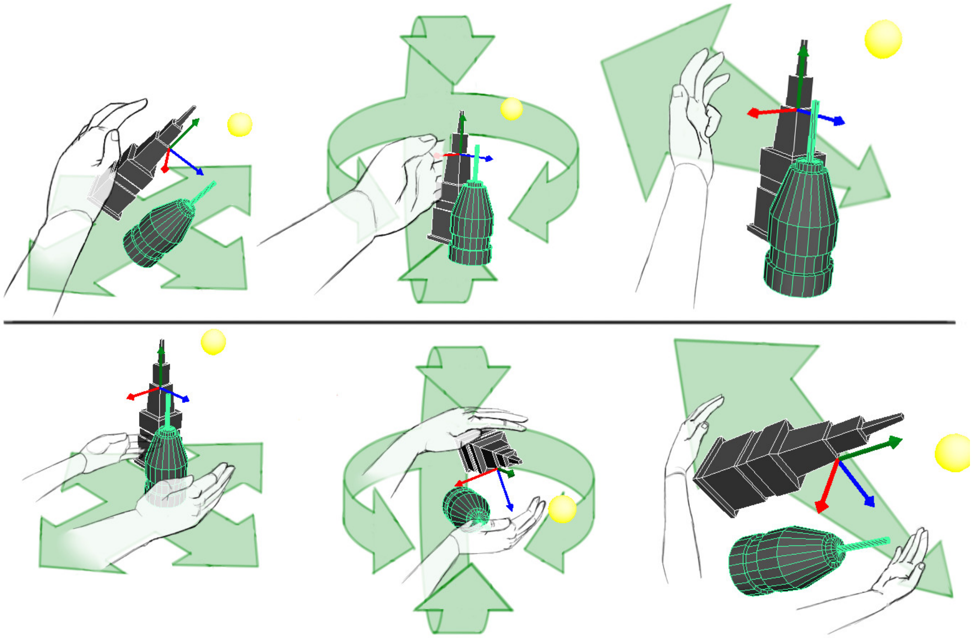


Figure 4.5. Navigation concept: the world origin is transformed to give the illusion of positioning the objects without editing. *Top:* Single handed navigation. “Zooming” is applied when the *Alt* key is held (right). *Bottom:* Bi-manual navigation. The pivot is the mid-point of both hands, and “zooming” is controlled by the distance of the hands.

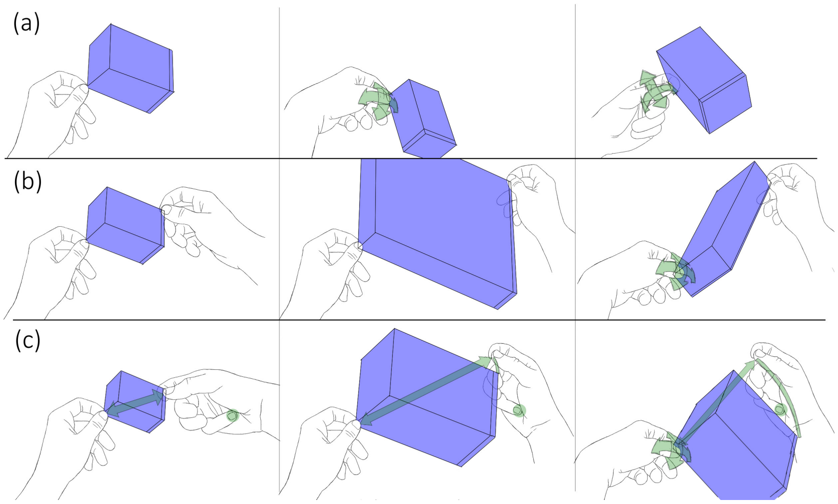


Figure 4.6. Utilizing the hands DOF for efficient and intuitive manipulation (editing) of objects. *Top:* A single hand offers intuitive 6DOF. *Middle:* The position of a secondary hand controls 3DOF of scaling. *Bottom:* When the *Alt* key is held, scaling proportions are preserved.

along. When a second hand is used, the tool offers all 9DOF at the same time by using the first hand (the one that started the interaction and therefore controls the previously described 6DOF of translation and rotation) as a frame of reference, and making the second hand control scaling in 3 dimensions. It is intuitive to understand when observing the case of grabbing two opposite extremes of an object: it can now be stretched or squashed by moving the pointers further away or closer to each other in each direction of the object's coordinate system. The underlying concept that makes the interaction intuitive is that the pointers always stay on the same place on the object, just like a real object would deform to follow the motion of hands pulling it apart. Figure 4.6 shows the usage of 9DOF bi-manual manipulation.

Since it is sometimes undesirable to control the three dimensions of scaling separately, if is possible to hold the *Alt* key on the secondary hand, thus locking the scaling. The mere distance of the pointers now defines the scaling in all three dimensions together.

A common problem with this approach is that users when they only want to change one or two dimensions of scaling will grab two points close to one plane. Therefore it is necessary to detect when both interaction pointers were grabbed along one plane of the object and then ignore the normal to this plane in the scale calculations. I found a simple threshold to be sufficient for this.

Of course, offering the artists 9 degrees of freedom in manipulation does not remove the requirement for the traditional limited tools. One reason being that sometimes only manipulation in a single dimension is preferred, and further degrees of freedom only worsen an otherwise precise result. Shaw [65] explored the stages of modeling from coarse placement—which is best done with a high number of DOF—and subsequent refinement, which is more feasible when using some form of restraints, limiting the degrees of freedom. Teather and Stuerzlinger provides guidelines for 3D positioning techniques regarding limiting the degrees of freedom of manipulated objects [70]. Another reason is that since the motion and rotation of the hand are partly interleaved: we move our fingers by extension or supination of the wrist, thus additionally rotating the hand intuitively. It can be somewhat tedious to move the hand in a correct way without adding involuntary rotations. Therefore, the usual tools for moving, orienting, or resizing objects provided by Maya are also available in the 3D UI. However, they now allow three DOF operation instead of the two DOF mouse, and both hands can be used simultaneously to increase precision and stability.

4.3.7 Improved animation time control.

Some specific tasks usually offered in the form of windows can be approached in a more optimized manner in 3D UIs. One such concept that is important to animation is the control of time. In traditional 2D animation, the artist commonly “flips” a small number of images. In current 3D animation, such a feature is commonly not implemented, and a window element is used to set the current time with the mouse or replay a short playback interval. With an “animation” key on the non-dominant hand, it is possible to offer the artist the ability to “flip” through the animation with a simple motion. Since the primary hand can still be used to manipulate objects simultaneously, the artist is able to create coarse animations very fast (see Figure 4.7). Walther-Franks et al. proposed a similar system for multi-touch surfaces [72]. In order to make the motion more ergonomic, rotation around wrist or elbow is used as an input variable to control time since it’s the most convenient motion to be performed. Playback over the whole time span can be started and stopped by clicking this “animation” button.

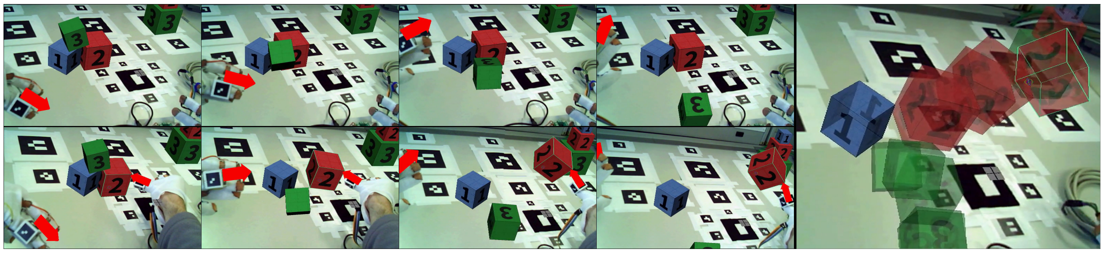


Figure 4.7. The user's view through the HMD while animating. *Top row:* The left hand controls the time. *Bottom row:* The right hand can be used to simultaneously manipulate objects, creating rough animation keyframes very fast. *Right:* Resulting animation.

4.4. Pilot User Study

4.4.1 Prototype

In order to conduct tests and further validate these concepts, I developed a prototype 3D UI. I decided to use Autodesk Maya as a basis for the implementation because it is one of the most commonly used 3D Design applications used in film production (both animation and visual effects) and games. This automatically helped to keep the UI closely related to professional production needs. For example, the vast feature set that Maya offers its users quickly made it clear that just implementing a few features by myself would not be sufficient, but that the UI had to be able to allow generalized access to the Maya toolbox. I developed the UI in the form of a plug-in which made it technically possible to access Maya's internal functionality.

I used Vuzix Wrap 1200 VST-HMD and made the plug-in render the scene in stereo over the images read from the front-facing camera. Since no depth information was provided with the color images, the virtual objects always appeared in front of any real object. Correct occlusion was not provided. I implemented HMD tracking through computer vision by using the AR-Toolkit library, which provides tracking of fiducial markers.

Looking for suitable input devices, I was not able to find gloves that were both commercially available and offered the features required to implement the planned UI. Therefore, I decided to create prototype pinch gloves myself. I used normal thin cotton gloves as they are often used when handling fragile objects and sewed conductive materials to them in order to register pinch gestures as closed electric circuits. Early versions were connected to the PC via the USB interface provided by a Tinkerblocks microcontroller. However, the cables turned out to be cumbersome. So I switched to an Arduino microcontroller with a BlueTooth modem that could transmit the pinch information wirelessly.

A first version of the gloves used fiducial markers attached to the thumbs for hand-tracking. However, I soon discovered that this was not practical, as the size of the markers never yielded satisfactory results. Small markers proved to be too imprecise and loss of tracking was too often a problem, while bigger markers inhibited the free motion of the thumb too much to do actual precision work. Therefore, I switched to attaching a single color LED on the thumb (different colors for each hand). This brightly colored light proved much more reliable in tracking and did not inhibit hand motion.

Throughout the development of the prototype, I stayed in contact with 3D artists and other media professionals to receive feedback, which in turn advanced my understanding of the topic as well as prototype development. When the prototype reached a usable stage, I performed a pilot user study with three 3D CG media production professionals. In the following, I first explain details about the tests performed and then present the results.

4.4.2 Procedure

The sample consisted of three participants, two male one female, all currently employed in digital media production, all with several years of experience.

I gave a short introduction to the system before letting the participants use it. The sessions were about 10-15 minutes long and did not contain specific tasks, letting the participants explore the system freely, for which I provided a demonstration 3D scene. Help and explanations were given whenever asked.

I used a Vuzix Wrap 1200AR as HMD for the user test and the first version of the gloves, which still required cable connections and fiducial markers on the thumb. Also, at this stage of development, the tracking of the HMD still required fiducial markers. I provided the prototype system at the respective environment of the participants instead of a controlled lab environment. A laptop computer was used to run the software, so nothing but a table was provided by the participants.

Immediately after the session, I asked each participant to fill out an anonymous questionnaire regarding the experience. In the first part of the questionnaire I asked general questions about the participant's expertise in various areas of 3D design, what 3D software they commonly use and if they had any experience with AR (all had little to no prior experience with AR). A second part contained 25 Likert Scale questions regarding the users' experience with the prototype. In the third part, I asked for the participant's comments, including asking for "the three greatest limitations", "the three greatest advantages" and a general statement on the system itself.

4.4.3 Results

The feedback was generally positive: all participants were interested in the system and the possibilities of AR UIs for 3D design.

While it was insightful to do the test under real-world conditions instead of a controlled lab environment, it also led to problems due to inconsistent lighting and space limitations. Loss of tracking was the most frequent critical comment I received during the test. This comes as no surprise since fast, precise, and reliable interaction is vitally important for productivity.

None of the participants tested all the features of the prototype exhaustively, because the short time frame was just enough to experiment with the most basic functions and get a feeling for the way the system worked.

Still, the participants were eager to give their opinion on the concept, and from the questionnaire, I was able to deduce three key insights.

Importance of depth perception: I found that the stereoscopic depth clues of the stereo HMD I used did not provide sufficient depth perception to allow efficient direct hand interaction, which greatly limited performance. In the Likert Scale questions, all participants agreed that "it was hard to tell where exactly in space the objects were" (one even strongly agreed). An inversely verbalized question to confirm the results received similar reactions. In the open questions regarding the limitations of the system, two participants mentioned the limited depth perception as the number-one disadvantage of the system and the third did so on place three. All participants raised this topic during the test and suggested possible improvements.

This shows that for 3D design work, the requirements for correct and precise depth perception are greater than for merely viewing AR elements, and can currently not be met with off-the-shelf HMDs.

Number and placement of the glove buttons is appropriate: While previous publications already suggested gloved based with a high number of buttons, few ever made user tests with non-scientists. For the user tests, I mirrored the buttons on both hands, removing the additional functionality on the second hand. Still, from test-runs with students from my university, I expected participants to be overwhelmed by the number of buttons to use. This, however, was not the case: all participants agreed that the buttons were intuitive, and not too many nor difficult to remember (two even strongly agreed). One participant mentioned the glove functions as one of the greatest advantages of the system, one participant suggested assigning different functions to each glove as my design intended. This is interesting as it sets a precedent for media UIs. Current mouse-type devices commonly only feature three buttons. It appears that artists are willing to learn a greater number of buttons when these are mapped to areas on their hands directly. This reflects the findings from the survey, that artists are used to learning a lot of hotkeys.

Gloves as input devices are not readily rejected: Considering the inherent problems with gloves and the fact that I used a primitive prototype with wires and metal meshed sewed to cheap cotton gloves by unskilled personnel, I expected participants to oppose the notion of using gloves as input devices. Instead, notions that the glove restricted the freedom of the hand or that it was unhygienic were disagreed on, even though all participants knew that they were not the only people using the gloves. When asked whether they felt the glove would still allow interaction with physical objects such as keyboards, one participant agreed and two disagreed. None of the participants mentioned glove related problems in the comments. It appears that even though totally glove-less user interaction is preferable in the future, using gloves may be a viable solution for the time being.

These results provided initial insights that guided the further development of the prototype UI. However, as a pilot user study, it had several limitations. Apart from the obviously small number of participants, the short use time did not allow for discovery of effects related to long-term use. Furthermore, the participants did not have experience with other 3D UIs to compare the prototype to. Nonetheless, their feedback was helpful in deciding the future direction of development.

4.5. Formative User Study

4.5.1 Experimental Platform

Extending upon the prototype UI described in the previous section, I exchanged the Vuzix HMD with an Oculus Rift DK2 to which I attached an ovrVision stereo camera rig to turn it into a stereo video-see-through (VST) HMD. This allowed the user to see the real environment, including his own arms and legs, while working with the virtual content. The pinch gloves stayed largely the same except for detail improvements. I had also experimented with LeapMotion [1] and different alternative glove technologies, but found that the precise tracking of the thumb in 6 DOF and the high number of pinch areas that my prototype required to identify reliably required a special design that I could not achieve with off-the-shelf components.

During use, the artists were seated in a chair and only operated in a work area within arms reach.

4.5.2 Procedure

In order to test my UI design concept, I performed a formative user study with eleven volunteer participants. Two of them had experience as professional artists, two were amateurs, and seven were design students who had previously taken classes on 3D modeling. The age of participants ranged from 19 to 35 (average 23.36 years). Five of the participants were female, six male. Nine were right handed, two were left-handed. Ten were Japanese, one was Brazilian but living in Japan.

In order to get a better understanding of the level of expertise of my participants with 3D design, I asked them to fill out a form on their previous experience. For a time frame of the past ten years, I asked the participant to estimate the number of hours spend each week working with 3D design software, averaged over the whole year. So if a participant worked on a project for an average time of 3 hours for half a year, then the average time spent over the whole year would be 1.5 hours per week. From this, I calculated an estimate of the total hours of experience by multiplying each annual estimate by 50 weeks per year (assuming that participants tend to neglect holidays). While this is naturally an extremely coarse estimation, it still gives us some informa-

---

[1] http://www.leapmotion.com

Figure 4.8. The virtual scene provided to the participants for use of the evaluation of the UI. The center figure is a fully working animation rig, the right figure just a polygonal model. The top cube on the left side (labeled “3”) was animated to fall down and tumble over the ground.

tion about the likely level of expertise of the participant on a logarithmic scale. Users whose experience is limited to tens of hours can be assumed to be beginners, those with hundreds of hours to be amateurs, those with thousands of hours to have achieved a professional level, and those above 10,000 hours to be experts. This is in accordance with prior findings from different fields of artistic endeavor [26]. my participants reported experience levels between 25 hours (least experienced participant) and 16000 hours (most experienced participant). The average experience level was close to 2500 hours ($SD \approx 5400$ hours).

After filling out this initial questionnaire, I introduced the participants to the concept of the user study and the prototype system. I encouraged them to mention any problems or inconveniences they might encounter and specifically instructed them to halt the experiment if they experienced cyber sickness or felt otherwise uncomfortable. All participants used the system while seated in a chair, and without any obstructive object within arms reach.

All participants were given the same example content, which consisted of two figures (one of them with a working animation rig) and two assemblages of three colored cubes each, one of which was animated to fall down and tumble over the ground level. See Figure 4.8.

In pre-trials, I found that trying to learn how to use the UI before use was difficult, because people had trouble remembering the functions. Therefore I implemented a tutorial that consisted of a virtual slideshow in the AR environment that explained each function in turn. The participants could go forward or backward through the slide show either by telling the experimenter to go a step “forward” or “back”, or by “clicking” on one of two respective buttons on the slide show with a pinch gesture. I did not give participants a time frame to follow for each step, but encouraged them to try out every function at least once. I did not encourage them to move on to the next slide when they decided to spend more time exploring any one function. At the end of the slideshow I displayed a slide explaining that this was the end of the tutorial, but that the participant was welcome to continue exploring the prototype UI on their own.

The use time consisted of 30 minutes because of time limitations. This would have been enough to try out the complete set of functions provided, however, most participants did not complete the full tutorial during that time because they took longer to learn and get used to some functions. However, all participants learned at least the basic operation of the UI including navigation, transforming objects, and using the palette. More advanced topics like animation could not be experienced by all participants during this session. No participant asked to stop the experiment, and only one participant reported slight cyber sickness after ending the tutorial. However, several participants reported various degrees of strain in the arms or hands due to the use of the 3D input glove.

After ending the use session, I asked each participant to fill out another questionnaire, consisting of three parts. In the first part, I asked participants to make a list of usability problems they have encountered, and to rate each problem in terms of gravity on a scale from on to ten, where one is “negligible” and ten is “critical”. I provided nine boxes for writing down problems but explained to the participants that they do not have to write exactly nine items. my participants wrote down between two and nine problems (about 5 items on average). The average gravity rating was about 5.15 (SD $\approx$ 2.33). In the second part, I asked participants to estimate their work performance using the prototype system, as compared to their usual UI. The questionnaire instructed them to state their estimation in form of a percentage and gave four examples from “0% - My work is impossible to perform with this user interface”, to “200% - I could work twice as fast with the new user interface than with my current software”. Participants were asked to give two estimates, one for the current prototype, and one for a hypothetical UI, where all usability problems would be solved. The current prototype UI was rated between 0% and 100% (mean $\approx$ 41%; SD $\approx$ 30%), and a hypothetical future version of the UI was rated between 50% and 200% (mean $\approx$ 123%; SD $\approx$ 53%). Finally, the questionnaire invited participants to write down some additional comments about their experience with the system.

I considered using a standardized questionnaire like task load index (TLX) [33] or System Usability Scale (SUS) [17], but these only evaluate the system as a whole and thus are only meaningful when different UI implementations are compared. Instead, I designed my questionnaire to gather feedback on implementation details.

4.5.3 Results

In order to quantitatively evaluate the feedback I received, I tried to categorize the reported issues. I found that all usability problems belonged to one or more of the following categories: problems related to the **glove** input device or the **HMD**, the **complexity** of the UI, the ability to **learn** how to use the UI, **vision** such clear visibility of objects, **rendering** of virtual objects, **depth perception** and spatial understanding, the **UI** itself (i.e., conscious design decisions I made), or more specifically **selection** and **rotation** of virtual content, **tracking** of the HMD or 3D input glove, and general system **performance**. Some items related to several categories (for example: when tracking was considered bad when rotating objects), and in these cases were added to all categories that they belonged to. See figure 4.9 for details how many items I received for each category. Criticisms related to the input gloves and UI decisions made up for the majority of problems, both in numbers and in the subjective gravity of the problems.

Technical problems. A great part of the feedback I received was not related directly to my prototype design, but instead concerned underlying technical issues such as the HMD or tracking. Six of my participants wrote about problems seeing clearly, two of which theorized that it may be related to their use of glasses, and four of the participants complained about the weight and comfort of the HMD. These problems are likely to be solved in the future by advances in HMD technology, but for the time being it emphasizes the importance of adjusting the HMD to the user for best vision and comfort and not relying on extensive head motions which could be tiresome and

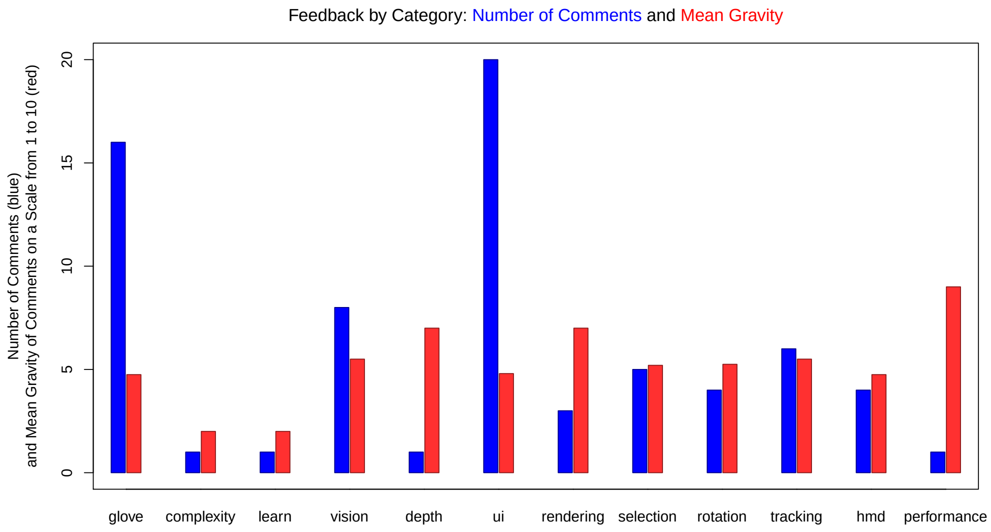


Figure 4.9. Categories of usability issues raised by my participants.

detrimental to vision due to the HMD moving on the users face.

Gloves. One of the most criticized components of my system were the gloves. Five of my participants complained about the lack of tactile feedback when pressing one of the buttons. I had implemented visual feedback where the cursor flashes brightly when a button is pressed and had assumed that this, in conjunction with the physical sensation of touching one's own hand should be sufficient, but discovered that users might prefer "clicking" sensation of a physical button giving way when pressed. However, this might also be a case of habituation to the mouse and might fade with increased use time. Seven of my participants reported difficulties performing certain glove gestures, especially those that involve the *"Alt"* button. The great differences in hand shape and range of motion appear to be a major and lasting problem for employing gloves as UI devices on a larger scale. Only one participant — the most inexperienced — complained about the sheer number of buttons on the glove being hard to remember, which gives a testimony of the UI complexity that 3D artists are used to and require for their work. None of the participants criticized the concept of an *Alt* key on the glove to increase the number of possible button combinations, indicating that it is easy enough to understand and that the hand gesture was not too hard to perform.

