VR Interaction Modalities for the Evaluation of Technical Device Prototypes

Patrick Harms

VR Interaction Modalities for the Evaluation of Technical Device Prototypes

Patrick Harms[1][0000−0002−9988−2028]

Institute of Computer Science, University of Goettingen,
patrick.harms@cs.uni-goettingen.de

Abstract. Virtual Reality (VR) systems have reached the consumer market which offers new application areas. In our paper, we examine to what extent current VR technology can be used to perform local or remote usability tests for technical devices, e.g., coffee machines. For this, we put virtual prototypes of the technical devices into VR and let users interact with them. We implemented four interaction modalities that are suitable for low-level smartphone-based VR systems up to high fidelity, tethered systems with room-scale tracking. The modalities range from a simple gaze pointer to interaction with virtual hands. Our case study and its counter evaluation show that we are able to detect usability issues of technical devices based on their virtual prototypes in VR. However, the quality of the results depends on the matching between the task, the technical device, the prototype level, and the interaction modality.

Keywords: Virtual Reality · Usability · Usability Testing · Virtual Prototype · Device Assessment · Interaction Pattern.

1 Introduction

Virtual Reality (VR) is of interest for performing usability evaluations of technical device prototypes. In this area, a virtual prototype of a technical device, e.g., a coffee machine, is put into VR. Then potential users of this device are asked to interact with it in VR. The problems that the users have when interacting with the device may be an indicator of usability issues of the device. This basic idea was already proposed and done with former generations of VR systems, e.g., [4]. These former technologies had a lower quality than those today and interacting with them was not always easy. Through this, former VR systems may have influenced the usability test and it was not always clear if an issue stems from the virtual device prototype or from the VR system.

Because of the improvement of their quality, we think it is worth evaluating if the current state-of-the-art VR systems serve better for usability evaluations of technical device prototypes. In addition, as VR devices are also affordable by consumers, it is worth to consider performing these usability tests in a remote fashion as it is already done for websites and mobile apps [22, 23]. This means, users perform the tests separated in space and/or time from the evaluator while using their own technical setup. This setup may differ from user to user and

from simple smartphone/cardboard VR to high fidelity, tethered systems with room-scale tracking. Hence, it is important to use interaction modalities that are easy to use by the user, match the users technical setup, and serve the purpose of usability evaluations for technical device prototypes. As a first step in this direction, we focus in this paper on the following research question:

RQ 1: To what extent can different types of current consumer VR systems and corresponding interaction modes be used for interacting with virtual representations of technical devices in VR?

To answer this question, we implemented four established interaction modes and checked whether they are applicable for interacting with virtual device prototypes. Each mode can be implemented with a different technical level of consumer VR systems. As our goal is to use VR for usability evaluations of technical device prototypes in VR, we also consider the following research question:

RQ 2: How and to what extent can different types of current consumer VR systems be used for usability evaluations of technical device prototypes in VR?

This second question is especially important as even the best mode for virtual prototype interaction in VR (RQ 1) would be worthless if it does not allow to detect usability problems of the virtual prototype in a usability engineering scenario (RQ 2). We executed a case study with 85 participants to answer both questions. We evaluated the interaction modes and used them in a usability evaluation scenario for virtual prototypes of two different technical devices.

The paper is structured as follows. First we introduce the main terminology in Section 2 and provide an overview of related work in Section 3. We describe the interaction modes in Section 4. Our case study description in Section 5 covers details on its execution, our findings, and a discussion of the results including potential threats to validity. We conclude the paper in Section 6.

2   Foundations

In this section, we introduce the terminology used in this paper, such as usability, VR, and VR system.

