VR Gesturing

Title: VR Gesturing
Authors: Andrin Gorgi & Stefan Lustenberger
Matrikel Nr.: 16-573-636 & 17-683-137
Date: June 2020

Mentors: Prof. Jürgen Späth, Florian Bruggisser
Typeface: Radikal, Inter UI

VR Gesturing

The identification, design and implementation of a user tailored VR gesturing system based on ergonomics research and user tests

VR Gesturing

Acknowledgements

This project could not have happened without the help of a lot of our friends and colleagues.

First of all, we want to thank our mentors: Jürgen Späth and Florian Bruggisser. They advised us about their experience and expertise and helped us handling technical difficulties. Also, we want to thank Luke Franzke, who agreed for an additional mentoring where we had doubts about uncertainties which came up during our process.

We are very grateful for our understanding collaboration partner, Sensoryx, where we had contact with Philippe and Hagen. We got the most imaginable freedom with our process but could still contact them every time we needed help.

This project would not have been the same without the inspiring and cognizant people we had the chance to get an interview with. Many thanks to Mike Alger, Oliver Sahli and Max Rheiner.

Patrick Lustenberger you were unbelievably helpful, thank you for proofreading our thesis and standing by for questions.

Thank you, Dan Büschlen, for your expertise in web programming. Many thanks to Nicola Fürer, who helped us create an awesome storyboard for our movie. Thank you, Claudia Buck, for your amazing linguistic proficiency. Thank you, Marcial Koch, for the motivation we got through you and for the constant encouraging feedback. Many Thanks to Mauro Gorgi who let us occupy his living room for days to set up a studio to film our movie, and also for the fantastic climate in your house, where I (Andrin) could work during the whole quarantine situation. Moreover, of course, a huge thank you to Camille Büsser and Celestino Gorgi for letting us film you and your beautiful hands for hours. Many thanks also to Mike Schönenberger, who could help us with excellent tips according to process management and his expertise in Adobe InDesign.

Also, we are unbelievably grateful for all our friends and family members who participated in our user tests and submitted hundreds of videos with many amazing ideas of different gestures.

Much knowledge we gained was through youtube. Many thanks to all the creators providing helpful content on youtube, especially to "Valem", "Brackeys" and "Gabriel Aguiar Prod.".

Furthermore, last but not least, we want to speak out a huge „thank you" to our whole class. IAD2017 forever! The uncountable hours we spent in zoom meetings, raging about the corona virus and the circumstances, motivating each other, sharing each other's pain and criticising each other's work to get the most of it.

Abstract

Virtual Reality is finding its place in the broad masses. Nevertheless, its potential has not yet been fully explored. Its standard of interaction still focuses on the use of controllers even though accurate hand tracking has become available recently.

We therefore investigated where gesture-based control systems can improve user interaction, and in particular which gestures are best suited to control these virtual worlds.

We analyzed daily interactions and found that scrolling, zooming and triggering actions are the fundamental components of most current interactions For these three actions, we created a novel gestures system based on state of the art ergonomics research and several user tests performed using an innovative website-based user-testing platform. Further we incorporated these gestures in VR demo environment allowing fast testing and evaluation of gesture systems.

With the designed interaction gesture system, we showcased how specific tasks can benefit from specifically tailored gestures to improve user experience.

Keywords

Gestures, Gesture Control, Oculus, Quest, Scroll, Trigger, Unity, Usertest, Virtual Reality, VR, Zoom

Table of Contents

Introduction

Almost 60 years after Ivan Sutherland invented head-mounted displays, Facebook's Oculus released its 4th Headset, and by the end of 2019, they announced hand tracking within their headset. In the past ten years, virtual reality (VR) has made tremendous progress in terms of usability and public adoption. It seems that the use of our hands without any controllers is currently also coming to a start, with companies investigating tracking gloves, for example, Senoryx or HaptX, with additional haptic feedback and Facebook's beta version of their hand tracking released in early 2020.

The current state of the art for VR applications is, however still the usage of controllers. Controllers offer excellent accuracy and a type of usage which most people are already familiar with from other interactions. The question is, however, where does all this technical advancement lead us? Will gesture control systems are successful, and will they provide a benefit in comparison to the use of controllers? What new possibilities will be enabled, and how can we, as interaction designers have an impact on how it turns out?

These questions were guiding our process, where we experienced many changes of direction due to new findings from deep diving into this exciting topic. When we started, we wanted to investigate the use of gestures in a virtual work environment. We ended up identifying and testing basic gestures which are needed as a foundation for all these future gesture-based applications. Since only a few studies were available that investigated gestures from a user-interaction point of view.

Early on, we reached out to experts in the field and started to collaborate with the Zurich based start-up Sensoryx one of the most promising companies in the field novel gesture enabling hardware. They are developing a glove with haptic feedback to enable more intuitive ways of interaction in the VR space. In the course of our study, we, however, found that before using this new state of the art tools, there must be done some more fundamental research regarding gesture identification. Therefore, we focused our efforts in that direction, keeping the current advancements of Sensoryx and their future products in mind, however.

On the following pages, we first outline the context within our work must be considered as well as our personal point of view on the topic. Then we present related projects and findings in discussion with key opinion leaders in the industry. This leads then to the precise research question and concept.

VR Gesturing

Research Field

Background and Context

In this chapter, the most important concepts are explained in order to provide the context around this project. This includes a review of the current state of technology as well as other vital topics.

Gesture Control Technology

Our most comfortable way for communication is, for most of us, to use gestures. This phenomenon goes way back. “While non-human primates almost completely fail to acquire spoken words, they are capable of learning various manual signs. For example, the chimpanzee Washoe acquired 132 manual signs (gestures) within 51 months of training and even combined the signs to form sequences.” (Gardner & Gardner, Comparative psychology and language acquisition, 1978). Derived from this knowledge, we know that this must be deeply encoded in our nature. In the book named “Silent Messages” (1971), Albert Mehrabian discusses his research on non-verbal communication.

> *“He concluded that prospects based their assessments of credibility on factors other than the words the salesperson spoke the prospects studied assigned 55 per cent of their weight to the speaker’s body language and another 38 per cent to the tone and music of their voice. They assigned only 7 per cent of their credibility assessment to the salesperson’s actual words.” (Ubiquity: The 7% rule, 2011)*

Now, 50 years later, we know that these numbers are not entirely correct. Different researches showed numbers between 7 and 50 per cent, depending on the research setting.

> *In the research paper published by Pomboza-Junez Gonzalo and Holgado-Terriza Juan A., they summarized nicely: “The interaction between the user and the devices through a natural gesture is relatively new. The gesture is defined as a mental concept of an idea associated with an action, response or a requirement that the user realizes intending to achieve a result.” (Control of Home Devices based on Hand Gestures, Pomboza-Junez Gonzalo and Holgado-Terriza Juan A.)*

The last decade was groundbreaking and laid the foundation for gesture control. Various products like LEAP Motion from Ultraleap or the MYO bracelet by Thalmic Labs have awakened public interest. Most research papers speak of a new field and promise many changes for the better.

> *Rafiqul Zaman Khan and Noor Adnan Ibraheem stated that: “Hand gesture recognition systems received great attention in the recent few years because of its manifoldness applications and the ability to interact with machines efficiently through human-computer interaction.” (Hand Gesture Recognition: A Literature Review, Rafiqul Zaman Khan and Noor Adnan Ibraheem)*

We humans can quickly, among us, have a basic conversation just by using gestures. Nevertheless, the communication methods we have to talk to computers and machines are limited. Buttons, keyboards, triggers, switches, mice, and so on were developed. But they all have one thing in common. These Methods were not quite fitting for humans. People speak, and people gesticulate. Voice and speech inputs, as well as gesture control inputs, allow designers in Human-computer interaction a more human-like design. To give a little bit of a context, we found this quote in one of the papers which played a leading role in our research process:

> *“Human-computer interaction (HCI) also named Man-Machine Interaction refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human. There are two main characteristics that should be deemed when designing an HCI system, as mentioned in functionality and usability.” (Hand Gesture Recognition: A Literature Review by Rafiqul Zaman Khan and Noor Adnan Ibraheem)*

The usability and functionality of all conventional input methods like buttons, mice and keyboards have been the subject of extensive research. There are also many theories, ways of thinking and research publications regarding the creation of good user experience with these conventional tools.

> *Donald A. Norman has coined the term user experience: “User Experience” is a collective term. Drawing for the overall experience an operator has with a system in terms of industrial or product design, interface design and interaction design.” (Rainer Dorau, Emotionales Interaktionsdesign, 2011, pp. 17)*

Usability and functionality are therefore essential, but the experience also incorporates a feeling, and this feeling is what counts the most in the end. By using gestures, this experience might be introduced, which would enable a more natural feeling way of human-machine interaction.

Gestures in Virtual Space

In the last ten years, the development in the VR field has made tremendous and rapid progress, and the devices have become affordable for private users and smaller companies, allowing them to access this new technology. The range of different devices and controllers has also increased significantly, showing that the industry has not yet settled for a uniform control and user interface. All providers are currently following their ideas. Especially when it comes to hand-based gesture control, there are many different directions. Simple actions, such as opening menus in a VR environment, are solved in many different ways. This, for example, is currently implemented by the commercially available systems by, opening and looking at the palm of the hand, making a fist or using a fixed virtual button.

Nevertheless, some gestures have established themselves for specific tasks. For example, scrolling is practically always done with a wiping motion. Be it with the whole hand, with outstretched fingers, or only with the index finger, the overall intuition is the same. Most of these gestures are established and intuitive because they are already used with touchscreen devices.

Stefan Liszio and Maic Masuch from the University of Duisburg in Essen published a scientific paper in June 2016 with the title “Gesture-based Virtual Reality Interaction Design - Development and empirical validation of hand gesture-based interaction concepts in VR applications” in which they developed one of the first gesture concepts for Virtual Reality. They used a handful of functions for which they developed specific gestures. These are: Context menu, close/cancel/exit, select, more precise selection, activate object and the zoom function.
It was fascinating to see the results which they came up with, and the thoughts behind the different gestures.

Haptics

Our body is in constant exchange with its surroundings and thus provides a multitude of impressions, which our brain assembles into a complete picture. This is beautifully described by Juhani Pallasmaa:

> *“I confront the city with my body; my legs measure the length of the arcade and the width of the square; my gaze unconsciously projects my body onto the facade of the cathedral, where it roams over the mouldings and contours, sensing the size of recesses and projections; my bodyweight meets the mass of the cathedral door, and my hand grasps the door pull as I enter the dark void behind. I experience myself in the city, and the city exists through my embodied experience. The city and my body supplement and define each other. I dwell in the city, and the city dwells in me.” (Juhani Pallasmaa, The Eyes of the Skin: Architecture and the Senses)*

One of these impressions is the haptic sensation. Even if we do not touch anything directly with our skin, we still perceive our environment through our haptics. Whether it is a light wind blowing, or the frosty temperatures of the air, all of this can only be felt through our haptic sense of touch. The term haptics comes from the Greek and means “ feelable” or “suitable for touching” and

Research Field

expands our understanding of contact with an enormous spectrum of possibilities. It allows us to perceive different surfaces, materials, textures, temperatures, pressure, vibration, pain and much more. All of this is made possible by a large number of receptors in the skin, which transmit information about an object or the environment to the brain where it is processed. The average adult has about three million pain receptors. This complex network of receptors is also known as the somatosensory system. We can sense whether an object is hard or soft, warm or cold, wet or dry, rough or smooth, sharp or dull and more. We can identify all these things without the use of our other senses. These complete the overall impression and make our world the place worth living in.

So, what influence does haptics have on the world of technology today? For smartphones, tablets, laptops, peripherals and many more, haptics play a key role. It should be comfortable and exciting to hold such devices in your hands and feel their surfaces and materials. This is, however, only a part of why haptics is essential for such devices. The second part is the haptic feedback—noticeable feedback, which confirms whether an input or action was successful or not. For example, on physical keyboards, this feedback is provided by the yielding of the keys, which allows the user to judge very well whether the input was successful or not. For digital keyboards for smartphones and tablets, this process is slightly different. As there are no physical keys on these devices, the feedback is in the form of vibrations. This form of haptic feedback is distributed in the digital world.
Robert Blenkinsopp stated that

> *"Haptic feedback is the use of touch to communicate with users. Most people are familiar with the vibration in a mobile phone or the rumble in a game controller - but haptic feedback is much more than that." (Robert Blenkinsopp, VP Engineering at Ultraleap)*

In addition to these classic examples of haptic feedback, there is potential for more unusual applications such as sensing presences, emotional connection and well-being, as well as the ability to interact and explore with things.
In Virtual Reality, haptic feedback is currently mainly about feeling the virtual world. To make the experience of such a virtual world more tangible and to intensify the perception, haptics will play a significant part in the experience.

Various companies such as DextaRobotics, Senso, Manus VR, and also our partner Sensoryx are currently developing gloves which enable haptic feedback. However, different goals are pursued. HaptX and DextaRobotics research and develop on force feedback as well as very accurate haptic feedback. This makes their gloves large in size and heavy, which is due to all the mechanics and sensor technology that is used. The other three, Senso, Manus VR and Sensoryx, are more interested in simple haptic feedback through vibration and accurate tracking of the hand and fingers. These companies offer a lightweight glove with little to no wirings for maximum user freedom.

Skeuomorphism

Skeuomorphism has been part of every designer's life since the digital revolution. People born after the year 2000 associate the icon of a floppy disk immediately with saving a word or excel file, even though they do not even know what a floppy disk is. In the settings of an iPhone, we use switch buttons to switch our Wi-Fi on or off, even though we barely use switch buttons in real life, except for pilots and people working in huge control rooms.

We ask ourselves if this is still the way to go, and we want to investigate what work has been done in this direction. Is it even possible to design an intuitive UI without trying to design with skeuomorphism?

Our Partner

In this section, we would like to introduce our partner that supported us from a hardware point of view. Further, we had several exchanges of ideas to understand where the VR industry is going and where current limitations and opportunities are currently seen.

Sensoryx (VR-Free Glove)

We work together with Sensoryx a tech startup from Zurich that was founded by an ETH and HSG alumni. They are developing the VR-Free, lightweight, fingerless gloves and a head module for virtual reality applications. The product is compatible with different headsets: Oculus, HTC Vive, Valve Index, MS Mixed Reality, Samsung Odyssey, GearVR, Daydream or similar.

The glove consists of 6 different, complementary sensor types that are fully integrated into the glove and the HMD (Head Mounted Display) mount. This portable sensor system detects every movement of its wearer with the highest precision in real-time (update frequency (120 Hz, 8 ms) from hand to HMD). Further, no external references, such as fixed cameras or beacons, are needed. This allows users the highest possible freedom of movement. It is currently the only mobile 3D hand and finger tracking system that offers motion detection beyond the field of view, no visible latency, millimetre-scale precision, multi-user capability and broad compatibility.

The high amount of precision of Sensoryx is achieved by combining multiple types of sensors to overcome the limitations of using only IMUs. Usually, IMU motion detection systems suffer from drift and therefore accumulate measurement errors over time. This results in an increasing difference between where the system thinks it is

located and the actual location. Overcoming this limitation allows Sensoryx to build precise, lightweight gloves.

With the patented mobile system of complementary sensors controlled by proprietary algorithms and a lightweight glove form factor, VR users will get their hands back. Multi-user capability and long battery life ensure excellent user immersion. It is Sensoryx's vision to enable VR/AR enthusiasts to interact with the virtual world intuitively and naturally, using the best input devices nature has to offer: Their own hands and ultimately their entire bodies for the best immersive VR experience.

This vision aligns very well with our goal to understand what a natural and intuitive interaction is all about and therefore, partnering allows for a mutual benefit. Sensoryx has extensive knowledge of the state of the art of haptic devices. In contrast, we are investigating what kind of gestures (eventually also tracked by these devices) allow for the ideal user interaction.

By being able to test the VR-Free System, we got a first impression of what kind of interactions could be possible in the future using glove-based systems. In a VR demonstration application, we could see for ourselves what the potential of this technology could be, and where currently the limitations are to make the vision of Sensoryx become a reality. Especially the tracking of the hand and fingers was very precise and therefore exiting as it holds the potential to enable also small gestures. We also appreciated the freedom of movement without any cables, since this allowed for a smooth immersion. The interaction with objects in VR showed us that one of the current limitations are the translation of haptic feedback to the user. For example, holding a bowl or a ball felt unusual since the hand was not feeling the object. In that sense, vibration or other means of haptic feedback as it is used for controllers will boost this technology even further. Regarding our project, we found that through the interaction using the own hand, the immersive feel/experience was very different and exciting in comparison to controller-based systems and felt more intuitive and natural.
The partnership with Sensoryx, therefore, helped us to understand and test a futuristic system that will allow in the future to use our results in a real-world setting. Understanding the limitations and potentials beforehand helped, therefore, to guide our gesture investigations.

Research Questions – Hypothesis

In this section, we summarized our main research questions along which this project was conducted. They were used as a guide to inform decisions along the way.

What does it take for a GCS (gesture control system) to be intuitive and fun to use?

Our main goal is to create a catalogue of hand gestures, allowing people to interact intuitively and joyfully with their surroundings, whether this is in Augmented Reality, Virtual Reality or real life. Therefore, we need to find the most intuitive hand gestures through observations and field research. In, Hand Gesture Recognition, a literature review by Rafiqul Zaman Khan and Noor Adnan Ibraheem they explained what has to be considered by designing a gesture for computer recognition:

> *"The essential aim of building a hand gesture recognition system is to create a natural interaction between human and computer where the recognized gestures can be used for controlling a robot or conveying meaningful information. How to form the resulting hand gestures to be understood and well interpreted by the computer considered as the problem of gesture interaction." (Hand Gesture Recognition: A Literature Review, Rafiqul Zaman Khan and Noor Adnan Ibraheem)*

How can the Use of gesture control have a positive impact on daily office tasks?

We believe that gestures also can have an increasing effect on the power of concentration. A lot of people, for example, start to doodle as soon as they pick up a ringing phone or start to dial a number. These actions, as well as other epistemic actions, are just performed in order to gain concentration and could be increased by subtle little gestures and movements.

Moreover, there is also the importance of joy in every action. In order to satisfy the user with joyful gestures, we need to perform different experiments and observations. We believe that besides all other positive effects, this technology has the possibility of bringing more fun into people's everyday life. It might be even possible to gamify the act of picking up a phone without extending the used time to do so.

Our methodologies

In this section, we give an overview of the different methods we used in order to tackle and answer our research questions.

Research and Literature

Virtual reality is currently a burning topic which means that there are many publications around it. This reaches from public opinions in forums to published research papers. Further, there are also applied use cases, games and other pieces of software, which can be tested to analyse different approaches or solutions to particular problems or topics.

