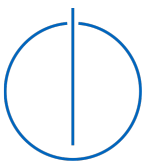
DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Informatics: Games Engineering

Smartphone-Assisted Virtual Reality Using Ubi-Interact

Michael Lohr



DEPARTMENT OF INFORMATICS

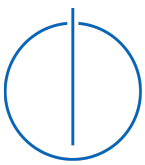
TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Informatics: Games Engineering

Smartphone-Assisted Virtual Reality Using Ubi-Interact

Smartphone-gestützte Virtuelle Realität mit Ubi-Interact

Author:	Michael Lohr
Supervisor:	Prof. Gudrun Johanna Klinker, Ph.D.
Advisor:	Sandro Weber, M.Sc.
Submission Date:	October 15, 2019



I confirm that this bachelor's thesis is my own work and I have documented all sources and material used.

Munich, October 15, 2019 Michael Lohr

Abstract

Virtual Reality is an emerging medium which enables immersive presence and interactivity in a three-dimensional space. Standard input devices like a mouse or keyboard are made for two-dimensional environments. Thus, new interaction devices have to be invented. This is not an easy task since the view is obscured by the headset, so all devices need a virtual representation. Most consumer-ready Virtual Reality peripherals, therefore, require an expensive tracking system.

Smartphones have a variety of different sensors built-in and also can run custom software. This makes them cheap general-purpose devices. Thanks to the orientational sensors and wireless capabilities, the phone can be represented in a virtual environment. Thus it is possible to use them as interaction devices for Virtual Reality.

To prove that the smartphone can be used as an input device for Virtual Reality, three interaction examples are presented. An orientation-based device, a pointing device and a virtual keyboard were implemented and evaluated in a SUS usability study. The UBI-Interact networking solution is used to make the system reusable and abstracted from device-specific environments.

Contents

Abbreviations

API Application Programming Interface

DOF Degrees Of Freedom

HMD Head-Mounted Display

IMU Inertial Measurement Unit

JS JavaScript

PC Personal Computer

Protobuf Google Protocol Buffers

SUS System Usability Scale

UBII UBI-Interact

UI User Interface

UID Unique Identifier

VE Virtual Environment

VR Virtual Reality

WLAN Wireless Local Area Network

2D Two-dimensional

3D Three-dimensional

1. Introduction

1.1. Motivation

Virtual Reality (VR) is an emerging technology, which provides new ways to present and interact with digital information. Best practices are not yet defined, which leaves much room for new methods and research. Viewing three-dimensional (3D) Virtual Environments (VEs) using a VR headset (or Head-Mounted Display (HMD)) is a great experience. But VR shines when interactivity comes into play. Since consumer HMDs are now available, the development of consumer tracked hand controllers (also known as VR/3D/hand/motion controllers) has begun. By mapping the "interaction devices" of the human body to a VE, the user is given a natural way of controlling and interacting with the virtual world. While common consumer VR headsets are rather similar to each other, motion controllers differ a lot from each other, as seen in Figure 1.1.1.

Figure 1.1.1.: A collection of different VR controllers. From left to right, top to bottom: HTC VIVE Controllers, Valve Index Controllers ("Knuckles"), VIVE Tracker, Oculus Touch Controllers, Samsung Odyssey Controllers.
Source: Lawrence Yang, www.steamcommunity.com/games/.../announcements/...

As of now, most controllers try to map the movement of our real fingers to the virtual world. The Leap Motion[1] sensor uses multiple infrared cameras to track hand poses, which is only possible in front of the sensor. Newer generations of VR controllers do this, too. The Oculus Touch[2] controllers track the distance of the fingers from the controller and the Valve Index[3] controllers even have pressure sensors built-in. However, for many interactions, hand inspired controllers are not ideal. This applies especially to productive VR applications, which require interactions like inputting text for labeling. Since conventional VR controllers are designed for some specific tasks (like hand movements), they are tedious to use for other tasks. They also require complex tracking systems.

The Google Cardboard[4] uses a smartphone as a display and for tracking. This makes it an inexpensive VR headset, but also demonstrates the versatility of a smartphone. Most people already have a smartphone, since they are general-purpose devices and not very expensive anymore. Thanks to Wireless Local Area Network (WLAN) and Bluetooth[5] it is easy to connect the smartphone to other devices. They have input devices like buttons, a touch screen as well as sensors and systems like a Inertial Measurement Unit (IMU)[6]. Also, output devices like vibration motors and speakers are present. This makes them similar to VR controllers.

One significant difference between smartphones and common VR controllers is that smartphones do not have positional tracking built-in. The position can be estimated using an IMU, but since the error accumulates over time, other tracking methods have to be used to correct the drift [Zha+15]. Sophisticated approaches using the phone camera are computationally expensive and still not as good as complete tracking systems. Besides the missing positional tracking, the advantages lead to the assumption that the smartphone can be used as an alternative VR controller.

[1] The Leap Motion controller is a hand tracking device, which is often used to display a hand avatar. Website: www.leapmotion.com

[2] The Oculus Touch controllers are hand tracking devices included with the Oculus Rift HMD. Website: www.oculus.com/rift

[3] The Valve Index is a HMD which includes its own set of controllers, called "Knuckles". Website: store.steampowered.com/valveindex

[4] The Google Cardboard is a HMD made out of cardboard, which uses a smartphone as a display and for tracking. Website: vr.google.com/cardboard

[5] Bluetooth is a wireless standard for exchanging data over short ranges between mobile devices.

[6] An IMU is an electronic component which is part of most smartphones and allows to measure a specific force, angular rate, and magnetic field.

1.2. Problem Statement

This thesis aims to explore the possibilities of using the smartphone as an interaction device in VR experiences. Different methods of implementing the phone as an alternative input device for typical VR interactions are going to be presented and evaluated. They are going to be implemented in experiments using the UBI-Interact (UBII) architecture, to create an abstracted and reusable system. The goal of those experiments is not to create a better system, but rather show that the smartphone can be used as well as common VR controllers for certain interactions. Additional tasks, to complete in the experiments, are going to be implemented to evaluate the presented methods in a usability study.

1.3. Outline

In the next chapter, previous work on this topic is presented. Before discussing the implementation of the experiments in the following chapter, UBII and the technology used to make the experiments possible, is introduced. The user study and its approach are presented afterward. After discussing the results of the study, the conclusion is drawn.

2. Related Work

2.1. Deller et al.

Deller and Ebert propose a modular framework to enable interactions between smartphones and large screen applications. They use a typical client-server architecture with an XML[2]-based protocol. One difference to the system presented in this thesis in terms of networking is that they differentiate between application clients (the large screen) and interaction clients (the smartphones). They provide different modules for the client app. Some modules offer similar functionality like the ones implemented in the experiments of this thesis: The text module enables the user to enter a text; The accelerometer/magnetometer module sends IMU data like acceleration and magnetic field data in the background to the server. They also implemented their framework into an application where users can navigate a map and toggle display settings [DE11].