Usability is the matching between user groups, their tasks, the context of task execution, and a product, e.g., hardware or software. If this is given, then users can perform their tasks with *effectiveness*, *efficiency* and *satisfaction* [1]. In addition, a high usability can increase the *learnability* for a system and decrease the number of *errors* users make [19]. To assess and improve the usability of a product, a diverse set of methods can be applied which is referred to as *usability engineering* [19]. Examples for such methods are *usability tests* [25]. Here, potential users of a product are asked to utilize a prototype or a final product for predefined and realistic tasks. During that, an evaluator observes the users and takes notes of the users' issues. Afterwards, the issues are analysed to identify potential for improvements. For a better analysis, the test sessions can be recorded, e.g., using video cameras or by logging the users' interactions. In addition, it is possible to determine the users' thoughts and opinions by, e.g., asking

them to think aloud during the test, by interviewing them, or by letting them fill in questionnaires after the test [11].

VR is a technique in which the user's senses are triggered by a *VR system* to pretend that the user is in another world, a *virtual world* [8]. The technical level of a VR system is called *immersion*. With an increasing immersion, usually the users feeling of being in the virtual world, the *presence*, increases. A term related to immersion and presence is *interaction fidelity*. It considers the similarity of the users physical actions for real world tasks compared to his or her actions to fulfill the same tasks in VR. A high fidelity means that the performed actions in the real world are very similar to the ones necessary to complete a task in the virtual world [3]. A high fidelity can increase the learnability for VR.

State-of-the-art consumer VR systems differ in their technical setup. For our work, we divide them into four categories according to their immersion and interaction fidelity. We list those categories with their basic characteristics and examples in Table 1. VR systems of Category 1, such as Google Cardboard [30], are smartphone-based. They only track the users head orientation and do not provide controllers. Hence, their immersion and interaction fidelity are rather rudimentary. Category 2 VR systems extend category 1 by the use of hand-held controllers providing buttons. An example is Oculus Go [20]. These VR systems track orientation movements of the controller and react on the usage of the controller buttons for allowing users to interact with objects in VR. Category 3 VR systems provide a room-scale tracking for headsets and controllers. This means, they do not only track their orientation but also their movement in space. All orientation changes and movements of the head set and the controllers are transfered to movements in the VR. This allows users to move freely, restricted only by the boundaries of the physical world. This also applies to virtual representations of the controllers that are usually shown in VR on this level. Overall, this level allows users to approach and touch virtual objects with the controllers. An example for a category 3 VR system is the HTC Vive [6]. Category 4 is characterized by its possibility to display parts of the users body in VR. These virtual body parts are used for interacting with the VR instead of a controller. For this, additional tracking systems are required. An example for such a tracking system is the Leap Motion [18]. It captures the users hands and their movements and

Characteristic	Category 1	Category 2	Category 3	Category 4
Head-mounted display	yes	yes	yes	yes
Head tracking	orientation	orientation	orientation, movement	orientation, movement
Controller	none	orientation	orientation, movement	none
User tracking	none	none	none	yes
Example	Google Cardboard [30]	Oculus Go [20]	HTC Vive [6]	HTC Vive [6] with Leap Motion [18]

Table 1. VR system categories and examples.

transfers them into a digital hand model in VR. With such a system, a rather high immersion and interaction fidelity can be accomplished.

Interaction with VR concerns mainly the *selection* and *manipulation* of virtual objects [14]. With a selection, users identify the object in VR for further interaction. With the manipulation, they change an object's position, orientation, scale, or other features. In our work, we further divide the manipulation of objects into the two actions *grab* and *use*. Grabbing an object means the same as in real world, i.e., selecting and moving it. Using refers to objects that cannot be moved but manipulated otherwise, like fixed buttons that can be pressed. Objects on which grab or use actions can be performed are called *tangible objects*.

Virtual worlds may incorporate *snap drop zones*. A snap drop zone is an area that detects objects in a defined proximity. If an object comes close to the snap drop zone, it is moved to a predefined position and locked there.

3   Related Work

The idea of using VR and Augmented Reality (AR) for usability evaluations of product prototypes is not new. There were attempts to assess a wheelchair-mounted robot manipulator [7], a microwave oven and a washing machine [4], as well as a phone [16]. Also the automotive sector utilizes virtual prototyping and evaluation [24]. While many of the studies focus on larger machinery, only few studies focus on consumer product usability [9] as we do.