Finally, there are also a few books on the topic of virtual reality and user-friendliness and user experience.
We leveraged all of these information sources to build up our expertise and reference the sources of our knowledge throughout the thesis.

Contact to Experts

We intend to speak to experts to get hands-on advice and insights into the topic. Therefore, we reached out to multiple experts.

The people at Sensoryx supported us with their considerable know-how in the technological field. When it comes to user experience or the interaction between the user wearing their glove system and the computer, they have not yet conducted extensive research, which is also the reason for them to partner with us.

Further, we reached out to Max Rheiner, a member of the Interaction Design Team at the Zurich University of the Arts, to learn from his experience in immersive applications. He has previously done very inspiring work in the area of VR.

Furthermore, we also contacted the "ZHdK Immersive Art Lab" which has vast experience with VR through the studies of game design. We especially hoped to get some more insights into human-computer interaction in virtual reality and technical help for our first prototypes.

Gabriel Bach also degreed his bachelor in interaction design, two years ago, by working with virtual reality techniques, using a leap motion mounted to the head-mounted VR device to track hands. We, therefore, reached out to him to benefit from his experience.

Experiments

To verify and answer our research questions, it was essential to perform multiple experiments. We, therefore, conducted experiments with ourselves, with our friends but most importantly, with strangers. This is where we would probably get the most honest feedback. It is also crucial to have people from many different target groups in order to get reliable results.

This is an excellent opportunity to make a cultural probe, finding a broad set of information as well as different points of views from different target groups. Of course, we will use various human-centred Design methodologies to gain alternative insights.

Research Field

Motivation and Intended Contribution

In this section, we would like to outline our motivation as well as the intended contribution of each of us.

Our Teamwork

We teamed up because we recognized that our two projects started to go in a direction which would make sense to combine. We then quickly brought a video and a prototype to life, also proofing that collaboration would work on a personal level. We quickly agreed on our thoughts, and we could build on each other's ideas and realized that we have the same vision. This was for us the most significant boost of motivation.

Applications of AR and VR in work environments are known from a lot of movies like "Iron Man" or "Minority Report". People are impressed by these scenes and would love to use AR the same way Tom Cruise and Robert Downey Jr. are using it in their movies. For the two of us, it is the same. We would love to have this happening in the near future. Furthermore, we think, it might never be possible to have the exact same experience like Iron Man. However, we believe it is possible to create the feeling to work with your hands, controlling the interface with gestures in an intuitive way.

Virtual Reality

The field of virtual reality is so rapidly advancing and exciting that we would love to be a small part of this. Moreover, the research that needs to be done in terms of usability and intuitive handling is not yet as advanced as the technology is.

Virtual Reality overall is a technology we both are heavily interested in, and we both see immense opportunities in various applications in the future. The following statement of Chris Milk summarizes our beliefs very well:

> "Virtual reality is the ultimate empathy machine. These experiences are more than documentaries. They are opportunities to walk a mile in someone else's shoes."
> - (Ted, Chris Milk, 2015)

Corona

Like most people in Switzerland, we were forced into lockdown on Friday the 13th of March in 2020. In order to slow down the spread of COVID-19, the "Schweizer Bundesrat" requested that leaving the house should just be done when really necessary. All non-system relevant stores, as well as the entertainment business, all the sports clubs and every school and university, had to shut down, also the Zurich University of the Arts. Most of our friends working in a regular office were never able to work from home before Corona; these friends experienced a struggle at this moment. Some of them struggle while trying to find the motivation to sit in front of their computer to work. Others have a lack of material resources, tools and knowledge which is not beside them when they are not at their desk in their office. Furthermore, last but not least the swiss internet experiences a bottleneck and has multiple outages every day, which means that people at home are disconnected from their terminal session or lose their connection to their company's VPN and files.

The use of virtual reality will not solve the problem of internet outages and most probably not the problem of lacking motivation. But it will solve the lack of knowledge; it will solve the problem of office workers suddenly working in different environments with several deflections they do not have to deal with when they are in their office.

We see a lot of potential in virtual reality because wherever you are with your headset and a stable internet connection, it will be possible to give workers the same experience as they are used to from working in their office. In this project, we did not investigate the specific topic of home office nor the context of an epidemic catastrophe. We believe, however, that the findings will provide a strong foundation to start working in these directions. We, therefore, see this crisis more as a motivation boost since it provides more significance and relevance to our investigations in creating an intuitive environment and letting people be as productive as possible.

Concept

In this section, we describe the process from scanning the field of VR interaction by gathering information from literature, projects, and expert interviews to formulate a refined, clear project goal and research question.

Related Work

In this section, work that is related to this project is presented in order to put the objectives and findings of this thesis in perspective and to provide an overview of the current state of the art.

Literature

From an academic perspective, there a lot of research that has already been done in relation to this topic of gesture interaction and VR. In most papers, people worked with grand experiments in order to find out how a user interface or user interaction has to be designed to be intuitive and aesthetically pleasing to work with. Others were trying to lay a foundation on the ergonomics in virtual and augmented reality. Furthermore, a few papers also investigated the use of hand gestures in virtual reality. A lot of these papers can be found in our bibliography.

Microsofts Windows Mixed Reality Toolkit

Microsofts Windows Mixed Reality Toolkit (MRTK) is a toolkit Microsoft created for developers working with the Microsoft Hololens 2. It provides the basic building blocks for Unity development on HoloLens, Windows Mixed Reality, and OpenVR. Examples of these blocks are various UI components like buttons, sliders, bounding boxes to interact with virtual three-dimensional objects or windows which allow scrolling by hand input. Nevertheless, some more experimental blocks like Spatial Awareness to make the virtual world interact with the real world are available. For example, walls and floors are recognized, and virtual objects align themselves with it or use it as boundaries as well as many other possibilities. For Eye Tracking, three functions are currently supported by the MRTK: Firstly, to make selecting and interacting with virtual components fast, accurate and secure, by combining eyes, voice and hands input. Secondly, to scroll text automatically when the eyes reach the point where scrolling is needed, as well as to zoom in on focused points. Furthermore, the third point relates more to the analysis. An example of this would be to track where users are looking most in order to process this data later.

It enables rapid prototyping via in-editor simulation that allows you to see changes immediately. Furthermore, it operates as an extensible framework that provides developers with the ability to swap out core components. (Microsoft, 2020)

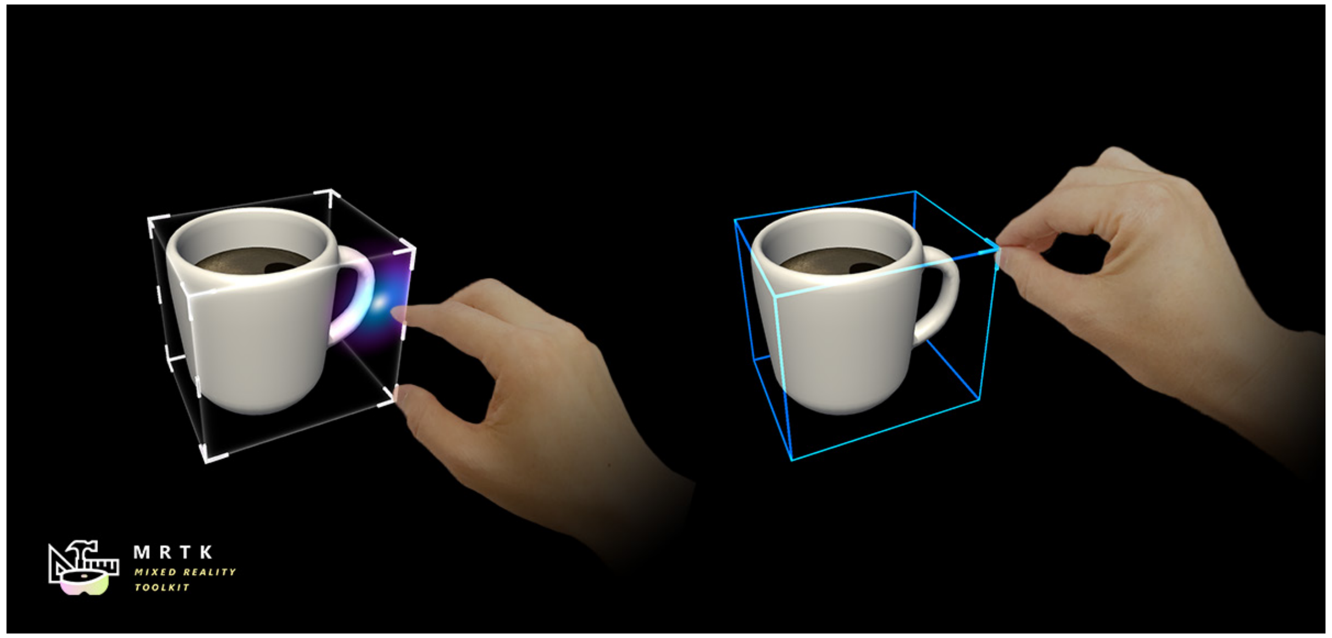


Figure 1. Screencast of MRTK Bounding Box. (n.d.). https://github.com/microsoft/MixedRealityToolkit-Unity/blob/mrtk_development/Documentation/README_BoundingBox.md

The Windows Mixed Reality Toolkit is essential for our work because Microsoft has already investigated intensively similar questions and has developed an already powerful solution. The ideas behind the MRTK and its functions are very inspiring for us and show us new possibilities of what can be realized with VR/AR/MR. Nevertheless, we also critically question this solution and try to create possible improvements.

Leap Motion SDK

Leap Motion Controller is an optical hand tracking system, created by a company called Ultraleap, using two infrared cameras and three infrared LEDs. It can track hands and fingers through these two cameras delivering raw sensor data to an advanced computer vision algorithm running on a computer. ("How Does the Leap Motion Controller Work?", 2015)) Before Facebook released the Oculus quest, which is able to track hand and finger through the inside out cameras on the headset, LEAP Motion was the only way for finger tracking in VR without the use of gloves. Attached Leap Motion Controllers to VR headsets were standard practice for all applications which were controlled just by bare human hands.

Together with the LEAP Motion Controller comes an SDK (Software Development Kit) for most known programming languages, IDEs (Integrated Development Environment) and Game Engines. This SDK is a very related project to our bachelor work because the designers of this SDK also created example environments. In these, there are examples of different hand types like for example hands with just bones, but also example environments, as soon as you face the palm of your hand, a small user interface appears, containing two buttons and a slider. Clearly, this is very intuitive. Also, before we had tested this example, we thought about something exactly like this. In the context of learning the basics of Unity, we tested this example of this small UI and also gave friends from our class the possibility to test this. Everybody immediately found the UI and could interact with it. However, also everybody found it quite a weird experience because of the lack of feedback. The buttons are designed the same way as buttons in real life, because of this, our brain knows what kind of feedback should happen, but it does not. We think through finding another form factor, and we can change this need for feedback. Still, the Leap Motion SDK is an inspiring related project and has a lot of beneficial solutions.

BMW gesture recognition

In 2015, BMW introduced its new 7 Series and with these new cars also its new gesture control, which was developed for this series. It was one of the first real applications for gesture control at this time. The idea behind it was to create an interaction between driver and vehicle that minimizes the potential distraction since the car is a place in which concentration is essential. According to BMW, the distraction of the driver can be significantly reduced by simple gestures.
A camera in the roofline makes this possible. It scans the area from the gear stick to the controlled display and recognizes the gestures. This area is particularly suitable, as many drivers rest one hand on the centre console anyway.

BMW has assigned the following applications gestures as standard:

- Changing the volume by a circular movement of the hand/arm with an extended index finger. Clockwise the volume is increased, and in the other direction, it is decreased.
- Incoming calls can be answered by tapping in the direction of the display with the index finger stretched out. To suppress a call, it can be rejected by swipe in the direction of the passenger.
- To interact with the parking assistant, the back camera can be controlled by a pinching gesture to show different angles.

Individual gestures can also be created and assigned to other functions such as changing the radio signal.

Many applications of gesture control today are just a "nice to have". For example, the new Google Pixel 4 smartphone that can switch a song with a single swipe motion or block incoming calls with a similar gesture. However, let us be honest, are these features essential? Although the BMW gesture control has the same functions, there is a crucial difference. Gesture control in the car creates an additional benefit, as the driver's concentration is much less affected than interacting with a screen while driving. This means that gesture control enables increased safety in road traffic.

This is precisely the reason why this project is crucial for us. Our goal is also to create an additional benefit, but not for road traffic but the operation in virtual environments through gesture control.

Virtual Reality Desktop

Virtual Reality Desktop was the first software letting users use their computer in VR. Users could customize the appearance of their desktop in their virtual surroundings, in order to watch 360 degrees video, browse the web generally as we know it, watch Netflix series in a virtual cinema, play games or just answer emails and use Microsoft Word and Microsoft Excel for homework.

Virtual Reality Desktop is an extended display which has the size, according to the users' wishes, between smartphone size and the size of a big house right in front of the user's face, almost every size is possible. Working in it is fun, but still, it is quite the same as working with a large and curved display. In VR, there is no need for limitations to square screens like there are in real life; this is, however, not explored in the product. While trying Virtual Reality Desktop, one thing was astounding; the capability of concentration was much higher. We did not use it for a long time but could feel that the level of concentration is quite high because there is no surrounding to distract your work and disturb your flow in writing. A few chapters of this thesis were written in Virtual Reality Desktop, and if the resolution of nowadays headsets were higher, many more would have been written in there.

Field Research

This section summarizes the findings from our field research, where we conducted multiple interviews.

Input Interactions Breakdown

What actions are we performing on our devices? Day by day we use our smartphone, our computer and various other digital systems like, for example vending machines. However, still, we are not performing multiple different interactions. For days we paid attention to what interactions we perform.

On our smartphone we tap, we swipe, we tap and hold, we tap and drag, and we pinch our fingers. These are all the movements we do on our smartphone, for all the tasks we fulfil on our smartphone. These are very few moves and ways to interact. For example, most phones are being picked up by literally physically picking up the telephone headset, the movement towards the ear, on our smartphone, we just tap on an icon. Our daily tasks were simplified and broke down to just five interactions. If we break down our interactions with our computer, it is even more drastic. We have a mouse with two buttons, and in most cases, a scroll wheel. The mouse is then fully controlled by sliding it over the table and clicking. In addition, we have the keyboard, a board with about 80 - 100 physical buttons. This leads us with the same amount of fundamental interactions, as in the case of the smartphone. We have just combined the actions from multiple input devices.

Of course, not just physical interactions can be broken down; the essential part is the virtual interactions. And by virtual we mean the interaction which happens virtually on the other side of the screen. There we trigger actions, we navigate and scroll horizontally and vertically, we zoom in, and we zoom out, and we write text.

Concept

Virtual Action	Physical Action, Computer with Mouse	Physical Action, Computer with Trackpad	Physical Action, Touchscreen
Trigger an action	left-click on the mouse	click on trackpad	tap
open context menu	right-click on the mouse	right-click on the trackpad (a matter of settings, mostly click with two fingers)	tap and hold
choose focus	move mouse	move one finger on the trackpad	never happens
navigate (e.g. maps)	click left and hold, move the mouse	click left and hold, move one finger on the trackpad	move one finger on the touchscreen
scroll up / down	Scroll wheel up / down	drag two fingers up / down	drag one finger up / down
zoom in	hold ctrl and use the scroll wheel	pinch thumb and index finger apart from each other	pinch thumb and index finger apart from each other
zoom out	hold ctrl and use the scroll wheel	pinch thumb and index finger closer to each other	pinch thumb and index finger closer to each other
move item	left mouse button hold and drag mouse	tap and hold then hold and drag	tap and hold then hold and drag
write text	use a keyboard, push on the buttons	use a keyboard, push on the buttons	use the on-screen keyboard, tap virtual buttons

Interviews

Interview with Oliver Sahli

Fortunately, the ZHdK has an Immersive Lab which is an artistic and technological research project of the Institute of Computer Music and Sound Technology. It is a media space that integrates panoramic video, surround audio with full touch interaction on the entire screen surface. Oliver Sahli is working in this lab in the area of VR. He is also doing his Master in Design, specializing in Game Design. We had the opportunity to talk with him about our intentions and to explain and discuss our work. This conversation was very enlightening and exciting.

We talked for example, about haptic feedback and that we consider it as an essential element for the whole experience in VR. Oliver agreed with us, but he believes that today's technology is not yet ready to simulate adequate haptic feedback which could have a positive effect on the immersion experience. Pure vibration feedback does not contribute much to improved immersion.

Another topic we discussed was locomotion in VR. Currently, the most common solutions are the curved pointer and the limited locomotion. Fascinatingly enough, for Oliver Sahli, locomotion in VR was an essential point that in his opinion, should not be underestimated. Nevertheless, he also asks himself if it is necessary to be able to move in VR. He agrees with Mike Alger (interview bellow), who asks himself the same question. Concerning working in VR, except in exceptional cases, we agree with both of them. Why should I be able to move in VR while working with a natural walking movement? What is the advantage? However, if we now look at the gaming industry, which wants to make the VR experience even more intense, it makes sense to make the movement more natural, because this way the consumer can immerse himself even more intensively in the digital world.

We had another exciting discussion about UI elements and environments in VR. In the VR systems tested by Oliver Sahli and us, many UI elements are not thought through but instead adopted from other platforms. For example, windows are still rectangular. However, in the struggle to establish oneself as the leading VR system, it would be essential to question these things and think differently. The same applies to the environments you encounter in VR. The very first place you are in after putting on the headset is a living room, or when you watch a movie in VR, it is played in an environment that reflects a movie theatre. Why are the environments in VR modelled based on real-life reality?

Interview with Mike Alger

We had the chance to interview Mike Alger, a Senior Interaction Designer at Google. He sees himself as an: "Interaction designer for digital eyewear, spatial computing, immersive wearables, mixed reality, XR, or whatever you want to call it. Designing and prototyping user interactions for current and future displays." (Mike Alger, Mike Alger) Mike did his master degree in Moving Images at Ravensbourne University in London. In his thesis, he investigated and visualised different methods of volumetric user interfaces and experiences within the enormous scope of a virtual reality operating system. (Visual Design Methods for Virtual Reality, Mike Alger)

The Interview with Mike was very insightful and gave us a great new perspective on the topic. After we explained our current state to him, we first asked for some related work and other projects, realised by himself or others, which inspired him or which he thought could help us. First, he spoke about the "Mixed Reality Toolkit" by Microsoft." The Mixed Reality Toolkit is a Microsoft-driven project that provides a set of components and features, used to accelerate cross-platform mixed reality app development in Unity. (explained in Related Work) This is mainly developed for Microsoft's Hololens, which is an augmented reality headset where the user interacts with the augmented content by just using his or her hands. We thought that the most intuitive interactions to humans are to use the bare human pair of hands. But Mike was not feeling the same way in this discussion. He argued with undeniable thoughts, some of them we had ourselves, some of them were new and highly interesting. So what is maybe the most important thing we also have looked into, because of an input Luke Franzke gave us after we had our first presentation, is that just using our hands and arms to interact through gestures is very tiring. Mike just took this one step further, he said, that we could use our hands and arms for a long time, but we do not want to because we are too lazy.