2.2. Benzina et al.

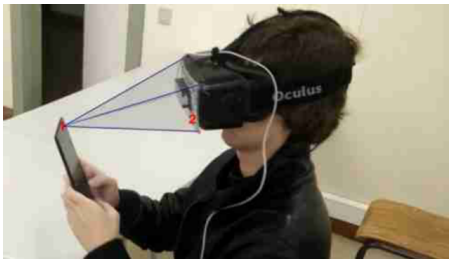
A similar problem is solved by Benzina, Toennis, Klinker, and Ashry. They introduce a system for flying through VEs in VR. A VR headset is used, which means the phone display cannot be used to display information because the sight is occluded by the HMD. Different methods of controlling the flight movement are presented. They came to the conclusion that the most accurate method for controlling the flight uses an approach, where an airplane metaphor (four Degrees Of Freedom (DOF)) is simulated [Ben+11].

2.3. Dias et al.

Dias, Afonso, Eliseu, and Santos propose a solution, where the smartphone has a visual representation in VR. The visual representation also shows information and User

[2]XML is a standardized data exchange format, that uses human-readable text.

Interface (UI) on its virtual screen. The camera in the smartphone tracks a marker on the HMD to track itself. Because the user interacts with the UI using the touch screen of the phone as he would in real life, the fingers have to be tracked. Otherwise, the user would not know where his fingers hit the touch screen, because the sight is occluded by the HMD. To solves this, they attached a Leap Motion sensor to the HMD, which displays a hand avatar [Dia+18]. The setup is shown in Figure 2.3.1.



(a) 1) The front camera of the smartphone tracks the 2) marker on the HMD.

(b) The virtual smartphone representation in the VE while interacting with the UI.

Figure 2.3.1.: The tracking system by Dias, Afonso, Eliseu, and Santos [Dia+18, pp. 4, 5].

2.4. Steed at al.

The approach by Steed and Julier also uses a smartphone and a VR headset. They implemented a visual representation of the phone. But since they do not have positional tracking for the smartphone, the position is fixed relative to the position of the HMD. There are two different possible positions, one in front of the HMD and the other one in front of the stomach. The position is switched if a hand raise gesture with the phone in the hand is detected. Gestures and orientation of the smartphone are detected using the data of the IMU. On the virtual phone screen, a virtual UI is displayed as seen in Figure 2.4.1b. This UI has control elements like buttons, which amongst others can be used to toggle a selection mode. In the selection mode, the phone casts a ray out of the top (similar to a laser pointer) as seen in Figure 2.4.1a. The ray direction can be changed by rotating the smartphone. As soon as a UI button is pressed, the objects intersecting with the ray are selected [SJ13].

(a) The virtual device in selection mode.

(b) The virtual UI and the cursor.

Figure 2.4.1.: The virtual smartphone representation by Steed and Julier [SJ13, p. 43].

3. Implementation

3.1. Ubi-Interact

UBII[1] is a framework for distributed applications, which enables to connect all kinds of different devices together. A centralized server is used to manage the system in a local network. The abstraction into devices, topics, and interactions allows decoupling the implementation of software from device-specific environments.

3.1.1. Architecture

The main components of the UBII framework are:

UBII Clients describe a basic network participant. For every UBII client registered on the server also exists one network socket address. Clients are an abstraction of a physical network device. They are defined by a Unique Identifier (UID).

UBII Devices can be registered by clients. A UBII-device groups different input and output UBII devices together. It is defined by a UID and a list of UBII components.

UBII Components contain the UBII topic name, UBII message formats for input/output UBII devices and whether it publishes input or receives output data. A data source for such an input UBII device, could be any sensor, for example, a button or a camera. Data output examples for input UBII devices are lamps and displays.

UBII Message Formats define the format of data published to a UBII topic. Even though it is possible to implement custom ones, most common data types are available. For example, `Vector4×4` (a four by four matrix), `Vector2` (a two-dimensional vector) or `boolean` (a binary value) are built-in.

UBII Topics are data channels which are addressed by a name. UBII Clients can publish messages to UBII topics, which are registered by a UBII device. They are able to receive messages, after subscribing to a UBII topic. Such messages

[1] UBII is currently developed and maintained by Sandro Weber, who is also the advisor for this thesis.

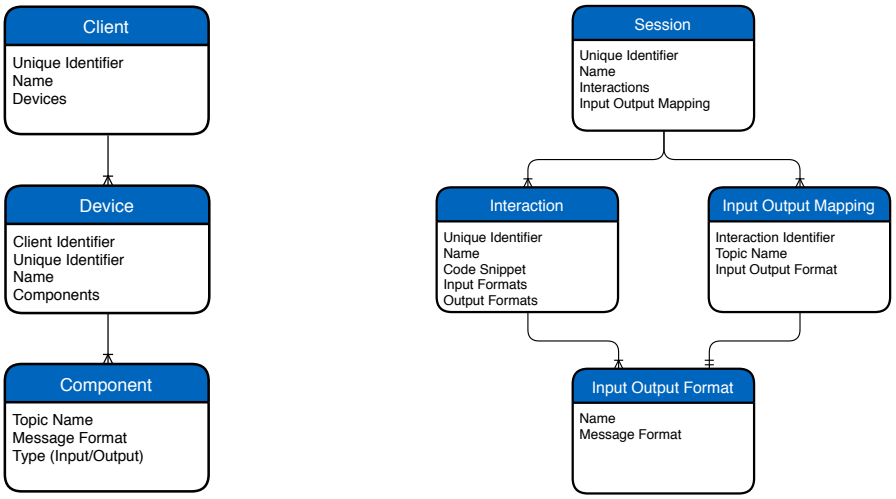
(also called "UBII topic data") are formatted as JSON[1]-string, whose structure is defined by the device.

UBII Sessions operate on the server but can be specified by the UBII client. They are defined by a UID as well as a list of interactions and **input/output mappings**. The mappings are defined by a UBII message format and UBII topic name.

UBII Interactions are reactive components. They operate on UBII topics and are defined by a source code snippet[2]. UBII Interactions are executed in a fixed interval on the UBII server. They can subscribe to UBII topics and use the received topic data as input, given an input/output mapping description. The output of the UBII interaction is published into another UBII topic. It is also possible to keep data to use in future executions (persistent state).

UBII Services are channels, used to send commands or requests to the UBII server. For example, they are used to subscribe to a UBII topic or list all available UBII topics.

The Figure 3.1.1 visualizes the relationships of the different components.



(a) The client components.

(b) The session components.

Figure 3.1.1.: Relationships of the core components in an entity relationship diagram.

[1]JSON is a standardized data exchange format, that uses human-readable text. It is often used for web-based data communication [ECM17, p. iii].

[2]Currently only JavaScript is supported as a script language.

3.1.2. Interactions

A powerful but optional core feature of UBII are interactions. As explained in the component overview (see 3.1.1), they are reactive components, which operate on UBII topics and regularly execute given code snippets (processing functions) on the UBII server. Interactions are isolated components, which just depend on topic data and nothing else. This abstraction introduces the possibility to reuse logic in other applications in a similar context. The data flow from a device to the interaction is visualized in the Figure 3.1.2.