Selecting objects. Three participants wrote that it was at times difficult to see which object was currently selected. Out prototype used the normal Maya rendering, which highlights the selected object by rendering the “wireframe” of the polygons in a brighter color with one single pixel width. The Oculus Rift DK2 HMD has a resolution of 960 × 1080 per eye which makes it easy to see single pixels. This suggests that distortion or incorrect fitting of the HMD on the head was the source of the problem. Still, 3D design software might consider novel ways of highlighting selected objects that are more suited for AR and VR UIs. One participant also suggested using novel methods for selecting objects rather than implementing the “click on it” metaphor used in 2D UIs.

Object rotation. Three participants reported problems with the “rotation” tool. The problem appears to be that there are universally two approaches to implementing rotation in 3D UIs. One way is to make the orientation of the object follow the position of the input device, similar to rotating a bolt via a monkey wrench. The other is to make the object imitate the rotation performed on the input device, similar to turning a screwdriver. Since both operations are natural, users easily become confused when first using the tool and expecting a different behavior. This is because it is unnatural if not impossible for the human hand to perform a rotation without change of position or vice versa and the system therefore receives both inputs, translational and rotational. However, it is likely that users will eventually understand how the system interprets their motions and will not be confused anymore.

Strain and fatigue. A more persistent and problematic issue that is unlikely to be solved by technological advances is the strain and exhaustion of using 3D UIs. Traditional 2D UIs allow resting one’s hands on a table surface, making work comfortable even when working for longer periods of time. In my user study, participants worked holding their arms in free space without a supporting surface. Of my participants only the most experienced artist commented on the issue of tired arms explicitly in the questionnaire. Further research is necessary on whether extended periods of work in 3D UIs causes discomfort or even medical problems, or on the other side alleviates the common problems of 2D UIs such as muscle deterioration and carpal tunnel syndrome.

Expected productivity. Regarding the estimated productivity when using the system compared to their current UI, I expected to see a correlation between user expertise and estimation, because beginners and amateurs may be more easily overwhelmed by a large number of functions, while professional artists might require more functions to be able to finish their work. On the opposite side, for professional artists speed of

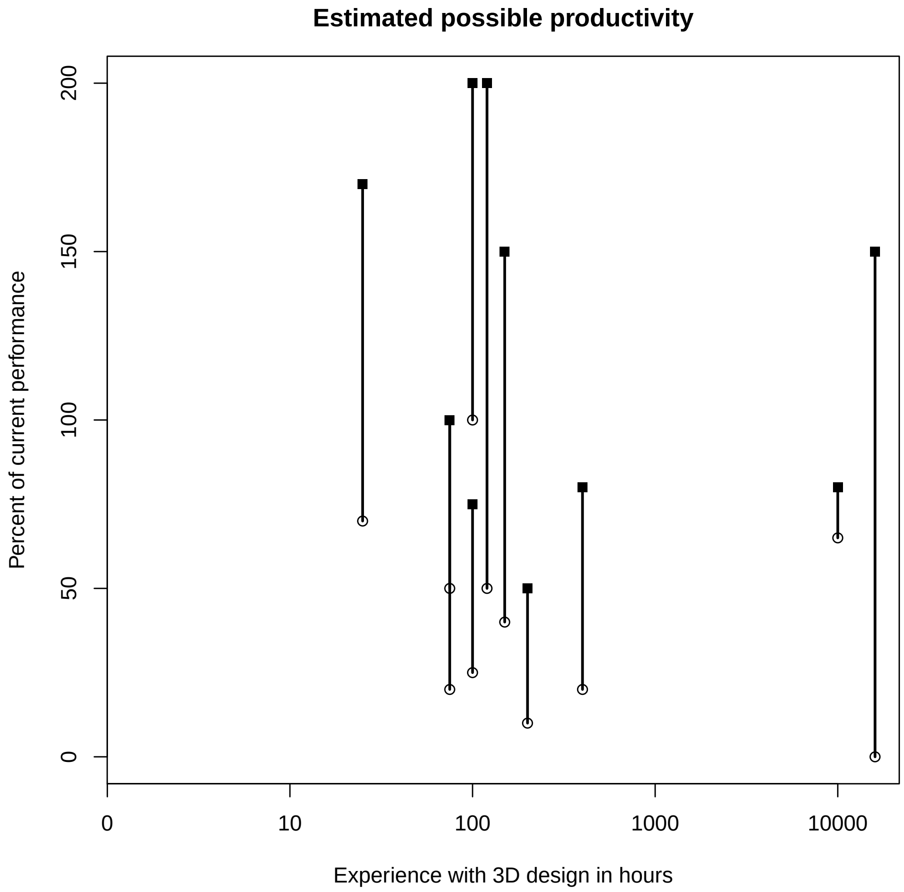


Figure 4.10. Estimated productivity of the UI for the current prototype (white circle) and an idealized future version (black square).

interaction is more critical for productivity than it is for amateurs. Therefore I expected to see a greater difference between the estimates of the current system and an idealized future system in the more experienced artists. However, no such correlation was found (see figure 4.10).

4.5.4 Discussion

The feedback I collected from 3D artists provides a better understanding of the expectations and requirements of 3D design work. With the constant development of AR and VR technologies, it is likely that in the future more and more professions will move towards virtual or augmented workplaces. Due to its inherently spatial nature, 3D design lends itself particularly well to such an adoption. However, since 3D UIs differ in many factors from traditional 2D UIs, I cannot simply copy what worked on the desktop into the HMD.

My study revealed a number of specific factors that can be overcome by improvements in the UI, such as problems with selecting and rotating objects. More importantly, it showed that artists are comfortable with a higher degree of complexity than previous prototype systems allowed and that their expectations from the new technology does not appear to correlate with the level of experience of the artist. As I found that gloves are problematic input devices, more research on how to provide a similarly high number of functions on a tangible input device is required.

While formative user testing can provide a better qualitative and quantitative understanding of general usability principles, it is not a means in and of itself, but should be employed in an iterative process to “zero-in” to the most useful UI as a combination of the best tried-and-proven metaphors. Therefore, further studies into the particulars of 3D UIs for artists are necessary.

The questionnaire that was presented to the participants is in Appendix B.

CHAPTER 5

User Study on Positional Head-Tracking

5.1. Introduction

In the previous chapter, I have introduced the prototype UI that I developed and evaluated with 3D artists. However, such insights may be only valid for the particular system with which they were generated. In order to gather more universally applicable information it is necessary to formulate generalized hypotheses in specific factors and test these with two-sample hypothesis testing in a controlled study. Since my prototype was sufficiently stable and usable, I was able to use it to perform such studies. As outlined in chapter 2, positional head-tracking is one of the important factors in current generation VR devices that add cost but are not clear whether they provide real value to professional use in the form of performance increase. In this chapter, I report on a user study that I performed on whether positional head-tracking provides significant performance benefits.

In chapter 3 section 3.3 I reported on a particular recurring behavior of "wiggling the viewpoint," wherein the artist rapidly and repeatedly changes the position of the virtual camera by small amounts, apparently in an attempt to gain a better understanding of the 3D shape of the virtual scene in the editor by emulating head-motion parallax. From this, I theorized that a full-fledged immersive 3D UI could offer increased effi-

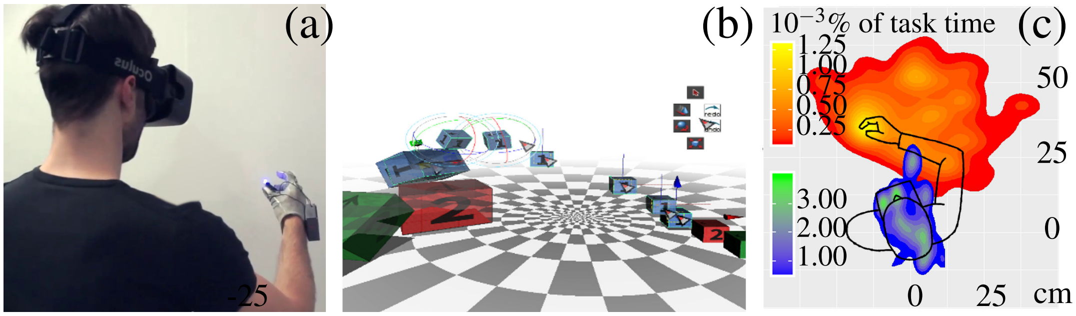


Figure 5.1. My study analyzed the effect of positional head-tracking on task performance in a 3D object placement task that was modeled closely after 3D design work and was performed by experienced designers. (a) The experimental set-up in the glove condition. (b) Illustration of the task: Translating, rotating, and scaling objects on the right to match objects on the left. Stages in the transformation of the blue object are shown in this composite. (c) Heat-map indicating the head (blue) and hand (red) motion of one participant (top view).

ciency, at least in part because positional head-tracking (also called head-coupling or viewpoint-dependent imaging) would provide the necessary parallax effect automatically and probably even without conscious action by the artist.

Klinker et al. identified five distinct scenarios for viewing virtual content in a product design context: “Turning”, “Overview”, “Detail”, “Discuss”, and “Compare” [45]. “Overview”, where the user is expected to walk around a virtual object to see it from all sides, is likely faster when performed with an input device rather than physical motion since the distance is too large for a quick head motion. “Detail”, where the user focuses on a specific detail of the virtual object, is where positional head-tracking may provide a significant advantage over input devices since head motions provide a quick and intuitive way to change the point of view. In the other three scenarios, the user is not expected to move, making positional head-tracking irrelevant.

Previous studies of the effect of positional head-tracking on user performance in various task settings have come to contradictory conclusions: some have found it beneficial [2, 3, 64, 73], some have found no effect [11, 70], and some have even found a negative effect [3, 5]. However, none of these addressed the challenges of 3D design. Since different tasks impose different requirements on the UI, studies performed with a generalized task and participants chosen from the general population may not apply

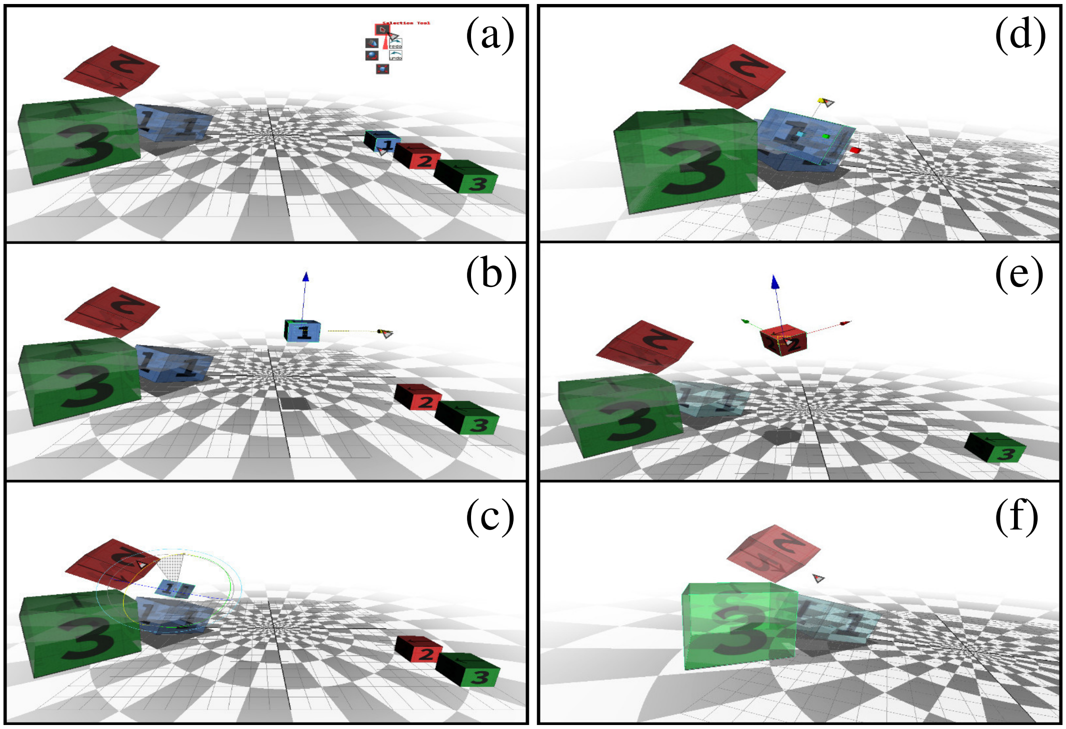


Figure 5.2. Study task as seen through the HMD. (a) The initial 3D scene. Source objects are on the right, and transformation goals are on the left. The tool menu appears at the top. (b) Translating the first object (blue cuboid labeled “1”). (c) Rotating the object. (d) Scaling the object anisotropically. (e) After the first object is correctly transformed, it is highlighted, and the goal disappears. The second second object is being transformed. (f) All three objects transformed correctly.

to a particular real-life work environment, just like performing a common ball-tossing task with graduate students may not produce usable results on making baseball gear for professional athletes. To make the results relevant to 3D design, I conducted a user study with 3D artists on a 3D object selection-and-transformation task and UI that are based on 3D design work (see Figures 5.1 and 5.2). My within-subject experiment examined performance while wearing an orientation-tracked stereoscopic VST-HMD, comparing the 3D pinch-glove described in chapter 4 with a 2D mouse, and the presence or absence of positional head-tracking.

This work was published in 2017 in Elsevier Computers & Graphics Journal [49].

5.2. Related Work

Several publications have analyzed the effects of positional head-tracking on task performance in VR and AR 3D UIs, but came to contradictory conclusions. See Table 5.1 for an overview of the publications I discuss and their respective results and limitations.

Arthur et al. [3] analyzed the effect of stereo vision and positional head-tracking on task performance in fish-tank (i.e., monitor-based) VR systems with a non-interactive cognitive task. Participants were shown two intertwined 3D tree structures and were asked to assign a given leaf to either of the two root points. The study compared task completion time and error rates for several conditions, including a monoscopic 3D image, a stereoscopic 3D image, a monoscopic 3D image with correct (head-coupled) perspective, and a stereoscopic 3D image with correct (head-coupled) perspective. Participants were instructed to move their heads and try to make as few errors as possible rather than optimizing time performance. They found that head-coupling alone (without-stereo) was slower than with stereo and even slower than static monoscopic and stereoscopic images, but that it decreased the error rate. In the stereo viewing condition, however, they found that positional head-tracking improved both time and error rate.

Similarly, Ware and Franck [73] performed experiments to determine which depth cues help participants perceive complex 3D graphs correctly in a fish-tank VR environment. They found that binocular stereo 3D improves performance by a factor of 1.6, parallax motion from positional head-tracking by 2.2, and a combination of both by a factor of 3. They consistently found that motion parallax has a stronger effect than stereo 3D, but recognized that the source of the viewpoint motion does not have to stem from head motion. Mouse input or a predefined slow rotational motion appear to work just as well.

These results provide some evidence for the effect of positional head-tracking, but are not directly applicable to 3D design work performance because the task was non-interactive and observational.

Boritz and Booth [11] also published a fish-tank VR study for a 3DOF (translational) target-pointing task. In this study, they found that binocular stereo allowed significantly faster performance than monoscopic imagery, but could not find a sustainable effect from positional head-tracking. During the first trials, positional head-tracking significantly improved performance in the monoscopic display condition but degraded performance in the binocular stereo condition. However, this effect quickly wore off as participants adjusted to the task. A second study by the same authors analyzed a 6DOF placement (object-docking) task, but positional head-tracking was not among the conditions tested [12].

Teather et al. [70] performed another fish-tank VR user study in which participants performed basic translational positioning tasks under different viewing and interaction conditions, including positionally head-tracked stereo 3D. They found that stereo 3D has a positive effect on error rates (though not on task completion time), but could not find evidence for any effect from positional head-tracking. While these results give evidence that positional head-tracking is not beneficial to 3D interaction task performance, they too are not directly applicable to 3D design work because the task was intentionally chosen to be simplistic. The authors focus on novice users and limited the interaction to selection and surface-registered translation, ignoring the complexity of object rotation and scaling, as well as the possibility that objects may float in space, thus making the task essentially 2D.

One issue that these studies have in common is that they used fish-tank VR systems, for which the possible range of head motion is limited with respect to the position of the virtual object relative to the viewer. It is not possible to "look around" the virtual object or alter the perspective significantly. Furthermore, in these studies, there was a local separation between the input device and the virtual objects. The participants would interact with the input device next to the screen where they would see the result, which might have affected the outcome.

Arsenault and Ware [2] performed a user study whose interaction took place behind a mirror in which the participant sees the virtual environment, thus achieving perfect alignment of real hand and virtual 3D cursor. Their results show significant improvements when positional head-tracking is enabled. However, in their experimental design, they force participants to change their viewing angle by about 18°, causing a misalignment of the hand and virtual cursor when positional head-tracking is disabled. So the effect may have resulted from this misalignment since users might not have naturally shifted their head position that far.

Bajer et al. [5] tested a selection task in a fish-tank AR system in which the participant was close to the screen and the 3D input device was aligned with the virtual pointer. They compared task performance with a control condition that used a 2D mouse. Their results show that positional head-tracking actually made participants slower when the perceived height difference of selection targets increased. However, this may have resulted from participants spending more time to move the pointing device "upwards" instead of just from side to side, indicating increased depth perception, which in this case happens to be detrimental to task performance.

Sandor et al. [64] performed a user study on object selection performance using a haptic device and a video see-through (VST) HMD. In half their conditions, they simulated a half-mirror–based VR system, displaying a virtual semi-transparent screen floating over the work area on which the virtual objects were visible. This condition (among other differences) provided no positional head-tracking and was found to significantly decrease performance. However, they did not explicitly investigate the effect of positional head-tracking and only mentioned it as one of three possible explanations for the observed effect. Furthermore, there was no virtual cursor displayed in the simulated half-mirror condition, which meant that any head motion from the center of the virtual camera would make it impossible for the participant to know precisely where they were pointing at in the virtual scene.

Using an HMD, Jones et al. [36] measured participants' ability to estimate the distance of objects in real, VR, and AR conditions, both with and without cues from head-motion parallax. Their study showed no benefit from lateral head motion for virtual objects, and only negative effects for real objects when wearing an inactive OST HMD, whose only intended effect was to add artificial inertia to the participants' head motions. However, objects in this study were positioned several meters away from the

Table 5.1. Overview of user studies on the effect of positional head-tracking, grouped by their result. The right column mentions limitations in the respective study design.

Positional Head-Tracking Improves Performance:	
Arthur et al. [3]	Only stereoscopic passive viewing task. Fish tank VR.
Arsenault and Ware [2]	Forced people to move their head (causing misalignment without positional head-tracking).
Sandor et al. [64]	Simulated half-mirror (includes other effects such as brightness).
Ware and Franck [73]	Only passive viewing task. Fish-tank VR.
No Effect from Positional Head-Tracking:	
Boritz and Booth [11]	Fish-tank VR.
Teather et al. [70]	Fish-tank VR. Focused on novice users.
Jones et al. [36]	Only tested depth-estimation ability.
Positional Head-Tracking Decreases Performance:	
Arthur et al. [3]	Only monoscopic passive viewing task. Fish tank VR. Decreased speed, but lowered error rate.
Bajer et al. [5]	Fish-tank VR. Effect possibly caused by 3D motion, not “on-screen” motion.

participant and not within arm's reach, as they would be in a workplace environment.

In addition, none of these prior studies on the effect of positional head-tracking used current-generation HMDs, which may have affected task performance due to the increased resolution, improved tracking performance, a larger field of view, increased freedom of movement, and a greater sense of immersion that modern devices offer over previous systems. They also considered only fairly generic tasks and recruited either subjects from the general population or engineering students with little or no experience in 3D modeling; thus, their results might not directly apply to 3D design.

5.3. Hypothesis

As described in chapter 3 section 3.3 I had performed examined the habitual 2D UI workflow one professional artist and one amateur prior to this study. In this analysis, I found several interesting actions in which both the amateur and the professional substantially engaged. One was the frequent and rapid change of viewpoint (camera position) on a very small scale. This camera motion did not allow the artist to see a different object or a previously hidden side of the object, and often ended very close to the original position. This behavior—where the artist returns to the original viewpoint and continues working on the same part of the object or scene as before—makes up about 42% of all viewpoint changes (3D scene navigation) and about 8% of total work time on average. The intended purpose of these actions seems to be to gain a better spatial understanding of the virtual 3D object. Since the monitor is monoscopic and no parallax effect from head motion is available, the visible image is ambiguous in its depth and 3D shape. A slight "wiggle" of the virtual camera produces a parallax motion that gives the artist a better understanding of the virtual scene. This finding is in accordance with the results of my survey described in Chapter 3 section 3.2, in which 3D artists reported that they used the camera controls of their 3D software extensively or even constantly.

Therefore, I theorized that in a VR work environment, positional head-tracking might make this operation unnecessary, possibly improving work performance. To validate this conception I formulated the following hypothesis:

H_1**: Enabling or disabling positional head-tracking has a significant impact on task performance in a 3D selection and transformation task.** Assuming that positional head-tracking in an immersive VR environment would provide these cues sufficiently to make wiggling the camera unnecessary, I hypothesized that disabling positional head-tracking would negatively impact performance. If this hypothesis would be falsified, it would allow for VR work environments that are less expensive and easier to set up because the rotational information from an IMU in the HMD is sufficient, and no additional hardware for positional tracking is required.

Some prior studies have found that depth cues can have different or even opposite effects based on whether a 2D input device or 3D input device was used [5]. Therefore, I decided to test this hypothesis with both a traditional mouse and a 3D input pinch-glove.

5.4. Experimental Platform

To test this hypothesis, I built upon and at the same time simplified the UI prototype described in chapter 4.

I performed the study with Oculus SDK 0.6.0, with dynamic pose prediction based on internally measured latency enabled. When I later updated the SDK to 0.8.0, I measured an average Motion-to-Photon Latency of 16*ms*.

I extended the prototype to not only support using the 3D pinch glove but also work with a traditional 2D mouse. The mouse was a wired laser mouse (a Dell MOC5UO), which is the most common input device in 3D design work (see chapter 3 section 3.2). When the mouse was used, a 2D cursor was displayed in the dominant-eye view only and thus was parallax free (similar to looking through a red-dot sight or reflector sight of a rifle). The mouse cursor was projected on a virtual plane that was at 10cm distance and perpendicular to the HMD. It always followed the HMD's motion, so a head motion would not result in a visual motion of the mouse cursor as long as the mouse remained stationary. Since the cursor was displayed to a single eye, it did not actually appear to be at a specific distance. Selection was performed by ray-casting from the dominant eye. The UI was the same as in Autodesk Maya with two exceptions: a marking-menu [50] on the right mouse button to select the tool (translation, rotation, and scaling), and viewpoint navigation on the middle mouse button. When the middle mouse button was pressed, the mouse controlled the viewpoint to support tumbling around the selected object. The mouse wheel allowed dollying forward and backward to focus on certain areas of interaction. This is similar to the camera motion used in Maya and other professional modeling software products, so I assumed it would be immediately understandable to the participating artists.

For this study, I prepared two different sizes of the glove both left-hand and right-hand versions, to accommodate differences in hand size and handedness. Each glove featured the eight buttons described in chapter 4, but in this experimental design only the most basic four were used in this study. These buttons were the main interaction button (similar to the left mouse button), a button to invoke the tool menu (similar to the right mouse button), an undo button, and a navigation button, which allowed changing the virtual viewpoint with a "grabbing-the-air navigation" metaphor [53].

In this experimental set-up I used a NaturalPoint OptiTrack Flex3 motion-capture camera system to track the position of the thumb, which acted as the 3D cursor. The set-up featured six OptiTrack cameras: four in front of the participant (two facing down from above and two facing up from below) and two behind the participant (viewing the work area over the participant's shoulder). The total area of the tracked space was approximately 1.5 × 1.5 × 1.5 meters.