Furthermore, he points out that it is not unnatural for humans to use tools. However, more about that in the section "Using a controller in VR is not that wrong". Another, significant, fact also is that we usually have a resting position for our arms which is not claiming a lot of muscular effort to hold the arms there, for example, we cross our arms, or we lay them on a table or an armrest. This comes into conflict when we design user interfaces because we are very limited in space. Considering that buttons need to be relatively large due to the low resolution, we will have in head-mounted displays for the next few years at least, as well as the limited space we have left for original content after taking away the uncomfortable and unreadable space explained earlier. We are not able to design extensive menus with a lot of different items, options or content in general without some sort of scrolling through this content. Back to our issue with the arm resting position, before and after an interaction with an item on top your comfortable site of view you have to cross other, potentially bulky, items in order to get back to your resting position. You have to pay attention not to press other items. Also, because of our laziness, we do not want to stretch our arms fully, so the ideal distance from eye to an item is about two-thirds of an arm's length.

Interview with Max Rheiner

In another interview, we had the opportunity to talk to Max Rheiner. He is a former member of the Interaction Design Team at the Zurich University of the Arts. He was Head of Masters in Interaction Design and taught specific modules in the Bachelor's degree. As head of masters, he mentored many projects in the field of virtual reality. Earlier, during his studies at the Zurich University of the Arts, he developed "Birdly". Max now continues to work on "Birdly" in the context of a startup called "Somniacs" and distributes these worldwide. Birdy is a full-body VR experience in which the user finds himself as a bird in a virtual world. The user can fly over different landscapes and through different cities using the inputs of his or her full body laying on this machine which is the "controller".

Accordingly, Max Rheiner has excellent know-how in the field of VR and was an ideal interview partner for this area.

At the beginning of the interview, we gave Max Rheiner an overview of our topic and explained what we have already done and tested. We talked about the different VR headsets and their finger and hand tracking. Max Rheiner pointed out to us that Facebook has its own Oculus Computer Vision department here in Zurich, which we could contact. This would undoubtedly be very exciting and instructive.

In the following conversation, we went into more detail about the comparison of interactions with the pure hand and a controller. When we talked about the idea of a ring as a simple controller, Max Rheiner remembered Padrone, a Swiss startup who exhibited their product at CES at the same time as he did. (see Figure 2) The product is a ring that was designed to replace the mouse completely. With simple gestures from the resting area, the user can interact with the computer.

Furthermore, Max Rheiner advised us to make a decision. According to him, we have already looked at and tested many different things and have now enough insights to decide. It is not advisable to tackle many different topics at once. He would choose a system and then use this as a starting point to build our work on. He also advised us to define our target area like, for example:

> *"Hand interaction in immersive space in relation to GUIs."*

It is important to do many simple tests and to work iteratively in these conditions, as this is very satisfying for us and results are obtained quickly. This topic area also has

Figure 2. Photo of Padrone Ring (n.d.). https://www.gamestar.de/artikel/mausalternative-padrone-ein-ring-sie-zu-steuern,3339182.html

a lot of potential for testing and research; for example, 3D UI is still very new and not fully developed.
Since Max Rheiner's area of expertise is more focused on experience and immersion, we naturally wanted to know from him what is necessary for an excellent immersive feeling. Surprisingly, according to him, not much is needed. If you are visually entirely absorbed, it is already quite sufficient. He quoted: "Es ist wie wenn man jemanden einen Eimer über den Kopf stülpt, man ist dann in diesen Eimer drin" Which is translated:" It is like putting a bucket over somebody's head. No matter what, this person then is inside this bucket." The question is then more importantly whether one wants to stay in the bucket, because it is well and pleasantly done, or if one wants to get out again immediately.

Known problems on a visual level are the lack of depth of field and too low resolutions. However, this is being worked on, and when the technology is ready, Max Rheiner also believes that it is quite possible that we will work in VR in the future. One reason for this is that in VR, the whole eye is covered by a screen surface. Called Ultimate Display, and is developed by Ivan Sutherland in 1965.

Figure 3. Photo of Sutherlands Ultimate Display (n.d.). https://www.roadtovr.com/wp-content/uploads/2016/05/ultimate-display.jpg

Concept

Another advantage of VR is that the workspace can be divided into different areas, which is extremely helpful for finding your way around.

However, the handling in VR is crucial, says Max Rheiner. Something similar to the mouse, which is ideal for 2D UIs, should be developed for VR.

Findings Interviews

In summary, we can say that these interviews have helped us extremely in terms of the path we want to take with our project. We received several very exciting inputs concerning different problems in VR.

The most important points which we take with us for our work:

- All three interview partners clearly criticized the current UIs in VR and believe that they need to be reconsidered.
- VR headsets need to develop further in different areas to become more versatile.
- Interactions from a resting position are essential to staying in VR for a long time.
- The type of locomotion is not essential for our work

VR/AR ergonomics research

Zones of Content in a virtual space for headset displays

In the context of Mike Alger's master degree, he did a lot of research and did sum them up in 2 manifesto videos. One finding which will be needed for further details in the interview was his research in defining zones where the content should be placed into. According to different levels of importance, the content should be categorized and finds its place accordingly.

Most VR headsets in 2020 have a field of view between 94 and 100 degrees. The field of view mostly is circular, because the lenses are circular. Mike Alger cites numbers from a talk by Alex Chu at a Samsung Developer Conference. When you are sitting in a chair, you can rotate your head horizontally 30 degrees comfortably and maximum 55 degrees. Vertically you can move your head between 20 and -12 degrees comfortably and between 60 and -40 degrees maximum. (see Figure 4) (Ravensbourne University London, 2016)

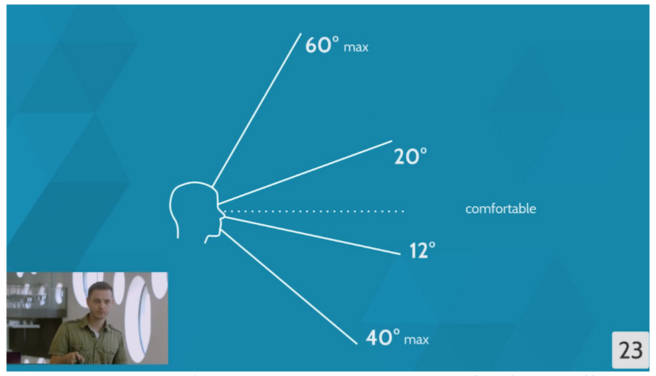


Figure 4. Still image from Mike Algers Thesis Video (n.d.). https://vimeo.com/153517639

In the following illustration a birds-eye view of the different zones is presented. (see Figure 5) Starting in the center, there is a zone where your eyes are not able to focus because things in this zone are too close. Everybody knows this from reading a book, and your book is coming too close to your eyes, so you can't focus on the letters anymore. Nothing should permanently stay in this no-no zone. Then there is the Main Content Zone, this is the zone you can move your head comfortably, and the field of view added. Things in the Peripheral Zone are still visible for the user but should not be used for essential and necessary content. What is behind your back is not visible unless the user is literally turning her or his body, that is what Mike calls the Curiosity Zone. (Ravensbourne University London, 2016)

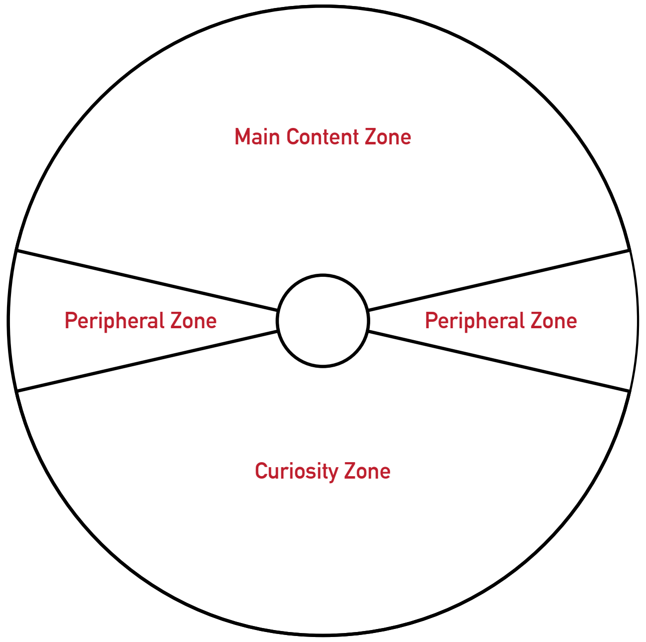


Figure 5. Illustration, defined zones According to Mike Alger. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Then there comes the next issue, the vergence-accommodation conflict. The vergence-accommodation conflict describes the phenomenon which happens in VR headsets that eyes are physically not able to focus on the actual focusing target because the lenses and the displays inside the headsets are too close to the eyes. Most drivers of headsets do solve this problem for us designers quite well. Still, we have to take this issue into account.

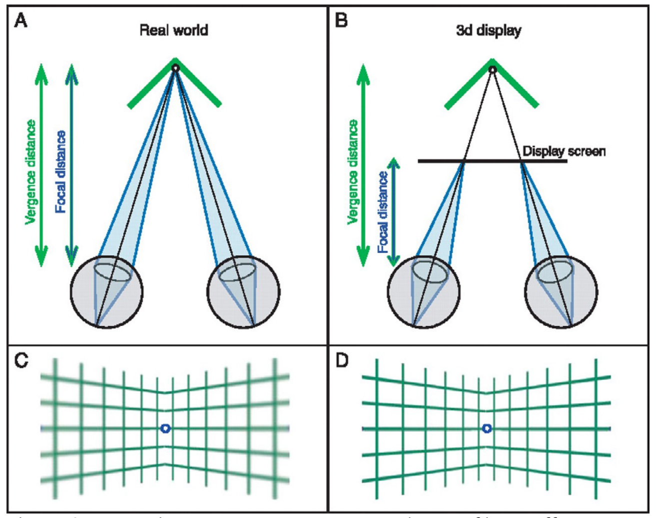


Figure 6. Illustration, vergence-accommodation conflict, Hoffman et. al. Copyright Journal of Vision, 2008.

Mike Alger found that the optimal distance to place text in VR is to go for a virtual distance of 1.3 meters. (Ravensbourne University London, 2016) We could test and verify this by reading newspaper articles in virtual reality in different distances and found that for us everything between 1.1 meters and 1.5 meters worked the best with an average font size of 12. Research in Office Ergonomics shows that the best angle of our neck is between a straight position and a position where the neck is bent down about 25° (Katarina Kacjan Žgajnar et al. 2016)

Therefore, taking in account that we see the sharpest at around 1.3 meters, adding the central content zone but subtracting the no-no zone as well as the length of our arms as well as the human laziness not to raise our arms above shoulder level, we end up with a quite limited range of motion to interact with the crucial pieces of user interface components. (Mike Alger, 2016)

Our natural resting position

We know that our natural instinct is to stay in a comfortable position at any possible moment. We realized this when we observed people using VR headsets if there is an arm not used at any given moment they let the arm sink or find the armrest in case of using a headset sitting in a chair.

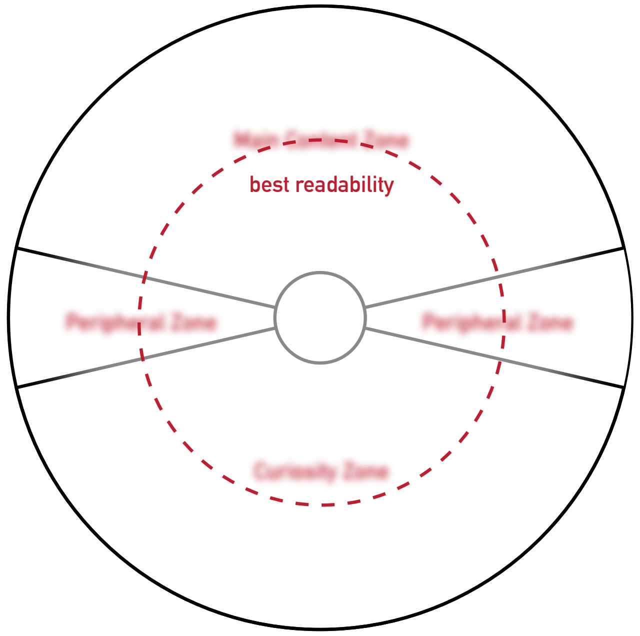


Figure 7. Illustration, best readability at 1.3 meters. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Having this in mind, we need to find ways of performing basic tasks even from our resting position in order to create an experience with no tiring aspect. Otherwise, we cannot use these virtual spaces for a long time without getting tired arm muscles.

Using a controller in VR is not that wrong

In our interview, we have done together with Mike Alger; we heard his statement about controllers. He told us that he is always in favor with some sort of a controller in the hands of users who use VR. Starting the interview, we both were convinced that the most humane way to interact with virtual or augmented content was by using just our bare human hands. By the end, we were both not that convinced anymore. Mike agreed with us that the most rational way, interacting with anything is to use just bare human hands. But still, also the first homo sapiens were using sticks, stones in their hands to interact with their surroundings comfortably. Therefore it might be a little less intuitive than using our bare hands in real life (for example picking up a mug in order to drink coffee). However, stills are way more intuitive to use with virtual and augmented content than having no haptic feedback or no feedback at all and in addition not being able to be free with arm movements because we do cross other UI components which might trigger an interaction not being intended by the user.

Key refinement decisions

Controller or not?

We analysed many different controllers, which are currently used for VR purposes. Most of them were designed for gaming. This shows, for example, the joystick which plays a central role on almost every controller. What we realised quite fast was that there is no such thing as a “one style fits for everything” kind of solution. To interact with different UI components needs completely different input methods than to play a game or to watch a movie.

From our interviews and mentoring, it became clear that we had to decide regarding the interaction medium. Do we work or design a controller, or do we just deal with our bare hands? To decide, we compared the two and discussed the advantages and disadvantages of each side.

Hands

Pro
The hands are not occupied by a controller and can therefore still interact with other objects, such as a pen or a keyboard. The hand is also the most intuitive way to interact with its environment. In addition to the visual sense, humans often feel the need to experience the world through the sense of touch. Also, the hand-eye coordination of the human being is so well developed and trained that work processes can be completed faster. In comparison to a controller, where the human must first memorize where each button is and what its function is. If we decide to use our own hands, this means that we do not have to develop a new type of controller, which means that we do not have to do a lot of product design, physical computing and engineering.

Contra
If the gestures and interactions are designed in such a way that the user has to make many and large movements with his hand and arms, this quickly becomes very tiring and exhausting. Most users are not conductors who are trained to move their arms too long away from a relaxation position. This problem is also known as Gorilla-Arm Syndrome. Apart from this, it is still not possible to cross your hands without interrupting the hand recognition process. In the future, this will certainly be possible.

Controller

Pro
A significant advantage that a controller has over bare hands is that it allows haptic feedback. This means that the hand does not simply float around in space and feel nothing, but knows exactly when an object is interacting with it. Evolution has also shown that humans like to use tools that they use with their hands (stones, branches, pens, measuring instruments, forks, etc.). Nowadays, it is no coincidence that we work with a mouse and keyboard on the computer. It merely has turned out that these two tools are the most efficient for working on such a system. It is the fastest way to get work done. Also, the input is very accurate. Probably more accurate than your own hand. (e.g. There might be a comparison to a mouse on a computer screen and the finger on a touch screen. Everybody knows the issue from touching the wrong button on a touchscreen, but this barely happens with a mouse, which controls a cursor on the screen.)

Contra
Depending on how the controller turns out, the hands are occupied and have no possibility to interact with other things.

One especially important point is that our work will be more speculative, because designing a controller that is designed to work in VR takes more time than we have available, in our opinion, testing can only be done once a product has been developed. The focus is more on the development and design of the controller than on the interaction in VR.

Decision making

Based on the listed pros and cons, as well as based on the skillset we two bring together, we have decided to focus our project on the interactions we can perform with just our bare hands.

Our goal is to find interactions with which we can perform basic tasks like selecting and opening different items, scrolling through lists and text and zooming in and out. This also leaves the advantage of free hands. Free hands allow us to use specific kinds of controllers like pens or keyboards to perform more specific tasks like, for example, exact drawing, writing text or model in three-dimensional space.

VR vs AR

Virtual reality and augmented reality are not fundamentally different; the difference between these two worlds lies in the degree of virtuality. While VR manages completely without the real world, in other words, is 100% virtual, Augmented Reality is a subclass of Mixed Reality. This means that the real world is mixed with the virtual world.
For our project, we must decide for one or the other because designing an environment for both is out of scope in terms of effort. A UI designed for VR can be well adapted to AR in terms of interaction due to the similarity of interaction. This means, for example, that a button can be used in the same way, no matter if VR or AR. The style of a designed UI could become problematic as it is more challenging to transfer into the AR space. As soon as a UI is embedded in the real world, components may merge with the background and are thus almost no longer visible. To prevent this, some adjustments would have to be made after the transfer of the environment.
Another point to consider in AR is where to place the interactive components because, in AR, work is almost exclusively performed with their own hands, which limits the distance for accessible UI elements.

Of course, both worlds have other advantages and disadvantages. We however settled with the decision to develop our project for VR environments primarily. AR is at the current point of timeless mature than VR, which means that the hardware available, the know-how as well as the community is stronger developed for VR. Looking from a customer perspective there are already great AR products available like the Microsoft Hololense 2, and they come however with a price tag of 3'500$ which prevents the general public still to adopt this technology on a large scale. For VR, there are multiple affordable systems available such as the Oculus Quest, developed by Facebook, which offers maximum freedom to the user by being wireless. Besides hardware and know-how, it is also essential that we can leverage a strong community that can support us in case of problems, but that will also be early adopters for our findings. This community is already developed for VR technology since the devices are widely spread.

Our partner Sensoryx also focuses on VR as a first application, out of the same reasons. Therefore it makes sense for us to start working on VR as well since this increases the chances of early adoption of our findings and it ensures that we can get support when needed.

What hardware to use?

We decided to use Facebook's Oculus Quest for our project. It is Facebook's first standalone headset with inside out tracking. While the most popular headsets like the Oculus Rift or the HTC Vive still used a computer with massive graphical power as well as sensors distributed in the room to track the position of the headset, the Oculus Quest integrates all of that in the headset without any cables or external computer. Four ultra-wide-angle sensors in the headset control Oculus Insight, which pinpoints your surroundings and instantly translates your movements into VR. The image has a resolution of 2560 x 1440 Pixels (QHD / WQHD), is powered by an ARM processor, 4 gigabytes of RAM and either 64 or 128 gigabytes of storage. (Wikipedia, Oculus Quest)

The main reason for using the Oculus Quest is, however, that it is since 2020 capable of tracking fingers and hands which is key to our project. Hand Tracking included in the headset allows us to use the device from the shelf and not having to integrate a Leap motion device to enable the hand tracking.