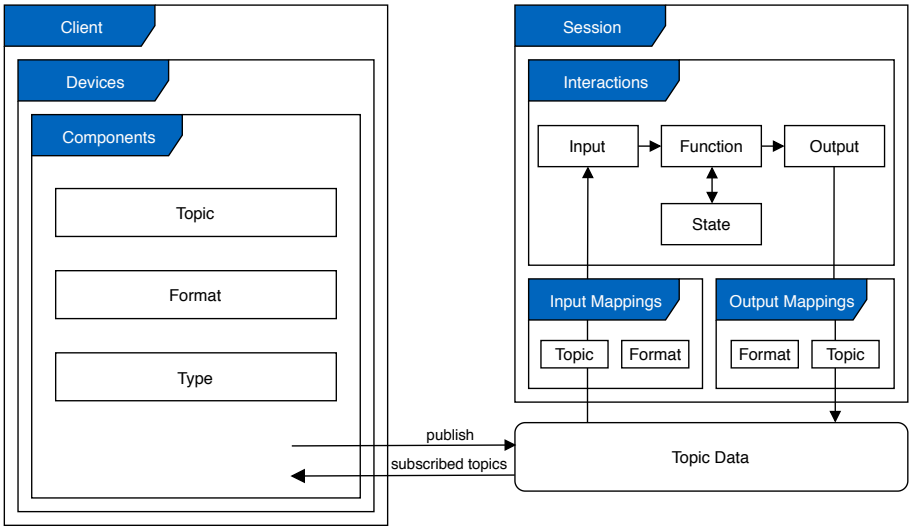


Figure 3.1.2.: UBII interaction processing overview. This graphic gives a rough overview of the dataflow when using an UBII interaction. Figure created with the help of Sandro Weber.

UBII Interactions should be designed generalized so that they are easy to reuse. They can be used to discretize data, converting data to other formats or just to outsource some logic from the application. Concrete examples include detecting button presses, transforming coordinates and evaluating data. An example implementation which detects position changes can be seen in Figure 3.1.3.

They are also useful if two UBII topics with different formats should be connected. An example of such a scenario could be an application, which consumes a rotation given in Euler angles. But some input UBII devices publish Euler angles in degrees. A UBII interaction, which takes Euler angles in degrees from one UBII topic and publishes Euler angles in radians to another one could be implemented.

```
1  // detect intentional movement by comparing the current position with a previous one
2  function (inputs, outputs, state) {
3    const threshold = 0.05;
4
5    if (state.lastPosition) {
6      const vector = {
7        x: inputs.position.x - state.lastPosition.x,
8        y: inputs.position.y - state.lastPosition.y,
9      };
10
11     const squaredDistance = Math.pow(vector.x, 2) + Math.pow(vector.y, 2);
12
13     outputs.moved = squaredDistance < threshold;
14   } else {
15     outputs.moved = true;
16   }
17
18   state.lastPosition = inputs.position;
19 }
```

Figure 3.1.3.: An example for an UBII interaction written in JavaScript. This UBII interaction calculates the squared distance of two points. One of the points is provided through the input, while the other one is stored in the state variable. To achieve this, the Euclidean vector norm of the subtraction of both vectors without the square root is calculated and compared with a threshold constant. The result is then written into the output as a boolean data type. This is used to detect intended changes of the input position.

The code snippet has to define a function, which accepts three parameters: `inputs` is a collection of values, which contains values which were published into a UBII topic. The UBII topic which was used, is defined by the input mappings of the UBII session. `outputs` is an empty collection, where values can be added. Those values are then published into a UBII topic, defined by the output mappings of the UBII session. `state` stores a persistent collection of values, which can be used in later executions of the same UBII interaction.

3.2. Technology Stack

Since most of the existing software for UBII was written in JavaScript (JS)[1] using a web-based architecture, I decided to use this approach for my application. This has the major advantage of platform independence. Most modern devices can run web-based software, which means they can also run my application. Also, the application is served by a web server, which means the user does not have to install the software onto his device.

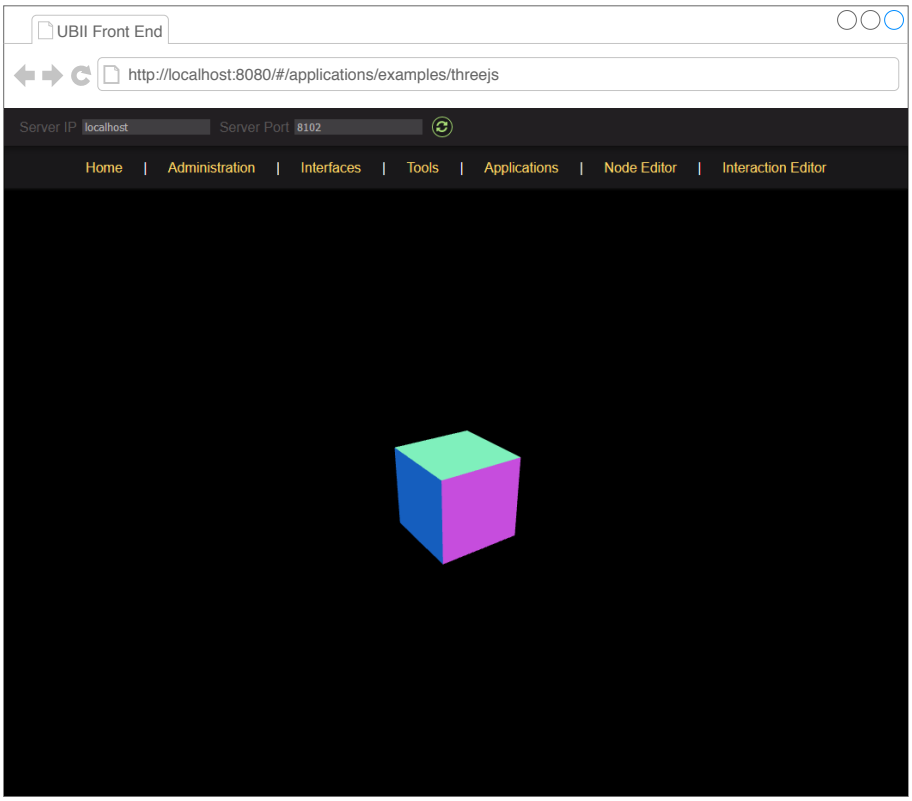


Figure 3.2.1.: A screenshot of the UBII front end rendering a 3D cube.

A web interface with some UBII content (the UBII front end), demos and debugging tools were already written[2], so I included my experiments in this application as well. The technology stack of the front end was built with the following technologies:

[1] JS is a just-in-time compiled scripting language, widely used in web technology. It is a dynamic prototype-based language, which supports object-orientated programming [ECM18, pp. 43, 47].
[2] The front end was developed by Sandro Weber, Daniel Dyrda and me.