The users were seated on an armless chair at the far end of this tracking volume, looking into its center. In the glove condition, the participants had no possibility to rest their elbows and I did not observe participants bending over to rest their elbows on their knees. However, it was possible to perform the task with bent arms which they naturally did. In the mouse condition, a small table was positioned beside the chair so that the participants could rest their mouse-hand on the table.

I used Tsai's hand-eye–calibration algorithm [71] to align the coordinate systems of the Oculus Rift DK2 and OptiTrack. The OptiTrack software ran on a separate networked computer, to ensure good performance. When using the glove, a 3D arrow was rendered in stereo in the virtual environment at the location of the participant's thumb. The UI was the same for both mouse and glove conditions, except for the addition of a "6DOF Tool" for the glove, which allowed simultaneous control of translation and rotation. The control/display ratio was 1:1 in the glove condition, and 1cm:14.7° FOV in the mouse condition. The total FOV of the Oculus Rift DK2 was about 106° vertically and 95° horizontally.

In the normal state, the system presented a fully immersive 360° virtual environment with both positional (lateral) and rotational head-tracking. Head-tracking was performed by the Oculus Rift which was very stable and free of noticeable lag. Positional head-tracking could be turned off and on by the experimenter at any time. When positional head-tracking was turned off, it produced the effect of looking at a virtual 3D monitor large enough to fill the complete FOV when looking straight ahead, and following all translational head motion. Further, without positional head-tracking, the correct alignment between the real hand and 3D cursor would be broken if the user's head moved. For example, if the user were to move their head sideways, the virtual screen would move with it to display the 3D cursor at the new position, even though the hand had remained stationary, thus resulting in a translational offset between the real hand and cursor. It should be mentioned that in a completely virtual environment, such an offset would not necessarily cause confusion, just like the mouse being in a different place than the on-screen mouse pointer would not confuse 2D desktop software users. Independent of positional head-tracking, rotational head-tracking was always enabled, so the user did not have to rely on peripheral vision to view any part of the work area. This is important because the Oculus Rift DK2 exhibits significantly more distortion toward the periphery of each panel.

The more complex features of the prototype described in chapter 4 such as 2D floating windows or animation time control were not available to the participants during the study tasks they performed; however, participants were allowed to use the whole prototype freely for 30 minutes prior to performing the tasks. This guaranteed that all participants had at least basic experience in working in an immersive modeling system and knew the basic UI used for the task performance trials.

Furthermore, the study was exclusively performed in VR instead of AR, because AR without any positional head-tracking is uncommon and can be quite confusing to the participants (giving the sensation of the whole room moving with you whenever move your head). Therefore, I did not use the cameras in all conditions and instead presented a simple virtual work area consisting of a circular base.

5.5. Procedure

I conducted the user study in December 2015 at Kyoto Saga University of Arts with nine participants: seven design students, one faculty staff member, and one professional 3D artist (five female, ages 19–35, mean age 22.2). All had at least one year prior experience with 3D design software.

In order to get a better understanding of the level of expertise of the participants with 3D design, I asked them to fill out a form on their previous experience. For a time frame of the past ten years, I asked the participant to estimate the number of hours spend each week working with 3D design software, averaged over the whole year. So if a participant worked on a project for an average time of 3 hours for half a year, then the average time spent over the whole year would be 1.5 hours per week. From this, I calculated an estimate of the total hours of experience by multiplying each annual estimate by 50 weeks per year (assuming that participants tend to neglect holidays). The participants reported experience levels between 75 hours (least experienced participant) and 10000 hours (most experienced participant). The average experience level was around 1200 hours. While this is naturally an extremely coarse estimation, it still gives us some information about the likely level of expertise of the participant on a logarithmic scale. Users whose experience is limited to tens of hours can be assumed to be beginners (two participants), those with hundreds of hours to be amateurs (six participants), those with thousands of hours to have achieved expert level (one participant). This is in accordance with prior findings from different fields of artistic endeavor [26].

Before the participants started using the prototype, I determined their ocular dominance with a Miles test (five right-eye dominant), asked them about their dominant hand (seven right-hand dominant), and informed them about the risk of cybersickness when using an HMD. They then tested the glove-based 3D UI for 30 minutes, following a tutorial. This allowed them to get used to immersive modeling and learn how to use the glove. During this period, I enabled the full AR-mode of the prototype described in chapter 4. This reduced disorientation and allowed the experimenter to point things out to the participant in order to help them. After this session, a rest period was given, during which the participant filled out some forms. The cameras were then turned off and remained off during the entirety of the timed study trials.

Participants were timed on a 3D object selection-and-transformation task with simple 3D objects. Two groups of 3D boxes were displayed, one set being the “goal” arrangement, the other being the “source” objects to place. The task was to transform each “source” object in the same way and to the same place as its corresponding “goal” object. This task involved nine degrees of freedom (DOF): three DOF each for translation, rotation, and scaling. The “goal” arrangement was chosen randomly from a set of ten manually prepared scenes in which the objects formed a pile. Each of these ten arrangements was very similar in that one cube was resting on the ground, one leaned against it, and one was balanced on top of the others; however, the exact place, orientation, scale, and order of cubes was different in order to avoid learning effects. The “source” boxes always started on the right side of the work area and were uniformly placed and scaled and aligned side-by-side with each other.

The task was designed to emulate artistic 3D modeling, in which the artist starts with a certain goal in mind (either provided by a concept artist or art director, or of the artist’s own imagination) and tries to reach this goal. I have chosen this task for this analysis because 3D selection and transformation of whole objects —together with tool selection which was also an element in my UI— is common in 3D design work, but also because it approximates a large portion of more complex 3D interaction tasks. For example, skeletal animation can be seen as the selection and manipulation of “joint” objects and animation handles, fluid simulation as creation, selection, and manipulation of “emitter” and “effector” objects, and sculpting as manipulation of a “virtual sculpting tool”. Of course, purely observational tasks may therefore not be represented correctly in my study.

I always grouped three 3D objects at a time, since artistic scenes are rooted in the relative arrangement of objects—an isolated single-object transformation task would be less representative of the tasks I was targeting. The work area was 70cm wide from side to side, initially displayed at a distance of 60cm in front of the participant and 35cm below eye level. Thus, it was possible to observe the whole area without requiring extensive rotational head motion (the work area consumed about 60° of the HMD’s 95° horizontal FOV). However, participants were still able to manually change the position of the work area to gain a better view of certain details. In the starting location, the cubes were approximately $5cm\ x\ 5cm\ x\ 5cm$ in size. In the “goal” configuration their size ranged between $10cm\ x\ 10cm\ x\ 10cm$ and $17.5cm\ x\ 12.5cm\ x\ 12.5cm$ (after anisotropic scaling). See Figure 5.2 for an example of the task.

Prior to starting the trials, the task was first demonstrated to the participant by having the experimenter briefly take over the UI from the outside and perform the transformation while the participant was wearing the HMD. This ensured that there was no confusion as to the goal. During this demonstration positional head-tracking was enabled. Each participant was informed that time was the critical factor in task completion, but was also told not to work any more quickly than they found reasonable to perform the task.

For each trial, the participant used either the 3D glove input device introduced in the first (practice) part of the user study or the 2D mouse. While the task was the same in both conditions, it was possible to perform it with fewer, more complex steps when using the glove, since it allowed 6DOF interaction which made it possible to perform translation and rotating simultaneously. During half the trials, positional head-tracking was switched off by the experimenter. However, rotational head-tracking was always enabled. The participants were not told during which trials positional head-tracking was enabled.

This study design yielded four different conditions: using a 2D mouse with positional head-tracking enabled, using a 2D mouse with positional head-tracking disabled, using a 3D input glove with positional head-tracking enabled, and using a 3D input glove with positional head-tracking disabled. The conditions were presented in a randomized order, but always with either both mouse conditions or both glove conditions first, never switching back and forth between mouse and glove. Switching the input device required the experimental set-up to be changed slightly, and this served as a brief resting period for the participant.

The first set in every block (three object transformations, during which head-tracking was enabled) was treated as a training set and removed from the sample. Some measurements were lost, due to technical problems or difficulties in the time schedule not allowing all conditions to be tested. The final analysis contains 30 three-object sets in the mouse conditions (performed by eight of the nine participants) totaling 90 object transformations, plus 14 three-object sets in the glove conditions (performed by five of the nine participants), totaling 42 object transformations.

I did not ask participants to fill out a standardized questionnaire such as SUS [17] or PSSUQ [52], since these provide measurements that are only relevant in comparison to other user interfaces. I did not intend to compare this user interface to any other user interface, but instead focused solely on measuring differences in task performance related to positional head-tracking.

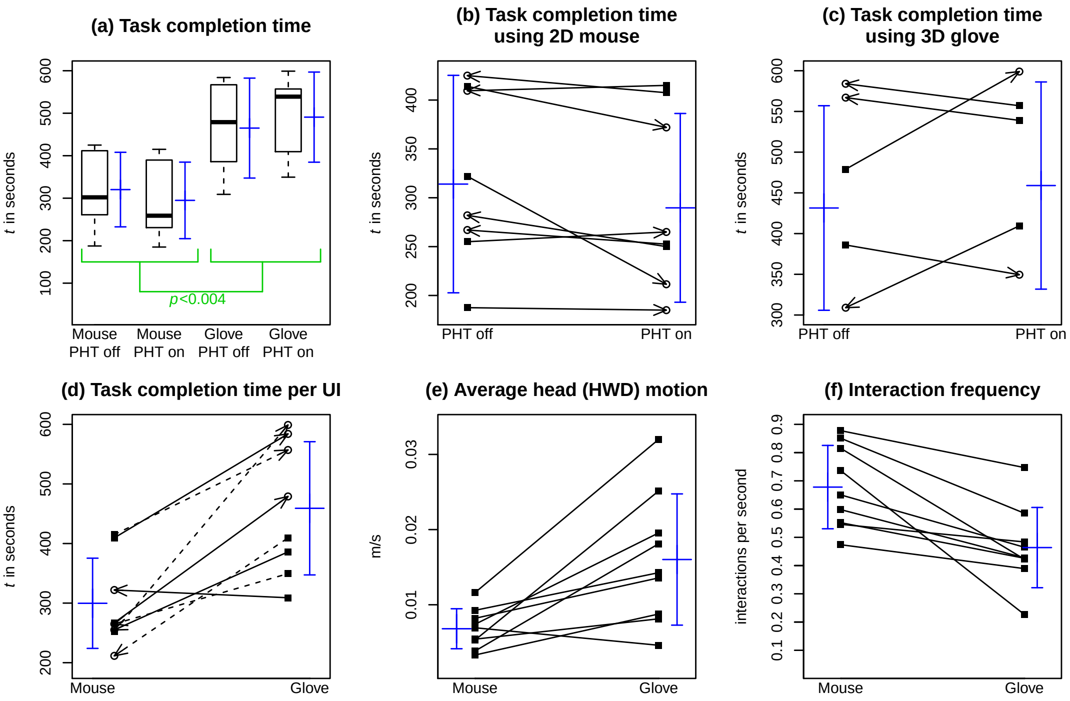


Figure 5.3. Experimental results. PHT indicates positional head tracking. Blue bars indicate mean and standard deviation. Data points are participant means. Dashed lines indicate measurements without positional head-tracking. Arrows indicate the order in which the conditions were performed.

5.6. Results

A summary of the recorded measurements can be seen in Figure 5.3(a). When using the mouse, mean task-completion time was 314s without positional head-tracking (SD=111s), and 290s with positional head-tracking (SD=97s). When using the 3D input glove, mean task-completion time was 431.4s without positional head-tracking (SD=125.6s), and 459s with positional head-tracking (SD=127.2s). Analysis of within-subject performance showed a significant difference (defined by an α of 0.05) in the task performance between using the 2D mouse and the 3D input glove (average improvement of 159.4s; $p < 0.0035$; Figure 5.3b), but no significant effect from positional head-tracking, neither for the 2D mouse nor for the 3D input glove. (Mouse: difference of means $\approx 25.4s$ (7.7%), $p > 0.1$, Figure 5.3(c); Glove: difference of means

$\approx -25.8s$ $(-6.4\%), p > 0.49$, Figure 5.3d). Following the experimentation guidelines developed by Keppel [40], I considered these analyses as planned comparisons between the conditions as they were stated as hypotheses. Therefore, no error correction for multiple comparisons was performed.

I considered extending the user study to find more minute differences in task performance, but a power analysis using my sample to estimate population variance (i.e., assuming that future participants would exhibit a similar variability to previous ones) indicated that this was impractical, as I found that I would need $n > 77$ and $n > 76$ respectively for a test of power 0.95. This indicates that the expected effect of positional head-tracking is small compared to other factors. My results show an effect size r of 0.1 (Cohen's $d \approx 0.2$), which is considered small. Conversely, they express 95% confidence that the performance improvement for positional head-tracking is $< 17\%$.

I further analyzed the recorded motion data and found that participants moved their head significantly less when using the mouse ($p < 0.008$ on a within-subject t-test; see Figure 5.3e). Figure 5.1c shows an example of the recorded motion data of one participant as a heat-map. The motion volume in which the participants moved their heads was about $0.0066m^2$ on average when using the mouse, and $0.0257m^2$ on average when using the glove.

Another possible cause for the imbalance between mouse performance and 3D input-glove performance could be the familiarity with the traditional device, allowing for faster interaction. I therefore analyzed the recorded data, measuring the frequency of interactions ("clicks" for the mouse, and "pinches" for the glove) and found that the mouse was indeed used more vigorously ($p < 0.0027$; see Figure 5.3f).

At any rate, the fact that the same participants also participated in the formative evaluation of my prototype allowed me to analyze their task performance time with their prior work experience. See Figure 5.4.

5.7. Discussion

Regarding my original hypothesis, I have found no evidence for H_1 and therefore cannot confirm the assumption that positional head-tracking affects artist performance in a 9DOF object-transformation task, both for the 2D mouse condition as well as for the 3D input-glove condition. In this section, I discuss possible explanations and implica-

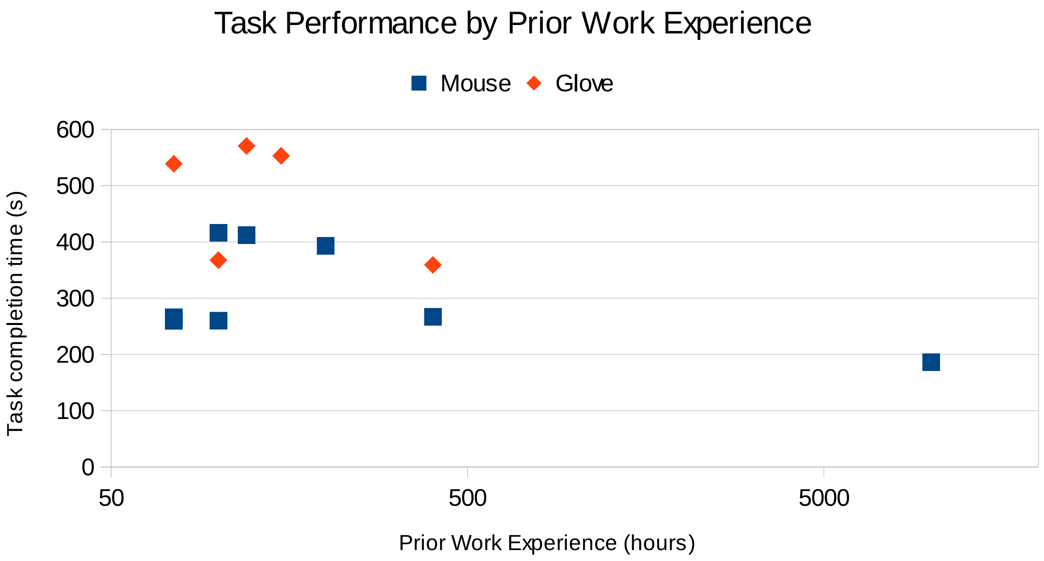


Figure 5.4. Relation between task performance and prior work experience.

tions in turn.

Not finding that positional head-tracking has a positive effect on task performance is in accordance with prior findings [3, 36, 70]. Their explanations for the lack of effect usually relate to the fact that while motion parallax can be an important depth cue, the natural range of motion of the head combined with the physical exertion make it an unattractive option to gain spatial understanding of the virtual object. The rather small amount of head motion observed in the participants supports this idea.

The difference in head motion between mouse use and glove use is likely to be the result of secondary motion. When using the 3D input glove, the hand has a wider range of motion in three dimensions, thus forcing the arm, shoulder, and subsequently, the head, to move more to support hand interaction. The difference in head motion could have had an effect on task performance, but no such effect was observed. It is possible that the advantage of motion parallax was still too small in both cases to be noticed.

One explanation for the limited range of motion may have been that participants felt hindered by the HMD. Three of the participants criticized the weight of the HMD after testing the prototype, giving respective estimations of the gravity of the problem of 4, 8, and 6 on a scale from 1 (negligible) to 10 (critical). It is possible that —

burdened by the weight of the HMD — a head motion to change one's point of view may actually be slower than a hand motion achieving the same effect, and thus not at all desirable by the participants.

Another possible explanation that is specific to our use case is that the "camera wiggle" has become so habitual that the user performs it even when the spatial relationships are clear. In out experimental platform, the users were able to use the middle mouse button or navigation button on the glove to change their viewpoint, in both positional head-tracking enabled and disabled conditions. While this feature was intended to allow major changes of the viewpoint necessary to perform the task, it may have been used by participants to perform the habitual "wiggling" even in the positional head-tracking enabled condition when a small head-motion could have sufficed. In order to test this assumption we analyzed the log files recorded during the experiment. Figure 5.5 shows a histogram with 5°bins of the viewpoint change operations performed by the user over the total rotational change achieved through the operation. Thus, the rightmost end of the graph (180°) means the user ended the operation looking at the objects from the opposite side of where the operation was started. While there was a slight tendency in the mouse condition for users to end close to where they started, this was not nearly as pronounced as in the 2D UI where "wiggling" made up 42% of all viewpoint change operations. Furthermore, no such tendency was present in the glove condition. Most viewpoint change operations fell into the range of 20°to 80°. The average was 50°in the mouse condition (SD: 43°) and 49°in the glove condition (SD: 36°).

An alternative interpretation for these result is that the moments of "wiggling" the camera in a 2D UI are used to consider the next steps. Thus, even when the wiggle is not necessary, the user might still pause the interaction regularly to consider what to do next, thus minimizing the possibility to improve task-completion time.

Finally, it is possible that the availability of stereo vision in my UI provided enough spatial understanding to make the parallax motion unnecessary. This explanation, however, raises the question why stereo glasses are not widely used in 3D design, since they would provide stereo 3D vision quite easily.

It is possible that positional head-tracking would prove more helpful in more cluttered scenes, where it may be necessary to move the head to gain a clear view of the area of interaction. However, in this case, the advantage of positional head-tracking

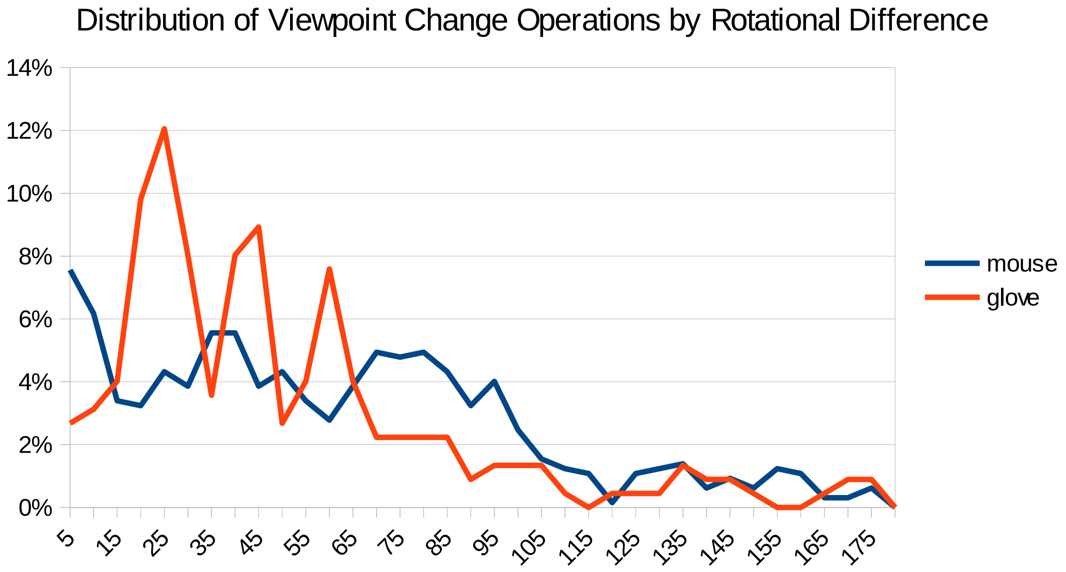


Figure 5.5. Frequency analysis of the viewpoint change operations by total rotational difference achieved through the operation.

would be artificial and not necessarily reflect the real-life work environment of artists accustomed to a 2D UI where clutter obstructing the line of sight is unacceptable.

My results clearly show reduced performance when using a 3D input device, similar to previous work [70]. Given the great efficiency with which we interact with physical objects in my everyday life, it can be assumed that this stems, at least in part, from the artists' familiarity with the mouse and maturity of the device. While the mouse has had a long time to mature into a reliable, precise, and universal interaction device, the gloves used in this study were research prototypes that all participants used for the first time. Further, the 6DOF tracking of the glove was not perfectly reliable, resulting in occasional jumping or misalignment of the virtual cursor. This is likely to have decreased performance in the glove conditions.

In fact, every single participant criticized the glove after testing the prototype UI. The estimated gravity of the reported problems on a scale from 1 (negligible) to 10 (critical) ranged from 1 to 8, with an average of 5.2. Common points of critique were the size and placement of the touch contact areas and the lack of haptic feedback.

It is also worth noting that the mouse was placed on a table, giving the participant's hand more stability when using the system. This might have resulted in higher precision and in turn improved task performance. In addition, the pinching motion used with the 3D input glove can alter the 3D cursor position involuntarily. In my study, I could often observe participants struggling with finalizing the placement of the virtual object, because they were unable to end the interaction at the intended point.

In conclusion, this user study made the contribution of questioning the common belief in the importance of parallax depth cues from positional head-tracking. While positional head-tracking can increase the sense of immersion and may be crucial for some tasks, some 3D design-related tasks may not profit significantly. This further points to the possibility that VR work environments for 3D design can be created more easily and inexpensively by omitting hardware required for positional head-tracking (which is commonly achieved either by additional environment-mounted tracking hardware or computationally expensive visual odometry algorithms).

CHAPTER 6

User Study Comparing Task Performance in AR and VR

6.1. Introduction

Historically, AR has been associated with interaction with real-world objects. Usually, the virtual content is attached to some physical object which is part of the interaction metaphor. However, interaction with real objects is not a strict requirement of AR. From the perspective of sole interaction with virtual objects the key feature of AR is the ability to observe the real environment.

This provides a number of advantages. Accidents can be avoided when the user is able to see the physical boundaries of the work area and all objects within it. While many VR systems provide a chaperon feature that warns the user when he is about to leave the designated "safe space", these are inaccurate and can not react well to changes in the environment such as other people entering the work area. It is also commonly believed that the ability to see one's own body has a positive effect on interaction based on foundational research on visuo-spatial perception. For example, Coello [22] reviewed a large number of publications showing the importance of different depth cues such as one's own limbs or textured backgrounds for correct spatial understanding, and Graziano et al. [32] performed neurological experiments on monkeys identifying key areas of visual perception related to seeing one's own arms.