The first tests with this new technology convinced us even further since the tracking performance was comparable to Leap Motion, which already has ten years of experience in tracking algorithms and the hand tracking features.

For our user tests and prototypes, we will, therefore, use the Oculus Quest and their hand-tracking. We also evaluated to use the glove developed by Sensoryx to perform the tracking. For our initial prototypes, we, however, considered the advantages of the fully developed SDK fo Facebook with its active community as a Kickstarter. IN any case, the findings and prototypes can be easily transferred to the Sensoryx system in a second stage, when the project has been successful.

What development environment to use?

There are different ways to create applications for VR. But so far, also due to the focus on gaming applications, the most established technique is by using a game engine like Unity, Unreal, CryEngine, Frostbite or one of many others. Game engines are designed for programming and creating environments in three-dimensional space, precisely what is needed for VR as well. Within the group of all the game engines, there are many different manufacturers with many different products, designed for various applications. Differences are programming languages, the support for different platforms (Windows, Apple OS, Android, ...), price plans and most importantly, their capabilities. Unreal is the preferred tool to design hyper-realistic games, whereas Unity focuses more on secure handling and easy access, resulting in a broader community.

Since both of these engines are new to us, an essential criterium in choosing the right engine was the amount of support we could expect from the corresponding communities in order to solve our problems.
This is why we decided to use Unity since the online

support and documentation will allow us to fast acquire the needed skills to conduct this project. Further, there are many open source libraries and example projects available, which can be leveraged to kickstart our own development of a VR application.

Therefore we will use Facebooks Oculus Quest together with Unity to create our prototypes, user test and our final exhibition.

Development Concept

After having taken all these necessary decisions and gathered the input from literature and our field studies, we finally came to a refined objective for this project. Our main goal is to investigate hand gestures for the three main interactions we have evaluated(scrolling, zooming and triggering). Identifying suitable gestures for these actions will build an excellent foundation and starting point for hand-based VR navigation. We would like to evaluate different gestures by leveraging user tests and combine them to create a logical and intuitive navigation system. It is for us essential to directly work with users and to have them experience different gestures to come to a conclusive result from a user perspective and not motivated by other factors. Therefore we want to create a demo environment, where we can quickly test different gestures with people, which could then be used in the future as a platform for gesture-based user tests.

Our focus on hand-based gestures is mostly motivated by the fact that currently, virtual reality environments, user interfaces and applications are mainly designed for interactions with controllers. Mostly these controllers come with the head-mounted display and try to aim for one style that fits all kinds of solutions. This is not bad per se, but it introduces complexity and is for many applications not intuitive for first-time users. Further gestures offer customization to the task, which is more versatile and can also incorporate the user's preference for specific movements.

Further currently available controllers are mainly used for VR gaming, which is also influencing their design. Our field research showed that there is no one style fits all kind of controller, there is, and most likely always will be, the need for diverse controllers for diverse applications. Product designers working in VR will need another kind of controllers than gamers, educators or drawing artists. Gestures, however, allow for a large variety and can be adapted to specific tasks. This is why we think designing interactions using only bare hands and not in combination with any kind of controller will add a significant benefit. Using gestures, users can interact with all kind of UIs, applications or systems and where specific controllers provide an advantage, they can be added to enrich the experience.

Further, we think it is time to rethink current ways of interaction which are mainly based on the technology legacy (computers, tablets, etc.). The current advancements in hand tracking allow for these new gesture-based interactions to be explored and have the potential to disrupt traditional interaction schemes. Especially adding the 3rd dimension also calls for the reinvention of our ways of interaction, to fully leverage the newly gained spatial freedom.

For these reasons, we want to make it an integral part of our project to question UI components in VR as well as the interaction with them, with the goal of user-centric reinvention.

Therefore, as a prototype, we want to create a VR environment to show people the potential of gesture tracking and the usage of gestures. We would like to build a platform where we can test gestures for the three most commonly used actions (zooming, triggering and scrolling). The focus should lie on the gestures their application, ergonomics as well as intuitiveness and not on the UI elements even though these will probably play a significant role during our design process as well.

Summarized, with our work, we critically question to days interactions in VR, which are mostly based on controllers and want to find out how the use of gestures in VR can have an impact. We want to find out how a gesture control system could look like and what it takes to design it intuitively and fun to use. We would like people who test our prototype to experience the nearly unlimited potential of working in VR and that they see the benefits that might lie in front of us.

Development

User Tests

In this section, we present the results from various user tests that were conducted in order to learn about, identify and select gestures for our key actions.

Ring Test

In our first user test, we initially concentrated only on the key actions scrolling and clicking, since these are essential to interact with traditional systems(laptops, phones, tablets, etc.).

To test these two actions, we created a web page with a running text to scroll, followed by a button which should be clicked. We placed a Leap Motion in front of the computer such that test subjects had the impression that their hands were tracked. We instructed the test subjects to scroll to the end of the text and click the button. In the first round, the participants were only allowed to use their hands to accomplish the task. In a second-round, they had to use a ring. The goal of the test was to identify different gestures that can be used to complete the given task. To motivate participants to use multiple gestures, the text was scrolled down only after the user used multiple gestures. By that, the tester had the impression that his first gesture was not working and therefore will try different ones until the task was completed (by us scrolling). The same was done for the clicking of the button. By that, a broad set of potential gestures could be collected.

Scrolling only with your hands

It was interesting to see that in most cases, the first gesture used by the testers for scrolling was a kind of swiping movement as it is used for interacting with touchscreens. The gestures that were used by the participants after trying the swiping motion were very diverse. For example, rocking with the open hand in different directions, sweeping the thumb over the index finger with the hand closed, as if there was a trackpad on the index finger, a kind of waving gesture to different sides, including a waving movement or pushing the text up and down in the air like on an iPad.

Scrolling with the help of a ring

With the ring, three groups of test subjects could be identified. A first group still used the whole hand to make gestures despite the ring. A second group concentrated more on the finger with the ring and performed gestures with that finger. A third group did not understand the ring as a finger ring, but more as a kind of controller. Especially group two and three got very creative in terms of gestures that they used, which was the goal of this exercise. Most of them used some sort of rotational movement. Be it to turn the ring on the finger, or to turn it between the fingers or to make a circular movement with the whole hand.

Click only with the hands

Most of the participants immediately searched for the haptic feedback of the table by tapping the tabletop with their index finger. When this method did not work, the test persons became creative again and tried out many different gestures. These gestures ranged from merely opening the hand, to pinching with the index finger and thumb, to snapping, to typing with all the different finger combinations.

Clicking with the help of a ring

Again, users were using the ring in very different ways. If a tendency had to be identified, most would tap the ring with their thumb or index finger. There were also more sophisticated solutions, such as tapping with the index finger through the ring or turning the whole hand while the ring is on the index finger, comparable to opening a lock with a key. In some cases, the second hand was added as an aid to interact with the ring.

Summary

This test was very informative and served its purpose to identify many gestures for these simple actions. This obtained collection of gestures for scrolling and clicking can be leveraged for the next steps of the project. We also noticed that it is not easy to introduce some kind of controller since a significant number of participants were confused about what to do with the ring. However, we are also aware that this test could be quite different by changing different parts. For example, we can safely say that the appearance of the ring can be used to control the result of the test in different directions. Nevertheless, it was an essential step for us to find out whether a ring has the potential to be a controller.

The intuitive need for haptic feedback by most of the users also confirmed our hypothesis that interactions of two-dimensional screens are heavily influencing the behaviour of the testers and that it could be beneficial to think about how to translate these into three-dimensional interactions.

Zoom User Test

Right after we started with the user tests on the campus of Zurich University of the Arts, the Corona quarantine situation started. The first and most obvious way to conduct our test out of the quarantine was to use videoconferencing (Zoom/Skype). We contacted multiple people to participate in our next user-test which took about 20-30 minutes., Here we asked the testers to perform gestures for the three critical interactions in front of their camera. We shared our screen with simple UI elements where they had to perform gestures to zoom, click and trigger actions. Similar to the tests in real life, we observed their hand movements and tried to find similarities within these movements.

This test led to unsatisfying results because it was very time consuming, and only a few tests could be conducted. This lack of data prevented us from gaining any insights other than that this method is not suited for our needs. That is why we are not discussing the results in detail.

The first website enabled user-test

Since Videoconferenceing was too time-intensive to gather vast amounts of test data we decided to create a Website, where people can participate in our tests and provide us with their insights, without leaving their home.. s. On the website, people should be able to record and upload videos directly from their webcam, and we could display instructions on the site in order to explain the different tests. The complexity of such a website is rather high due to the implementation of webcam access and secure data transaction and storage. It is, however, crucial that the user-test procedure is as simple as possible from a user point of view since a complex process would prevent people from participating and sharing their data. The website was implemented using HTML, CSS, PHP and javascript, which is not stock standard for web applications working with video on web servers, but served our purposes and matched our hardware. As a video framework, we used videojs-record. (Github, Collab-Project - Videojs-record)

The visual design of the website was implemented using a template by “styleshout” which we adapted for our needs.

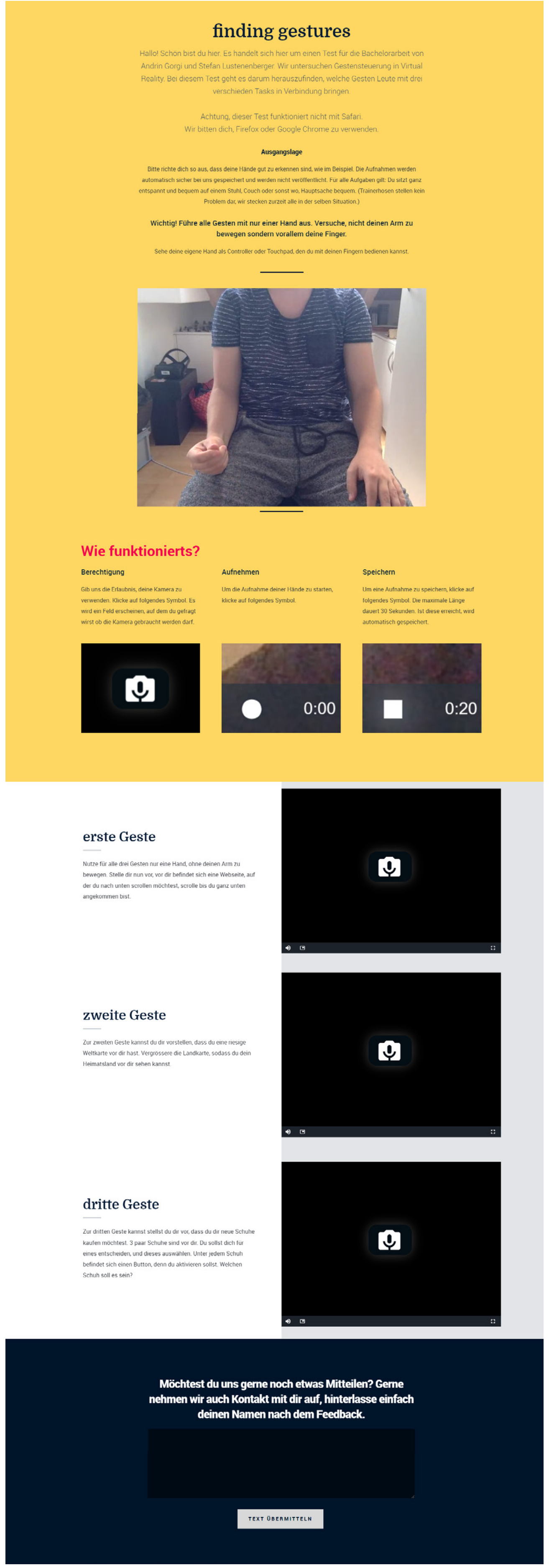


Figure 8. Screenshot of findgestures.andringorgi.ch. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

When entering the website, the user first faces a welcoming text which briefly explains the goal and purpose of our bachelor thesis. Then some general instructions are given to instruct the users on how to position their hands in the video (showed in the picture below), to ensure that their gestures can be recognized. Then it is explained how users can record and save videos. Users are asked to repeat their gestures three times to make the evaluation process much precise.

Figure 9. Screenshot of instroduction image on findgestures.andringorgi.ch. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

We recruited user testers by reaching out to people in the Interaction Design Department, Friends and Family to share the website and to participate in the user-tests. The website is still online: https://findgestures.andringorgi.ch and tests can be conducted to support our research.

The first round of user tests

It was decided to explain the three actions of interest (scrolling, zooming and triggering) using small scenarios around the action. This was done to prevent users from being biased from their experience with these actions from tablets and smartphones.

Scroll
The scenario exlaining the task of scrolling was the following:" Stelle dir nun vor, vor dir befindet sich eine Liste mit allen Ländern dieser Welt. Jedoch ist nur ein kleiner Teil davon auf einmal ersichtlich. Navigiere durch die Liste, bis du Zypern erreichst, welches das letzte Land der Liste ist." Translated this means:" Now imagine that there is a list in front of you with all the countries of the world. However, only a small part of it is visible at once. Navigate through the list until you reach Cyprus, which is the last country on the list."

Zoom
TThe scenario to explain zooming was the following:"Zur zweiten Geste kannst du dir vorstellen, dass du eine riesige Weltkarte vor dir hast. Vergrössere die Landkarte, sodass du das Land deiner Träume vor dir siehst." Translated this means:"As a second gesture, you can imagine that you have a huge map of the world in front of you. Enlarge the map so that you see the land of your dreams before you."

Trigger Actions
Zooming was explained using the following scenario:"Du hast nun das Land deiner Träume vor dir und siehst nebenan die Nachbarländer. Damit wir uns sicher sind, dass du das richtige Land vor dir hast, bestätige deine Wahl mit einer Geste." Translated this means: "You now have the land of your dreams in front of you and see the neighboring countries next to it. That we are sure that you have the right country in front of you, confirm your choice with a gesture."

Evaluation and Findings
An unforeseen incompatibility with a Google Chrome update led to the loss of some user-test as it prevented the saving of the videos on the webserver. People were then instructed to use Mozilla Firefox in order to prevent further data loss. This is because there was no error message whatsoever.

However, still, we are happy and thankful for everyone that participated in our test. 36 user-tester were nevertheless successfully submitted video footage of their test. Some participants even did submit multiple videos for some gestures. From these 36 videos, we were able to conclude the first set of gestures.

The analysis is following, divided into three different gestures we were looking for. In each of them, there are different numbers of submitted videos because of the creativity of our participants. Some participants did submit multiple videos for some gestures. For example, somebody submitted three different gestures of scrolls but just one trigger gesture. That is why there are not even numbers for all the gestures.

In the following, the observed gestures are explained using static images taken from the videos. To illustrate/explain the observed gestures for each gesture and equivalent commonly known situation is described where such a movement is used.

Scrolling action

Tester used, on 13 submitted videos, a natural scroll gesture by waving their whole hand up and down. Most of them communicated the direction to scroll to by the speed of their hand. If they want to scroll upwards, they waved upwards faster than downwards. Some testers also stretched their hand on their way up and loosened their hand when waving back down to indicate direction.

Figure 10. Still image of "Full-Hand-Waving-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

This very obvious gesture was expected to occur a lot as it is similar to what people use in conversations when they, for example, want someone to turn up the volume of a radio. The gesture is easy to perform, but since the movement is significant involving the complete arm, the user will get tired by the movement very quickly.. Since the movement is significant it is also challenging to scroll accurately with this gesture. It is sometimes helpful to just scroll a little bit, time by time. Due to this quite significant movement, it requires a well trained gross motor skill. Therefore we think that this gesture is more suited for quickly scrolling over content but not well suited for precise scrolling motions.

Another original scroll gesture also was performed by a group of 13. Tester used a scroll movement with just one finger, the index finger. We all know this gesture. We perform it so many times daily. It is the gesture we use to scroll on our smartphones if it is used with two hands. It is the gesture used to scroll on tablets. It is the gesture used to scroll on any vertical touch screen like ticket vending machines or ordering screens at restaurants like Mc' Donald's.

Figure 11. Still image of "One-Finger-Scrolling-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

This gesture has one significant advantage in comparison to the first gesture, which uses the whole hand. The gesture is just using one finger. This means it can be used while having the hand staying in the resting position without even moving it. Further moving the finger is a very accurate and precise motion. Keyboards, Mice and touchscreens have taught us to have precise fingers. This leads to a more precise gesture if you use just one finger instead of the whole hand, which is moved by the arm.

Four additional videos were submitted with a very similar gesture, using two fingers instead of one. People using two fingers added the middle finger to the index finger, and all of them had their fingers close to each other. Interestingly people using two fingers, slightly turned their hand towards their camera when performing the movement. On the videos of the people just using one finger, we mostly just saw the top of the users' index fingers, from the nails on upwards. On the videos of people using two fingers, we just saw the palm of their hands.

Two of the submitted videos contained spinning movements where people were spinning their fingers. They were holding their index finger horizontally and rotated it around its axis to indicate scrolling. The direction of spinning indicated the direction of spinning. This gesture looks like the gesture we perform on the side of our forehead to tell somebody that he is crazy. This gesture also allows people just to use their fingers without moving their whole hand or arms. It also allows controlling the distance to scroll accurately by the amount and timing of spinning.

Also, on two of the submitted videos, people were spinning their whole hand instead of just the two fingers. We know this gesture from, winding up charging cables around the palm of our hand. In comparison to spinning just one finger, this is a more significant movement involving the complete hand, which in most cases is also

moving the arm. This forces the person to leave the resting position leading to fast tiring of the hand and arm.

Figure 12. Still image of "Two-Finger-Scrolling-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

On five of the submitted videos, people did swipe sideways. Most of the videos the participants were using their arms quite heavily. It was a surprise that swiping sideways was used to scroll downwards. It is even more impressive that five videos were submitted this way. These people had the intuition to swipe sideways, and we would like to know if that was their first intuition or if this is a second submission. This shows that intuition is not the same for every person. But to analyse the gesture completely unbiased we also have here the problem of leaving the hands resting position.