The existing studies utilize a wide range of VR technologies from simple 2D screens [15] via 3D video walls [5] to Cave Automatic Virtual Environments (CAVEs) [26]. For interaction, the studies use, e.g., the Wizard of Oz technique [29], tracked objects [21], or special tracked gloves [4]. To the best of our knowledge, there are no studies focusing on current state-of-the-art consumer VR systems as done in our work. In addition, the related work does not consider performing remote usability testing on the side of the end user as we do. Furthermore, none of the existing studies had a test sample as large as ours.

The existing studies showed that with their corresponding technical setup, diverse problems are introduced. For example, test participants have problems in perceiving depth in a CAVE or request for haptic feedback [17]. Also tracked gloves can be problematic when trying to perform usability evaluations with many participants as they need to be put on and off again [17]. Such issues can have a strong influence on the evaluation results [27]. In our work, we analyze if similar issues arise with the current state-of-the-art consumer VR systems or if the typical interaction concepts, although being not fully immersive, are anyway sufficient for usability evaluations of technical device prototypes.

A similar study as ours was executed by Holderied [13]. In her student project, the author asked participants to use a virtual coffee machine and evaluated its usability as well as four different interaction modes. In addition, she had a control group of seven participants which used the real world exemplar of the coffee machine. In our work, we analyze two devices and, due to our case study setup, also assess learning effects. In addition, our interaction modes are different and

base on what was learned in the study of Holderied. The former study, furthermore, did not evaluate a hand mode as we do, used a less efficient highlighting of the tangible objects, and utilized other and less intuitive controller buttons.

4 Interaction Modes

We intend to use VR for usability evaluations of technical devices prototypes. In these evaluations, end users shall interact with the virtual devices while using their own VR systems to allow for a potential remote usability evaluation. As the VR system category per user may differ, we implemented four different interaction modes, one for each of the VR systems categories mentioned in Table 1. The modes are extensions and improvements of the interaction modes defined by Holderied [13] as referenced in the related work (Section 3). In this chapter, we describe these interaction modes in a generic way to allow for an implementation with different concrete VR systems. The descriptions follow the same pattern in which we mention the modes name, the aimed VR system category, how the selection, grabbing and using of objects works, and further important details such as the highlighting of tangible objects for user feedback.

4.1 Interaction Mode 1: Gaze Mode

VR system category: Category 1.

Object selection: Looking at objects, i.e., positioning a cross hair in the view center on them.

Object grabbing: After selection, leave the cross hair on the object for a defined period of time. Afterwards, the object becomes attached to the cross hair and can be moved by rotating the head. The object can only be released in close proximity to snap drop zones.

Object usage: After selection, leave the cross hair on the object for a defined period of time.

Further details: If the cross hair is positioned on a tangible object, the object's color changes to a highlighting color. If the cross hair stays on the object, the color blurs to a second highlighting color. After a predefined time, the color blur ends and the object is manipulated, i.e., grabbed or used.

4.2 Interaction Mode 2: Laser Mode

VR system category: Category 2.

Object selection: Pointing a virtual laser beam on the object using the orientation tracked controller.

Object grabbing: After selection, press and hold a button on the controller. The object stays attached to the tip of the virtual laser beam. To release the object, the controller button must be released.

Object usage: After selection, press a button on the controller.

Further details: When the laser beam points on a tangible object, the object's color changes to a highlight color. If the controller has a virtual representation in VR, a tooltip next to the required button can be displayed indicating the possible action. An example for this is shown in Figure 1a. This figure contains a virtual representation of an HTC Vive controller emitting the virtual laser beam. The tooltip is shown if the user points on a tangible object. We change the tooltip text while grabbing an object to indicate how the object can be released.

4.3   Interaction Mode 3: Controller Mode

VR system category: Category 3.