Web APIs are Application Programming Interfaces (APIs) available in modern web browsers to provide access to functionality or data outside the browser. The WebAPI is an additional layer of abstraction of the APIs of an operating system. While this has the advantage that the API is the same on every device, this also prevents the access to the raw sensor data[1]. In this thesis, the WebVR API and the device orientation API were used. The former enables to render to external VR headsets. The latter gives access to the data of the Inertial Measurement Unit (IMU).

Vue.js is a modern open-source JavaScript web framework[2,3] [You19]. Being released in 2014 and developed by Evan You, it is a relatively young framework [Koe16, p. 17]. But it quickly gained traction and is quite popular now [Koe16, pp. 12 sq.]. Packages like Vue.js itself, Vue.js plugins and other JavaScript libraries are managed using the package manager npm[4].

Three.js is a lightweight open source library which utilizes WebGL to render 3D computer graphics[5] [Cab19]. It can be used to render scenes to the display as well as to a VR headset using WebVR. This high-level library comes with a lot of features, similar to a game engine, like scenes, effects, lights, animation, geometry and much more.

UBII Client is an JavaScript client for the UBII system. It abstracts the protocol and provides high-level functions to register devices as well as send and receive UBII topic data. The UBII system uses Google Protocol Buffers (Protobuf)[6] to serialize the data.

Figure 3.2.1 displays a test view, which uses Vue.js to manage the views and Three.js to render a cube.

[1] The specification is available on www.w3c.github.io/deviceorientation

[2] A web framework is a software framework which provides a standard way to build web applications. It comes with tools and libraries to automate and make the development of web applications easier.

[3] Vue.js: Website: www.vuejs.org; Source code: www.github.com/vuejs/vue

[4] "NPM" stands for "Node Package Manager" and is also used in the UBII server itself. Website: www.npmjs.com

[5] Three.js: Website: www.threejs.org, Source code: www.github.com/mrdoob/three.js

[6] Protobuf is a mechanism to serialize data. The data is defined in a platform-neutral language, which compiles as a library to all commonly used programming languages [Goo19b]. Website: www.developers.google.com/protocol-buffers/

3.3. Smart Device

The "Smart Device" is a part of the UBII front end. Because it is web-based, only data which is available through the **Web APIs** can be obtained. Since it was not designed for a specific use case, it is thought as a general purpose or testing device. Only touch positions, touch events, orientation, and acceleration are sent to different UBII topics using the **UBII Client**. For more specific scenarios, the smart device can not be used and a custom interface has to be implemented. For the experiments in this thesis though, the smart device client was sufficient, after implementing some improvements.

The data which is published is also displayed on the screen for debugging purposes. It is possible to set the view to full-screen mode, to prevent unintentional interactions with control elements of the web browser or the operating system. Since the reference system for the orientation is fixed to the earth [Dev19, Chapter 4.1], a calibration system was implemented. With the press of the "Calibrate" button, the device is calibrated to the new orientation.

3.3.1. Topic Data

The orientation is provided by the **Web API** through the `DeviceOrientation` event. It is defined by three euler angles named `alpha`, `beta` and `gamma`, as seen in Figure 3.3.1. While `alpha` returns values in the range $[0, 360)$, `beta` only returns the range $[-180, 180)$ and `gamma` $[-90, 90)$ [Dev19, Chapter 4.1]. This limitation entails that no full orientation tracking is possible with this event.

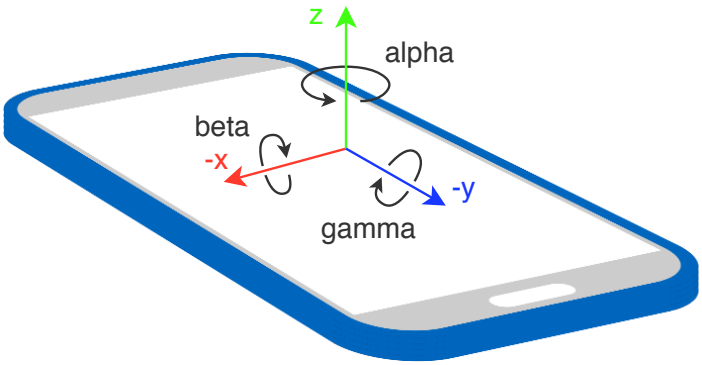


Figure 3.3.1.: The specification of the orientation values visualized. The x and y axes are inverted for the sake of clarity in this graphic.

The **Web API** also provides the `MotionEvent` which returns multiple vectors, one being

the acceleration with the gravity (`accelerationIncludingGravity`). Since the gravity vector always points down, this vector can be used as a reference vector. Together with the values from the `DeviceOrientation` event, the full orientation can be derived. The resulting orientation, then has to be smoothed, because the acceleration vector uses the raw IMU acceleration output.

The data from the `DeviceOrientation` event already provides all three euler angles and is smoothed. Implementing the same for the data from the `MotionEvent`, would be outside the scope of this thesis. Because of this consideration, the `DeviceOrientation` event data is used in the experiments.

The touch position on the display is normalized to a range from zero to one. This removes the influence of the display resolution and size. Events for start and stop touching are sent on different UBII topics. The acceleration of the smartphone is also sent to a UBII topic but is not used in the experiments of this thesis.

3.3.2. UBII Device Definition

The smart device is registered as a device in the UBII network. The definition in JS can be seen in Figure 3.3.2. The structure of a UBII device was described in Chapter 3.1.1.

```
const ubiiDevice = {
  name: 'web-interface-smart-device',
  components: [
    {
      topic: clientId + '/web-interface-smart-device/touch_position',
      messageFormat: 'ubii.dataStructure.Vector2',
      ioType: ProtobufLibrary.ubii.devices.Component.IOType.INPUT
    },
    {
      topic: clientId + '/web-interface-smart-device/orientation',
      messageFormat: 'ubii.dataStructure.Vector3',
      ioType: ProtobufLibrary.ubii.devices.Component.IOType.INPUT
    },
    {
      topic: clientId + '/web-interface-smart-device/linear_acceleration',
      messageFormat: 'ubii.dataStructure.Vector3',
      ioType: ProtobufLibrary.ubii.devices.Component.IOType.INPUT
    },
    {
      topic: clientId + '/web-interface-smart-device/touch_events',
      messageFormat: 'ubii.dataStructure.TouchEvent',
      ioType: ProtobufLibrary.ubii.devices.Component.IOType.INPUT
    }
  ]
};
```

Figure 3.3.2.: The smart device definition in JavaScript. It is defined by a name and a list of UBII components. The structure of a UBII device is further described in Chapter 3.1.1.

A UBII device and all UBII topics must be registered with an UID for each client because it should be possible to read the data from different devices. This allows for using multiple devices at the same time so that they can be differentiated in UBII interactions. If the UBII topic names would not include the `clientId`, each connected device would publish to the same UBII topic, which would make the data unusable.