The last gesture we got submitted is in our eyes the most exciting scroll gestures. We got two testers submitted a gesture, where the participant used their index finger as a touchpad for their thumb. This was very similar to some of the patterns we observed in the ring experiment. There the ring was used on the index finger, and the thumb was accurately controlling the movement of the ring through a touch interface. A limitation to the usage of such gestures is, however, the current accuracy of finger tracking since the movement is very subtle and therefore difficult to be recognized. The accuracy of the tracking system would have to be very precise in order to create a smooth experience for the user because the users' movements also are very subtle and precise. In this case, we would have to assume that hand tracking works perfectly fine. Nevertheless, the gesture is from a usability point of view a perfect mix between, the hand staying in the resting position and a very accurate way of indicating speed and distance to scroll. The subtle and precise movement could, however, be challenging to perform by motorically disabled people (e.g. trembling hands), which could be seen as another

Figure 13. Still image of interesting "Smartphone-Swipe-Like-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Zooming action

Sixteen of the participants performed the gesture where the index finger and thumb are pinched together or spread apart in order to zoom. This is the commonly used gesture on trackpads, tablets and phones to perform this action and is, therefore, most likely inspired by these devices. This gesture works well since it is possible to use it within the resting position of our hands, and since we can perform very accurate movements with our fingers leading to a precise gesture.

Figure 14. Still image of "Traditional-Two-Finger-Zoom-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Another 16 submitted videos we got showed the classical zoom gesture using two hands (separating and bringing them together to zoom out or in). From movies like "Iron Man" or "Minority Report," this gesture is widely known and associated with AR and large screen interfaces. Also, Science Fiction movies used this gesture extensively to interact with holograms or gesture-based interactions, which might have influenced the participants. In movies, these gestures are over-dramatised. These two hand movements are. However, very large movements making them not comfortable in the long term. Finger positions are not intuitive and weird, with a little bit of "Doctor Strange-Esque". This is why we believe this kind of gestures should only be used to perform tasks on an occasional basis where the ample interaction space is needed.

Figure 15. Still image of "Two-Handed-Zooming-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Only one participant submitted a gesture that was not already widely known from other applications. One very interesting gesture, we did not think about this in advance, but it really makes sense, and now we even wonder why just one participant submitted this gesture. We call it the "Come to me" gesture. The gesture was similar to the gesture people perform if they want to tell somebody to come closer (an opening and closing of the hand with the palm facing upwards). Some people, perform this gesture just with the index fingers, others use all the fingers but the thumb to perform the movement. This gesture has like the smartphone inspired zoom gesture the advantage that the accuracy is high due to the accuracy of the finger movements. Further, it is not necessary to leave the resting position, just a flipping of the hand would be needed (the resting position is for most people facing the palm downwards).

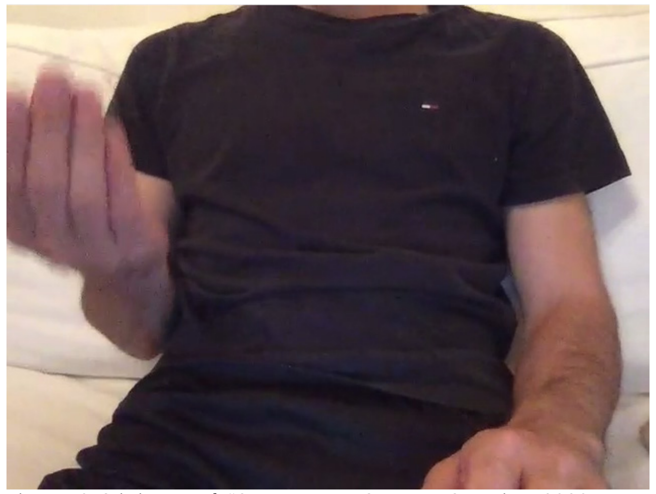


Figure 16. Still image of "Come-At-Me-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Triggering action

Last but not least, the trigger gesture was analyzed. This is used to trigger actions, like pressing a button, opening a file or confirming previous steps. We observed that before triggering an action, the people were searching for that action to be triggered using their hand. This raised the question what the "cursor equivalent" in VR space could be, which is covered at a later stage.

The first type of gestures was the most obvious one or at least our participants were most biased on this one. Twelve participants used a simple tap or poking motion with the index finger to perform the triggering action, in the same way as it is performed on trackpads and touchscreens. On another twelve videos, people were either tapping in the air, performed some a poking kind of gesture or used a surface to tap on similar to a trackpad. Two of the participants made the trigger gesture by using their table or their legs as a trackpad the others performed the action in the air. Comparing these two versions, they both have advantages and disadvantages. The poking gesture has the advantage, that if there are multiple options and one to choose, it is easy to determine which of all these options was meant to be selected. The "trackpad-clicking-version" is easier to perform and more comfortable to use out of a resting position. It is, however, more challenging to communicate which option should be triggered if there is more than one.

Figure 17. Still image of "One-Finger-Poking-Trigger-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

The last assignment for our participants was to confirm their previous zoom action where they had to "select" a country. Therefore Six testers used the thumbs up hand sign to confirm their action, and another two performed the hand sign standing for "OK" (making a circle with thumb and index finger). These are widely used hand signs in everyday life to confirm things and therefore, self-explanatory choices of gestures when it comes to just confirming previous actions. They are however not useful if an option out of a catalogue has to be chosen.

Therefore they can more be understood as a confirmation action following on a selection action.

Figure 18. Still image of "OK-Approval-Gesture". Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

The same tester with the innovative scroll gesture also submitted a new kind of triggering gesture. This was an inspiring moment because the submitted gesture was close to the one where the tester scrolled. In these two gestures was a pattern which could be used for a system of gestures. The gesture was very similar to pressing a button on a smartphone with the thumb, as it is usually done when using the smartphone single-handed. The submitted video showed precisely this motion but using the side of the index finger instead of the smartphone (taping with the thumb on the side of the index finger).

Findings

Summarising our findings from the first website-based user test, the most prominent finding is that people were very biased from gestures they perform all the time on touchscreens and trackpads. This was not very surprising, even though we tried to minimize the direct associations as much as possible by using scenarios. We learned that there is a desire to have familiar gestures. This makes sense, in order to not force people to relearn individual interactions. But there are also disadvantages to this. By using gestures based on 2D screens, it might be the case that the full potential of 3D enabled gestures is not leveraged. This is similar to the situation where multi-touch displays where introduced, that allowed for new possibilities in terms of user interaction gestures. Therefore, it still makes sense to investigate new gestures in order to tap into the potential of 3D enabled gestures. That is allowed by the currently available accurate finger tracking with head-mounted displays. If we just reuse the gestures from multi-touch displays, we believe, we lose a significant potential.

On most of the submitted videos, testers were using their arms quite heavily, and people did either not read the top part of the website or did not take it account when performing the tasks.

Another important finding was the importance of detecting the start and end of a gesture. On touchscreens, pinching is used for zooming. Therefore, the gesture is activated whenever the index finger and thumb are touching the touchscreen simultaneously. In VR there is no physical screen that can be used to decide on the starting of an action, which makes this topic important and challenging at the same time, with the clear need for further investigation.

Gesture definition 0.5

After the analysis of the first big user test, we tried to define gestures for the first version of gestures we can use in VR. There was one major issue with the first user test, in any case. In the analysis, it is already pointed out to the problem of the very biased gestures which were submitted. Therefore we struggled hard, finding a system of three gestures which was in the frame of what the desk research told us to be correct for the ease of use but also the ergonomic aspects we read about and learned within our user tests. We could not manage to find a system of three gestures which met all the requirements we wanted to meet as well as have seen enough users use these gestures in our website user tests.

In order to get compatible gestures in our website user test, we needed to relaunch the website with more exact instructions in what is allowed. For example, we needed to make sure that it is not allowed to move entire arms but use subtle finger movements. So we started to work on the website test 2.0.

Second website-enabled user-test

In a second test, we expanded the action set to drag and drop of items, to copy and paste text and even to write text. This showed how versatile our user-test platform was and how it can be diversified towards any task where a gesture needs to be found. By considering the results of the first user test we, however, decided to refocus on the first actions since we wanted to dive deeper in order to find a set of gestures that work as a connected system supporting user intuition and ergonomics at the same time.

Even though this test gave us no insights which we could use to proceed with the primary goal of our thesis, we could learn from it, and it definitely led us to future steps, where we plan to investigate further gestures.

Third website-enabled user-test

We confirmed in the first two user tests that the website was a very efficient and user-friendly way to collect user test in relation to specified tasks.

In the first user test, we were able to gather many different gestures but also found that many were inspired by the interaction with 2D devices and therefore not exploring the full potential of the 3D space. Further, the user's intuition led to many gestures that were intuitive at a fist sight but not ergonomically if they have to be used over a long duration of time, since the movements performed were large and leaving the resting position. Therefore, no set of gestures could be created on one side aligns with the state-of-the-art research on gesture ergonomics as well as with the tester's intuition. Therefore, we redesigned our test to collect gestures with more precise guidance towards ergonomics.

We adapted the text on the website describing the three actions and added keywords to clarify the task. We also wrote added a restriction that the gestures have to be performed by the fingers to motivate gestures that are not leaving the resting position and that are therefore more effective from an ergonomic point of view. Moreover, because we learned in the first test that the intro text is now read by most of the users, we again pointed out on this in the first assignment.

Overall the test setup using the website structure was similar to the other user tests. We also used the chance to adapt the code to enable users with Google Chrome again, which was necessary from a usability point of view, since most participants were using this browser.

The new instructions on the website were the following :

- Intro: "Use only one hand for all three gestures without moving your arm."
- Scrolling: "Now imagine that you have a web page in front of you on which you want to scroll down, scroll until you reach the bottom."
- Zooming: "As a second gesture, you can imagine that you have a huge map of the world in front of you. Enlarge the map so that you can see your home country before you."
- Triggering: "For the third gesture, you can imagine that you want to buy new shoes. Three pairs of shoes are in front of you. You are asked to choose one. Under each shoe, there is a button, perform a gesture to activate it. Which shoe should it be?"

The third scenario was changed since we wanted to investigate not only the confirmation but also the selection process. Investigating the option of a gesture-based cursor for VR, where in the first website based user test, participants had to confirm their previously done choices. In the case of controller-based VR, a motion-controlled "laser pointer" is used to point and select things. To investigate the same using gestures, we asked the testers to select a shoe and to trigger the action then. We also specified that pressing a button is necessary for confirmation to prevent confirmation-only gestures like the thumb up movement.

Evaluation and Findings

In comparison to the previous test, fewer testers submitted videos, however with still enough to come to conclusive results. By changing the test instructions, the submitted gestures, were now in line with the requirement from an ergonomics point of view to not leave the resting position. To invest in a second test was worth it, we saw that we finally found words, which led people to perform gestures in the "allowed boundaries". For example, not getting out of their resting position. And second, we were happy to receive quite a lot of submissions we really liked. By receiving a large variety of gestures that fitted the state of the art of a comfortable gesture that can be used over a long period of time, we were able to start combining gestures into systems that met all the requirements we had from the desk research, resting position, not much arm movement, high accuracy etc..

Gorilla Arm Test

The "Gorilla Arm" or the "Gorilla Arm Syndrome is a key finding from user tests and studies for the usage of touchscreens which is highly relevant for our VR research. Here the definition from Technipedia:

> *"Gorilla arm is what happens when the user interacts with a vertical touchscreen for a long period of time. The arm becomes tired, and it becomes more difficult to interact with the interface. One excellent example is the use of a floor-standing kiosk, the kind you might find in an airport library. Short-term use is relatively easy for most users — but as time goes on, the burden of raising the arm and making selections causes a certain kind of fatigue, since the arm is not physically supported in any way." (Technipedia: Gorilla Arm)*

Many studies are observing and investigating the Gorilla Arm and the fatigue that is experienced in the arms in general. Apparently, it is also the reason why Apple decided not to introduce a touchscreen to its Macbooks. In a press conference on Wednesday the 20th in October 2010, Steve Jobs said:

> *"We have done tons of user testing on this, and it turns out it does not work. Touch surfaces do not want to be vertical. It gives great demo, but after a short period of time, you start to fatigue, and after an extended period of time, your arm wants to fall off." (Wired: Why 'Gorilla Arm Syndrome' Rules Out Multitouch Notebook Displays, 2010)*

Having these studies and findings in mind and knowing that most knowledge can be translated to the usage in VR, we still wanted to test it for ourselves. We wanted to find out how capable our arms are working in uncomfortable positions to get an impression on what timescale the Gorilla arm is occurring. Since the issue of Gorilla Arms is not an issue which just occurs in virtual reality we wanted to make it as easy as possible for our friends at home to test it as well, which means that there is no possibility for a test we can actually perform in VR. We searched for a solution so everybody we ask can realise the user test at home and send us a video of him and her testing his arm capabilities. We, therefore, instructed our testers to find a level surface between the chest and shoulder height with the size of at least one DIN-A4 paper. This surface is supposed to be as flat as possible but could be realised by whatever they have available at home. On this surface, our testers should sort cards that they shuffled before. They were supposed to record a video of them performing the task. Sorting should be done alternating between sorting by class and sorting by value until they fatigue. Most importantly, the testers were not allowed to take down their arms. By that, we could investigate the time window within which the task was perceived as being comfortable. Since the testers were also instructed to comment about their feeling while performing the task, the border between comfortable and tiring could be identified.

We observed that for most testers after 40-60 seconds, the arm starts to become very tired. Between two and three minutes, it started to hurt. Pain starts to rise in the upper arms as well as in the shoulders. After three minutes also, the last participant wanted to quit.

This was about the same as we could find in the publications about Gorilla arms and meant for us that these large gestures were not suited for our VR application.

Figure 19. Still image from submitted video of "Gorilla-Arms-Usertest". Copyright 2020 by Marcial Koch.

The identified Gesture System

After having finalized all user-tests using the website interface, we investigated the combination of the three gestures into a system. It was important that the gestures would fit together from an intuition point of view as well as that they satisfy the ergonomics criteria from our literature research. Intuition was mainly assessed by the number of submission of the same gesture. We had our favourite user submissions accordingly to our research as well as through hours of moving our fingers the same was our testers did.

The need for a cursor

On touchscreen applications, fingers are used to indicate with which item it should have interacted. When it comes to computer screens, a cursor is used. With trackpads and computer mice, the cursor is moved on the screens, pointing on the items where interaction is desired. So far, in most VR applications, there is the usage of a "laser pointer" kind of cursor to determine with which item-specific actions like zooming, triggering or scrolling should be performed. The action is performed where the pointer is aiming.

In order to use VR environments without controllers, we need to be able to point on things with just our hands. The most intuitive thing is to use the index finger to point on things. This is also where the index finger got its name from. But we also want to be able to point on things without moving our hands too profoundly, at least without forcing the user to leave his resting position. Most testers solved this problem by using only their index finger to point on things.

Figure 20. Still image from Project Video, Cursor Gesture. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Scrolling

More than a third of the testers brushed with their thumb over their index finger to scroll. The user is using the surface of his index finger as a trackpad where he performs the same movement to scroll as we do on the touchscreens of mobile phones. This is also in line with the ergonomics criteria of staying in a resting position. When introducing this gesture in the gesture system together with the already mentioned pointing of the index finger as a cursor, we had to perform slight adaptions. Since index finger was already used as an indicator of the cursor, the "tracking surface" from the index finger had to be moved down from the tip of the finger more towards the middle to allow scrolling while pointing. The brushing motion is still as intuitive as on the video just the location has been moved. This is as comfortable to perform as on the index fingers, and a self-test showed no tiring signs even after a long time scrolling up or down.

Figure 21. Still image from Project Video, Scroll Gesture. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Zooming

Having a gesture system combining scrolling and pointing the next step is to integrate zooming. But when it comes to zoom, it was a bit more complicated. Considering the user-tests it was found that indicating the start a zooming gesture is rather straightforward, but signalling the end was not trivial. Therefore, we decided to use a three-step process leveraging the traditional zooming movement that most users were performing inspired by trackpads and smartphones:

1. First, the middle finger and thumb are pinched together to select the item to be zoomed and to indicate the starting of a zooming action (the index pointing gesture is disabled at that point in time)
2. Stretching out the index finger is leading to zooming in reaction while moving the index finger closer to the tip of the thumb is resulting in zooming out the reaction. By that, the user can precisely control the amount of zooming.
3. When the pinching of the middle finger and the thumb has released the end of the zooming action is indicated (index pointing is again enabled).

This scheme has the advantage that it leverages the most intuitive gesture. However, it ensures clear triggering and ending of the action that can also be detected with currently available tracking systems. Further, the ergonomics requirements are met, and the gesture fits well in the developed system using only one hand. Using this gesture in the first zooming prototype, we learned that the human hands excellent motor skills are amazingly accurate, and it is a lot of fun to zoom this way.

Figure 22. Still image from Project Video, Zoom Gesture. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Triggering Actions

Triggering actions to for example press a button is convenient on touchscreens due to the haptic feedback. Users tap on the position of the item on the touchscreens to trigger actions. We kept the taping motion, since it was used by many testers in the user-tests but integrated it into the gesture system. A user will choose an item from a list by pointing at it with his index finger. The triggering is performed by double tapping the tip of the middle finger and the tip of the thumb. Since this uses the same overall hand posture as the scrolling as switching between the two gestures is very convenient, allowing the system of gestures to be intuitive.

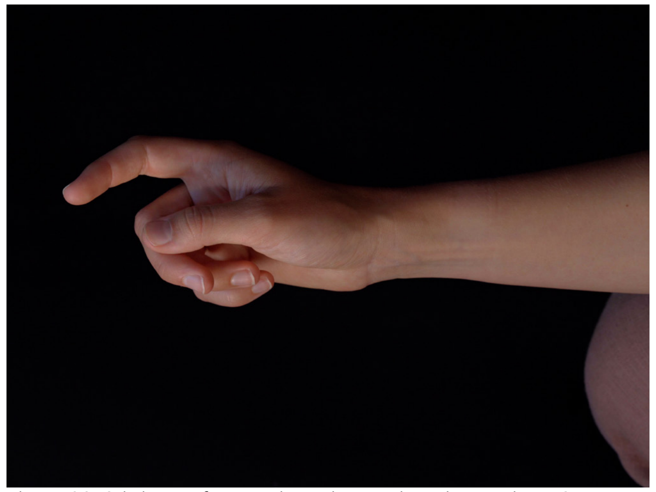


Figure 23. Still image from Project Video, Triggering-Actions-Gesture. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Prototypes

Hands-on Gesture Recognition

After getting in touch with Unity and learning how to set up hand-tracking in Unity using the Oculus Quest properly, we started our first gesture recognition experiment. The goal was the creation of a simple scene where the code for gesture recognitions can be tested. The user, wearing the headset, is seeing a grey cube laying on the floor. The tracked hands of the user were visualized in the standard Oculus black mesh look.

We recorded simple gestures like “thumbs up”, “victory” or “single stretched fingers”. For recording the gestures, we performed these gestures and ran a function which saves the positions of all the bones of the hand. We then measure the distance from all these bones to the root of the hand to make sure that the recognition works by using just the hand itself and does not rely on the hands’ total position. These distances were stored together with the name of each gesture.