Object selection: Colliding the virtual representation of the room-scale tracked controller with the object (see object usage for further details).

Object grabbing: After selection, press and hold a button on the controller. The object stays attached to the virtual representation of the controller. To release the object, the controller button must be released.

Object usage: For object usage, colliding the controller with the object may be problematic in case the object is small. Hence, object usage in this mode is similar to the laser mode. The object selection is done using a virtual laser beam. In contrast to the laser mode, the beam in this mode must first be enabled. For this, the users use a second, pressure sensitive button on the controller. With a slight pressure on this button, the laser beam is enabled. To use an object, the users point the laser beam on the object and then fully press the button.

Further details: The mechanism for object usage can also be used for object grabbing. We implemented this for the purpose of consistency. We implemented a color highlighting if an object is selected using the laser beam as in the laser mode. Also in this mode, tooltips can be used to describe the controller buttons functionality. An example for this implemented with an HTC Vive controller is shown in Figure 1b. As multiple buttons are used in this mode, the tooltips should become partly transparent to indicate which button is pressed. In addition, they should change their caption depending on what can be triggered with the buttons (use or grab) in a certain situation.

Fig. 1. Screenshots of HTC Vive based example implementations of three of our interaction modes: a) laser mode with controller attached tooltips, b) controller mode with tooltips, c) hand mode.

4.4   Interaction Mode 4: Hand Mode

VR system category: Category 4.

Object selection: Touching an object with the virtual representations of the user's hand(s).

Object grabbing: Closing the hand(s) around the virtual object.

Object usage: Touching an object with the virtual representations of the user's hand(s).

Further details: In this mode, users see virtual representations of their hands as shown in Figure 1c. Every hand and finger movement is transferred to this virtual hand model. If a grabbable object is touched with a virtual hands, its color changes to a highlight color. If a usable object is approached, its color changes already on close proximity of the virtual hand.
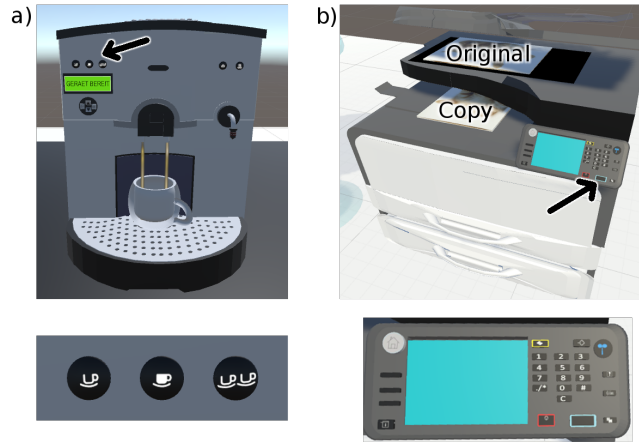
5   Case Study

To answer our research questions, we performed a case study. In this case study, we asked users to interact with two virtual technical devices in VR, a coffee machine and a copier. In the following subsections, we describe the technical setup of the case study, the execution, and our findings. Furthermore, we discuss the results and identify threats to the validity of our analyses.

5.1   Technical Setup

In the case study, we implemented two VR scenes, one for each technical device. In the first scene, we asked users to brew a coffee. Therefore, we call it the *coffee scene*. The coffee scene contains a cup and a virtual coffee machine. For executing the respective task, users need to put the cup under the outlet of the coffee machine and then press the button on the machine of which they think it produces coffee. This setup is similar to the case study described in [13]. In the second scene, our participants were asked to copy a sheet of paper. Hence, we call it the *copy scene*. The scene contains a copier and a sheet of paper lying next to it on a table. In this scene, the participants need to open the copier top, put the paper on the copier screen, and close the copier. Then they need to push the button on the copier which to their opinion creates the copy. Screenshots of both scenes are shown in the upper parts of Figure 2 (a: coffee scene, b: copy scene). In both scenes, the cup and the paper have already been put in place. The arrows mark the correct buttons to use for the respective tasks. In the copy scene, there is an additional sheet of paper being an already created copy.