On the other hand, AR is usually associated with higher cost. In the most simple case, suitable cameras must be attached to the HMD. However, the requirements to the cameras are not easily matched. A high field of view combined with a high frame rate and minimum latency makes these cameras quite expensive. On top of that, they need to be carefully calibrated in order to match the rendered virtual objects. Using an Optical See-Through (OST) HMD is often even more expensive due to the complexity of the arrangement and calibration.

For private use where personal preference is the main factor in decision making this may not be an interesting comparison, nor is it in cases where interaction with real objects is required. For professional 3D content creation however it is actually possible to measure and calculate an optimal solution, since work speed improvements may raise income, making the additional expenses worthwhile. For this, however, it is necessary to know how large and in what shape the differences in performance are.

Until now, 3D UIs have only been studied in task settings that differed greatly between AR and VR, due to the different technology and focus of both technologies. Direct comparisons were hardly performed or reported on. My hypothesis was that seeing the real environment (in AR) has a significant effect on task performance even when working with solely virtual objects, due to a more direct understanding of spatial relations. Prior research had not produced conclusive evidence for this hypothesis in the setting of 3D design work.

In 2015, I performed a user study to directly compare task performance in AR and VR. The task consisted of selecting and transforming a "source" object to resemble a "goal" object in position, orientation, and scale in three dimensions (9DOF; Figure 6.2). Different from the previous user study, this was not performed with 3D artists, but instead with graduate students, since I was unable to recruit professionals. However, the greatest share of work time in 3D design are spent on simple 3D transformations (see Section 3.3) which are relatively easy to learn. Thus, the results obtained from novices are likely to be universal and apply to professional artist as well. However, since the participants in this study learned both the 3D UI and 2D UI for the first time, seasoned artists with long experience in using the 2D UI may exhibit some bias before getting sufficiently familiar with the 3D UI as well.

Figure 6.1. Illustration of a user using either the 3D input device (left) or the mouse (right).

In order to further shed light on the possible reasons for performance differences, I asked participants to perform the same task with both a 6DOF 3D input device and a traditional 2D computer mouse (Figure 6.1). When using the 3D input device, the user's hand and the virtual cursor are perfectly aligned, providing additional visual feedback to 3D interactions. In the mouse condition, the only additional feedback was seeing an empty work area, which should not provide much benefits on task performance other than a general sense of orientation, spatial limits to movement, and possibly a sense of connectedness with the real world. On the other hand, these positive effects might be counterbalanced by increased sensory load in AR vs. VR, reducing or even reversing the overall effect.

Thus, the experiment consisted of four conditions: AR with 3D input device, VR with 3D input device, AR with mouse, and VR with mouse.

Additionally, I asked participants about their subjective level of comfort in each condition in order to find whether people preferred AR or VR, or if using a 3D input device resulted in increased strain from arm motions in mid-air.

The results showed a statistically significant increase in performance in AR over VR when a 6DOF 3D input device is used: in the VR environment, it took participants on average almost 22% more time to perform the task. To my own surprise, I found a similar, albeit reduced effect when participants used the mouse: about 12% reduction of task performance in VR compared to AR. While most participants expressed a pref-

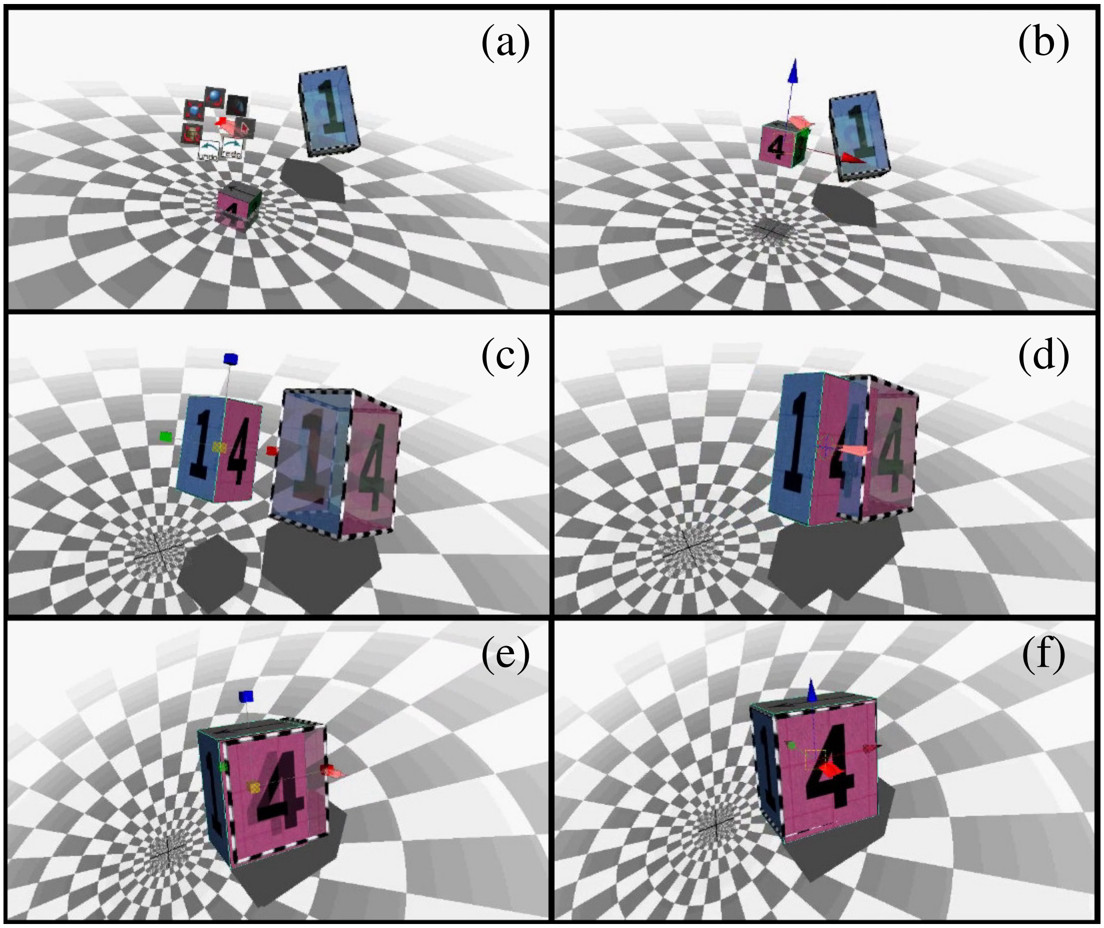


Figure 6.2. The task in the VR + 3D Input Device condition, as seen through the HMD. (a) The 3D scene at the beginning with the source object in the center, and the tool menu visible on the left. (b) Translating the object. (c) Scaling the object in 3DOF after rotation. (d) Moving the object into the goal. (e) A look from the side reveals that the object length is not correct. (f) The object lies within tolerance thresholds. The task is completed.

erence for either mouse or 3D input device, I could not find a statistically significant overall trend of either device being perceived as more comfortable to use.

In this chapter, I will explain the details of the experiment as well as discuss possible explanations and implications.

This work was published in 2017 in IEEE Transactions on Visualization and Computer Graphics [48].

6.2. Related Work

Only little prior work had been published that directly compares AR and VR in the same setting.

Kiyokawa et al. [44] performed two experiments comparing a shared augmented environment with a shared virtual environment for use as a face-to-face collaborative virtual workspace and found AR to be superior to VR. However, they only measure collaboration efficiency on a selection task. No further interaction with the objects such as 3D transformations were performed, and the experiments focused on collaboration efficiency, not on individual performance.

Boud et al. [14] compared several systems as a training tool for assembly tasks, including VR and AR settings. They find AR to outperform several VR variants, which in turn outperform conventional instructions by a great margin. However, their AR system was what they describe as “context-free”. This means the AR graphics were not registered with the real world and could be described as a Heads-Up Display (HUD) of a static diagram image. The measured performance was in the assembly of real objects, for which the various AR and VR conditions were only used as prior training for the task. Our work focuses on 3D interaction tasks with virtual objects in AR or VR.

Jones et al. [36] compared AR and VR systems in their effect on human depth perception using an HMD. Based on prior work that had shown that depth is consistently underestimated by users in a VR setting, they tested whether a similar effect existed in AR. Their results showed that no such underestimation of depth occurs in AR. This indicates that the additional spatial cues of the real environment may help spatial understanding. However, no interaction with virtual objects was required in their task. Furthermore, they only analyzed depth estimation at a range of 2-8m, which is far beyond the usual work area for most tasks.

Cidota et al. [20] performed a similar study with a focus on serious games. They measured subjective usability and task performance in AR and VR under various visual effects such as blurring and fading on a simple hand-based select-and-move task designed to measure depth perception performance. The pair-wise within-subject comparisons found no statistically significant effect in mean task performance for neither the different visual effects nor when comparing AR and VR. Only when they removed all data except those participants who got their best score in only one sub-condition they found significant differences. However, whether AR or VR performed better depended on the visual effect and partially contradict their results on measured performance.

Juan and Pérez [37] studied differences between AR and VR in acrophobic scenarios. Instead of performing a task, the objective was to expose participants to anxiety or phobia provoking situations. They find both AR and VR to be effective for creating anxiety at appropriate moments in the simulation but find no statistically significant advantage of either AR nor VR.

Arino et al. [1] compared AR and VR directly using an autostereoscopic display instead of an HMD. Their participants were children (8 to 10 years) and the task they performed resembled more of a game in which children were asked to passively count specific objects in the scene. The children could only interact with the scene by rotating a single fiducial marker around one axis, which would rotate the virtual object on it in order to see it from a different angle. They did not find significant differences in the mean task completion time, nor in post-use questionnaires regarding the experience. However, AR was generally preferred by the children over VR in direct comparison.

Botden et al. [13] compared two systems for laparoscopic surgery simulation, the LapSim VR and ProMIS AR. The AR system was found to be more realistic, to have better haptic feedback, and to be more useful for training purposes. However, this study mostly describes differences between two competing systems. This does not necessarily imply general differences between AR and VR, since a better VR simulator could easily be build by improving haptics and so on.

Sandor et al. [64] performed a user study on object selection performance in both AR and a simulated half-mirror VR condition. They simulated a mirror-based VR system by displaying a virtual semi-transparent screen floating over the work area on a video see-through (VST) HMD. Comparing user performance on an object selection task, they find that the AR condition was superior to the simulated VR condition. However, their conditions differed in several factors such as head-tracking, object visibility and occlusion. Therefore, the results do not necessarily indicate a general advantage of AR over VR in all settings.

Irawati et al. [35] created a 3D edutainment environment for falling domino blocks, which can be used both as an AR and a VR environment. However, they did not perform any evaluation on advantages or disadvantages of either method.

Similarly, Rhienmora et al. [63] created a simulator for dental surgery that can be used either in AR or VR. A preliminary evaluation by an expert dental instructor indicated that the AR version better resembles a real clinical setting. However, only the AR version was used with an HMD, while the VR version was displayed on a 2D screen. Furthermore, no quantitative evaluation was performed.

Lee et al. [51] investigated the effect of visual realism of the environment on task performance in a search task, by comparing an AR system with VR simulations with varying degrees of rendering realism. They recreated a real outdoor environment and found some indication that visual simplification in level of detail, texture, and lighting may have some positive effect on task performance. However, their virtual environment differed from the augmented environment in several areas, such as added objects and changes in vegetation. Most of the performed tasks did not show significant differences between AR and VR performance.

Möller et al. [56] performed a user study on indoor navigation with an AR guidance system on a hand-held device, which had an alternative “VR” mode. They found that users navigated a path about 15% faster with VR than with AR. Furthermore, VR appeared to be more robust to errors than AR. However, the VR mode differed greatly from the AR mode. In the VR mode, the device could be held at a low angle and allowed manually changing the view direction of pre-recorded panorama images in a drag-and-pan fashion, resembling more of a panorama picture viewer than actual VR. This may have had a great influence on the result, as participants reported that holding the device upright in their field of view (in AR mode) felt both straining and awkward (since they were worried about the opinion of passers-by).

Khademi et al. [41] compare projector-based tabletop AR with non-immersive monoscopic screen based VR regarding performance on a “pick and place” task designed for rehabilitation of stroke patients. They used healthy subjects for their evaluation and found that they performed better in the AR condition than in VR. In both conditions, they interacted with a physical object. The AR or VR gear was only used to display target object placement areas, which means that in the VR condition participants had to perform a mental transformation from the on-screen computer graphics (where they only saw the placement object and target area, not their own hand) to the table surface where they had to perform the task.

Bowman et al. [16] explored the possibilities of using a high-end VR system to simulate different VR and AR systems of equal or lower display fidelity in VR in order to analyze the effects of specific design factors of those systems such as field of view (FOV) or latency. They argue that comparisons between existing VR and AR systems are inherently ambiguous because any two systems will differ in a number of factors such as FOV, weight, form factor, and so on, making it impossible to isolate the effects on any one single factor. They further provide evidence for the viability of their approach by recreating prior studies with real systems in their simulation and achieving similar results. However, they admit problems due to the technical limitations of the simulator such as delay and lack of photorealistic rendering capabilities in VR. We go the opposite way by using an AR system to simulate VR by artificially blocking out the live video stream. This removes the limitations of not being able to simulate AR sufficiently while staying true to the concept of simulating one technology with another in order to isolate certain factors for analysis.

Howlett et al. [34] analyzed differences in task performance and eye movements on the same object sorting task performed in a real environment and a VR reproduction of the same environment using back-projection and a haptic input device. They found that their VR reproduction of reality — while quite close to the original — had some effects on the participants. In VR, people took longer to perform the task, had a longer average eye fixation duration, and tended not to look ahead to plan their next steps. The average saccade amplitude was similar in VR and reality within each task. This points to the possibility that even slight deviations in reproducing a real environment in VR can have a significant effect. However, since the study compared two groups of only four participants (between-subject), and no statistical analysis of the measurements was performed, the results may not be generalizable to other applications.

Werkhoven and Groen [74] performed a study on task performance as measured in both speed and accuracy on an object placement task in VR, comparing hand-tracking to a table-bound input device. Unlike this study, the mouse was a 3D SpaceMouse. They found that the correctly aligned virtual hand input metaphor performed significantly better, but admit that this may have been influenced by technical factors and task design. In their study, the task was strictly divided into separate sub-tasks, first rotating the object before positioning it. It thus did not resemble a natural work-flow where rotation and translation are used together or in alternation to zero-in on the desired result. Furthermore, the hardware used in their study appears quite outdated by today's standards, and therefore results may differ when using modern hardware.

Different to these prior publications, I intended to perform a direct comparison of interaction performance in AR and VR, using the exact same device, set-up, and task. By doing so, I could isolate the key factor (the ability to see the real environment) and was able to provide quantitative evidence for its effect on task performance.

6.3. Hypotheses

In theory, the ability to see the real environment could have several effects on task performance, which could be either advantageous or detrimental.

The most important effect is that the visual feedback of seeing one's own hand is often considered helpful to perform tasks that require some form of hand-eye coordination. In an AR environment however, the key features of this feedback system such as occlusions, shadows, or direct and indirect lighting effects may not be presented physically correct. This could result in confusing the user instead of improving performance. For this study, I decided to ignore all of these factors and designed the system to render the virtual content exactly the same in both the AR and VR condition.

Another effect is that seeing the boundaries of the work area may give users more confidence to move around freely and swiftly, without worrying about bumping into physical objects. Furthermore, it may provide a better sense of direction and reduce disorientation, which may be helpful to some users by reducing confusion and cybersickness. Again, these could also turn out negatively, when VR causes users to become more daring in their motions and less distracted due to the removal of the real environment.

For the VR condition, it is obvious that the type of the virtual environment affects task performance. It was therefore important to carefully consider the strategy from which to conduct the analysis. The options can be categorized into *(A)* attempts to approximate the real environment in VR; *(B)* creating a fictitious environment; and *(C)* providing as little environmental cues as possible. The same categories naturally correlate to different approaches to representing the user's body in VR, where one can *(A)* attempt to capture detailed information on the user's body and represent it faithfully; *(B)* generate a virtual body based on sparse information (such as the position of the 3D input device); or *(C)* omitting any display of the user's body, only providing a virtual cursor to indicate the device position.

Option *(A)* is usually connected to additional efforts, since common VR or AR HMDs do not provide out-of-the-box solutions to scan and track the environment and the user's body. It is not well researched what efforts end-users are typically willing to take in order to improve their VR experience. Furthermore, it also raises concerns towards the degree of fidelity since the stated end-goal of approach *(A)* is the elimination of differences between AR and VR. Practically, a Video See-Through (VST) AR system that uses depth reconstruction can be technically seen as a VR environment created from ad-hoc reconstruction, because the image and depth buffers are just one form of 3D representation of the environment created from video images which is then used in the rendering process. In such a perfect VR reconstruction of reality, AR and VR become synonymous and any difference found between AR and VR conditions can be seen as a failure to successfully recreate the real environment. While it is interesting to study which failures result in performance differences and which are tolerable, such an analysis is application specific and may be better suited to be performed from a framework of diminished reality.

Option *(B)* appears to be the most practical since it requires no additional effort from the user. However, it bears the risk that certain design choices unduly affect performance measurements. We could no longer be certain that the same task performed in a different virtual environment would not produce different results. For example, the performance of claustrophobic users may decline in larger environments while nyctalopic users may perform better in brighter environments, unrelated to whether the environment is virtual or real. A middle ground between option *(A)* and option *(B)* would be a coarse approximation of the real environment, for example by letting the user choose from a selection of virtual environments the one that most resembles the real environment. However, for the purpose of this research, such a selection would only increase complexity and uncertainty as to what was the original cause of measured effects. Furthermore, regarding the representation of the user's body, hand-tracking may actually decrease performance if precision and reliability fall below a certain threshold. Since common VR hardware only tracks the position of the controllers, the position of the elbow and shoulders must be guessed completely.

Option *(C)* is the most abstract approach. It therefore has a greater potential to yield reproducible results, but may not allow conclusions to be transferable to real-life use cases. However, when we look at common current VR design applications such a Google TiltBrush, Oculus Medium, Adobe Project Dali, or the Unreal Engine VR Editor, we see that neither of them makes an effort to provide an artificial virtual environment other than a horizon, nor do they attempt to display the user's hands and arms.

Therefore, I decided to follow approach *(C)* to display a completely empty virtual environment, since it not only allowed me to avoid unwanted influences on the result from particular design choices but also is an acceptable approximation of common current real-life applications. Thus, I designed the system to display only a ground plane in order to provide a sense of orientation and limited depth cues, similar to the real environment in the AR condition. Only a simple arrow was rendered to indicate the position of the 3D cursor in the VR condition. I decided not to depart from this purely abstract representation to avoid any advantage that the user might gain from seeing virtual hands. With this, I match current VR design applications such as Google Tilt Brush [1], Oculus Medium [2], Kodon [3], and Tvori [4], which do not show the users hands but only an abstraction of the input device. This allowed me to measure the full effect to which visual impressions (including hands, arms, and environment) can positively impact user performance. In interpreting the results of this study it is important to note that the more information we have about the real environment, the more we can start to imitate the AR condition by displaying objects of the real world in the VR condition. I ignored hardware differences such as factors in Optical See-Through (OST) HMDs or latency which are arbitrary to the used devices and change rapidly with every gen-

---

[1] https://www.tiltbrush.com

[2] https://www.oculus.com/medium/

[3] http://store.steampowered.com/app/479010

[4] http://www.tvori.co/

eration. Thus, I was able to assume that the closer the VR condition would resemble the AR condition by providing visual feedback about the environment and the users own body, the experiment would yield more and more similar results. Therefore, this study can be seen as a measurement of the maximum difference between VR without any knowledge of the real world, and AR, which is a full visual representation of the real world. Of course, we do not know whether any measured difference between the conditions will disappear gradually as we add more visual information. It is possible that even showing one single polygon indicating the opposing wall could lead to exactly equal task performance in AR and VR. However, since the VR condition is based strictly on the information available to VR developers, the findings produced in this study are of interest. That is to say that VR developers cannot indicate the opposing wall because they have no information where the opposing wall is. As discussed before, one may choose to “stimulate” possible advantages of the AR condition by guessing the position of walls, objects, or the user’s limbs. But for this study, I abstained from simulating arbitrary guesswork, as there is no accepted framework for estimating details about real-world VR installations.

Since I asked participants to perform exactly the same task with a 3D 6DOF input device as well as a traditional 2D mouse, I was able to shed some light on what single effects may have an influence on task performance. In the case of the mouse, seeing the real environment should not have the same effect, because the cursor is two-dimensional and not directly aligned with one’s real hand (which is below the work area and therefore usually outside the field of view). Only the general effect of the virtual environment, including artificial lighting, emptiness and thus, possibly, a sense of isolation and disorientation could affect the user.

Based on the design of the study I formulated 3 hypotheses:

H_1**: Task performance with a 6DOF input device will be significantly improved when the user is able to see the real environment (AR vs. VR).** If this hypothesis were to be supported, it provides an argument for attaching cameras to VR HMDs in order to increase users performance in professional 3D design work when using 3D input devices (such as Oculus Touch[5] or HTC Vive controllers[6]) or hand tracking (such as Leap Motion[7]).

---

[5] www.oculus.com/en-us/touch/

[6] www.htcvive.com

[7] www.leapmotion.com

H_2**: There will be no difference in performance between AR and VR when using a 2D mouse**. If this hypothesis would have been falsified, then there is a possibility that the effects related to H_1 are at least in part dependent on a general sense of spatial awareness instead of task specific visual feedback. For example, purely virtual environments might improve performance by reducing visual cognitive load.

H_3**: Subjective measures of comfort will differ between AR and VR environments, as well as between 2D mouse and 3D input device.** I hypothesized that the isolation of VR environments may have a negative effect on the users' comfort, because the lack of visual feedback of the boundaries of the work area and one's own body may induce a feeling of unease. Supporting evidence for this hypothesis could, on one hand, be seen as an explanation in the case that H_2 is rejected, and on the other hand provide an alternative incentive for using either AR or VR, which is not related to task performance but to user satisfaction. A similar argument could be made regarding the mouse vs. input device conditions, and whether users actually want to use a 3D input device solely for its novelty factor.

Although the study design could also allow me to compare the relative performance of the mouse and the tangible 3D input device, it would not provide useful general information. Several arguments can be made regarding the various advantages and disadvantages of each device. Some stress the advantages of 2D mice, like high familiarity of most users, and its stability and precision when placed on a flat surface. Others favor 3D input devices for their higher number of DOF or the correct spatial correlation of the device and the cursor in space. However, because many technical or experimental factors contribute to overall performance of either device, a direct comparison is difficult. The 3D input device used in this study was a prototype, which I specifically designed for the experiment, and the results may not apply to other devices.

6.4. Experimental Platform

To test my hypotheses, I updated the prototype 3D modeling UI introduced in earlier chapters.

Again, I used the Dell Precision Notebook (Intel Core i5 CPU with 2.90 GHz, 8GB RAM, and a Nvidia Quadro K4100M graphics adapter) running Windows 7 and Maya 2014.

For this user study, I again attached the ovrVision stereo cameras to the front of an Oculus Rift DK2 to turn it into a VST-HMD. The cameras have a resolution of 800×600 pixels and a frame rate of 25 fps. The cameras were running in both the AR and VR condition and the exact same image processing (image rectification and undistortion) was applied, in order to achieve the same computational load and avoid one condition performing faster than the other. The only difference was that in the VR condition, the prototype software cleared the frame buffer with a white color instead of using the image from the camera as a background for rendering. No environment was rendered neither in the form of polygonal models nor textures. However, in the VR condition, I displayed a circular pattern to indicate a "ground plane" in order to give the user a basic sense of orientation even when the target objects were not visible (see Figure 6.2). I indicated that this pattern was indeed the "ground level" by displaying simple object shadows on it. In the AR condition, no ground plane was displayed, and therefore no shadows were visible.