For the recognition, the finger bone’s distances to the root of the hand were compared in real-time with the ones we have stored from the saved gestures. If the values are not differing more than the tolerance values, the gesture is recognized and triggers a Unity function. This implemented test function was changing the colour of the cube in the VR scene. Each gesture was triggering a different colour. This gesture detection algorithm was inspired by a youtube tutorial from Valem, see in the acknowledgements section for further information.

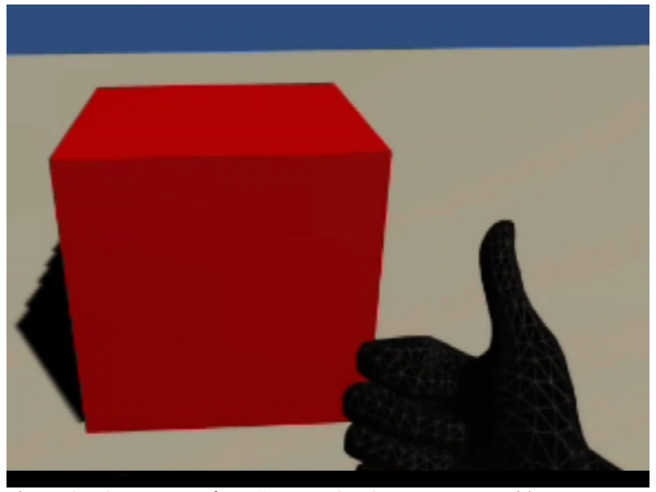


Figure 24. Screenshot from “Hands-On-Gesture-Recognition” Prototype. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Findings

This first gesture recognition worked well, proofing that the selected development environment is suited for the project. The accuracy of the Oculus Quest hand tracking also satisfied our requirements". For recognition, it was crucial to fortune the tolerance values find the best setting.

First working gesture in Unity

In the next step towards our prototype, we worked on integrating the user input into the Unity functions to interact with objects in the VR space. For that, we had the goal to change the size of the previously mentioned grey cube using a two-handed gesture. With the left hand, the user could start and stop the gesture by pinching his index finger and thumb. While the index finger and thumb is pinched, the user could set the size of the cube by the distance between his right hand's index fingertip and thumb tip. If the user's right hand's index fingertip was very close to the thumb, the cube was small, and if the user spreads his fingers, the cube will get bigger. As soon as the user is happy with the size of the cube, he can release the left-hand pinch, and the size of the cube is set.

Button Arrangement Test

We have learned from the research of Mike Alger, that the arrangement of content around the user is of high importance. Especially when navigating in the VR space from one object to another, the content has to be arranged such that no other objects are in the way, to minimize wrong selections and actions. We also got aware of the problem. That while moving the hand from the resting position to the place where a button, for example, is placed and back to the resting position, there is supposed to be nothing in the way. Because of the risk of touching or interacting with items which were not meant to be interacted with.
To test content arrangement, we created a second VR scene to experiment with different arrangements of buttons. Like in the first gesture recognition scene, the tool to work with was a cube which is changing colours. Buttons were used to change the colour of the cube, and we investigated how different arrangement of these colouring buttons affect the accuracy time it takes to perform recolouring.

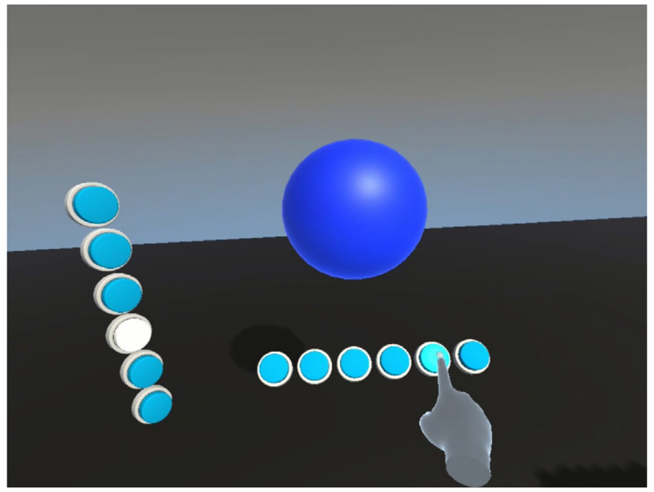


Figure 25. Screenshot from "Button-Arrangement" Prototype. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Findings

Due to the conditions in March 2020 we, unfortunately, could only test it with ourselves as well as two people living with us. But still, all the ones who tested it had one common denominator: we preferred button arrangements in which buttons were arranged horizontally. But it was not just an intuitive preference but also allowed for a faster usage of the buttons. Further horizontal arrangements also showed much higher accuracy (fewer buttons pressed by mistake while navigating to the buttons) Therefore we found that horizontal arrangements should be preferred if possible from a usability and efficiency point of view.

Zones Test

In the interview with Mike Alger, we talked about the zones he defined in his master thesis, in which a virtual reality environment should be divided, to improve the comfort while working in VR. To experience these zones and the conclusions drawn from them, we set up a simple test scene in Blender and Unity. We separated the different zones by walls and defined the zones with a colour code:

- green, the central content zone,
- blue, the peripheral zone
- red, the curiosity zone.

We decided to make the division with walls so that the user gets the feeling of looking into different corridors. This division was made both horizontally and vertically to cover the whole spectrum of head movement. We also used the No-No Zone to transform it into an interactive panel that can be operated from a resting position, as suggested by Alger. All these defined zones were then imported from Blender into Unity so that we could test it effectively in VR.
To be able to test this accurately, we also wanted the horizontal divider walls to rotate with the head (turning of the head) so that the division is always in view. The vertical dividers had not to be moved as they were always large enough to be in view. The fact that the partitions, which define the horizontal zones, move with the tilt of the head makes no sense because firstly the vertical movement spectrum of the head is smaller than that in the horizontal and secondly the walls are high enough that they are still visible.

To achieve that we to map the position and rotation of the OVRCameraRig (name of the Camera Object in Unity which is provided by the Oculus integration of Facebook to create VR applications for your Oculus HMD glasses) to the position and rotation of the partitions. We had tried many different codes, methods and examples which we found during our research and adapted to our scene. Among them were functioning like "LookAt" or "FollowMe" which were promising but still did not work. We soon realized that the calculation of rotations in Unity had to be solved with the help of quaternions.

> *"The quaternions are members of a noncommutative division algebra first invented by William Rowan Hamilton. The idea for quaternions occurred to him while he was walking along the Royal Canal on his way to a meeting of the Irish Academy, and Hamilton was so pleased with his discovery that he scratched the fundamental formula of quaternion algebra, $i^2=j^2=k^2=ijk=-1$, into the stone of the Brougham bridge." (Quaternion, 2020)*

By that, a functional prototype was created to test the different zones proposed by Alger.

Development

Findings

The horizontal Main Content Zone is a little larger than the standard field of view of a human and is designed such that everything that is not in the field of view can be seen by comfortable and minimal movements of the head. In the test, this could be confirmed, and it was enjoyable to operate in this zone.

In order to get the peripheral zone in view, the head had to be turned more to the side, which leads to a stretching of the neck and cervical muscles, which reduces the comfort. This can be reduced by a chair with a rotating function by simply rotating the whole chair.

The Curiosity Zone is the counterpart of the Main Content Zone and is entirely located behind the user. It can only be accessed by adjusting the complete sitting position, for example by turning the entire body by 90°, by rotating the chair to one side, or by turning the head extremely into one direction. Even in these cases, only a small portion of this zone can be seen.

For the vertical subdivision, we found that the weight and its balance of an HMD (Head Mounted Display) influence these zones. For example, in the Oculus Quest, the three different zones are shifted more downwards, since the headset puts more weight on the user's nose compared to other VR headsets. This makes looking straight ahead or upwards more tiring.

As a result of this test, we can confirm that the zones defined by Mike Alger work very well also for our setup, but that the user's comfortable field of view can be extended by external aids such as office chairs without losing comfort.

Considering the balance of the used hardware device can also add a benefit by adapting the zones accordingly.

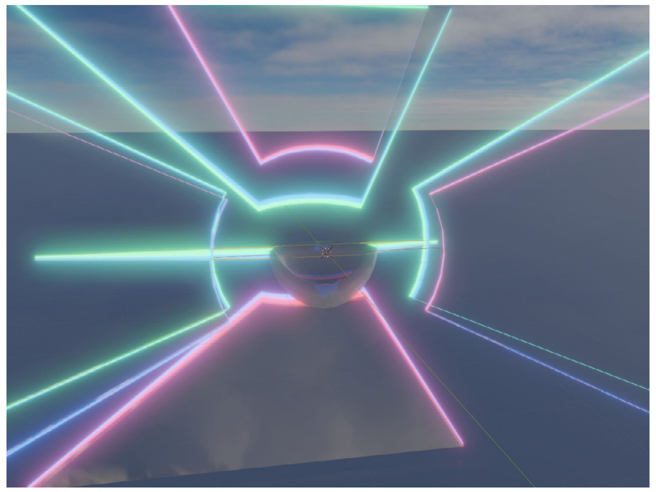


Figure 26. Screenshot from "Content-Zones" Prototype.
Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Demo

First Iteration: Zone-based work environment

We have tested and confirmed in which zones it is pleasant to work and where it is less. Since our demo also has an educational purpose, we decided to include this division of zones in our planned demo to explain to the testers the different zones and their properties.

For the initial test, the main focus was to test the functionality. The final demo should, however, not only serve its purpose but also be visually attractive and a pleasure to test. For these reasons, we decided to improve the demo environment in that regard. 3D modelling tools like Blender, allow to create freely using different objects and materials.

To support our creative process, we created a mood board. This is a good starting point to create visual changes. We believe that walls still create a visual division of zones, but should still be less present. Therefore we created a small mood board to inspire us.

Finally, we used glass panels to separate the zones and a glowing effect on the walls to indicate the colour code. This way, the partitions are still visible but are much more subtle to the user. The control panel, as also fitted to the walls. We also decided to add a floor so that the user does not float in the air, and a horizon was created. As a design element, the floor was designed with a waved structure. With a 360° HDRI image, the scene got a seamless and uniform background and illumination it at the same time. In order to stay focused on the task, a neutral cloud image was used for the background. Finally, several dummy UI elements were placed in the different zones of the environment, like windows of different sizes and formats and 3D objects. To show the potential of such a system. Before we covered all these windows with an image using UV maps, we wanted to know if the import of materials from Blender to Unity is as easy as with the 3D models.
Unfortunately, it turned out that it is not so easy to import materials with unique settings like transparency or light effects into Unity.

In Blender, there is the option to bake textures. “Texture baking is the process of transferring details from one model to another. The baking tool starts a certain distance out from the model (usually a low-resolution model for game use). It casts rays inwards towards another model (usually a high-resolution sculpt)”.(Render Backing)

This did not help in our case. So the best way was to reconstruct the material/shader in Unity. Meaning the full implementation of the materials and light effects were directly performed in unity since blender textures (bake texture function) were not compatible with the interface we were using. The glass material was implemented, due to time constraints we did not implement the colour coding, which is however documented in the form of sketches.

to improve the UI dummy element’s appearance several screenshots of programs like Microsoft Word, Spotify, Adobe Reader, GoogleMaps and folders were used to simulate a real working environment. In Blender, we mapped the different screenshots to the corresponding placeholders and imported them into the Unity scene. For the PDF, we decided to use a long, rather slim window, in which three pages can be placed under each other, and the text is still readable. The Word was mapped to a large surface to display several pages side by side. Therefore it was possible to test different settings and identify their respective advantages. We also placed the windows in the different zones, so that the user notices immediately what zones are suitable for working and which are not. The zooming action was implemented as a first gesture test.

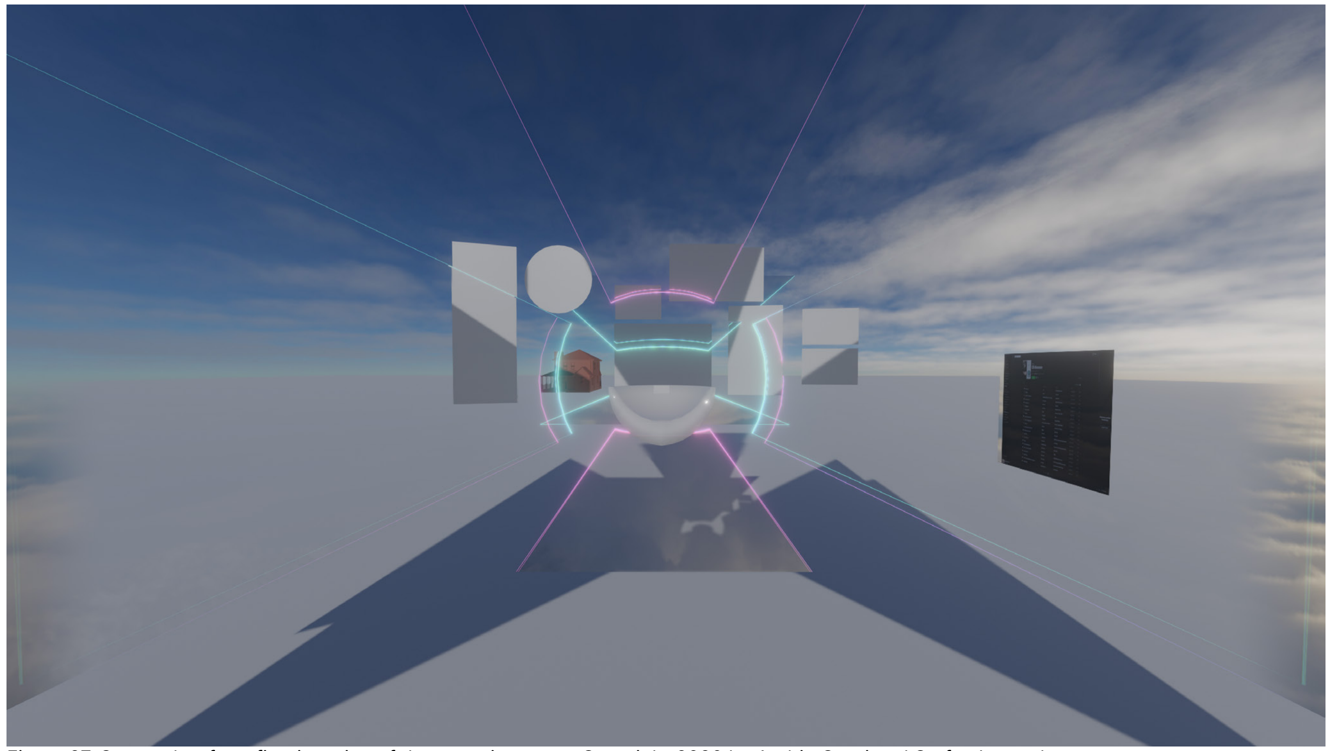


Figure 27. Screenshot from first iteration of demo environment. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Second iteration: Functionality reduced environment

The first version of the demo enabled the user to envision how a VR enabled work environment could look like but distracted from the goal of testing different gestures in such an environment. Especially the background and the photorealistic windows distracted the users from the actual goal.

For this reason, we reduced the demo to the essential elements. These are One element to zoom, one to trigger and one to scroll. Also, the division of the Zones of Content was kept. Further, a button was added to switch between the different applications. The applications were organized in a control panel which was equipped with buttons for each application. We not only reduced the number of elements but also simplified the layout. The goal was to move away from concrete examples and to abstract the environment to the essentials. In the first level of abstraction, we used a globe for zooming, a content-wise matching poem for scrolling and a TicTacToe game for triggering. In a second step, we broke the globe down to a section of a three-dimensional city map.

Our mentors and we still felt that the VR demo environment should be further abstracted so that nothing distracts from the gestures we developed. So, in the third iteration of abstraction, we went more in the direction of icons. The text became just lines, and the city became the typical icon for a picture (two mountains and the sun). The TicTacToe remained because it can be designed with very simple elements and adds a fun factor for the user. Next, we adapted the different elements in order to harmonize the VR environment better and to express the association to the individual components more clearly. We also created a new background consisting of a blackboard, with the benefit of making applications stand out better from the background.

Furthermore, we created a colour scheme. In a first draft, we defined the background colour. Then we selected four more colours, blue, green, yellow and red. Every application and the corresponding button got, therefore, its own colour to distinguish them from each other. We also investigated a two-colour system with blue and white but concluded that the four-colour system supported the distinction between the applications better.