We decided for this scene setup as it covers typical actions for interacting with technical devices. This includes moving parts of the device, moving utility objects, and pressing buttons. For both scenes, we implemented our four interaction modes. For moving the cup, the copier top, and the paper, we used the mode specific grab action, for pushing a button on the virtual device, the mode specific use action.

Fig. 2. Screenshots of both VR scenes of our case study including enlargements of the three coffee machine buttons that brew coffee and the interaction pad of the copier.

The scenes were created using Unity 3D [28]. The virtual devices were modeled using Blender [10]. They are based on a real coffee machine and a real copier. The functionality of the virtual devices was implemented as provided by the real devices. For the coffee machine, the three buttons on the top left produce coffee (enlarged in bottom part of Figure 2a). Per button, there is a different type of coffee produced, being light coffee for the left button, strong coffee for the center button, and two light coffees for the right button. The functionality of these three buttons was implemented so that coffee flows into the cup if one of the buttons is pressed. The other buttons on the coffee machine can be pressed, but no functionality is triggered and no other feedback is given. The display of the coffee machine shows "Gerät bereit" which is German for "Device Ready".

The functionality of the virtual copier was also partially implemented. The copy button is the large one on the bottom right of the interaction pad (enlarged in the bottom part of Figure 2b). When pressed, a copy of the paper is created at the copier outlet as shown in Figure 2b. As in the coffee scene, the other buttons can be pressed but nothing happens, and no feedback is given. In contrast to the coffee machine, the copier's display is empty and does not show anything.

In both scenes, the destination of the objects to move, i.e., the cup and the paper, are defined by snap drop zones. In the coffee scene, the coffee buttons work no matter if the cup is put into the snap drop zone. In the copy scene, the copy button only works if the paper is snapped correctly in the snap drop zone.

As VR system, we utilized an HTC Vive [6]. It comes with a head-mounted display and two controllers all being tracked at room scale. A visualization of a controller is shown in Figure 3. A controller has multiple buttons of which only some are of relevance for our case study. Those are the trigger button at the lower side of the controller and the trackpad on the upper side. The trigger button is pressure sensitive. We used the trackpad as a simple click button. In addition
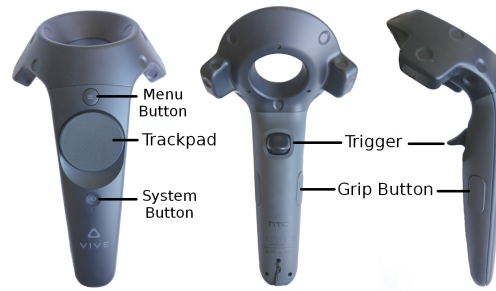
Fig. 3. Different views on an HTC Vive controller with its buttons.

to the Vive, we used a Leap Motion [18]. This device is a small hand tracker. It can be attached to the headset of the Vive so that the tracking of the hands is performed from the head of the user. Both devices allowed us to implement the four categories of VR systems and our interaction modes as follows:

- **Gaze mode**: Usage of Vive headset only (category 1 VR system).
- **Laser mode**: Usage of Vive headset and one controller (category 2 VR system); grab and use action are performed with the trigger button on the controller.
- **Controller mode**: Usage of Vive headset and one controller (category 3 VR system); as pressure sensitive button required for this mode, the trigger button is used; as click button the trackpad is used.
- **Hand mode**: Usage of Vive headset and the Leap motion (category 4 VR system).

For both scenes, we had a concept for resetting them. This helped to overcome unexpected situations. For the gaze mode, a reset was possible by restarting the scene. For the laser and the controller mode, we put a reset functionality on the grip button of the controller (see Figure 3). For the hand mode, we put a large red push button on the rear side of the scene.