The touch position is published multiple times per second, but only sends the current position of the first touch on the smartphone display. Using this data, it is not possible to detect whether the display was just touched or released. A new UBII topic using the `Boolean`-type could have been used, but the position has to be obtained from the touch position UBII topic. To remove this dependency on the other UBII topic, the new type `TouchEvent` was implemented. The Protobuf definition can be seen in Figure 3.3.3. It contains the two-dimensional position and the binary type `ButtonEventType`, which can

be reused in other events, too. `ButtonEventType` is an enumerated type which defines whether the touch interface was just touched or released.

```
1  syntax = "proto3";
2  package ubii.dataStructure;
3
4  import "proto/topicData/topicDataRecord/dataStructure/vector2.proto";
5
6  enum ButtonEventType {
7    UP = 0;
8    DOWN = 1;
9  }
10
11 message TouchEvent {
12   ButtonEventType type = 1;
13   ubii.dataStructure.Vector2 position = 2;
14 }
```

Figure 3.3.3.: The definition of the touch event, sent by the smart device client when a user touches or releases the touch screen. It is defined by a position and wether the touch pad was touched or released.

3.4. Experiments

Three main experiments were implemented to demonstrate how the smartphone can help with common interactions when using VR software. To achieve consistency amongst all experiments in terms of look and basic functionality, a parent class was implemented. The parent class implements utilities, which are required by each experiment. It also sets up a basic scene, which contains a sky, a floor, and lights. Also, the connection to the UBII server is handled.

3.4.1. Model Viewer

Virtual Reality offers a new way of experiencing 3D content. It is more convenient to view a model from different angles and gives a feel of a real presence of the object. Model viewers like Sketchfab[1] have implemented VR support a while ago [Den16]. But this experience can be enhanced with a smartphone. Models can be rotated without changing the position of the headset or using s expensive hand motion tracking system.

[1]Sketchfab is an online platform to publish and view 3D content. Website: www.sketchfab.com

Katzakis and Hori implemented this without VR. His approach uses a smartphone to rotate a model which is displayed on a conventional display. He uses a similar setup, in which the phone is wirelessly connected to a computer where the model is rendered. The orientation data comes from the magnetometer and, once calibrated to the screen position, is directly mapped to the model [KH10, p. 139]. In the evaluation of a mouse, a touch pen and the smartphone, the latter wins in terms of the time it takes to rotate the model to a certain pose [KH10, p. 140]. Since this approach turned out to be very successful, it was used in this experiment as well.

To feature how easy it is to view a more complex model using VR and the smartphone as a manipulator, a human skeleton model is used. This experiment is the only one supporting more than one smartphone client at the same time. For every client that connects, a new skeleton is created. The position is fixed and arranged around the position of the VR headset. A scene with multiple connected clients is shown in Figure 3.4.1.

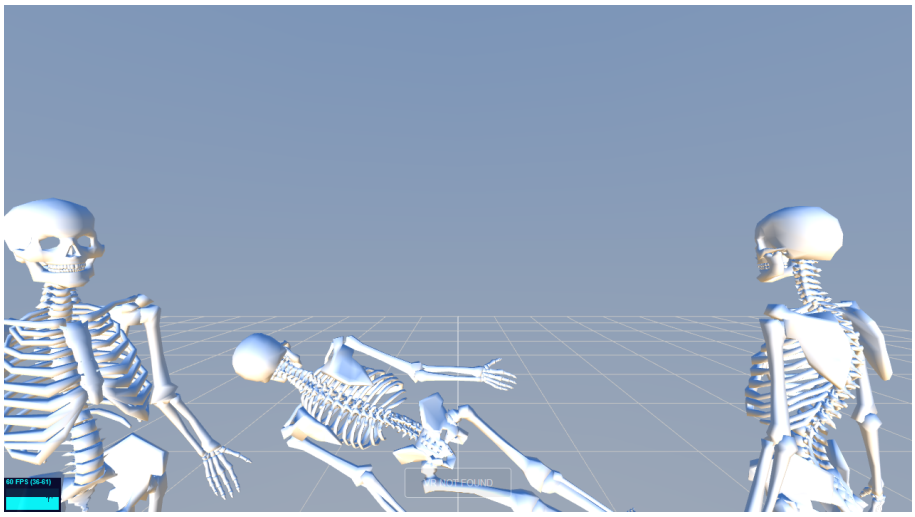


Figure 3.4.1.: A screenshot of three devices being connected and controlling the rotation of the models.

The implementation of this experiment listens for new clients. As soon as one connects, a new UBII interaction is published and the resulting UBII topic subscribed. Since the smart device (see Section 3.3) publishes the orientation data in a different format than ThreeJS needs for rendering, a reusable UBII interaction was created. The UBII interaction converts the angles from radian to degrees, changes the coordinate system and publishes them to the `[client id]/SAVRLaserPointer/orientation` topic. The code for

the UBII interaction is shown in Figure 3.4.2.

```
function (input, output, state) {
  if (!input) {
    return;
  }

  const deg2Rad = function(v) {
    return v * Math.PI / 180;
  };

  output.orientation = {
    x: deg2Rad(input.orientation.y),
    y: deg2Rad(input.orientation.x),
    z: deg2Rad(-input.orientation.z)
  };
}
```

Figure 3.4.2.: This UBII interaction is used to convert the orientation data sent by the smart device to the format ThreeJS needs for rendering. The values are converted by multiplying with an approximate of the number Π ("PI") and dividing by 180.

As in Section 3.3.1 described, the current implementation does not provide full 360-degree angles. This means that the model cannot be rotated upside down. This is very impractical for a model viewing application, but can be fixed and is not critical for a proof-of-concept.

3.4.2. Laser Pointer

Selecting elements in a virtual world is a basic interaction most VR applications use. The selection of elements in a two-dimensional (2D) environment with standard input devices like a mouse or touch screen is trivial. But the selection of elements in a 3D environment is problematic because the element might be too far away from the user or the cursor. Ray casting[1] is used to solve this problem: A ray, with the tracked device as origin, is created. Then, the element first hit by the ray is selected. Implementations without a tracked device, often use the position and orientation of the headset. The ray is fixed to the head of the user and cast along his viewing direction [Kam18, p. 23].

[1] Ray casting describes a technique to determine the objects which intersect with a ray, cast from a given point into a given direction.

This forces the user to keep the head still and look at a certain object to select it until a button is pressed or a certain time has passed.

A better solution is the use of handheld controllers, where the position of the controller is used as origin for the ray. Since the smartphone provides orientation data, it can be used for this task, too. But most handheld controllers have also positional tracking, which allows them to represent the hand of a user by displaying a virtual phone. This emulates the use of a handheld laser pointer in the real world. Since a smartphone does not have positional tracking, the origin has to be somewhere else. Again, the head could be used, but then the user would have no visual representation of the rotation of the phone. To give the user a better feel for the direction he is pointing, a visual representation is needed. The user has to see where the ray is going, even when rotating it in the opposite direction of the view direction.