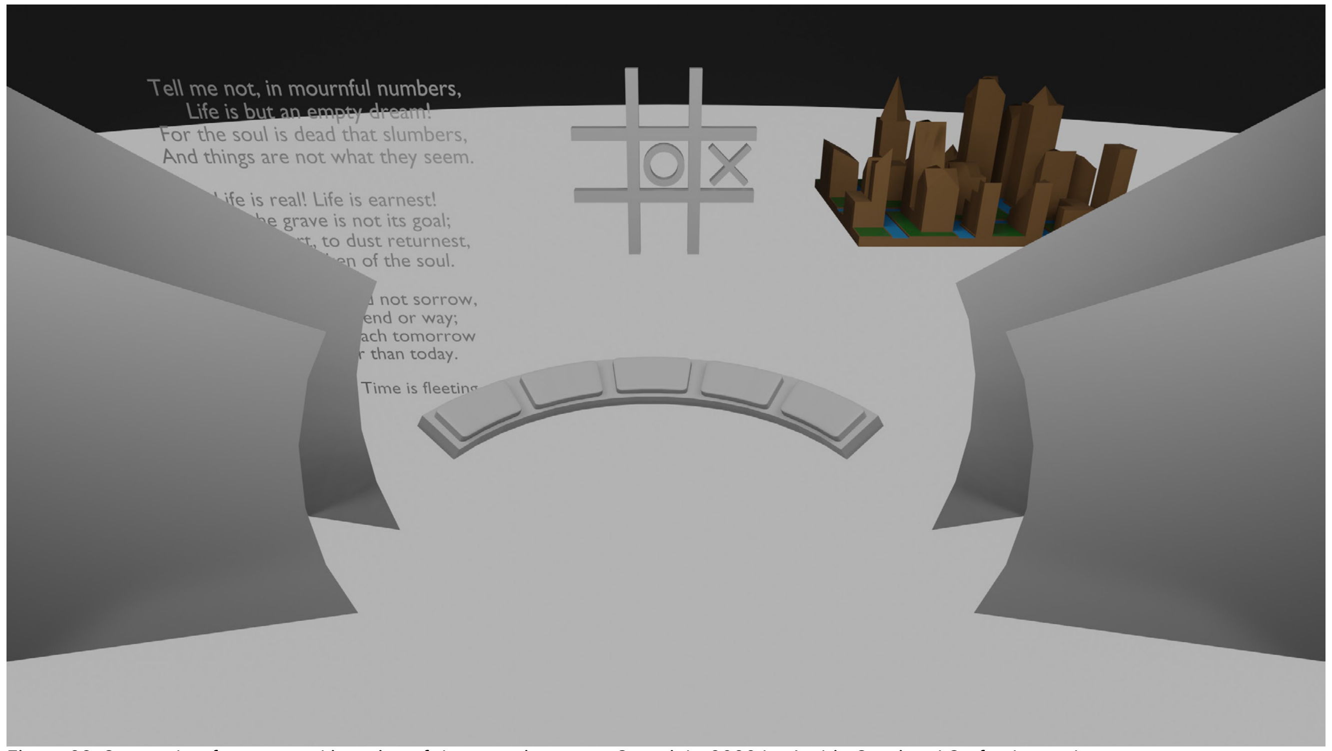


Figure 28. Screenshot from second iteration of demo environment. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

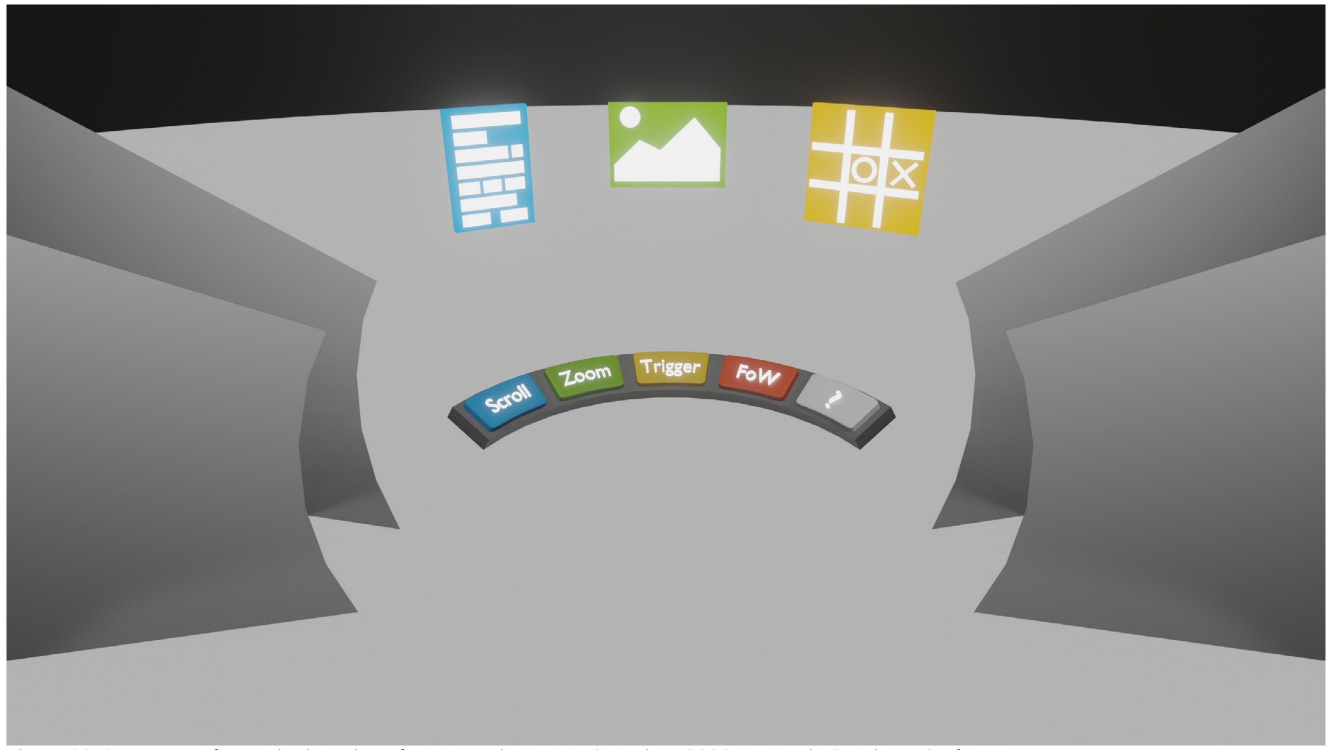


Figure 29. Screenshot from third iteration of demo environment. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Third iteration: Functionality reduced environment

Based on the external feedback, we decided to create another iteration of the demo by reconsidering the layout as well as the colour concept. Further, we investigated different evolutions of buttons.

In order to make the design more modern, we adjusted the line thicknesses as well as the control panel. The used thick lines gave the complete environment a fresh look, which was not our intention. Therefore we investigated different line styles and settled for thinner lines with a more delicate look. Further, the basic structure of the panel was removed to reduce complexity. The buttons were also reorganized and placed freely floating in the air.

Since we question with our study the conventional way of interaction in VR with the use of our emerging gestures, it would be wrong not to question the conventional usage of buttons as well. Why should a button be pressed if there is no haptic feedback? We, therefore, intended to develop a button which is designed to use gesture control and uses the strengths of VR. One of these VR strengths is undoubtedly the three-dimensionality. It was essential for us to explore this. In Blender, we designed different versions with different operating modes. We made designs where the button still must be pressed, and others where the hand must be held in a particular area to trigger an action.

For the classic pressing buttons, we were also inspired by the buttons designed by Mike Alger and the Microsoft Mixed Reality Toolkit. Mike Alger was investigating the missing haptic feedback. The idea of his designs was that the fingers do not touch the button at all, but that the button moves away from the finger and is triggered when a specified distance is reached.

In the case of area triggering, we started with simple shapes such as a box with an open side into which the hand can be held. We also tried other geometries with the idea of using this as a button. As an example, we can imagine a simple sphere that is triggered by touch. Further, we investigated lights to mark the activation zone.

We integrated three different variations into our environment, with the idea to test them in VR and to decide which one is best suited for our purposes. The first button type was close to the conventional design, but with the idea that the button and the fingers never touch each other when the button is pressed. The other two types were conceptually the same; they both marked an area where the hand can trigger an action. They, however, differed in terms of design. One was a frame, which gets filled if it gets triggered. The other war more like a coloured plane which changed its colour as an indicator to be triggered. The colour concept was also revised again. We used turquoise and orange tones, which gave the world an elegant look. On the five buttons, we applied the colours in different variations to discuss which one works best and

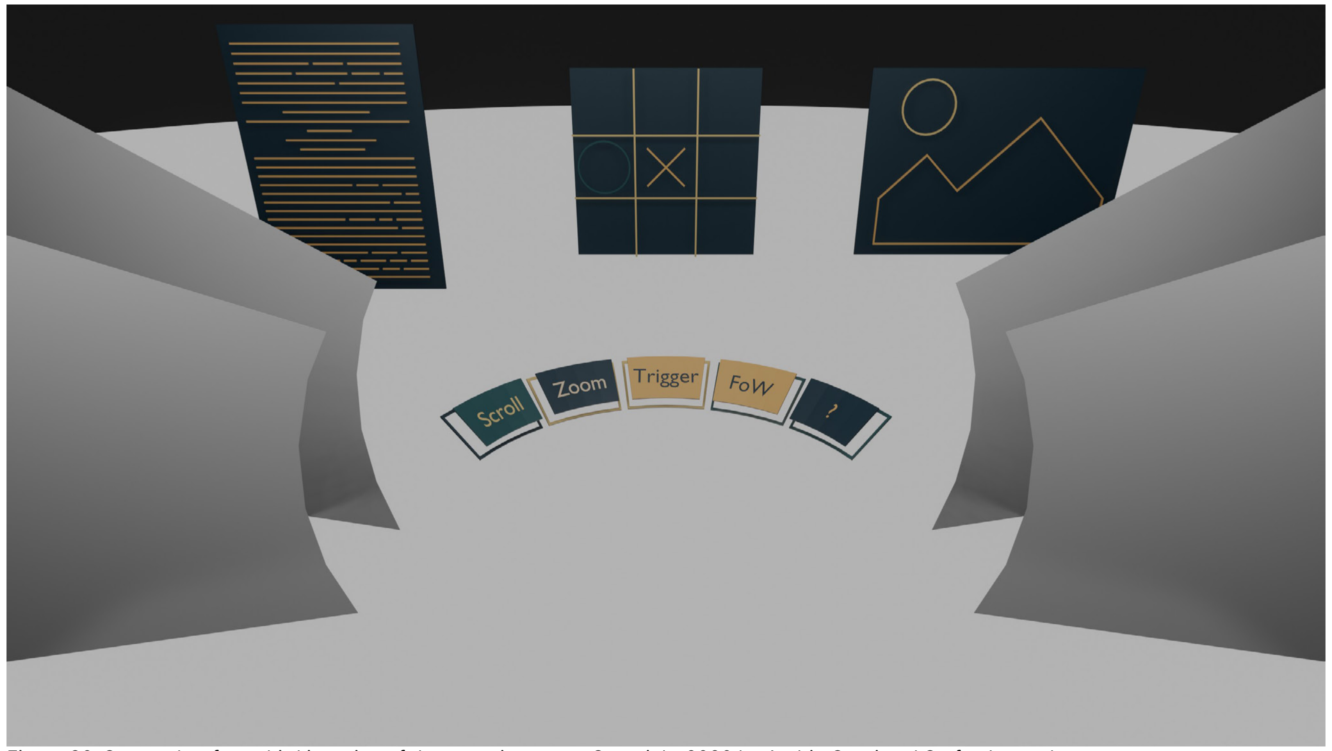


Figure 30. Screenshot from third iteration of demo environment. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Final Demonstration Environment

The difference between the third and the final version of the demo is mainly in the evolution of the colour concept and buttons. Through our investigations and test, the buttons developed into activation areas that can be triggered by placing the hand into the area. All these buttons/activation areas were very much focused on function and simplicity, and the handling was rather classical. Since we are questioning the current ways of interactions with VR with the project we wanted for our demo also buttons that were an evolution into this new direction, for this reason, we discarded the previous versions. We created a UI element that fits better into our demo and gesture control setting. In a mood board, we collected pictures of different UI and button designs that were inspiring us. One of the most important criteria was to make the components stand out from the dark background. The mood board contained a spectrum that ranged from futuristic Hollywood UIs to typical smartphone UIs. Inspired by the more futuristic examples, we were creating our button designs. For this, we first had to use VFX (Visual Effects) in Unity and Blender. In several steps, we explored the limitless possibilities of VFX. In the research phase, we let ourselves be inspired by many different effects.

With particle systems and VFX-Shaders in Unity, we created prototypes of possible activation areas. In movies and video games, people often play with the intensity of VFX, but our goal is to use these effects rather pure and subtle to give the VR world a particular dynamic. After trying different variations and multiple feedback cycles, a solution slowly crystallised that was exciting but also functional. We liked the aesthetics of slightly pulsating rings very much. You could also play with the speed of the pulsation to change the dynamics. For example, when a hand approaches, they could pulsate faster. This is similar to a hover effect on websites which indicate that these components can interact.

This is still two-dimensional, and as already explained, we want to take advantage of the power of VR and design components that are as three-dimensional as possible. With another particle system, we entered more into the third dimension. These particles were mapped onto a sphere and set in motion with different turbulences. We experimented with various parameters of the particles and turbulences to find the best solution, fitting our aesthetic expectations as well as functional requirements. Examples of these particle parameters are number, size, lifetime, traces, intensity, and some more.

Another change we implemented was to replace the label of the buttons, which was not necessary. Instead, we modelled objects that illustrate precisely what it is triggered when interacting with the corresponding activation zone. Three of these objects are hands that show the gesture to

be used in the application. Since the gestures look very similar statically, we animated the hands so that they perform the individual gestures. The other two objects are a scaled-down version of the Field of View zoning and a question mark to explain the demo to the user. These objects were then placed in the centre of the pulsating rings and the associated particle systems. By that, the purpose of each activation area is evident without the need for a label.

Finally, we focused on further improving the rings. It bothered us that they only moved on two axes not exploring all the third rotation axis. Assigning a random number to the start rotation was not enough to get a satisfying result. We solved the issue by rotating the rings in two axes and by that got a constant movement in the activation zone. After a few adjustments, we were thrilled.

Furthermore, we created a plateau around the zone in which the user moves. The different colour of the plateau offers an even better feeling of the zone where he or she could interact with the environment.

As already mentioned, bright colours and glowing effects were essential for the design, also to support the dynamics of the rings. That is why we adopted the colour scheme. We chose violets and blues as the colour space because we feel that they are related to the meaning of our work and reflect what we wanted to achieve. Furthermore, these two colours suit our taste. To underline the lighting effects, we kept the surroundings dark. With different reflective materials, we set colour accents and illuminated the demo. Together with all the other components, we found the VR environment to be harmonious and with a clear focus on the gestures developed.

Development

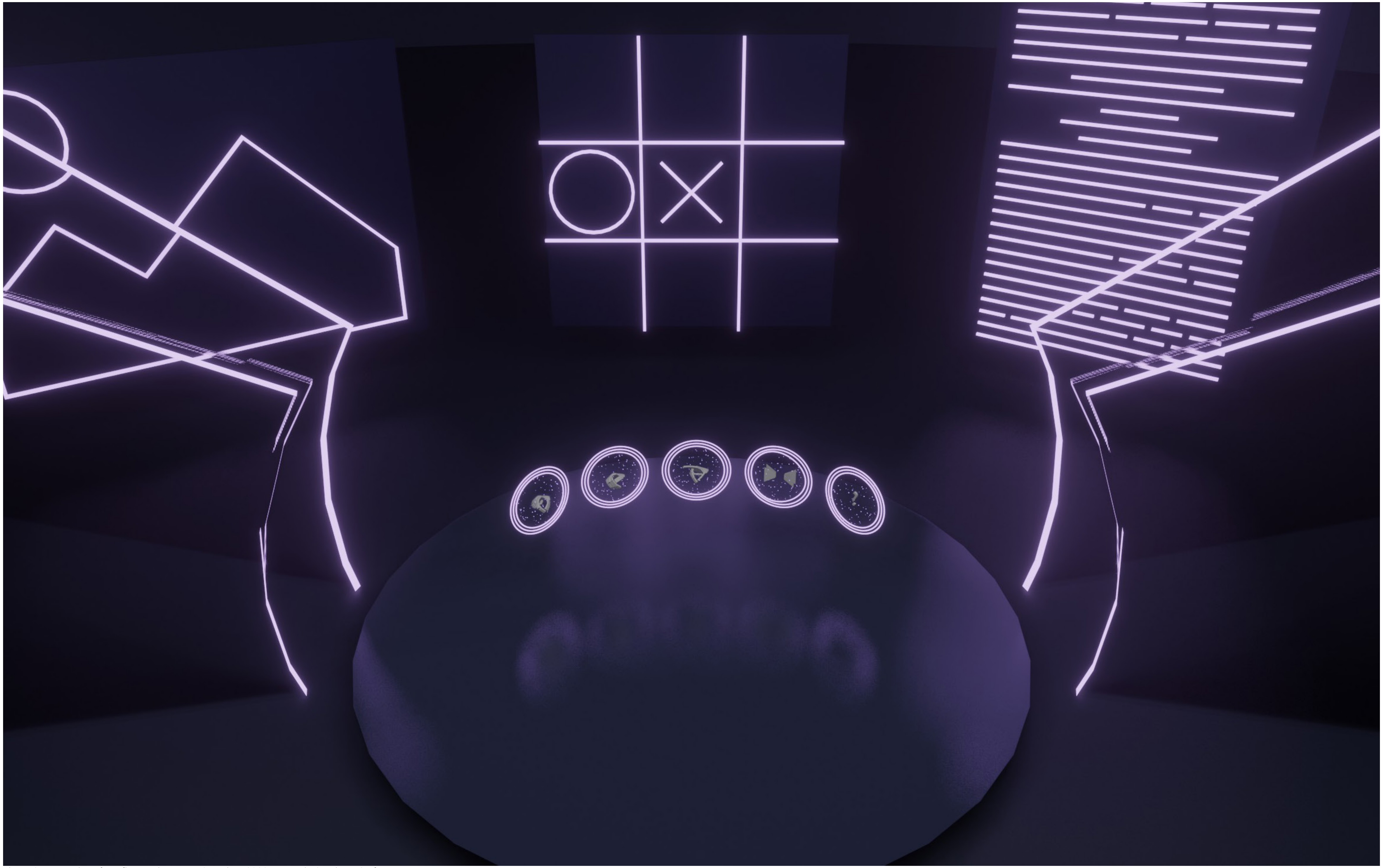


Figure 31. Rendering of the final Environment. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Development

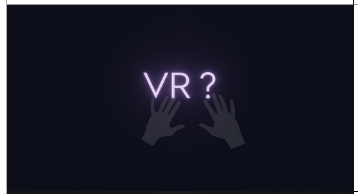
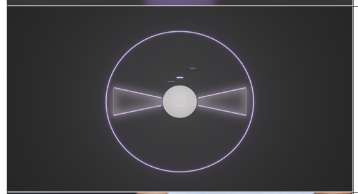
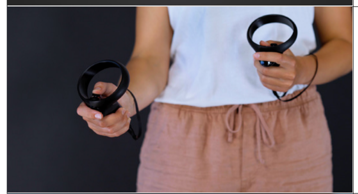
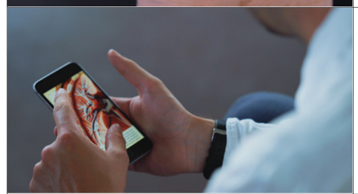
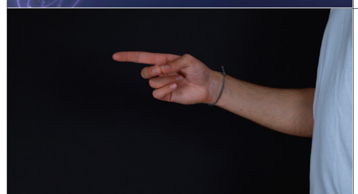
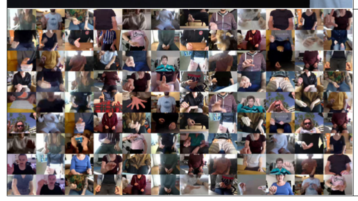
Video

In this section, we cover the creation of the video that is published on the diploma exhibition website.

Intention

Since the circumstances, given by the corona epidemic, the annual diploma exhibition can not take place at the Zurich University of the Artss, which is the reason why the video gained a lot of importance since it will be the only way of direct interaction with our audience. We live in the day and age of moving images, where people want to be teased, but also informed, through videos. But since this video will be shown on the diploma exhibition website, together with more than a hundred other projects, we needed to create a video which is short, on point and sparkles the excitement that we experienced on this journey. This is a big challenge. In our work, there was no distinct story to tell since it was guided by trial and error and multiple feedback cycles. Therefore we wanted to show the people how our process looked like and how we proceeded to come to the findings we had and the learnings we learned. It seemed apparent to us to focus on a well-written text which will be spoken as a distant voice. But to spice the off voice and to improve the viewer's understanding, we wanted to have simple images. Cleanly shot clips of our gestures and research as well as the use of screencasts from our VR Demo.