Again, I decided one condition to be performed with a traditional 2D cable-bound laser mouse (DELL MOC5UO). The mouse UI was the same as described in chapter 5. A cursor was displayed in the dominant eye view only. The UI was identical to Autodesk Maya with two exceptions: a markup-menu that appeared upon pressing the right mouse button to select the tool (translation, rotation, and scaling, as well as an "undo" and "redo" button), and viewpoint navigation by dragging with the middle mouse button and using the mouse wheel. When the middle mouse button was pressed, the mouse controlled the viewpoint in a tumbling motion around the selected object. The mouse wheel allowed moving forward and backward. I implemented this style of navigation, which is similar to the camera motion used in Maya and other modeling software products, in order to keep the results closely related to real-world applications.

Based on the experience from the prior studies where I noticed participants having problems with the gloves, I decided to change my approach towards the input device. I created a custom-made 3D printed ergonomic case in the shape of a pistol grip as a 3D input device. Four buttons were attached to the device. The first button was the main interaction button (similar to the left mouse button). Two of the other buttons would bring up the tool menu (similar to the right mouse button in the mouse condition). I created two buttons only for convenience since some test users found one button location easier to reach than the other. The last button on the device was a navigation button (similar to the middle mouse button in the mouse condition) which allowed changing the virtual camera's viewpoint in a "grabbing-the-air" navigation fashion without editing the objects. An Inertial Measurement Unit (IMU) was used to track the orientation of the input device, and an LED was attached to track its position in the HMD-mounted cameras by computer vision and triangulation. This meant that the device would only work when the user was looking at it. It was not possible to use it outside the field of view. The combination of IMU and LED tracking provided 6 DOF (translation and rotation). An Arduino microcontroller and a Bluetooth modem were used to transfer the IMU data and button interactions to the computer over a wireless connection. When using the input device, a 3D arrow was rendered on top of it to clarify where exactly the interaction would take place.

The menu and UI were the same for both mouse and 3D input device conditions, except for an additional "6DOF Tool" input metaphor which allowed controlling both translation and rotation of the object at the same time, similar to holding a real object.

The work area was about 1×1.5 meters and draped with patterned cloth in order to achieve constant lighting conditions and facilitate tracking. I prepared the real environment to be deliberately empty (thus highly similar to the VR environment) in order to measure the effect of AR in and of itself, without possible side effects from helpful or hindering objects in the real environment. Thus, I also avoided any possibility for participants to accidentally bump into objects in the VR condition, or distract the user in the AR condition.

In the mouse condition, a plastic board (45×31 cm, 3 mm thick) was placed on the participant's lap as a surface for the mouse.

6.5. Procedure

The study included four conditions: using the mouse in an AR setting, using the 3D input device in an AR setting, using the mouse in a VR setting, and using the 3D input device in a VR setting, all of which were performed seated (see Figure 6.1).

The conditions were performed in a Latin square balanced order. The first time either the mouse or 3D input device were used, a tutorial was displayed in the AR or VR environment that explained the complete set of functions available. This gave the participant some time to practice and ensured that the task was correctly understood.

This practice trial was not used in the later task performance analysis.

In the subsequent trials, participants were measured on task completion time on a 3D object selection and transformation task of primitive 3D objects. See Figure 6.2 for an example execution of the task. A textured 3D box and a semi-transparent “goal” object were displayed. The task was to position the “source” object in the same way and at the same place as the “goal” object, by manipulating translation (x, y, z), rotation (yaw, pitch, roll), and scaling in each dimension (width, height, depth). Thus the task required the participant to manipulate the object in 9DOF. The source object was set at the scene origin, aligned with the world coordinate system, and at unit scale at the start of each trial. The goal was positioned at random for each trial with the following constraints: position was always above the ground plane and between 9 and 10 units away from the origin, the scale in each dimension ranged from 0.5 to 3 times the size of the source object. The rotation was randomly chosen without any restriction. The scene was automatically positioned in a 70cm wide (side to side) work area with the source object at the origin in the center, 60cm in front of the user and 35cm below the users’ head. Thus the source object appeared around 4cm in size with the goal around 35cm away. The task was completed when certain precision thresholds were met. These thresholds were 0.15 units in Euclidean distance (in any direction), 8 degrees of rotation (around any vector), and 0.5 units difference in scale (sum of the 3 dimensions of scaling). When all conditions were met a sound would ring to inform participant and experimenter that the task was completed and the next task could be started.

The sample consisted of 24 volunteer participants (one female, 23 male; ages 22 to 43, average 27.9 years; all right-handed), which were selected among university students and staff members. Different from the previous study on Positional Head-Tracking, I did not have access to experienced 3D artists as test subjects, so I made sure that the UI was sufficiently understood by demonstrating the 3D input device, and providing a tutorial for each UI, which was completed in the AR or VR task environment. Before the participants started using the prototype, I determined their ocular dominance with a Miles test (12 right), explained the basic concept of the user study. Before starting each condition, I reminded the participants that time was the critical factor in task completion.

Immediately after the participant had completed the final placement task for each condition, I asked him or her to rate the current feeling of comfort on a scale from one to ten, where one would mean least comfortable and ten would mean most comfortable. These ratings were only intended for a within-subject comparison of relative comfort from one condition to another, so I gave no further guidelines on how to use the rating system.

I recorded six trials per condition for every participant. Throughout the study, I slightly optimized the process of conducting the measurements and fixed technical issues that caused problems, however, the task and criteria always remained the same. The first trial in each condition was discarded as training. Every participant completed all conditions in one session, with one single exception where technical difficulties caused a delay of 30 minutes, in which I allowed the participant to temporarily leave the experimentation area. On several occasions, I encountered problems, either of technical nature or because the participant got confused and reached a state from which he or she could not easily return or reach the goal. In these cases, I reset the condition to a new random state and asked the participant to repeat the task. On some occasions, I accidentally took less than or more than four (non-training) measurements. In order to satisfy all requirements for our statistical analysis strictly, I therefore had to exclude certain data. In those cases where I accidentally took additional measurements, we discarded all measurements after the fourth. In those cases where I did not take enough measurements, I excluded the participant's data from the statistical analysis altogether, which was the case for three participants. Thus the final data set used for the statistical analysis contained 336 measurements from 21 participants.

Since I was working with time as our main metric, the immediate measurements were not producing perfectly normally distributed residuals. In order to meet the requirements for ANOVA, I performed a logarithmic data transformation with base ten [39]. The resulting residuals were approximately normally distributed, which I ascertained with both a Shapiro-Wilk test ($W \approx 0.99, p \approx 0.85$) and an Anderson-Darling test ($A \approx 0.26, p \approx 0.69$), as well visually by generating a histogram and a QQ-Plot (Figure6.3). I further performed a Barlett's-Test for equal variances ($\chi^2 \approx 3.92, df = 3, p \approx 0.27$).

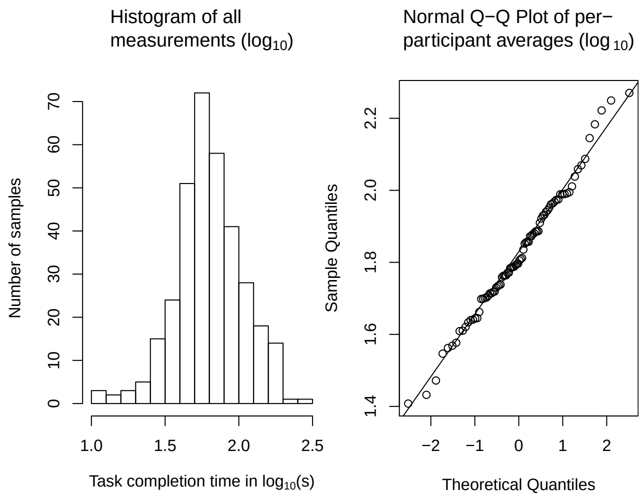


Figure 6.3. Normality tests for the resulting data. All data points used are the base ten logarithm of recorded task completion time.

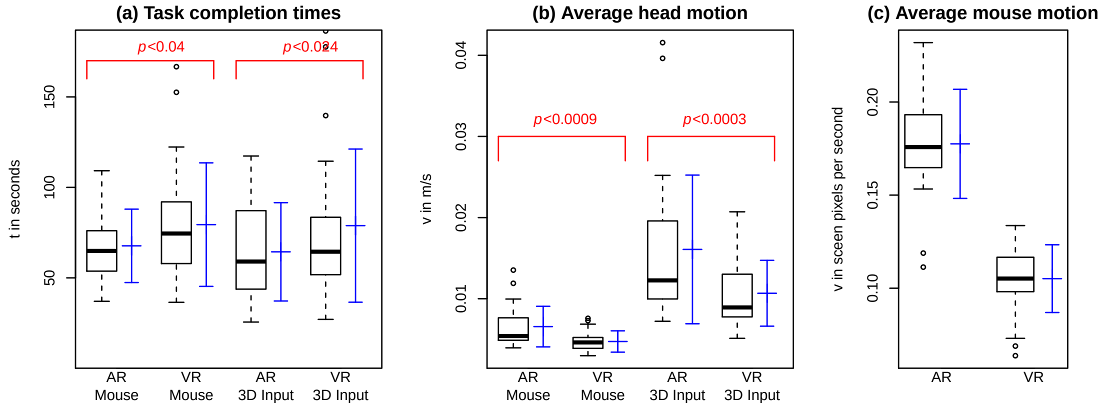


Figure 6.4. Experimental results. Blue bars indicate mean and standard deviation. Circles indicate outliers. The *p*-values were obtained by a paired (within-subject) *t*-Test, where each data point is one participants average in that condition.

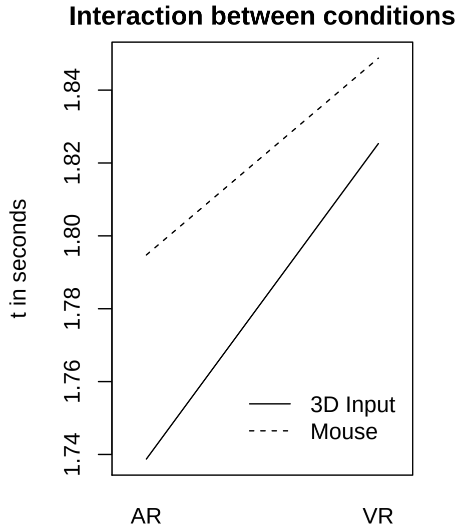


Figure 6.5. Interaction effect between the conditions.

6.6. Results

A summary of the recorded measurements can be seen in Figure 6.4(a). A two-way repeated measures ANOVA of participant performance showed a significant effect of the environment (AR or VR) with F (1,328) $\approx 4.1, p < 0.044$ (effect size $\eta^2 \approx 0.023$) but not for the input device used F(1,328) ≈ 0.17, $p > 0.68$). I did not find evidence for a significant interaction effect between the conditions ($F(1,328) \approx 0.4, p > 0.52$; Figure 6.5). In the AR condition, average task completion time was reduced by about 14.5 seconds ($\approx$ 18%; $p < 0.024$ on a paired (within-subject) t-Test). Thus I have found supporting evidence for H_1, and we can accept the hypothesis that AR has beneficial effects on task performance compared to VR when using a 6DOF input device. However, I also found similar (though reduced) effects when participants were using a mouse and therefore have to reject H_2: participants performed in fact ≈ 11.7 seconds or $\approx 14.8\%$ ($p < 0.040$ on a paired (within-subject) t-Test) faster in the AR condition. Following the experimentation guidelines developed by Keppel [40], I considered these analyses as planned comparisons between the conditions as they were stated as hypothesis. Therefore, no error correction for multiple comparisons was performed.

This raises interesting questions as to what the observed speed-up in AR over VR can be attributed to. The ability to see ones' own hand was only relevant in the 3D Input Device condition, so a naive estimate could be that $\approx 40\%$ of the observed performance improvement in the 3D input device condition stems from the improved hand-eye coordination due to visual feedback. However, I have no means of validating this assumption. A mundane explanation for this difference in task performance which I cannot rule out is that participants simply preferred the different background. The VR environment was mostly white, except a base that resembled a circular checkerboard. The AR environment was in a gray pattern with faintly saturated color spots. Therefore, some participants may have found the virtual objects to be more clearly visible against the darker background, even though no participant mentioned anything similar.

In order to investigate the reasons behind the differences in performance, I performed three post-hoc analyses on additional data retrieved from the log files of the experiment. I analyzed the recorded translational motion data of the HMD (Figure 6.4(b)) and found that participants moved their head more when using the 3D input device than when using the mouse (on average $\approx 50\%$ increase; $F(1,489) \approx 200$, $p < 10^{-15}$ on a repeated-measures ANOVA). This is to be expected, since the increased degrees of freedom and range of motion of hand movements may cause secondary motion in the head. However, I also found that even in the mouse conditions, participants moved their heads significantly more in AR than in VR (on average $\approx 40\%$ increase; $F(1,489) \approx 50$, $p < 10^{-11}$ on a repeated-measures ANOVA).

This raises the question whether participants in the AR condition were looking at the mouse. In order to verify this possibility, I further analyzed the ratio between horizontal head rotation (yaw) and vertical head rotation (pitch) and found that on average, horizontal rotation was prevalent in all conditions. Interestingly, the dominance of horizontal rotation over vertical rotation was even more pronounced in the AR conditions ($\approx 19\%$ increase in both the mouse and 3D input device conditions) than in the VR conditions ($\approx 14\%$ increase when using the 3D input device, only $\approx 8\%$ when using the mouse), meaning that participants on average even reduced their relative vertical head rotation in the AR conditions in favor of more horizontal rotation. However, these findings were not statistically significant, indicating that the ratio between horizontal and vertical head rotation was less based on experimental conditions than on personal preference. Since the cameras attached to the HMD had a 75° vertical field of view and the mouse was positioned on a board on the participants' lap, I estimated that in order to observe the mouse, a 50° downward angle would have been required. However, only one single participant achieved such a low angle in only three measurements of one single condition (AR, using the 3D Input Device). The median lowest angle in all measurements was 18° (the mean lowest angle was 19.6° with a standard deviation of 10°). This strongly indicates that the increase in performance and vigor in the AR conditions was unrelated to the ability to look at the mouse.

Another possibility is that participants were more effective simply because they were more engaged, without having a better strategy to solve the task. In order to verify this theory, I analyzed the recorded mouse motion, and again found a significant difference between AR and VR. In AR, participants on average moved the mouse more than 60% faster than in VR ($p < 10^{-13}$ on a within-subject *t*-Test; Figure 6.4(c)). This could either mean that the VR environment had a "stifling" (due to spatial unawareness) or "calming" (due to reduced visual load) psychological effect, or that the AR condition was just "more exciting" to the participants. It should be noted that since these three analyses were non-planned post-hoc comparisons, the required significance level per test was corrected to $\alpha \approx 0.01667$ (Bonferroni) or $\alpha \approx 0.01695$ (Šidák) in order to keep the Type I error rate at $\alpha = 0.05$ overall.

Regarding the subjective level of comfort, I found no significant difference between all four conditions. While it was obvious during the execution of the user study that most users found some conditions more comfortable than others, in the overall analysis these individual preferences did not produce a significant overarching tendency towards any of the conditions. A two-way repeated measured ANOVA showed only non-significant differences ($F(1,93) \approx 0.18$, $p \approx 0.67$ for mouse vs. 3D input device, $F(1,93) \approx 0.56$, $p \approx 0.46$ for AR vs. VR; Figure 6.6).

Therefore, H_3 was not supported, neither for differences regarding the input device and regarding the environment. One reason for this could be excitement over the novel input method combined with the rather short use time. Effects that are detrimental to comfort such as arm strain might not be felt immediately by participants who are excited about the ability to directly interact with virtual objects. Longer sessions might yield different results.

The same can be theorized about the missing personal preference between AR and VR. In longer work sessions, people might prefer the ability to see their surroundings in order to interact with real objects or communicate with their peers more naturally, but in this short task, no general trend became apparent.

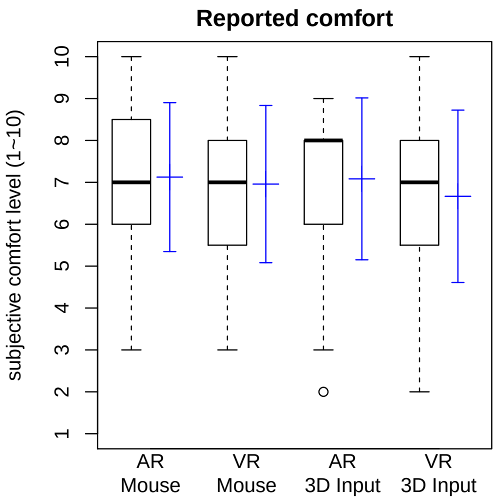


Figure 6.6. Subjective level of comfort. Blue bars indicate mean and standard deviation.

6.7. Discussion

In this study, I investigated the difference between AR and VR in an object selection and transformation task and found that AR consistently outperforms VR, even when using a 2D mouse as input device. While I found some indications as to why this might be the case, further research is required to determine specific factors and their effect.

One important question is how different physical environments affect users. The test environment was a separated area of approximately $1.5m^2$ in one corner of the room, held in a neutral gray color and without the possibility of people getting close to the participant. Real-life conditions that differ from this set-up may produce different results. A wider area might encourage users to move more, even in a purely VR environment, or clutter in the background might distract users in the AR condition. People walking around in the area could also have an effect, as it may make users feel uneasy not being able to see them in a VR set-up. The possibility to rest one's elbows on a table surface may improve comfort and precision when using 3D input devices and may thus affect performance as well.

Another important factor in AR is the quality of the video images. The cameras that I used only provide a low frame-rate, low resolution, and suffer from visible delay. Better cameras may influence the results even more in the favor of using AR. When using optical see-through HMDs, different factors come into play, such as differences in accommodation depth between the virtual objects and the real environment.

I also did not consider long-term effects. One possibility is that users suffered from a temporary disorientation when entering the VR environment, which might have waned during longer use sessions.

Finally, I have to acknowledge that in normal applications VR can be realized much easier than AR and will usually be more responsive and less computationally expensive since the system does not have to wait for camera images and perform undistortion and rectification algorithms. In the prototype, the performed algorithms were exactly the same, with the only difference being that I discarded the video image in the VR condition. I did this in order to eliminate effects from differences in system performance. In real-life applications, however, this very difference may be of importance.

This work represents a step towards understanding differences between AR and VR, and opens up several new venues for future research work.

The most interesting question that this study raised is the fact that even when using a 2D mouse, seeing the real environment has a positive effect on task performance, even though it adds no immediately relevant information for the task. This points towards the possibility that AR increases a users engagement, or that reduction of environmental complexity has a slowing effect on users. The former could be tested with a simulated AR system [16] that displays alternate versions of the real environment, on the hypothesis that any virtual environment is preferable to a completely empty environment, even if it does not represent reality. For the latter possibility, one factor might be the users' trust in knowing the boundaries of the workplace. A future study where the system displays a simple bounding volume could shed more light on this possibility. In this study, I only considered a 2D mouse and a 3D wand which was co-located to the real hand position, but not a desktop 3D input device. Such a device is a third alternative that — while allowing 3D interaction — is not co-located with the real hand. The differences in task performance between AR and VR conditions when using a desktop 3D input device could be close to that of a mouse — stressing the importance of seeing one's own arm — or to that of the 3D wand — indicating that AR somehow may provide better spatial understanding when performing 3D transformations.

This study only gives an estimate of the effect of all visual stimuli in AR combined and can therefore not predict which specific stimuli are important to task performance.

Especially visual feedback of the own arms and hands may be a critical factor. Future studies are required to dissect the measured performance difference which I presented in this chapter into its main factors.

I also only considered the case of users sitting in a chair. Given that the work area was small enough to make walking unnecessary, it is possible that the results may be similar when users are standing upright. However, further studies are required to validate this assumption.

Finally, I only analyzed participants' head motion as an indicator of the source of the performance difference. Future studies could also gather data on eye movement, dilation of pupil size (cognitive load), heart rate, and other indicators of subjective experience.

CHAPTER 7

Lessons Learned from Real-Life Adoption

Throughout my research, I continued to develop my prototype which was eventually named “MARUI”. MARUI is an abbreviation, where “AR UI” stands for Augmented Reality User Interface, and the “M” can interchangeably stand for “Modeler”, “Maya”, “Max”, or “Marui”.

As the prototype grew more and more sophisticated, I eventually decided to release it to the public. This allowed me to test the insights that I had gained through my lab research on whether they are valid for real-life application by professional users. On some occasions, I personally demonstrated MARUI to potential users and was thus able to observe their immediate reaction and first attempts at using it. However, most of the information I gathered over time was from people who downloaded MARUI from the website and used it on their own hardware and projects, without me being able to gather direct information on the user or how MARUI was used. In these cases, I rely on the feedback which was sent to me by users voluntarily, as well as the anonymous usage statistics that I gathered from active users. Since the UI is based on and required Autodesk Maya—which is a fairly expensive and complicated program aimed at professional production—I can be certain that all users at least had a minimum level of expertise in the area. In this chapter, I share some key experiences so far as they relate to my research on 3D UIs for professional design.

7.1. Release history

7.1.1 First demo

The first prototype version of MARUI was demonstrated to participants or the International Symposium for Mixed and Augmented Reality (ISMAR) September 2014 in Munich. This version was still working with a Vuzix Wrap1200 VST HMD and PTAM to do the required tracking, as well as with a cable-bound glove prototype.

7.1.2 Consumer VR version (v1.3)

In July 2015, the prototype had fully evolved into a stable and functional user interface using the Oculus Rift DK2 (which, at this point, was fairly popular and widely distributed). Therefore, I decided to release MARUI to the public by creating a website and putting up compiled binaries for people to download and use for free. User interaction still required a custom-made glove which I was unable to deliver to potential users, but MARUI could also be controlled with a mouse. Thus, everyone who owned an Oculus Rift DK2 could have a similar experience as participants in the mouse condition in the user studies described in Chapters 5 and 6. Downloads soon began to pick up and first feedback arrived.

7.1.3 HTC Vive support (v1.4)

In June 2016 I was able to acquire an HTC Vive VR HMD and added support for the device to MARUI. This meant that owners of the HTC Vive could now use the 3D UI developed during my research for the first time. This was originally released as v1.3.3, as I was still just counting up build numbers, but then jumped to v1.4 to signify the change. Furthermore, in order to learn more about how users were using MARUI, I implemented anonymous usage data collection starting from v1.4.6.

7.1.4 Sales version (v1.5)

One important concern in developing business-to-business (B2B) software is that it is unclear whether the software actually solves any real production purpose or whether it is just kept out of curiosity, as is often the case with free software. Evaluating how well a production need is addressed can be done by setting the price. Professionals will only pay for solutions to the extent that the solution provides benefits—or at least perceived future benefits—to their work. Thus, in November 2016 I decided to move away from the free beta to a monthly subscription system where users had to pay 2000 JPY to keep using the software. This allowed me to be sure that customers were continuously using MARUI and did not just buy it out of curiosity without adopting it.

7.2. User Feedback

Throughout the development, users of MARUI kept getting in touch with me to report problems, request new features, or just to share their opinion on 3D UIs. While I was never able to actually visit the user and see for myself how and for what work they used MARUI, their questions and comments gave some insights into the potential applications and challenges associated with novel 3D UIs. For confidentiality reasons, I refrain from providing any information that could be used to identify individuals or companies that use MARUI. Both well-known international organizations and individuals are using or have used MARUI for some time.