Argelaguet and Andujar evaluated more than 30 different selection techniques for virtual environments, but there is no technique that uses the orientation but not the position of the pointing device [AA13, Table 1]. To work around the missing position data of the device, the ray origin is set to a fixed location relative to the users head where the phone could be in the real world. The ray origin is represented by a 3D phone model, which orientation is synchronized with the one from the last connected smart device (see Section 3.3) client, similar to the first experiment (3.4.1). To keep the virtual phone inside the view frustum of the user, it rotates relative to the user on the y-axis. A line is attached to the front of the phone (the "laser") to indicate the direction of the ray.

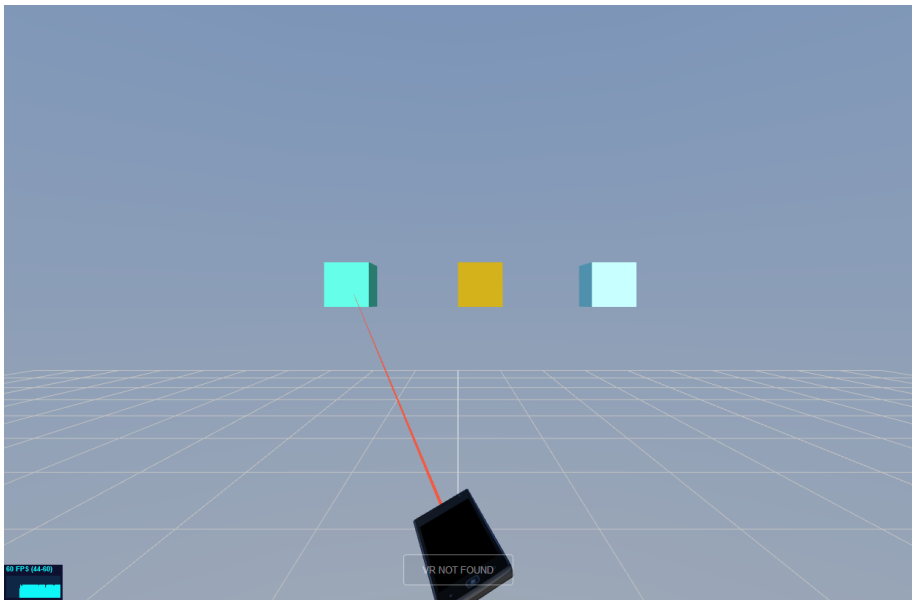


Figure 3.4.3.: A screenshot of the virtual laser pointer and selectable cubes.

In addition to the orientation UBII topic, this implementation subscribes to the touch events topic. The `touch down` event is needed to trigger the actual selection. In this experiment, cubes float in front of the user. If he points the laser at one and touches the display, the cube will change its color. This works not only with cubes but with any mesh. Also, the system can trigger any kind of event or action. A screenshot of this setup can be seen in Figure 3.4.3.

3.4.3. Virtual Keyboard

Text input is not an easy task to perform in VR. But it is often required for labeling, annotation, entering filenames for saving operations and setting parameters in visualizations and other productive VR software [RBP02, p. 2154]. This is why a lot of applications try to avoid it.

Tilt Brush[1] solves this by identifying their scenes with a screenshot of the scene rather than a filename. To save a scene, the users get a virtual camera attached to his hand motion controller, which he then uses to make a thumbnail of the current scene [Goo19a]. Others use a laser pointer either attached to a hand controller or to the headset to select keys on a 2D image of a keyboard [Wee17]. A more recent approach is the frequently

[1]Tilt Brush by Google is a tool for three-dimensional painting in VR. Website: www.tiltbrush.com

called “drum keyboard”, which attaches drum sticks to the hand controllers which are then used to hit the keys [Wei17]. There are also approaches which do not use hand controllers, like hand gloves [ESV99; RBP02], a real keyboard [McG+15; Wal+17] or other peripherals [Gon+09, pp. 111 sq.]. Other experimental methods are speech recognition [RBP02, pp. 2154 sqq.] and handwritten character recognition [Gon+09, p. 113].

Similar to the approach by Dias, Afonso, Eliseu, and Santos to implement user interfaces using a real smartphone and a virtual representation visible in the VR headset [Dia+18, p. 5], this experiment displays a virtual keyboard in VR as seen in Figure 3.4.4. The surface of the virtual keyboard is mapped to the touchscreen of the smart device (see Section 3.3). The cursor, represented by a blue circle, visualizes the position of the finger on the touchscreen. The cursor is only visible when the smartphone sends touch position data, which is only the case when the user is touching the screen. To select a key, the user has to move the cursor on top of a key and then keep the finger there for roughly a second. As long as the user is holding his finger on a key to select it, the blue circle fills up with another circle to indicate the running selection progress. When executing the key press on the first touch and without a cursor, the user would not know which key he is going to hit with his finger, because his sight is obscured by the VR headset. Dias, Afonso, Eliseu, and Santos work around this problem by using a Leap Motion sensor to visualize the finger positions [Dia+18, p. 4].

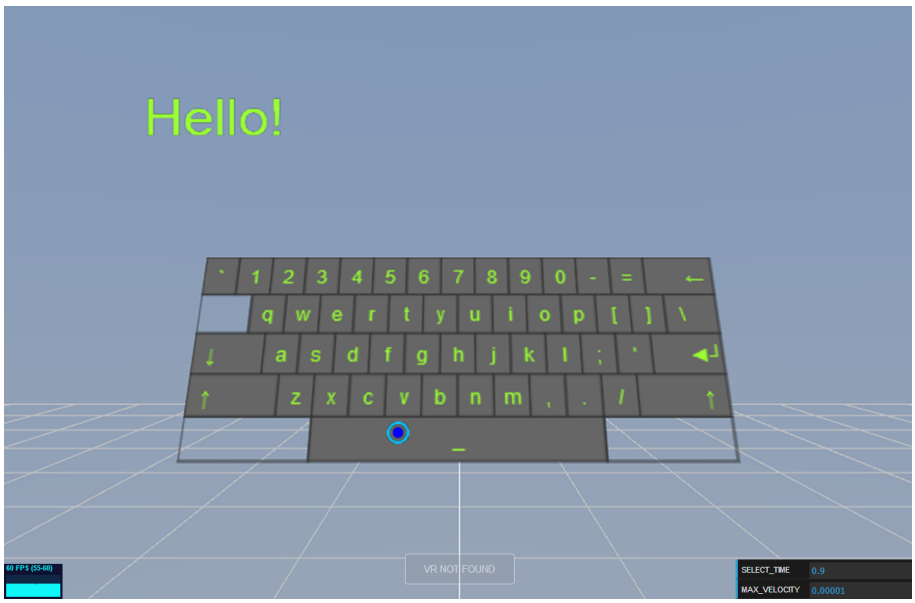


Figure 3.4.4.: A screenshot of the virtual keyboard with the blue cursor and the previously typed text.