Storyboard

Sketch	Description	Voice-Over Text
	glowings words on a black back-ground appearing time-matched to the voice	What does it take to design an intuitive gesture control system for VR?
	Blender rendering of the readability of text.	Considering the readability of text,
	Blender Rendering of Zones in which content belongs to.	zones in which content belongs to
	Closeup studio shots of hands using a controller.	and the usage of controllers.
	Closeup shots of differrent people using smartphones, tablets and trackpads.	We started analizing interactions with smartphones, tablets and computers. Therefore, we observed a lot of people and kept record of our own usage.
	Frame divided in 3 sections, one to represent the interaction of scrolling, one to represent zooming and one to represent triggering.	Broken down, we basically just scroll, zoom and trigger actions.
	VR screencast of our gesture demo	We built a VR Environment to demonstrate the 3 gestures we have evaluated.
	clean studio shots of the 3 gesture, each gesture with a different back-ground, shot either witrh different paper backgrounds or with a green-screen	The gesture to scroll, the one to zoom and the one to trigger actions.
	transition from 1 user to a total of multiple users on the screen, ending with as much people on screen as possible	Well we didnt come up with the gestures all by ourselfs. We got help from a lot of you!

	glowing words on a black back-ground appearing accordingly to the voice	Now its your turn again. Test our VR Demo by yourself.

Figure 33. - 42. Stills from Project Movie. Copyright 2020 by Andrin Gorgi and Stefan Lustenberger.

Realisation

The initial plan was to use the studio at the Toni Areal to shoot the images we had in our storyboard. But the circumstances allowed us to just visit the studio for one day, which was not enough to cover all the footage. Also, our hand models were not available on that particular day. Setting up the studio and filming with multiple models without a rush would also not have been possible. Therefore we were forced to find another solution. Fortunately, we could set up a studio at home, using a black paper background, lights from the Zurich University of the Arts LEIHS, a camera and a tripod. We managed to set up a professional-grade studio at home within days. There we could take enough time to film the shots exactly how we wanted to.

We filmed with a Sony A7Riii in 4K with S-Log2. For the shots where the focus lies on the models' hands, we used a Sony G-Master 85mm lens with an f-stop of 1.4, for the shots where the model's full bodies are shown we used a Sigma 24mm Art Lens with an f-stop of 1.4.

The Screencasts used in our video were recorded directly in Unity, using Unity's own recording function. When the user in VR performed the gestures, Unity recorded the game view, which is significantly better than the built-in recording function in The Oculus headset which is lagging and recording with only 720 pixels (a square image).

For the part where we show the importance of the readability of text as well as the zones in which content belongs to, we created an environment in Blender. We defined a camera drive to record and render an optimal image, looking similar to our demo environment.

The off voice is spoken by Jon T Coleman whom we found on Fiverr. We provided him with the text he had to speak together with the rough cut of our video. In the rough cut, we had a distant voice spoken by ourselves to show Mr Coleman the style of intonation we wanted to have. The first off voice he delivered was not satisfying, and we went through another iteration in order to reach a result that made us excited.

Figure 32. Studio Set-Up. . Copyright 2020 by Marcial Koch

Conclusion

Contribution

The scientific contribution which we have made with our VR Gesturing project covers different topics. In a long and detailed research phase, we collected several related studies. We based many of our decisions on these studies and leveraged their solutions whenever it was supporting us in reaching our own goals. It was, however, essential for us to question and test these theories and results ourselves to understand them and to adapt them whenever needed. In VR Gesturing we could confirm, improve or disprove the solutions of different problems.

Furthermore, due to the COVID-19 situation, we have developed new forms of data and information retrieval. Because of Corona, it was no longer possible to meet other people and to obtain information in traditional surveys and interviews. It was also no longer possible to conduct user tests in person. The latter, in particular, kept us busy, as user testing was an essential part of our planned process. We had to find a new method to carry out video-based user tests. We investigated video chats in which the screen is recorded, and where we could instruct our users in detail. However, this turned out to be very time-consuming and not very flexible. We wanted to take advantage of the fact that practically every computer has a connection to the internet and a webcam. With these basic requirements in mind, we designed and developed a website that autonomously provides the user with the necessary information and sends us the individual test results. This new method is very flexible in its adaptation. Also, for the user, there are multiple advantages, such as flexible access to the test and 24/7 availability. We see great potential in this website-based approach, especially for long-term studies since, after the release of the website, no further work is required apart from adaptation and evaluation. Further studies can be rolled out with participants distributed all over the world. This platform is a significant achievement and could support many other projects to gather user feedback simply and efficiently.

In our opinion, the most significant and most important contribution is the development of our three gestures system for an essential control of virtual reality systems. Research has also been done in this area, but always without a concrete implementation. This is precisely the difference to VR Gesturing. We did not only want to develop three good, intuitive and suitable gestures but also to integrate them into an application so that other people can also experience this kind of navigation. We see our project as basic research that can be extended by others towards a gesture-based VR future.

The fact that Facebook had implemented more gestures in their latest update for the Oculus Quest about two weeks before the end of our work showed us that this topic has an essential place in the VR community and that we are currently at the starting point of something very exciting to come.

Reflection

Project

We started with the idea to design a user interface for virtual reality which should be tailored to make work steps of a typical office job more productive. After several interviews with VR experts, we developed the idea into developing a control system that was not designed for gaming. We, therefore, investigated controllers as well as gestures. Finally, we decided to evaluate a gesture control system for VR. This way shows how versatile such a project can be and that it is not necessarily wrong to question and develop your project, even if the original idea does not match the final result.

A highlight of our project was the research on gestures. We were surprised at how much influence today's devices have. In our user test of the tasks, scroll, zoom and triggers, recent results were influenced by the use of touchscreens and trackpads. More exciting was the user test on unconventional tasks like drag-n-drop, copy-paste and writing without a keyboard. We were impressed by the creative ways in which the test subjects performed these tasks.

We see VR Gesturing as primary research. With our three gestures, we offer a foundation which should be further developed or built upon. With our gesture trio, the most common actions can be controlled and are sufficient to surf, watch videos and read, just for the things that an average consumer must be able to do.

As soon as more specific work needs to be done, our gestures reach their limits. It is like in real life or work, there are products that can do everything in their field, but they cannot do everything equally well. In order to achieve excellent results, no matter in which area, we rely on products that can only do one thing, but are the best at doing it.

We are convinced that for various other tasks in VR, gestures or controllers are needed which are specialised in these tasks. We hope that VR Gesturing will be used to develop further individual gestures, controllers and other steering elements.

Looking back on our project, we are delighted with what we have achieved. Of course, there are still elements here and there which could be implemented differently or better. We are still proud of our gestures, the VR application, our user-testing website and the video.

Process

The process we went through during this time was not the same as we knew it from previous projects. We imagined evaluating the ideal gesture for VR with many different user tests. In a further step, these gestures should be recognised by a camera with the help of an algorithm.

Conclusion and Reflection

In the best case, it is a Leap Motion or even from an HMD itself. When looking back, the procedure was the same, but from around mid-March, we had to resort to other methods. Considering that this was the most significant and most extensive work we both have done so far, we are satisfied with the process we were able to develop ourselves. We would have liked to do more small experiments and user tests during the research and implementation phase to get quick inputs and feedback. This was on the one hand not possible due to time constraints, and on the other hand, the focus of our thesis changed multiple times during the research phase. When the coronavirus via Italy came to Switzerland, and we were all forced to move to a home office, the complete process changed immediately, of course. It was no longer possible to do quick and easy user tests or get feedback from outsiders or fellow students.

As mentioned, several times in other chapters above, we found a way to develop a user test that delivered adequate results without ever meeting a participant in person. The process we went through, and the decisions that had to be made were essential for the further development of VR Gesturing.

Lessons Learned

In these approximately four months we have gained much new knowledge. We have intensively deepened our knowledge and got to know various technologies. We probably spent most of the time in Blender and Unity. These two programs harmonise very well with each other, especially when it comes to mesh objects. These can be exported and re-imported with a few clicks. But we had a lot of trouble bringing more complex materials from Blender into Unity. We have tried several methods with some success. The lesson we learn is that Blender is perfect for creating and adjusting 3D content for Unity. If the Blender file is saved directly into the assets folder of the Unity project, it is even possible to make live changes that Unity will take over directly. Also, when working with UV maps, materials can be imported into Unity without any problems. Otherwise, we have made the experience that it is not easy to reconstruct materials created in Blender in Unity. We believe that it takes much experience to understand materials and shaders in Unity and to be able to implement your ideas correctly.

Another site we had to struggle with was the Oculus integration for Unity. For some unknown reason, this integration made Unity extremely unstable. In a project that ran without problems, the import could cause Unity to crash always. To avoid this risk, we use the Oculus integration only at the last possible moment.

It is also not easy to work with technologies that are still in the beta phase. In our case, this concerns the hand tracking and link function of the Oculus Quest. We were often confronted with situations that we could not explain to ourselves. For example, it was possible that the tracking was suddenly not working correctly, and therefore the screen in the glasses turned black every two to three seconds, like a sleep mode. Only by pressing the Oculus button on the controller, the system could be woken up again. In another case, the hands were shifted in a specific direction, which felt very strange as a user.
The link function of the Quest is there to connect the glasses directly to the computer via a USB-C cable so that the created Unity projects can be streamed to the HMD. This feature is fantastic for creating VR applications for the Oculus Quest when it works properly. However, from time to time, some errors terminated the link. Here a restart of this function was enough to continue working, which is annoying when this happens every half hour. The second error froze the whole glasses and could only be fixed by restarting the Quest. Fortunately, this does not happen very often, but it was a real hassle every time. Therefore we conclude that working with beta phase applications is exciting since it enables a new state of the art features, but with the drawback of instabilities and little support.

The cooperation with Sensoryx was very exciting and provided us with a view on the state of the art device developmentOur interactions with their experts were very supportive and educational. We tested their device several times using our computer and got a lot of support from their side. We, however, realised in the research phase of our project that the foundations for our initial goal of a VR environment for the workplace using gestures were not there Therefore we refocused our efforts to the development of a gesture control system, which was designed without the glove since the integrated hand tracing of the Oculus Quest provided a platform for speedy progress. In any case, the result we created are still highly relevant for Sensoryx, and we are looking forward to discussing our project with them in detail and to get their feedback. Looking at our project, we also saw that hand-tracking has an intrinsic limitation due to the lack of feedback. Therefore, we are excited to see that companies like Sensoryx are moving in this space, which will allow for even more possible controlling schemes by leveraging haptic feedback in the future

The cooperation between the two of us was initially more difficult than expected. Since we both like to work in groups and get along very well with each other, this was a surprise. The problem was on one hand that we had different ideas about how to work, and on the other hand, we both have different working styles. For these reasons, we decided to fill out a team canvas and to speak clearly about our expectations regarding the bachelor thesis and teamwork. This sounds like a small crisis, but it was not. There were only a few points of disagreement on both sides. For the rest of the cooperation, we agreed to respect these and to take these points into account. From that point on, we harmonised. At least once a week, we discussed what had to be done and divided the tasks. So everybody knew what tasks were necessary and could plan their work individually. With this method, both of us could work in our own style. That meant that both of us selected their own workplace and time schedule such that we were able to complete our tasks in the best and most efficient way. For example, Stefan preferred to do research, read and write at home because he was less distracted and could concentrate better. Andrin, on the other hand, much appreciated the variety in the atelier, although he had no problems concentrating.

When the Toni Areal and therefore the ZHdK were closed due to the Corona measures of the Federal Council, the situation changed again. From one day to the next, the school was closed, and we were advised to take home everything we needed for our work. For us, this meant that we had to take home all the hardware we had borrowed. With full backpacks and bags, we went home on Friday, 13 March 2020. It felt like this was the last day of our studies.

Considering that we were forced into home office, we were relatively lucky, because we were not dependent on the workshops or other infrastructure of the ZHdK. The most significant limitation for us was that we were not allowed to meet any more. At home, we created a permanent workplace and installed the necessary hardware and software. Instead of meeting each other, we skyped or zoomed every day. This way, we stayed in contact and could discuss all our further work steps.

Furthermore, it was always a nice change to talk to someone else who was not locked up in the same household. Our meetings often lasted the whole day, although it was often just quiet because both of us were working. You could only hear the click of the mouse or the typing on the keyboard. In between, we asked for each other's opinions and showed each other our progress via screen sharing. This was also very helpful in the mentoring sessions and in solving problems, as everyone involved could see directly what was happening.

The loosening up of the restrictions that were introduced towards the end of our work suited us very well, as we were busy with the final video at that time. So we could meet again legally, which made the recording of the individual scenes much more comfortably and more efficient. One of us could stand, and the other could adjust the camera settings as well as the lights.

Looking back, we think that it was the right decision to work together. We complemented each other very well and had a positive influence. We were able to benefit from each other's strengths and solved various problems together. We were also very thankful that we worked in a group because this way we could motivate each other in hard times, especially helpful during the lockdown. Despite minor initial difficulties, we are all in all proud of our fruitful cooperation.

Future Steps

What could be done in the future? What could be improved and how? Where could we do more research? Which other angles does the project offer?

As a further step, we would certainly work on the functionality of the VR demo. Since at the current time, not all gestures are recognised, not all of them can be tested.
As we would like to extend our gesture system, it is evident that the development of further gestures is the logical continuation for this project. We have already started with our user test for drag-n-drop, copy-paste and gesture-based text input, and of course, many other tasks could be translated into the gesture language. We are very curious about what will happen in the future in this area.

Further, we see significant potential with our new way of gathering user feedbacks though the website. This is something that could benefit many projects at ZHdK and others.

Further, it would be interesting to discuss our results with Sensoryx, to see how they could be adapted to fit their vision of haptic feedback enabling glove.

We both are still big fans of the idea of using a ring as a controller. For this reason, we can imagine doing further research in this direction because a subtle hardware controller would take gesture control to a higher level. It would make it possible to give haptic feedback, which in turn would contribute a lot to the experience.

We are both convinced that our gesture system is just the beginning and that gestures will gain in the coming years much more importance in VR and AR.

Web Presence

See our video and download our alpha-version on vr-gesturing.ch

Bibliography

References

Books and Academic Papers

- Dorau, R. (2011). Emotionales Interaktionsdesign Gesten und Mimik interaktiver Systeme. Berlin: Springer.
- Gardner, B. T., & Gardner, R. A. (1978). Comparative psychology and language acquisition. Annals of the New York Academy of Science, 309, 37-76.
- Maria Eugenia Cabrera, Juan Manuel Bogado, Leonardo Fermin, Raul Acuña, Dimitar Ralev. Glove-Based Gesture Recognition System. Mechatronics Research Group, Simon Bolivar University.
- Rafiqul Zaman Khan and Noor Adnan Ibraheem. (2012). Hand Gesture Recognition: A Literature Review. Department of Computer Science, A.M.U. Aligarh, India
- Stefan Liszio, Maic Masuch (July 2016). Gesture-based Virtual Reality Interaction Design Entwicklung und empirische Validierung handgesten basierter Interaktionskonzepte in VR Applikationen. Universität Duisburg-Essen Fakultät für Ingenieurwissenschaften Abteilung für Informatik und Angewandte Kognitionswissenschaft.
- Mike Alger (September 2015). Visual Design Methods for Virtual Reality. Master of Arts (MA) Moving Images, Ravensbourne University London
- Katarina Kacjan Žgajnar, Kaja Govekar, Martina Oder (2016), Ergonomic burdens and harms at the workplace of accountants, Inštitut za sanitarno inženirstvo
- David M. Hoffman; Ahna R. Girshick; Kurt Akeley; Martin S. Banks. (2008, März 1). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue, Vision Science Program, University of California, Berkeley, CA, USA

Web Articles

- Ubiquity: The 7% rule. (2011, October). Retrieved February 17, 2020, from https://ubiquity.acm.org/article.cfm?id=2043156
- UltraLeap: What is Haptic Feedback?. Retrieved February 21, 2020, from https://www.ultraleap.com/company/news/blog/what-is-haptic-feedback/
- GIGA: Haptisches Feedback - Was ist das überhaupt? . Retrieved February 21, 2020, from https://www.giga.de/apps/android/specials/haptisches-feedback-was-ist-das-ueberhaupt/
- Medium: CryEngine vs Unreal vs Unity: Select the Best Game Engine. Retrieved March 21, 2020, from https://medium.com/@thinkwik/cryengine-vs-unreal-vs-unity-select-the-best-game-engine-eaca64c60e3e
- How Does the Leap Motion Controller Work? (Aug 27, 2015) Retrieved April 4th, 2020, from https://medium.com/@LeapMotion/how-does-the-leap-motion-controller-work-9503124bfa04
- Wired: Why 'Gorilla Arm Syndrome' Rules Out Multitouch Notebook Displays (October 21, 2010) Retrieved April 20th from https://www.wired.com/2010/10/gorilla-arm-multitouch/

Web Sites

- Microsoft. (2020, March 7). Microsoft/MixedReality-Toolkit-Unity. Retrieved 10 March 2020, from https://github.com/microsoft/MixedRealityToolkit-Unity
- What is the mixed reality toolkit? . Retrieved March 10, 2020, from https://github.com/microsoft/MixedReality-Toolkit-Unity
- HoloLens 2 pricing and options. Retrieved March 22, 2020, from https://www.microsoft.com/en-us/hololens/buy
- Wikipedia, Google Chrome Version History. Retrieved April 22, 2020, from https://en.wikipedia.org/wiki/Google_Chrome_version_history
- Github, collab-project/videojs-record. Retrieved April 2nd, 2020, from (https://github.com/collab-project/videojs-record
- Wikipedia, Oculus Quest. Retrieved Mai 20th, 2020, from https://en.wikipedia.org/wiki/Oculus_Quest
- Wikipedia, Xerox-Alto, Retrieved Mai 20th, 2020, from https://de.wikipedia.org/wiki/Xerox_Alto
- Mike Alger, Mike Alger, Retrieved March 12th, from https://mikealger.com/
- Render Baking. Retrieved April 24th, 2020, from http://builder.openhmd.net/blender-hmd-viewport-temp/render/blender_render/bake.html
- Quaternion. Retrieved April 24th, 2020, from https://mathworld.wolfram.com/Quaternion.html
- Technopedia: Gorilla Arm. Retrieved April 20th, 2020, from https://www.techopedia.com/definition/31480/Gorilla-arm

Videos

- [Ravensbourne University London]. (2016, April 18). Mike Alger PG student; VR Interface Pre-Visualisation Methods [Video File]. Retrieved from https://www.youtube.com/watch?v=iR4iRyLoJlg
- [Made by Google]. (2019, October 15). Made by Google '19 [Video File]. Retrieved from https://www.youtube.com/watch?v=XKmsYB54zBk
- [BMW USA]. (2015, October 22). Gesture Controls │ BMW Genius How-To [Video File]. Retrieved from https://www.youtube.com/watch?v=wqvAPskg
- [TED]. (2015, April, 22). Chris Milk: How virtual reality can create the ultimate empathy machine [Video File]. Retrieved from https://www.youtube.com/watch?v=iX-Hil1TPxvA

List of Figures

vr-gesturing.ch