7.2.1 Look-Through-Selected and Saving Viewpoints

In June of 2016, a 3D CG artist working at a Japanese game company contacted me with regards to trouble in the constraint system. Constraints are a feature of MARUI that allows users to attach virtual objects to VR devices (such as the HMD or controllers) and vice versa. Thus it can, for example, be used to make a virtual camera in Maya follow the users' view (i.e. HMD transformation) as he walks through the scene.

It quickly became apparent that is was not a bug, but a usability issue. The artist wanted to assume the position of a camera in the scene—presumably that of the viewpoint where the game character would later stand—while still preserving some freedom to look around. This was a mathematically different problem because in order to be usable it dictates that the cameras "up" vector should become the "up" vector in the real world. Therefore, we decided to implement a new tool: "look-through-selected" which allows the user to temporarily assume the position and orientation of a selected object. However, a sudden "jump" to the new position creates confusion even when the user knows which object he will move to. To compensate this, a transition with a one second transition time was implemented, where the user "flies" into the new position. In order to make it a smooth experience, ease-in and ease-out was adopted with $t_{smooth} = cos((1-t)*\pi)/2+0.5$, where t is a linear interpolation between zero and one over the time of one second, and T_{smooth} is the value that was actually used to calculate the linear intermediate transformation at any point in time. Rotation was interpolated with the use of quaternions, which allow more simple calculation of intermediate steps along the shortest possible rotation from one pose to another.

In March or 2016 a large automobile manufacturer requested an extended and simplified version of this feature where one could save the current position from where one was looking at the scene and restore it at a later point in time. Multiple of these "save-points" would be used and re-used during the design process. This feature was fairly easy to implement since I only had to save the current navigation data, which is a simple transformation matrix. Again, I used a smooth transition to avoid disorienting the user when switching from one viewpoint to another.

7.2.2 Customization and scripting API

While there were numerous detail improvements on the UI in order to make it more useful in the common use cases of modeling and animation, it became clear that users very much wanted to customize their experience by mapping different functions to the keys of the controllers.

This showed once again the great need in professional design to customize the software to fit specific production needs. This again reflected the findings from my survey (see Chapter 3 Section 3.2), where over 95% of participants used scripts and plug-ins to extend or change their 2D UI and over 50% even developed their own scripts and plug-ins to do so, Similarly, users of my own 3D UI asked me to implement ways for them to customize their experience by changing the assignment of functions to buttons and building their own menus.

While consumer UIs usually are expected and required to provide one implementation that fits all users, it seems to be an absolute necessity for production UIs to allow customization, even at the expense of making the adoption of the UI a much more technically challenging undertaking.

7.2.3 Performance

One critical point in consumer VR systems is rendering performance. Since the image has to be rendered twice (one for each eye) and at a higher framerate (60fps in order to avoid jittering and cyber-sickness) the demands are greatly increased compared to desktop applications. In the case of VR games and experiences, the developers have full control over the environment that the user sees, allowing all forms of optimizations and limiting the complexity of the virtual environment. In modeling apps, however, scene organization and complexity is decided by the user, who can load arbitrarily large and complex scenes into the modeling application. For desktop applications, this is not a big problem, since the UI remains responsive even if the framerate drops significantly. However, in VR or AR applications, a drop in rendering performance quickly leads to the application being unusable. Making MARUI more efficient was therefore a common request and a constant effort.

7.2.4 Desktop

While the great advantage of 3D UIs lie in the 3D interaction, there seems to be a great need to provide support for the 2D desktop UI at least for some time into the future. Several users requested a feature to see and use the Windows desktop in VR and—once implemented—the reaction in live demos was usually surprisingly positive.

7.2.5 Learning the UI and Cyber-sickness

In those cases where I demonstrated and introduced MARUI to potential users who had no prior experience with VR, I commonly encountered difficulties with explaining basic UI concepts and cyber-sickness. Even concepts that are most basic for 2D UIs such as "clicking" a button with the index finger in order to select an object below the cursor are not immediately understood by all users. Furthermore, since participants in these demonstrations usually did not have enough time or experience to calibrate the system optimally to their bodies—such as setting the interpupillary distance correctly—they often complained about experiencing first signs of cyber-sickness, such as a sense of disorientation or eye strain. While these symptoms usually subside with length of use and proper calibration, it was difficult to convince first-time users to adopt using MARUI, even when they saw the potential of 3D UIs positively.

7.3. Limitations

While MARUI supports both the Oculus Rift and HTC Vive, it only supports Autodesk Maya on Microsoft Windows operating systems. As one potential user pointed out–and I know from my own prior employment in the industry—a larger portion of Visual Effects companies are using some form of the Linux operating system. This limits the scope of MARUI as an experiment in professional application, as only those industries and companies that use Windows can use it.

"Mobile-VR" HMDs like the Samsung GearVR are not supported, mainly because they are stand-alone independent systems that require a different approach to engineering in order to connect the smartphone app to the PC running Maya.

The AR functionality including video capturing, image undistortion and rectification, and stereo depth reconstruction are still available in MARUI. However, using this functionality is very complicated and requires some understanding of camera geometry as well as external software to perform the calibration. Judging from the feedback that I received, only very few users make use of this functionality as of the time of this writing. Future AR HMDs that make the process easier may lead to a more widespread use.

7.4. Conclusion

At the time of this writing, MARUI is still under active development with new releases every month. Thus, new knowledge on how 3D UIs can be useful to professional designers is generated constantly. During the free beta version period, approximately 1000 people from all over the world downloaded MARUI, and while the sales version was not granted a similar success, the number of subscribers kept climbing steadily. Many users have written me or commented on the Facebook page or Youtube videos to express their excitement about the new technology. Since the goal of my research was to bridge the gap between laboratory research and real-life adoption, seeing artists use the technology I developed and hearing their feedback has been the most rewarding part of my work.

CHAPTER 8

Conclusions

In this dissertation, I have summarized the research work I performed towards enabling 3D UIs for 3D design work. Part of this research was already performed and published in my Masters Thesis in 2014, but continued and extended upon to produce novel insights and data. Focusing on professional 3D design work allowed me to keep my research efforts closely application oriented and produce results that can easily be transferred to real life.

My original hypothesis was threefold in that a 3D UI that is more efficient than current 2D UIs must be possible, that the factors for these must be deducible by empirical research, and that the 3D design professionals would readily adopt such a 3D UI if available, even at a non-negligible price. While it is too early to speak of a wide adoption, the early success and feedback gained through the process outlined in this thesis seem to validate this hypothesis.

In this chapter, I will reiterate the contributions I made through this research, show where my research showed limitations, and which future work is necessary in order to bring 3D UIs to media professionals worldwide.

8.1. Contribution

Chapter 3 gave an insight into how artists currently work. In addition to the survey, which was already published in my Masters Thesis, I performed the time-based analysis of the break-down of recordings of artists at work. This gives some detail information on how 3D professionals actually use their software. Both of these sources of information combined—survey and detail analysis—can allow researchers and developers to make informed decisions on how what to focus on.

Chapter 4 provided both a description of the experimental 3D UI which I created in the course of this work, as well as the results of the formative evaluation I have performed to improve its design and find shortcomings in the technology used to build it. The insights gained through this process may help others who are attempting to design 3D UIs.

Chapter 5 provided new evidence and quantitative date in the ongoing debate about positional head-tracking. While the study itself could not provide conclusive evidence for any effect, the statistical analysis provided some insight into how large performance improvements might be and how large and experiment to determine the precise effect size would have to be. Researchers attempting to establish the precise benefits from positional head-tracking can use this data to make decisions in their experimental design and execution. Furthermore, it provides positive evidence that the effect of positional head-tracking on task performance may not be large enough to justify the adoption of technologies enabling the effect.

The user study comparing AR and VR task performance (Chapter 6) provided not only a quantitative estimate for the long-held belief that seeing one's own body may be beneficial to task performance, it also discovered a novel effect that was not previously researched: participants were performing faster when they saw the real world, even when using a mouse and thus not actually seeing their own body. I have provided secondary analyses to provide possible explanations for this effect. This result is both of interest to researchers—who may be interested in discovering the precise reasons and form of AR performance benefits, as well as to developers who may be considering whether or not to support or adopt AR technologies.

Finally, Chapter 7 provided insights into real-life adoption of a novel technologies that most lab-based research cannot provide. While it is highly specific to the case the the MARUI project and its respective users, it can be useful to user interface developers in general who desire to bring novel technologies to the market.

8.2. Limitations

In order to be able to explore the topic, I had to limit the scope of factors that I wished to analyze.

I only considered a sub-set of 3D UI systems: those that are realized using HMD systems. CAVE systems or hand-held devices were omitted. And even in the subset of HMD systems, I relied on off-the-shelf VR devices and cameras for VST AR experiments. I did not evaluate custom made HMDs and did not concern myself with the hardware factors of HMDs such as resolution, weight, and field of view.

This also led to abandoning the concept of a pinch glove as main input device, since no such gloves are readily available to the potential users of MARUI and my experiences outlined in Chapter 4 Section 4.5 and Chapter 5 have revealed the difficulties in creating usable glove-based input devices. While I still believe that direct hand interaction provides benefits over controller-based interaction, my research in this respect showed little success.

Regarding the input devices which I used for my research, I mostly relied on self-made devices, which may have affected measurements to some degree. It is definitely not advisable to see our results as a definitive comparison between mouse-based and 3D-device-based user interfaces. Conclusive studies that compare novel off-the-shelve 3D input devices to traditional mouse-based user interfaces are required.

Finally, due to the limited availability of different types of hardware on the current market, most users of MARUI can follow neither of the suggestions derived from the user studies in Chapter 5 and Chapter 6, that is to either reduce cost by using mobile VR systems or improve efficiency by using AR systems.

8.3. Future Work

One factor that is constantly changing and requires us to reevaluate prior research is technological progress in the form of new devices. Next-generation HMDs may be more convenient to wear, offer higher resolutions and better tracking. This may affect the effects that I found in my research.

It is also important to remember that at the time of this writing, 3D user interfaces are novel to most consumers and professionals. This means on the one hand that they are more challenged in understanding and efficient use of 3D UI concepts. However, on the other hand, it also means they are less set in their ways and more open to exploring alternative interaction concepts. This is sure to change over the next years as more and more people will be exposed to 3D UIs. It is of great research interest how the perception and efficiency of 3D UIs changes over these years.

Finally, the adoption of future versions of MARUI that support native AR and mobile VR systems may validate the findings of the user studies in Chapter 5 and Chapter 6.

Since I first started working on MARUI, the concept has changed and evolved over time, but always with the goal of providing a complete and convenient 3D UI for professional design work. At the current stage, first advantages over the 2D UI become apparent, however the associated cost and effort combined with current limitations in the UI itself limits its attractiveness. Some more research and development work is required to “get over the hill” to where the novel 3D UI is clearly superior and attractive for adoption in daily work.

Publication List

Journal Articles

1. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato. Augmented Reality vs Virtual Reality for 3D Object Manipulation. In *IEEE Transactions on Visualization and Computer Graphics*, 14(8): 50–59, 2017.

2. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, Hirokazu Kato and Steven Feiner. Evaluating the Effect of Positional Head-Tracking on Task Performance in 3D Modeling User Interfaces. In *Elsevier Computers & Graphics Journal*, 65: 22–30, 2017.

3. Jaakko Hyry, **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, Hirokazu Kato and Petri Pulli. Design of Assistive Tabletop Projector-Camera System for the Elderly with Cognitive and Motor Skill Impairments. In *ITE Transactions on Media Technology and Applications*, Volume 5(2), 57–66, 2017.

Peer-Reviewed Conference Publications

1. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, and Hirokazu Kato. Towards Augmented Reality User Interfaces in 3D Media Production, In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 23–28, Munich, Germany, September 2014.
2. Goshiro Yamamoto, Jaakko Hyry, **Max Krichenbauer**, Takafumi Taketomi, Christian Sandor, Hirokazu Kato and Petri Pulli. A user interface design for the elderly using a projection tabletop system. In *Virtual and Augmented Assistive Technology (VAAT)*, pp. 29–32, 2015.

Other Conference and Workshop Publications or Presentations

1. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, Hirokazu Kato and Steven Feiner. Evaluating Positional Head-Tracking in Immersive VR for 3D Designers., In *the 15th IEEE International Symposium on Mixed and Augmented Reality*, Merida, Mexico, September 2016.
2. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato. User Study on Augmented Reality User Interfaces for 3D Media Production, In *Doctoral Consortium of the 14th IEEE International Symposium on Mixed and Augmented Reality*, Fukuoka, Japan, October 2015.
3. **Max Krichenbauer**, Goshiro Yamamoto, Takafumi Taketomi and Hirokazu Kato. MARUI — A novel AR user interface for high-level 3D design, In *Proceedings of the 7th Korea-Japan Workshop on Mixed Reality*, Seoul, South-Korea, April 2014.

Acknowledgments

First and foremost, I want to express my sincerest gratitude to Professor Hirokazu Kato, who accepted me first as a master's student and then doctoral student. Over the course of a total of five years he guided and supported my efforts in every way, while still giving me the freedom to pursue my own ideas and dreams. Thanks to his kindness and patience, I was able to learn about conducting research and state-of-the-art User Interface technologies.

Furthermore, I want to thank all the other committee members. Firstly Associate Professor Christian Sandor, who's vast knowledge was of critical help in many situations along the way. Thanks to Professor Kiyokawa and Professor Klinker for their immensely helpful feedback on this thesis. I also want to thank Assistant Professor Taketomi Takafumi whose sense of duty and overwhelming work ethic was a major inspiration at all times, as well as a much needed help.

Special thanks also go to Assistant Professor Goshiro Yamamoto who supported me greatly in the day-to-day work required of me. I also would like to thank my former co-supervisor Professor Naokazu Yokoya who gave valuable feedback on my Master's thesis before. I am also very thankful for all the help and support I received from Makiko Ueno, secretary of Interactive Media Design Lab who has helped me greatly to meet the challenges of studying at a Japanese University.

I want to express my appreciation to all participants in my studies. In particular I want to thank Noriko Shimonishi of Kyoto Saga University of Arts.

I want to thank all current and past members and students of the Interactive Media Design laboratory, who often helped me with productive feedback and encouragement.

Finally, I would like to express my deepest gratitude towards my parents, my family, and my friends back home in Germany, who have supported me on my journey.

Bibliography

[1] Juan-J Arino, M-Carmen Juan, Jose-Antonio Gil-Gómez, and Ramón Mollá. A comparative study using an autostereoscopic display with augmented and virtual reality. *Behaviour & Information Technology*, 33(6):646–655, 2014.

[2] Roland Arsenault and Colin Ware. Eye-hand co-ordination with force feedback. In *SIGCHI conference on Human Factors in Computing Systems*, pages 408–414. ACM, 2000.

[3] Kevin W Arthur, Kellogg S Booth, and Colin Ware. Evaluating 3d task performance for fish tank virtual worlds. *Transactions on Information Systems (TOIS)*, 11(3):239–265, 1993.

[4] Ronald T Azuma. A survey of augmented reality. *Presence: Teleoperators and virtual environments*, 6(4):355–385, 1997.

[5] Bartosz Bajer, Robert J Teather, and Wolfgang Sturzlinger. Effects of stereo and head tracking in 3d selection tasks. In *Symposium on Spatial user interaction*, pages 77–77. ACM, 2013.

[6] Martin S Banks, Tandra Ghose, and James M Hillis. Relative image size, not eye position, determines eye dominance switches. *Vision research*, 44(3):229–234, 2004.

[7] István Barakonyi and Dieter Schmalstieg. Augmented reality in the character animation pipeline. In *SIGGRAPH 2006 Sketches*, page 75. ACM, 2006.

[8] Devesh Kumar Bhatnagar. Position trackers for head mounted display systems: A survey. *University of North Carolina, Chapel Hill TR93*, 10, 1993.

[9] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and applications*, 21(3):6–8, 2001.

[10] Gunnar Borg. Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian journal of work, environment & health*, pages 55–58, 1990.

[11] James Boritz and Kellogg S Booth. A study of interactive 3d point location in a computer simulated virtual environment. In *Symposium on Virtual reality software and technology*, pages 181–187. ACM, 1997.

[12] James Boritz and Kellogg S Booth. A study of interactive 6 dof docking in a computerised virtual environment. In *International Symposium on Virtual Reality*, pages 139–146. IEEE, 1998.

[13] Sanne MBI Botden, Sonja N Buzink, Marlies P Schijven, and Jack J Jakimowicz. Augmented versus virtual reality laparoscopic simulation: what is the difference? *World journal of surgery*, 31(4):764–772, 2007.

[14] AC Boud, David J Haniff, Chris Baber, and SJ Steiner. Virtual reality and augmented reality as a training tool for assembly tasks. In *International Conference on Information Visualization*, pages 32–36. IEEE, 1999.

[15] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley, 2004.

[16] Doug A Bowman, Cheryl Stinson, Eric D Ragan, Siroberto Scerbo, Tobias Höllerer, Cha Lee, Ryan P McMahan, and Régis Kopper. Evaluating effectiveness in virtual environments with mr simulation. In *Interservice/Industry Training, Simulation, and Education Conference*, 2012.

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

[18] Jeannette Unge Byström, Gert-Åke Hansson, Lars Rylander, Kerstina Ohlsson, Gabriella Källrot, and Staffan Skerfving. Physical workload on neck and upper limb using two cad applications. *Applied ergonomics*, 33(1):63–74, 2002.

[19] Ana Amélia Amorim Carvalho. Usability testing of educational software: methods, techniques and evaluators. *Actas do 3^{o} Simpósio Internacional de Informática Educativa*, pages 139–148, 2001.

[20] Marina A. Cidota, Rory Clifford, Stephan G. Lukosch, and Mark Billinghurst. Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 172–177. IEEE, 2016.

[21] James H Clark. Designing surfaces in 3-d. *Communications of the ACM*, 19(8): 454–460, 1976.

[22] Yann Coello. Spatial context and visual perception for action. *Psicologica: International Journal of Methodology and Experimental Psychology*, 26(1):39–59, 2005.

[23] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.

[24] Laura Dipietro, Angelo M Sabatini, and Paolo Dario. A survey of glove-based systems and their applications. *Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(4):461–482, 2008.

[25] Renee Dunlop. *Production Pipeline Fundamentals for Film and Games*. CRC Press, 2014.

[26] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3):363, 1993.

[27] Peter Fiala and Nicoletta Adamo-Villani. ARpm: an augmented reality interface for polygonal modeling. In *International Symposium on Mixed and Augmented Reality*, pages 196–197. IEEE, 2005.

[28] Pierre Fite-Georgel. Is there a reality in industrial augmented reality? In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 201–210. IEEE, 2011.

[29] Bernd Fröhlich and John Plate. The cubic mouse: a new device for three-dimensional input. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 526–531. ACM, 2000.

[30] Bernd Fröhlich and John Plate. The cubic mouse: a new device for three-dimensional input. In *SIGCHI conference on Human Factors in Computing Systems*, pages 526–531. ACM, 2000.

[31] F Girbacia. An approach to an augmented reality interface for computer aided design. *Annals of DAAAM*, 2010.

[32] Michael SA Graziano, Dylan F Cooke, and Charlotte SR Taylor. Coding the location of the arm by sight. *Science*, 290(5497):1782–1786, 2000.

[33] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52: 139–183, 1988.

[34] Sarah Howlett, Richard Lee, and Carol O'Sullivan. A framework for comparing task performance in real and virtual scenes. In *Symposium on Applied perception in graphics and visualization*, pages 119–122. ACM, 2005.

[35] Sylvia Irawati, Sengpyo Hong, Jinwook Kim, and Heedong Ko. 3d edutainment environment: learning physics through vr/ar experiences. In *International Conference on Advances in Computer Entertainment Technology*, pages 21–24. ACM, 2008.

[36] J Adam Jones, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. The effects of virtual reality, augmented reality, and motion parallax on

egocentric depth perception. In *Symposium on Applied perception in graphics and visualization*, pages 9–14. ACM, 2008.

[37] M Carmen Juan and David Pérez. Using augmented and virtual reality for the development of acrophobic scenarios. comparison of the levels of presence and anxiety. *Computers & Graphics*, 34(6):756–766, 2010.

[38] Hannes Kaufmann. Construct3d: an augmented reality application for mathematics and geometry education. In *International Conference on Multimedia*, pages 656–657. ACM, 2002.

[39] Oliver N Keene. The log transformation is special. *Statistics in medicine*, 14(8): 811–819, 1995.

[40] Geoffrey Keppel. *Design and analysis: A researcher's handbook.* Prentice-Hall, Inc, 1991.

[41] Maryam Khademi, Hossein Mousavi Hondori, Lucy Dodakian, Steve Cramer, and Cristina Videira Lopes. Comparing "pick and place" task in spatial augmented reality versus non-immersive virtual reality for rehabilitation setting. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 4613–4616. IEEE, 2013.

[42] Aarlenne Z Khan and J Douglas Crawford. Ocular dominance reverses as a function of horizontal gaze angle. *Vision research*, 41(14):1743–1748, 2001.

[43] Kiyoshi Kiyokawa. 3d collaboration using mixed reality technology. In *International Symposium on Universal Communication*, pages 100–109, 2007.

[44] Kiyoshi Kiyokawa, Hidehiko Iwasa, Haruo Takemura, and Naokazu Yokoya. Collaborative immersive workspace through a shared augmented environment. In *Photonics East (ISAM, VVDC, IEMB)*, pages 2–13. International Society for Optics and Photonics, 1998.

[45] Gudrun Klinker, Allen H Dutoit, Martin Bauer, Johannes Bayer, Vinko Novak, and Dietmar Matzke. Fata morgana a presentation system for product design. In *International Symposium on Mixed and Augmented Reality*, page 76. IEEE, 2002.

[46] Brian Knep, Craig Hayes, Rick Sayre, and Tom Williams. Dinosaur input device. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 304–309. ACM Press/Addison-Wesley Publishing Co., 1995.

[47] Max Krichenbauer, Go Yamamoto, Takafumi Taketomi, Christian Sandor, and Hirokazu Kato. Towards augmented reality user interfaces in 3d media production. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 23–28. IEEE, 2014.

[48] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, and Hirokazu Kato. Augmented reality vs virtual reality for 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 14(8):50–59, 2017.

[49] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, Hirokazu Kato, and Feiner Steven. Evaluating the effect of positional head-tracking on task performance in 3d modeling user interfaces. *Elsevier Computers & Graphics Journal*, 65:22–30, 2017.

[50] Gordon Kurtenbach and William Buxton. User learning and performance with marking menus. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 258–264. ACM, 1994.

[51] Cha Lee, Gustavo A Rincon, Greg Meyer, Tobias Hollerer, and Doug A Bowman. The effects of visual realism on search tasks in mixed reality simulation. *Transactions on Visualization and Computer Graphics*, 19(4):547–556, 2013.

[52] James R Lewis. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Human Factors and Ergonomics Society Annual Meeting*, volume 36, pages 1259–1260. SAGE Publications, 1992.

[53] Daniel P Mapes and J Michael Moshell. A two-handed interface for object manipulation in virtual environments. *Presence: Teleoperators & Virtual Environments*, 4(4):403–416, 1995.

[54] Mark R Mine, Frederick P Brooks Jr, and Carlo H Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Con-*

ference on Computer graphics and interactive techniques, pages 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[55] Zhiyong Ming and Nina Zaproudina. Computer use related upper limb musculoskeletal (comrulm) disorders. *Pathophysiology*, 9(3):155–160, 2003.

[56] Andreas Möller, Matthias Kranz, Stefan Diewald, Luis Roalter, Robert Huitl, Tobias Stockinger, Marion Koelle, and Patrick A Lindemann. Experimental evaluation of user interfaces for visual indoor navigation. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 3607–3616. ACM, 2014.