Three components were implemented as JavaScript classes for this experiment. The `SmartphoneCursor` component uses the touch events and position data to display a blue circle (the cursor) on a given area in the scene. The cursor is only shown when the touch screen is pressed. If it is pressed, the position of the circle is synchronized with the position of the finger on the touch screen. To detect intentional movements, the current position (vector) is subtracted by the position (vector) of the previous frame. If the length of the resulting vector is smaller than a certain threshold value, it is assumed that this was not an intended movement. This calculation depends on the current framerate, which is not a good practice because the threshold value will behave differently depending on the current framerate. But as the target framerate is 60 frames per second and the testing system does not have issues in reaching this target, it is not a critical problem. As long as intentional movements are not detected, a timer counts up to a certain value (with the default settings, roughly one second). After this, the counter and a select event containing the cursor position is sent to the main program. To visualize this, a second blue circle gets larger until it fills up the whole cursor and finally disappears again.

The second component renders a virtual QWERTY[1] keyboard to the scene. The

[1]The name QWERTY describes the US layout for computer keyboards.

keyboard layout can be changed pretty easily as shown in Figure 3.4.5. Every key has a character or an action assigned as well as properties which influence the look. Special keys like the caps, caps lock, enter and delete key are fully functional as known from a real keyboard. If caps lock is activated, the key is drawn in blue and all characters are display in upper case. The `VirtualKeyboard` class draws the keyboard with just the keyboard layout, height, and width as input. If the `onPress(coordinates)` function is called, for example by the cursor component, the pressed key is calculated and returned using the provided position. The main program then applies the key to a string and sends the result to the third component.

```
// rows
[
  // columns
  ...
  [
    // keys
    ...
    {
      key: '=', // the returned character if no action is present otherwise just a label
      keyCaps: '+' // the returned character if in caps mode
    },
    {
      key: '←',
      action: KEY_ACTIONS.DELETE_ONE, // a special key action; in this case, it deletes
          the last character
      width: 2, // with of the key
      align: KEY_ALIGNMENT.RIGHT // the alignment of the label on the key
    }
    ...
  ],
  ...
],
```

Figure 3.4.5.: The definition of the virtual keyboard layout in JS. It is defined as a 2D array, where the first dimension corresponds to the key rows and the second one to the keys column wise.

The `TextDisplay` component just renders a given text inside a given area to a texture. If the text is changed it automatically updates and redraws the texture.

4. Evaluation

To show that the smartphone is a suitable interaction device in use with VR, the results of the usability study of the three experiments are presented. Simple tasks were implemented into the experiments, to measure the usability. After completing the task, they have to answer System Usability Scale (SUS) questions.

The general procedure of the experiment is as follows:

1. Introduce the user to the topic
2. Let the user fill out the consent form
3. Randomly choose an order for the experiments
4. For each experiment:
    a) Brief the user for the experiment
    b) Let the user play around in the experiment
    c) Start the task as soon as the user feels confident with the interactions
    d) Save the anonymized task results
    e) Question the user the usability questions

For the evaluation, two devices, connected to the same WLAN, were used:

- Personal Computer (PC)
  - Operating System: Windows 10
  - RAM: 32 GB
  - CPU: Intel(R) Core(TM) i7-6700K
  - GPU: NVIDIA GeForce GTX 1080
- Smartphone
  - Operating System: Android 9
  - RAM: 8 GB
  - CPU: Snapdragon(TM) 845

The PC was able to run the application smoothly with an average of 60 frames per seconds, which is enough to run a smooth VR experience. The WLAN connection and devices were capable of 20 Mbps, which is enough for synchronizing data without a visible lag.

4.1. Model Viewer

The model viewer experiment, presented in Section 3.4.1, allows users to view 3D model from different angles. To benchmark extensive usage, a second model instance in another color (the target) is spawned after starting the task. Since in the current implementation, the model cannot be rotated upside down (as mentioned in), the target is spawned with random reachable orientations. As soon as the task is started, the user has 30 seconds to match as many orientations as possible. After one orientation is matched, the target is rotated to a new random orientation. Because it is hard to match the rotation exactly on all three axes, it is enough to pose the model similar to the target.

Acknowledgments

Foremost, I would like to thank my supervisor Prof. Gudrun Klinker, at Technical University of Munich for giving me the opportunity to write my bachelor's thesis at her chair Forschungsgruppe Augmented Reality.

I would also like to express my sincere gratitude to my advisor, Sandro Weber for the close supervision and helpful advice.

A special thank goes to all my friends who took the time to participate in the evaluation.

Last but not least, I would like to thank my family for their continuous support during my studies and the writing of this thesis.

List of Figures

List of Tables

Bibliography

[AA13] F. Argelaguet and C. Andujar. "A survey of 3D object selection techniques for virtual environments." In: *Computers & Graphics* 37.3 (2013), pp. 121–136. ISSN: 00978493. DOI: 10.1016/j.cag.2012.12.003.

[Ben+11] A. Benzina, M. Toennis, G. Klinker, and M. Ashry. "Phone-based motion control in VR." In: *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. Ed. by D. Tan. ACM Digital Library. New York, NY: ACM, 2011, p. 1519. ISBN: 9781450302685. DOI: 10.1145/1979742.1979801.

[Cab19] R. Cabello. *Three.js: JavaScript 3D library*. 2019. URL: https://github.com/mrdoob/three.js/ (visited on 06/17/2019).

[DE11] M. Deller and A. Ebert. "ModControl – Mobile Phones as a Versatile Interaction Device for Large Screen Applications." In: *Human-computer interaction - INTERACT 2011*. Ed. by P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler. Vol. 6947. Lecture Notes in Computer Science. Berlin: Springer, 2011, pp. 289–296. ISBN: 978-3-642-23770-6. DOI: 10.1007/978-3-642-23771-3_22.

[Den16] A. Denoyel. *Virtual Reality evolved: Sketchfab VR apps and WebVR support: New on Sketchfab - 16 May 2016*. 2016. URL: https://sketchfab.com/blogs/community/announcing-sketchfab-vr-apps-webvr-support/ (visited on 06/23/2019).

[Dev19] Devices and Sensors Working Group. *DeviceOrientation Event Specification: Editor's Draft, 15 April 2019*. Ed. by Devices and Sensors Working Group. 2019. URL: https://w3c.github.io/deviceorientation/#dom-deviceorientationevent-alpha (visited on 06/17/2019).

[Dia+18] P. Dias, L. Afonso, S. Eliseu, and B. S. Santos. "Mobile devices for interaction in immersive virtual environments." In: *AVI '18: Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. Ed. by T. Catarci, F. Leotta, A. Marrella, and M. Mecella. New York, NY, USA: ACM, 2018. ISBN: 978-1-4503-5616-9. DOI: 10.1145/3206505.3206526.

[ECM17] ECMA International. *Standard ECMA-404: The JSON Data Interchange Syntax*. 2nd ed. 2017.

[ECM18] ECMA International. *Standard ECMA-262: ECMAScript 2018 Language Specification*. 9th ed. 2018.