[57] Sev V Nagalingam. *CIM justification and optimisation*. CRC Press, 1999.

[58] Wayne Piekarski. 3d modeling with the tinmith mobile outdoor augmented reality system. *IEEE Computer Graphics and Applications*, 26(1):14–17, 2006.

[59] Wayne Piekarski and Bruce H Thomas. Tinmith-hand: Unified user interface technology for mobile outdoor augmented reality and indoor virtual reality. In *Virtual Reality*, pages 287–288. IEEE, 2002.

[60] Wayne Piekarski and Bruce H Thomas. Using artoolkit for 3d hand position tracking in mobile outdoor environments. In *Augmented Reality Toolkit, The first IEEE International Workshop*, pages 2–pp. IEEE, 2002.

[61] James Pitkow and Mimi Recker. Results from the first world-wide web user survey. *Computer Networks and ISDN Systems*, 27(2):243–254, 1994.

[62] Chris Raymaekers, Tom De Weyer, Karin Coninx, Frank Van Reeth, and Eddy Flerackers. Icome: an immersive collaborative 3d object modelling environment. *Virtual Reality*, 4(4):265–274, 1999.

[63] Phattanapon Rhienmora, Kugamoorthy Gajananan, Peter Haddawy, Matthew N Dailey, and Siriwan Suebnukarn. Augmented reality haptics system for dental surgical skills training. In *Symposium on Virtual Reality Software and Technology*, pages 97–98. ACM, 2010.

[64] Christian Sandor, Shinji Uchiyama, and Hiroyuki Yamamoto. Visuo-haptic systems: half-mirrors considered harmful. In *EuroHaptics Conference, 2007 and*

Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint, pages 292–297. IEEE, 2007.

[65] Chris Shaw and Mark Green. Two-handed polygonal surface design. In *Symposium on User Interface Software and Technology*, pages 205–212. ACM, 1994.

[66] Y Shen, SK Ong, and AYC Nee. Augmented reality for collaborative product design and development. *Design Studies*, 31(2):118–145, 2010.

[67] Andreas Simon and Bernd Fröhlich. The yoyo: A handheld device combining elastic and isotonic input. In *INTERACT*, pages 303–310, 2003.

[68] Ivan E Sutherland. A head-mounted three dimensional display. In *Fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[69] Akinari Takagi, Shoichi Yamazaki, Yoshihiro Saito, and Naosato Taniguchi. Development of a stereo video see-through hmd for ar systems. In *International Symposium onAugmented Reality (ISAR)*, pages 68–77. IEEE, 2000.

[70] Robert J Teather and Wolfgang Stuerzlinger. Guidelines for 3d positioning techniques. In *Conference on Future Play*, pages 61–68. ACM, 2007.

[71] Roger Y Tsai and Reimar K Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *Transactions on Robotics and Automation*, 5(3):345–358, 1989.

[72] Benjamin Walther-Franks, Marc Herrlich, and Rainer Malaka. A multi-touch system for 3d modelling and animation. In *International Symposium on Smart Graphics*, pages 48–59. Springer, 2011.

[73] Colin Ware and Glenn Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *Transactions on Graphics (TOG)*, 15(2): 121–140, 1996.

[74] Peter J Werkhoven and Joris Groen. Manipulation performance in interactive virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):432–442, 1998.

[75] Valarie A Zeithaml. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *The Journal of Marketing*, pages 2–22, 1988.

APPENDIX A

Online Survey on 3D Design Work Environments

The following pages contain the complete online survey that was discussed in Chapter 3 Section 3.2, including the complete set of results.

Survey on Professional 3D Design Work: Final Results

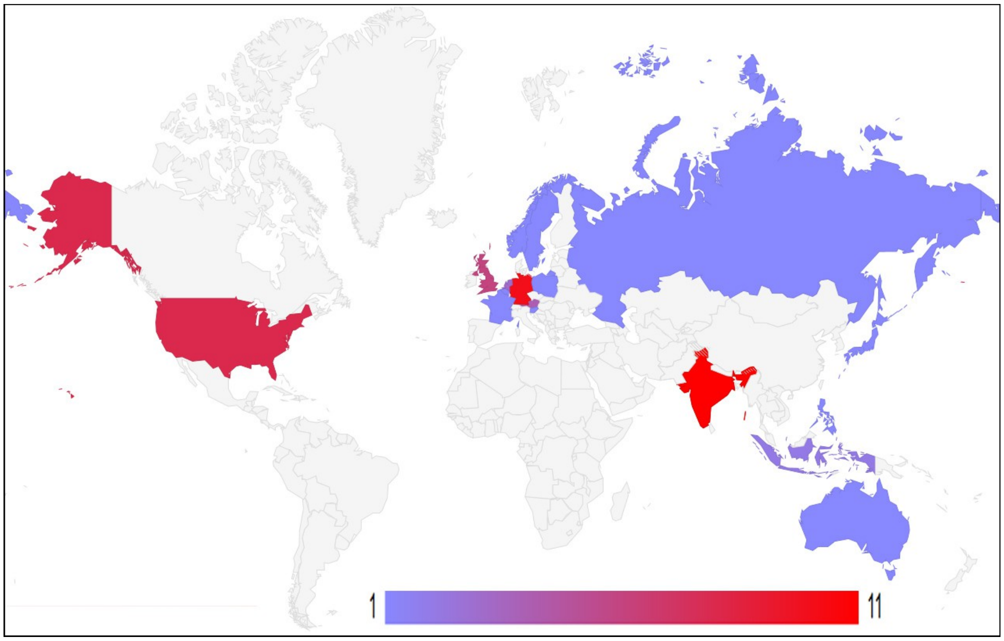
1: Introduction

In this document we provide additional information on the survey that we performed for our paper: "Towards Augmented Reality User Interfaces in 3D Media Production" (ISMAR Submission # 160).
The survey was performed using the SurveyGizmo online service (http://www.surveygizmo.com).
The logo of our laboratory was displayed on every page of the survey, and a contract address was provided.
On the following pages we present the exact questions and explanations (in *blue*) as well as the summary of answers.
Upon starting the survey, the participants were greeted with the following message:

> *Thank you very much for participating in this survey!*
>
> *About this survey:*
> - *It is part of a university research project, at IMD-Lab, Nara Institute of Science and Technology, Japan.*
> - *We are trying to capture the requirements and expectations of 3D CG media production professionals.*
> - *This survey is performed only to advance academic research – it is not used commercially.*
> - *We hope to publish the results at an international research conference, where it will reach a wider audience.*
> - *The aim is to draw attention to your work situation, making future research and development more application oriented.*
> - *Thus, we hope that this research will some day be of service to you by providing you with better user interfaces for your work.*
> - *It takes about 15 minutes.*
>
> *If you have further questions or comments, please contact us via email: max-k@is.naist.jp*

1: Number and distribution of participants:

India	11
Germany	10
United States	8
United Kingdom	6
Austria	4
Netherlands	3
Indonesia	2
Japan	1
Poland	1
Sweden	1
Slovenia	1
Philippines	1
Belgium	1
France	1
Australia	1
Russian Federation	1
Norway	1
Total:	**54**



3: Self-assessment questions:

Questions 1 through 7 were used assess the participants expertise in 3D design work,
in order to limit the observed sample to media professionals.

1. What is your connection to 3D content / media production
such as Movies (Animation, Visual Effects), TV or Games?

- *Currently working in media production.*
- *Previously worked in media production, but changed occupation or retired.*
- *Training or studying to work in 3D production.*
- *Hobbyist or amateur enthusiast.*
- *No connection.*

Please rate your expertise on the following production departments:
2. Modeling / Sculpting / Texturing (No experience, Basic knowledge, Hobbyist, Professional, Expert)
3. Rigging (No experience, Basic knowledge, Hobbyist, Professional, Expert)
4. Animation (No experience, Basic knowledge, Hobbyist, Professional, Expert)
5. Simulation / VFX (No experience, Basic knowledge, Hobbyist, Professional, Expert)
6. Lighting / Rendering (No experience, Basic knowledge, Hobbyist, Professional, Expert)
7. Production / Supervision (No experience, Basic knowledge, Hobbyist, Professional, Expert)

Participants were rejected if they chose either of the options “Hobbyist or amateur enthusiast” or “No connection” to question one.
Furthermore, participants were rejected if they failed to rate themselves with at least “Hobbyist” level expertise in at least one of the areas #2 (Modeling / Sculpting / Texturing), #3(Rigging), #4 (Animation), or #5 (Simulation / VFX).
Those participants who were rejected were presented the following message:

Thank you very much for participating in this survey.
Sadly, all our further questions are aimed at professionals who are experienced in the work with 3D software.

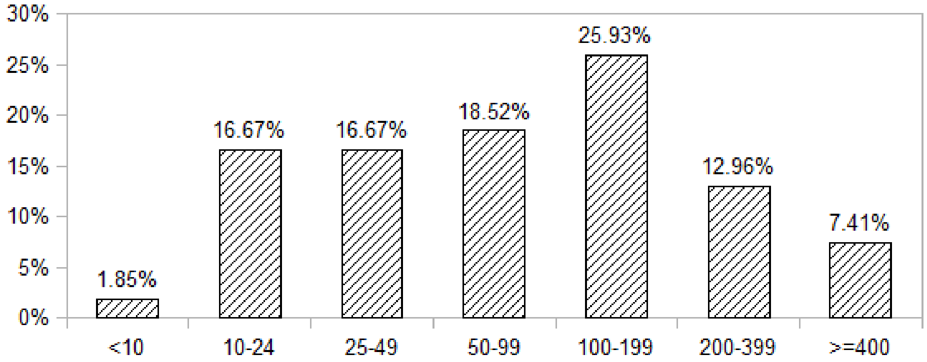
If you have colleagues who might want to share their opinion on professional 3D content creation,
please allow them to participate via: survey.makx.org.
Have a nice day!

4: Use of 3D software functionality:

Questions 8 though 14 asked about the use of 3D content creation software (for example: Autodesk 3D Studio Max / Maya, MODO, Blender) in a professional work context.

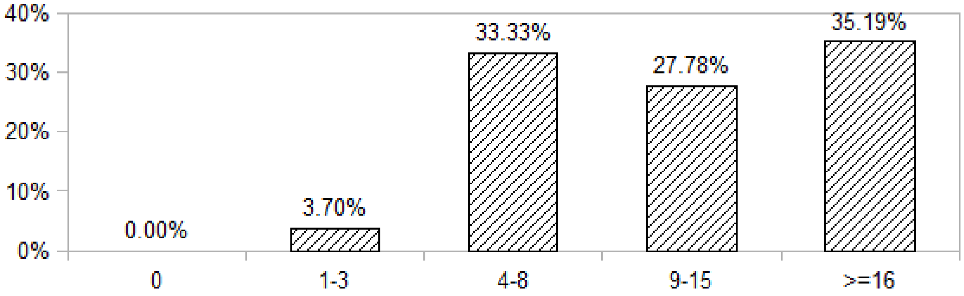
8. Please estimate how many different tools and functions of 3D CG software you regularly use.
Examples for distinctive tools and functions would be: Move Tool, Rotation Tool, Freeform Deformation Tool, Connect Joints, Set Animation Key, Bake, Add Particle Emitter, Add Force Field, Assign Material/Shader, Place Point-Light, Render, Batch Render …
As a reference: the complete set of functions in Maya consists of about 1000 functions.

- *less then 10*
- *10-24*
- *25-49*
- *50-99*
- *100-199*
- *200-399*
- *400 or more*



9. Please estimate how many keyboard hot-keys for your 3D software you use regularly?

- *0*
- *1-3*
- *4-8*
- *9-15*
- *16 or more*

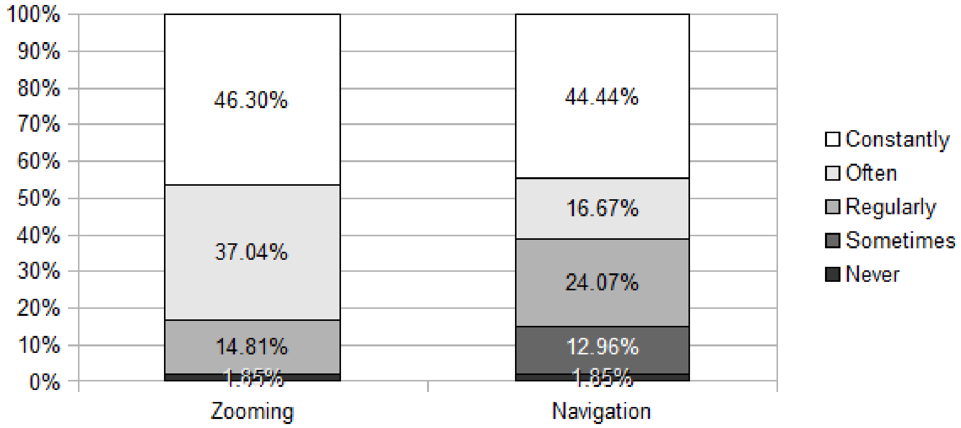


10. Please estimate the complexity of content you are working with at the following statements:

- *I usually only work at a scale / zoom level that captures the whole content.*
- *I sometimes have to zoom / change scale, for example: when starting to work.*
- *I regularly zoom in and out / change scale during work, for example: to review the result.*
- *I often zoom in and out / change scale while working.*
- *I constantly change zoom levels, or work with separate views at different scales.*

11. Please estimate the size of content you are working with at the following statements.

- *I usually work at the whole scene at once.*
- *I usually work at the same area of the scene, but have to navigate my point-of-view sometimes. (For example: when starting to work).*
- *I regularly have to navigate through the scene to get to another part, for example: to review it from another perspective.*
- *I often have to navigate through the scene while working.*
- *I constantly have to navigate through the scene or use multiple views from different parts of the scene.*

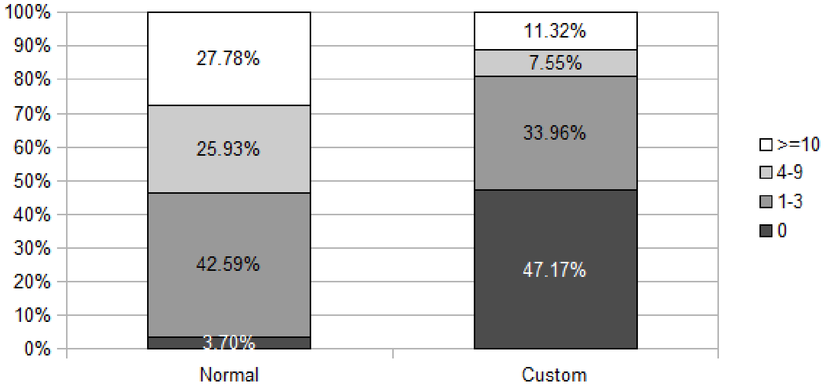


12. Please estimate how many (if any) additional scripts or plug-ins do you use with your 3D software that where acquired separately (not shipped with the software itself)?

- *0*
- *1-3*
- *4-9*
- *10 or more*

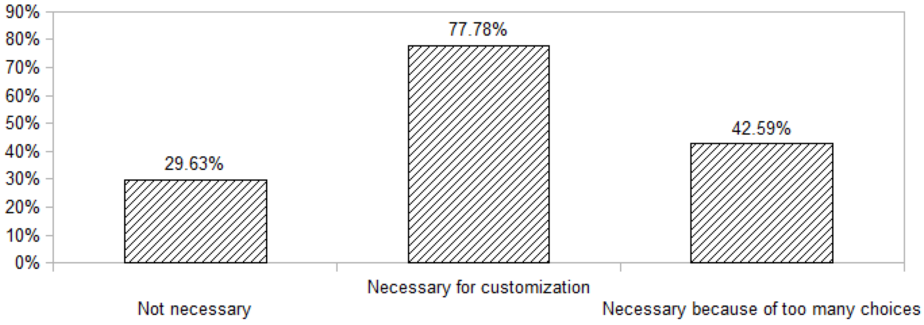
13. Please estimate how many (if any) custom scripts or plug-ins do you use with your 3D software that you developed yourself (in-house) or ordered exclusively (for example: developed by an external freelancer specifically for your studio).

- *0*
- *1-3*
- *4-9*
- *10 or more*



14. How do you feel about employing plug-ins and scripts in your work-flow?

- *Plug-ins or scripts would not be necessary if the base software was more sophisticated (ie: had more features)*
- *We will always need custom scripts or plug-ins to adjust the software to our needs (example: special pipeline requirements or to stay technologically ahead of the competition).*
- *Plug-ins and scripts offer many more choices than the software could realistically anticipate. (example: too many different choices, the software would become to big and expensive).*



4: Collaboration in 3D design:

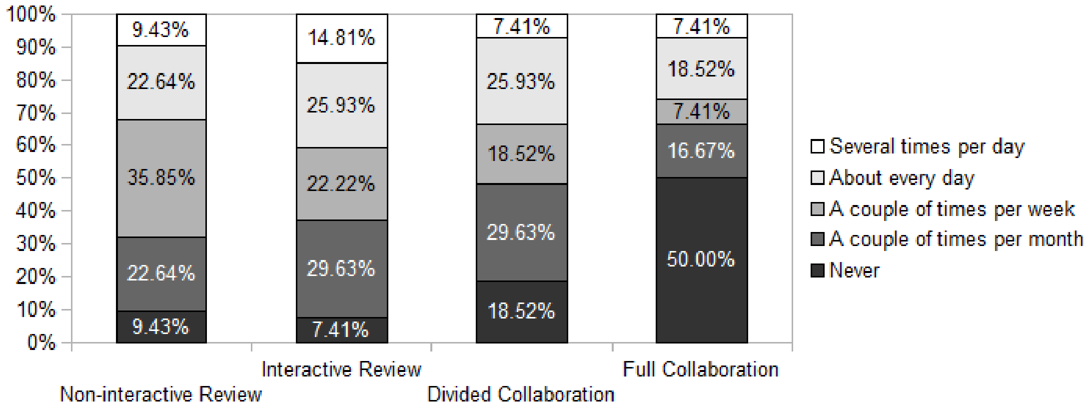
Questions 15 through 22 presented four different types of collaboration, and asked the participant how often each type of is performed, and how important it is (or: would be) to production.

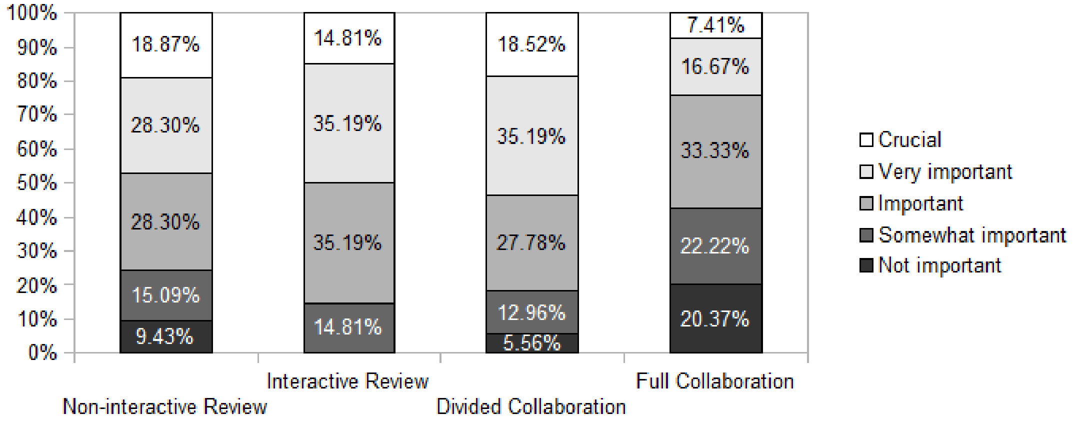
Non-interactive review
The work is presented, without being able to change it at the same time.
For example: sending content to supervisor / client, or group reviews in a cinema / review room.
(15. Regularity of Non-interactive Review; 16. Importance (or estimated usefulness) of Non-interactive Review)

Interactive review
The work is viewed by other people, while someone can edit it.
For example: supervisors / clients visit the artist desk during work.
(17. Regularity of Interactive Review; 18. Importance (or estimated usefulness) of Interactive Review)

Divided collaboration
More than one artist is able to work on the same content, by means of temporal or spacial separation.
For example: merging different versions of a file or working on a shared file in quick succession.
(19. Regularity of Divided Collaboration; 20. Importance (or estimated usefulness) of Divided Collaboration)

Full collaboration
More than one artist is able to work simultaneously on the same content / object.
(21. Regularity of Full Collaboration; 22. Importance (or estimated usefulness) of Full Collaboration)

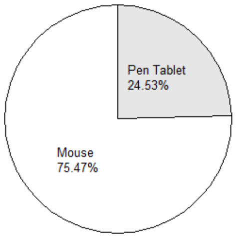




4: Work and User Interfaces:

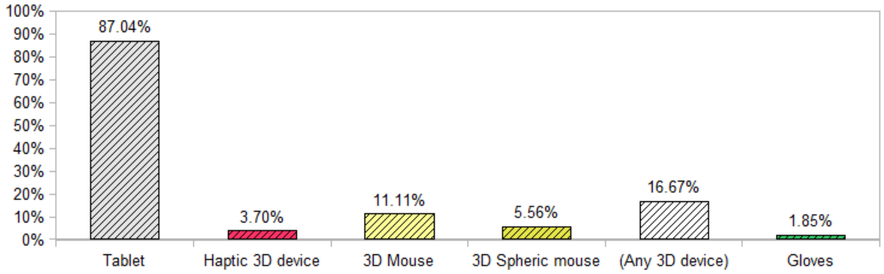
23. What is your main input device that your are currently using for your work?

- *Mouse (including vertical mouse and track-ball)*
- *Tablet (Wacom, touch-pad etc)*
- *Haptic 3D Input Device (Geomagic Touch etc)*
- *3D Mouse (Flying Mouse, Bat, 3D Ring Mouse)*
- *3D Spheric mouse*
- *Gloves (Data gloves, Pinch Gloves, hand-tracking)*
- *Other device.*



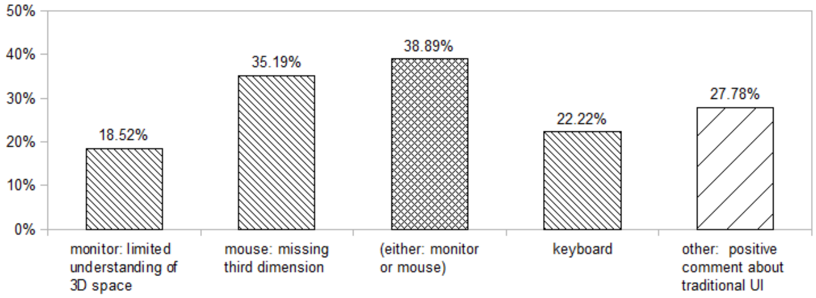
24. Which alternative / special input devices have you used for your work before (including testing them)?