[ESV99] F. Evans, S. Skiena, and A. Varshney. "VType: Entering Text in a Virtual World." In: (1999).

[Gon+09] G. González, J. P. Molina, A. S. García, D. Martínez, and P. González. "Evaluation of Text Input Techniques in Immersive Virtual Environments." In: *New Trends on Human-Computer Interaction*. Ed. by P. M. Latorre, A. Granollers Saltiveri, and J. A. Macías. London: Springer-Verlag London, 2009, pp. 109–118. ISBN: 978-1-84882-351-8. DOI: 10.1007/978-1-84882-352-5_11.

[Goo19a] Google LLC. *Protocol Buffers*. 2019. URL: https://developers.google.com/protocol-buffers/ (visited on 06/19/2019).

[Goo19b] Google LLC. *Tilt Brush Help: Saving and sharing your Tilt Brush sketches*. 2019. URL: https://support.google.com/tiltbrush/answer/6389651?hl=en (visited on 06/26/2019).

[Kam18] C. Kamm. "Precision of Pointing with Myo: A Comparison of Controller- and Gesture-Based Selection in Virtual Reality." Master's Thesis. Munich: Technische Universität München, 2018.

[KH10] N. Katzakis and M. Hori. "Mobile devices as multi-DOF controllers." In: *IEEE Symposium on 3D User Interfaces (3DUI), 2010 ; Waltham, Massachusetts, USA, 20 - 21 March 2010*. Ed. by M. Hachet. Piscataway, NJ: IEEE, 2010, pp. 139–140. ISBN: 978-1-4244-6846-1. DOI: 10.1109/3DUI.2010.5444700.

[Koe16] J. Koetsier. "Evaluation of JavaScript frame-works for the development of a web-based user interface for Vampires." PhD thesis. 2016.

[McG+15] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. "A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays." In: *CHI 2015 crossings*. Ed. by J. Kim. New York, NY: ACM, 2015, pp. 2143–2152. ISBN: 9781450331456. DOI: 10.1145/2702123.2702382.

[RBP02] C. J. Rhoton, D. A. Bowman, and M. S. Pinho. "Text Input Techniques for Immersive Virtual Environments: an Empirical Comparison." In: *Proceedings of the Human Factors and Ergonomics Society: 46th Annual Meeting, Baltimore, Maryland, September 30 - October 4, 2002 : Bridging Fundamentals & New Opportunities*. Ed. by Human Factors and Ergonomics Society. Annual meeting. Santa Monica, Calif.: SAGE Publications, 2002, pp. 2154–2158.

[SJ13] A. Steed and S. Julier. "Design and implementation of an immersive virtual reality system based on a smartphone platform." In: *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. Ed. by A. Lécuyer. Piscataway, NJ: IEEE, 2013, pp. 43–46. ISBN: 978-1-4673-6098-2. DOI: 10.1109/3DUI.2013.6550195.

[Wal+17] J. Walker, B. Li, K. Vertanen, and S. Kuhl. "Efficient Typing on a Visually Occluded Physical Keyboard." In: *Explore, innovate, inspire*. Ed. by G. Mark, S. Fussell, C. Lampe, m. schraefel m.c, J. P. Hourcade, C. Appert, and D. Wigdor. New York, NY: Association for Computing Machinery Inc. (ACM), 2017, pp. 5457–5461. ISBN: 9781450346559. DOI: 10.1145/3025453.3025783.

[Wee17] Weelco Inc. *Unity Asset Store: Keyboard VR Pro*. 2017. URL: https://assetstore.unity.com/packages/tools/input-management/keyboard-vr-pro-83708 (visited on 06/26/2019).

[Wei17] M. Weisel. *An open-source keyboard to make your own – Normal*. 2017. URL: http://www.normalvr.com/blog/an-open-source-keyboard-to-make-your-own/ (visited on 06/26/2019).

[You19] E. You. *Vue.js*. 2019. URL: https://vuejs.org/ (visited on 06/17/2019).

[Zha+15] K. Zhang, H. Hu, W. Dai, Y. Shen, and M. Z. Win. "Indoor Localization Algorithm For Smartphones." In: *CoRR* abs/1503.07628 (2015).

Appendices

A. User Evaluation Form

This evaluation is part of the Bachelor's thesis

„Smartphone-Assisted Virtual Reality Using Ubi-Interact“

of Michael Lohr.

If you have any questions, feel free to ask.

Section A: Preliminary Questions

A1. **What is your age?**

A2. **To which gender identity do you most identify?**

- Female
- Male
- Other

Other

A3. **What is the highest degree or level of school you have completed?**

If you are currently enrolled in school, please select the highest degree you have received.

- ☐ High school degree or equivalent
- ☐ Bachelor's degree (BA)
- ☐ Diploma's degree (Dipl.)
- ☐ Master's degree (MA)
- ☐ Doctorate (PhD)
- ☐ Other

Other

A4. **What is your main discipline?**

If you are a student, please select your current field of study. If you are employed, please select your area of work.

- ☐ Informatics
- ☐ Mathematics
- ☐ Law
- ☐ Medicine
- ☐ Other

Other

A5. **Please rate how much you used the following technologies in the last six months.**

	Never	Less than one to three times a month	One to three times a month	One to three times a week	More than three times a week	Daily	More than once per day
Smartphone	☐	☐	☐	☐	☐	☐	☐
Computer for Work/Studies	☐	☐	☐	☐	☐	☐	☐
Computer Games	☐	☐	☐	☐	☐	☐	☐
Virtual Reality Headset	☐	☐	☐	☐	☐	☐	☐

A6. Please rate the following statements in a range none (1) to a lot (5).

1 = Nothing/none, 5 = A lot

	1	2	3	4	5
How much do you know about virtual reality?	☐	☐	☐	☐	☐
How large is your interest in virtual reality?	☐	☐	☐	☐	☐
How much experience do you have with motion controllers (WII Remote, Oculus Touch)?	☐	☐	☐	☐	☐

Section B: Experiment: Model Viewer

System Usability Scale study reagarding the model viewer experiment.

B1. Please rate the following statements in a range from strongly disagree (1) to strongly agree (5).

1 = Strongly disagree, 5 = Strongly agree

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

Section C: Experiment: Laser Pointer

System Usability Scale study reagarding the laser pointer experiment.

C1. Please rate the following statements in a range from strongly disagree (1) to strongly agree (5).

1 = Strongly disagree, 5 = Strongly agree

	1	2	3	4	5
I think that I would like to use this system frequently.	☐	☐	☐	☐	☐
I found the system unnecessarily complex.	☐	☐	☐	☐	☐

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

Section D: Experiment: Virtual Keyboard

System Usability Scale study reagarding the virtual keyboard experiment.

D1. Please rate the following statements in a range from strongly disagree (1) to strongly agree (5).

1 = Strongly disagree, 5 = Strongly agree

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

Section E: Final Questions

E1. **If you have further critical or positive feedback or just a comment, please fill it in here:**

Thank you for participating in this study!