- *Pen Tablet (Wacom etc)*
- *Haptic 3D Input Device (Geomagic Touch etc)*
- *3D Mouse (Flying Mouse, Bat, 3D Ring Mouse)*
- *3D Spheric mouse*
- *Gloves (Data gloves, Pinch Gloves, hand-tracking)*
- *Other device.* [No additional user input devices were entered by participants]



25. How do you feel about the currently common 2D user interfaces (mouse, keyboard & 2D screen) for 3D CG work?

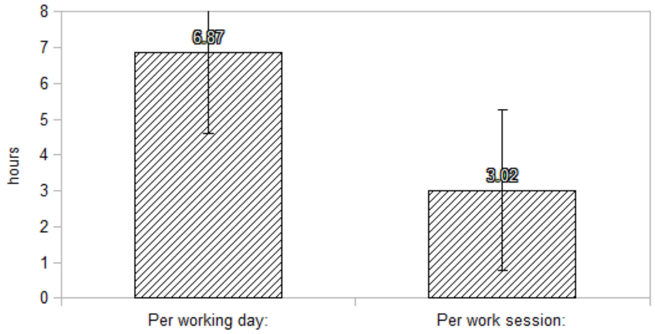
- *Working with a 2D monitor is making it difficult to work efficiently because of the limited understanding of 3D space.*
- *The keyboard is not suited for this type of work (because of layout, size of keys, missing icons or other reasons).*
- *Working with a mouse is inefficient because of the limitation to 2D interaction (missing third dimension).*
- *Other (comment).*



Comments:

Different software has different commands, which often makes it unfriendly to switch between software.
Why only negative options? I don't mind the 2D UI.
It's fast, responsive, and precise. That's what is most important.
Using specialized navigation tools offers little improvement over navigating like a FPS or with camera targets, but adds a separate tool which means I would have to move one of my hands off of the keyboard or mouse.
They do the job just fine.
Appropriate but could be enhanced.
I have no problems using a 2D interface for 3D work.
Mouse & keyboard combo works really well with an interface that knows and uses their strengths
No problems working in 3D space with current interfaces.
Work as expected.
I use a 3D monitor and mouse and feel that it is sufficient for creating 3D models.
Not sure.
They work fine, really.
Currently common user interfaces do a good job about 80 to 90% of the time. At specific use cases additional interfaces could close a gap.
No problems here.
It's fine.
Perfect.

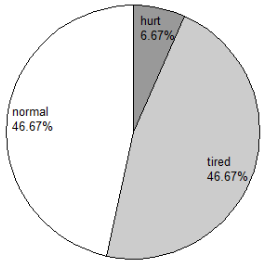
26. How many hours on a working day are you usually working with 3D software?
27. How many hours in one session (continuously, before taking a break) are you usually working with 3D software?



When mouse or tablet were chosen as main input device (which was the case for all participants), the survey form additionally presented two questions regarding the health condition of the arms, wrists, and hands.

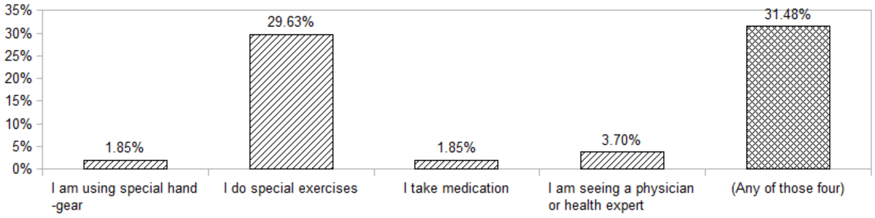
28. How do your arms (including wrists and hands) feel after several days of working with 3D CG software?

- *My arms feel good or normal.*
- *My arms feel tired or slightly strained, but not badly.*
- *My arms hurt from strain and physical stress.*



29. Are you currently taking any active measures to deal with work-related strain or pain in the arms, wrists or hands?

- *I am using special hand-gear*
- *I do special exercises*
- *I take medication*
- *I am seeing a physician or health expert*
- *Other (comment).*



Comments:

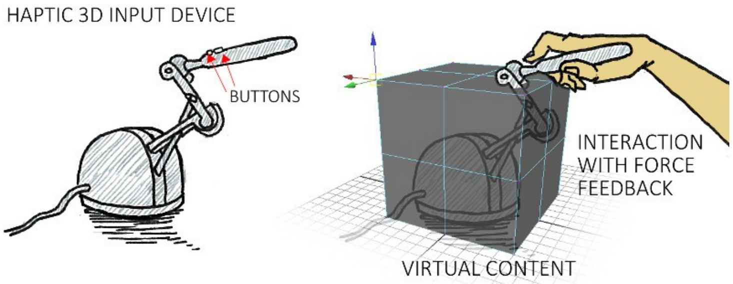
Just a mouse pad with cushion.
No, but i should.
I exercise with two 40 pound weights.
I'm using a standing desk at home to encourage better posture.
No special measures taken.
Quick break.
Being careful.
I work out.
Stretches.
I go rock-climbing.
Transitioning to using pen tablet as main mouse.
I am still young and fit.
I set my mouse to really sensitive so I don't need to move my hands as much.
Improved desk and chair height and positioning.

5: Novel User Interfaces:

For questions 30 through 52, we presented the participants with a number of novel UI concepts as they have been previously been proposed by scientists. For every concept we asked participants for their agreement with common arguments concerning these concepts. At the end, we also asked the participants to rate the proposed systems in regard to their usefulness for professional 3D design, comparing them to each other as well as to the traditional mouse as UI device.

In the following we present you a number of alternative User Interface concepts,
by use of Virtual Reality / Augmented Reality technologies.
Please give your opinion on these in the context of professional 3D content creation.
Don't worry about the precise details of the implementation and focus on the general concept (idea) instead.

Haptic 3D Input Device

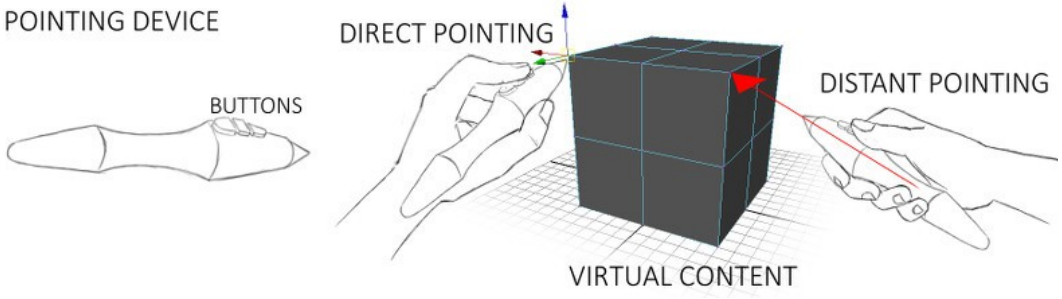


The artist is using a 3D input device which is overlayed with the virtual scene.
The device can give haptic feedback in the form of physical resistance when touching an object in the scene.
However, the range of motion is limited to about 30-40 cm.

30. Physical feedback (sense of "touch") would make the work more efficient. (a1)
31. The limited range of motion would be hindering when working on my content / scenes.(a2)
32. Comment (optional):

I've used a haptic wacom device and it didn't improve the workflow at all.
I cannot answer as I have not used one before.
I'm not familiar with that tool, doesn't look really intuitive.
I'm not sure it will help. Creating content in 3d is easier because you can ignore many physical limitation and problems. I'm not sure there is any point of bringing them back, but I never tried "Haptic 3d Input input device"
A Haptic 3D Input Device actually helps working close on the object in 3D space, but is very tiring for the hand.
Never had the chance to try one.
I can't know if the limitations would be hindering without trying it. I used to think that my tablet would hinder me as it's smaller than my screen, but once I used it I discovered that the benefits far outweigh the minor frustrations.
One of the biggest drawbacks of a Cintiq is your own arm obscuring the view. Using a small sized cursor would be a better solution than rendering the whole tool in 3D. The range of motion isn't as much about distance as it is about rotation/degrees-of-freedom.
The haptic feedback may be of some use to artists who have come from a practical background – clay sculptors or the like. For the regular 3d artist I'd speculate it wouldn't be much more than a fun to try out for a while before switching back to the more practical keyboard mouse/tablet setup.
It seem logical, but to keep your hand and arm up in that position would be very tiring.
Current tools work well. No need for haptic feedback.
Never tried that, so I can't tell.

3D Mouse / Pointer



*The artist is using a small physical device, similar to a mouse or pointing stick,
that is detected in the 3D workspace and thus allows 3D input.
It is held in mid-air and not limited in range.*

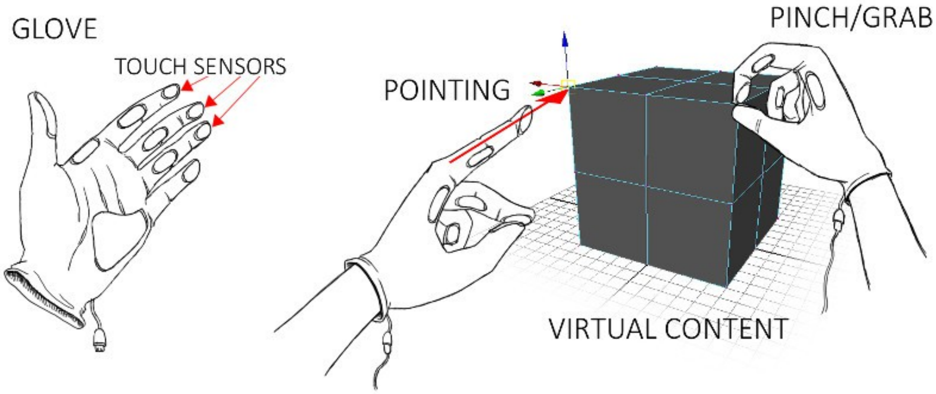
33. Being able to pass the device from one person to another quickly would be useful for work.(b1)

34. Having to hold an object in free space for extended time is too tiresome for work.(b2)

35. Comment (optional)

Again, this is redundant. Pen tablets do this better already.
For review purposes with management standing around a computer, this may potentially be of value, but I can't imagine it being all that practical for general art development.
Never tried that, so I can't tell.
This seems to offer more control than the haptic 3D tool. It offers more range of motion, and would be more comfortable in the hand.
The lack of tactile feedback would make this one more difficult to use.
It would be useful, but taking over someone's KBM is equally useful.
No tactile feedback.

Glove



The artist is wearing a thin cotton glove with touch sensors in various places. The position of the glove is detected by the system.
Thus the artist is able to interact with the scene either by pointing at objects out-of-arms-reach or by "touching" objects within arms reach.

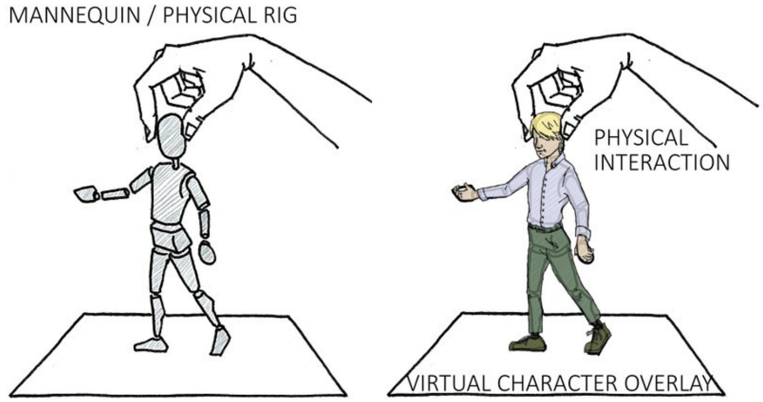
36. Not having to hold anything in my hands would make the work more convenient.
(For example: still being able to use mouse and keyboard or holding a coffee cup) (c1)

37. A glove is inconvenient, as it too warm for hot weather, unhygienic or otherwise physically undesirable.(c2)

38. Comment (optional):

I've never seen one of these in action. I suspect like the rest that it's something that the average artist might play with for a while and then forget about when they had to get back to serious work.
Unless the glove was specifically designed to stretch in a way that seems unfeasible for something requiring so many specifically-placed sensors and wires, it'd be a nightmare to find a pair that actually fit every specific artist that needed a pair.
A glove can be convenient, but not for every task.
Never tried that, so I can't tell.
A glove could have a thermoelectric cooler (Peltier) to keep it from being a problem in hot weather. One side of the device turns cold and the other turns hot when a current is run through it (run the current in reverse and you can reverse which side does what). You'd also have to get a specialty fabric that keeps sweat from accumulating and a large range of sizes so that they can be comfortable for people with large hands and those with small hands.
I don't think that the problems mentioned in question 37 would matter in an office setting. But, as you mentioned it being unhygienic, I would not want to wear a glove that other people were wearing. Also, it appears that if it contains 10 sensors, these sensors might be accidentally activated by unintended movements. Also, being hard wired to a computer system might be cumbersome. I would incorporate wireless technology to something like that. Then you might be able to zoom in and do work from a distance also. Look on the net for a remote device called "Fin". It is a single device that has many uses. Its sensors use position sensitive modes.
The gloves are for manipulation and navigation just like a mouse is, why would I use both? *Question 37: "Gorilla arms".*
No tactile feedback either.
Fingers are too big for accuracy.
It's important to get some form of physical feedback like you get from a mouse button.

Mannequin / Physical Rig



The artist has a small mannequin (figurine) that can be moved.
It is overlayed with an image of the actual character.
This is purely an input device and not able to move by itself (for replaying recorded animations etc).

39. The direct physical interaction with a representing object can make the work more efficient.
(For example: manipulating several joints at the same time.) (d1)

40. The content that I am working with is too complex for physical representation.
(For example: complex skeleton, hair, soft deformations, ...). (d2)

41. The content that I am working with is too diverse to make physical representations feasible:
we could not use the same rig again.
(For example: humanoid characters of different shape, animals or monsters, abstract shapes, ...) (d3)

42. Comment (optional).

That type of rig could be fitted with sensors similar to the glove and used in a similar way. The mannequin rig could be connected directly to a computer and its positions captured at certain frames. That would make a leap in 3D modeling and animation of action figures and would be used over and over in the movie industry! It would be like building a miniature robot, That would be the best use of that idea. Fantastic. I think that it could even have hair made of fiber optic sensors that would even sense the movement of the hair for frame capture modes.
90% of my work is done on cars. Moving cars could be quite useful.
This seems far more efficient than selecting rig controls with a mouse. A humanoid figurine would still be useful even if the proportions are different from the virtual character.
Question 41 could be an issue, but it could easily be solved by making the mannequin a modular thing rather than a fixed puppet. It would be more usable if it was just sticks and connectors without a volumetric aspect.
I work in environments – I would need an infinite variety of physical shapes to make that work.
Far too limited for creature work or abstract and extreme character designs.
I'm not an animator, but thinking there might be certain instances where something like this could be of some value. I work in animatronics a bit and it seems like something that might be interesting to experiment with as a control rig.

Voice Input

The artist is using a microphone to issue voice commands to the system.
Possible voice commands could include choosing a tool or operation
(eg. "split polygon tool", "set key", "assign material X", ...),
numerical and text input (eg. "set x-rotation to 45 degrees", "file: C:\project\texture1.tif", ...),
or interaction with the scene ("select PolyCube1", "align PolyCube1 and PolyCube2").

43. This would be a possible replacement to current mouse-and-keyboard input. (e1)

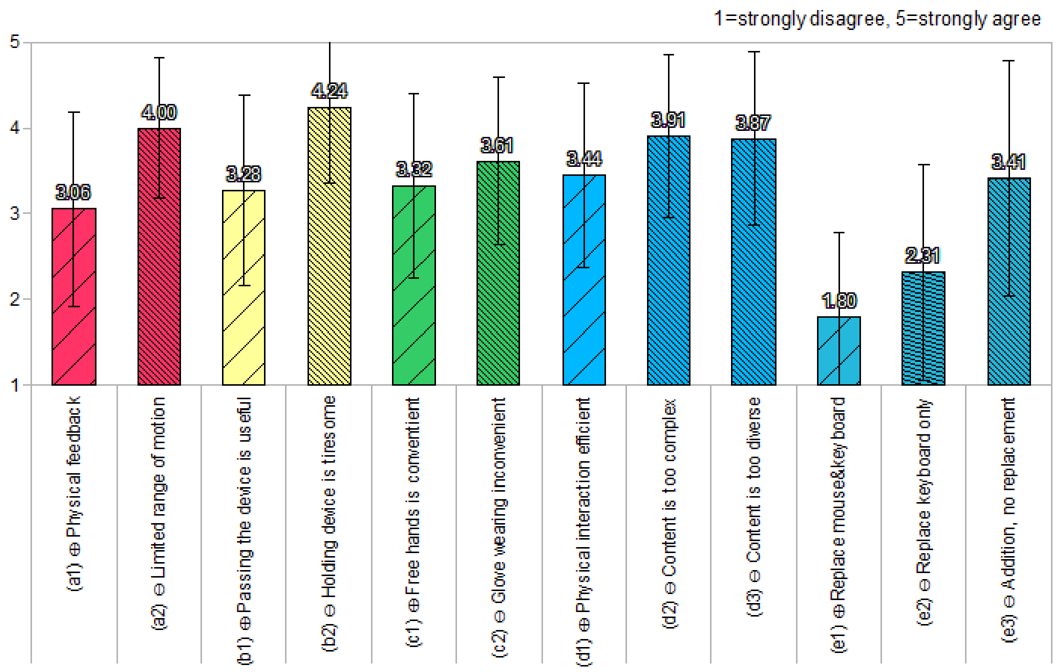
44. This would be a possible replacement to using a keyboard (text input, hotkeys),
but not a pointing device (mouse). (e2)

45. This would be a useful addition to mouse and keyboard, but cannot replace either. (e3)

46. Comment (optional)

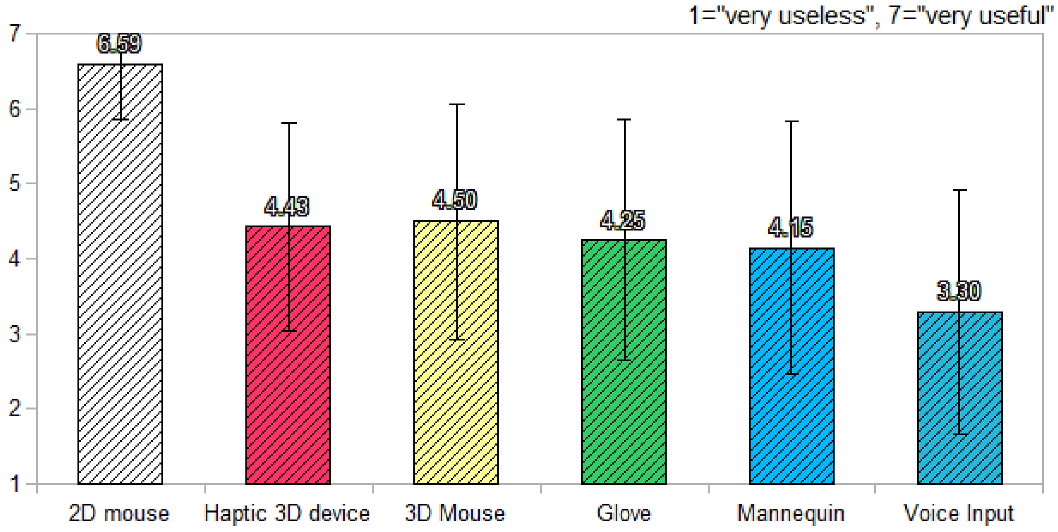
A keyboard & mouse offer a great speed advantage. No need to look only to CG software, see competitive games, such as Starcraft, Dota 2 or similar. Professional players in Starcraft can achieve 200 APM (actions per minute), even casual players can do around 50 APM. There is no other device in use that can offer such input speed from humans to computers.
Have you worked in an office? Do you really want people TALKING to their computer?
This would be incredibly distracting in an office environment. I might see it being useful for people with limited manual dexterity or a missing hand.
Due to the speed that the average artist works at, and the fact that they're typically working alongside numerous other individuals, I can't imagine voice being of particular value. Maybe in the review scenario with management standing around the computer, but even that seems a little dubious.
This would be obnoxious in an office or collaborative environment
Those type of devices are susceptible to all the background noises. In an office setting there would be more problems than advantages. If you wanted to listen to music as your worked it would interfere with the device. Also, conversations and telephone calls that are outside the use of the device would interfere with the voice commands of the device.
For complex applications like such that are used for 3D work, I wouldn't be comfortable with voice input. I need to trust that my input will be recognized correctly and the right action is being executed. Also I think it would make work much slower.
Your coworkers would want to murder you.
Not a feasible solution in average working environments.
It has some potential, but there are some people who have voices that computers don't recognize. Also, there's not much that could be done here that isn't already possible with other software for those truly interested in it.
Not very conducive to working in a room full of other people.
The best use case for this would be a search function. Not for changing values or accessing common tools like translation, but for anything that's hidden in three layers of menus.
Who wants to have talking people around in an office?

Results Q30-Q45:



Rating

Please rate the previously presented devices in respect to their usefulness,
comparing them to each other as well as to the traditional mouse as user interface device.



5: Authenticity check:

At the end of the survey, we asked name and company of each participant. This information was only used to ensure authenticity that the participant was in fact a media production professional and no one would insert multiple or bogus information which would distort the results.

Entries from participants who could not be identified were removed from the sample. Of 73 data sets we removed 19, leaving 54 data sets which we considered authentic.

APPENDIX B

Questionnaire for the Formative User Study on the Prototype User Interface

The following pages contain the questionnaire that was presented to the participants of the Formative User Study on the Prototype AR UI discussed in Chapter 4 Section 4.5.

Interactive Media Design Laboratory
Nara Institute of Science and Technology

User Study : Novel 3D User Interface for 3D Design and Animation

Thank you for participating in this user study.

It will take approximately 60 minutes to complete.

You will test a prototype user interface for 3D modeling and animation.

The prototype contains a number of usability problems.

Your task is to go through the tutorial and identify these problems.

Try to relate the use of the prototype to your 3D design work.

You will be guided through the use of the system by a tutorial.

If you have trouble using the system, please inform the experimenter.

You are encouraged to talk about your experience during the experiment.

We will record your use of the prototype for later analysis.

All data collected will be treated confidential and only used for scientific analysis.

If you feel uneasy or unsafe at any time during the experiment, please tell the staff.

You can abort the experiment at any time.

After you have completed the tutorial, you are free to keep on experimenting.

After the session is over, we will ask you to fill out a questionnaire and answer some questions in a quick interview.

If you have any questions or concerns, please tell the experimenter.

Questionnaire:

Please make a list of **usability problems** that you identified in the user interface.
Assign a gravity rating from 1 to 10, where 1 is “negligible” and 10 is “critical”.

Problem	Gravity (1-10)

How would using the user interface affect your **work performance**?
Please estimate a percentage, comparing the prototype use to your current system.
Examples:
“0%”: My work is impossible to perform with this user interface
“50%”: My work performance would be reduced by half when using this user interface.
“100%”: My work performance with the new user interface would be the same as with my current software.
“200%”: I could work twice as fast with the new user interface than with my current software.

Judging from the current state of the prototype:	%
Considering a future version of the prototype where the usability problems have been resolved:	%

Please feel free to provide **additional comments** about your experience below.

Personal Information:

Your personal information is collected for statistical purposes only.
It will be treated confidential, will not be published, and will be destroyed after evaluation.

Age:	
Gender:	○ male ○ female
Nationality:	

Please write about your **background, related to 3D design work**,
including employment, work on hobby projects, and studies (including self-study).
In the last column, please **estimate how many hours per week** you have worked on
3D design projects, **averaged over the whole year**.

Example: if you worked half a year on a hobby project for 8 hours per week, and the other half of the year employed full time with 40 hours per week, then your whole-year-average would be (40 + 8) / 2 = 24 hours per week.

Year	**Activity related to 3D design / animation / CAD** *Examples: "Worked at company", "worked on hobby project", "took classes on 3D animation", "self-study using online tutorials".*	**Average hours per week (over whole year).**
2015		
2014		
2013		
2012		
2011		
2010		
2009		
2008		
2007		
2006		

If you are not currently employed in 3D design related work, do you aim to work as a professional artists in the foreseeable future?

◎ *Definitely working towards it.* ◎ *It's an option but not my main goal.* ◎ *No plans.*