5.2 Usability Test Setup

Using the technical setup we performed usability tests of the coffee machine and the copier in VR. For this we recruited test participants and asked them to use the virtual devices. All participant sessions consisted of the following steps:

1. Welcome and introduction
2. Usage of first device
3. Interview on interaction mode
4. Usage of second device
5. Interview on device usage
6. Discharge

During the usage of the devices and the interviews, we took notes of the issues the participants had as well as of any statements they made with respect to the VR system, the interaction mode, and the virtual technical device. Through this, we ensured to cover both our research questions, i.e., if the interaction modes can be used for interacting with technical device prototypes (RQ 1) and if usability defects of the technical device are found (RQ 2).

Each participant used both, the coffee and the copy scene, but only one of our four interaction modes (Section 4). This means, our case study had a between-subject-design considering the interactions modes. The case study design allowed us to measure, how the participants can learn and apply a certain interaction mode. For identifying learning effects, we flipped the order of the scenes between participants. This means, half of the participants first used the coffee scene and then the copier scene, whereas the other half used the scenes in opposite order.

The welcome and introduction (step 1) included asking for demographic information, describing the case study setup, and what data is acquired. Then, we let the participants put on the VR headset. Initially, they were in the center of a virtual sports stadium, a default scene provided by the Vive. For the laser and controller mode, we gave a brief introduction into the controller usage. This started with showing the controller in front of the participants. We then asked the participants to take the controller into their hand as well as feel and try out the important buttons. For the controller mode, we gave the additional hint that the trigger button is pressure sensitive. For the gaze and the hand mode, we gave no additional introduction.

Afterwards, we started the first scene (step 2) and asked the participants to perform the respective task. If they were unsure how to proceed, we gave them a mode specific hint. For example, for the gaze mode, we asked them to look around hoping that an accidental highlighting in this mode caused the participants to detect the gaze pointer. If a participant required any additional help, e.g., a repeated description of the controller buttons, we took a note.

After the task in the first scene was completed, we asked the participants to put down the VR headset. Then we interviewed them on their experience with the interaction mode (step 3). For this, we used four guiding statements and the participants had to assess whether they agreed to a statement or not. We also took notes of any additional comment. The guiding statements were

– I appreciate this type of interaction with a VR.
– I found the interaction with the VR unnecessarily complex.
– I would imagine that most people would learn this interaction very quickly.
– I found the interaction very cumbersome.

Then, we asked the participants to put on the headset again and started the second scene (step 4). In addition, we mentioned the scene-specific task. We gave no further help regarding the interaction to see whether the participants learned the interaction mode after the first scene. Only for few participants, we had to provide additional mode specific help in the second scene. After the task in the second scene was completed, we again asked the participants to put down the

VR headset for the second interview (step 5). Here we also used four guiding statements to which the participants had to agree or disagree. These statements were as follows and focused on the technical device only:

- I thought the device was easy to use.
- I found the device unnecessarily complex.
- I would imagine that most people would learn to use this device very quickly.
- I found the device very cumbersome to use.

Again, we took notes of any additional comments of our participants. Finally, we asked our participants to provide us with any further feedback that came to their minds. Then we thanked for their participation and closed the test (step 6).

For the solving of the tasks in both VR scenes, the participants had to press one of the correct buttons on the devices. If they initially tried one or more wrong buttons, we gave them a hint to try other buttons. If a participant required a reset for the scene for some reason, we instantly mentioned that this is possible, and either performed it by restarting the scene or by describing to the participant how the reset can be triggered.

During the usage of the VR scenes, we recorded the screen of the computer, i.e., what the participant saw and did in the scenes. In addition, both scenes were equipped with a logging mechanism. Through this, we additionally got technical recordings of which controller buttons were used, the orientation and movements of the user's head, and the actions relevant for our case study, i.e., grabbing and using an object. The individual user actions were logged together with a time stamp in the order in which they occurred.

5.3   Execution and Data Analysis

We executed the case study in two separate sessions in June 2017. The first session was done at a central location between multiple lecture halls at our university. The second session took place at a shopping mall. Overall, we had 85 test participants (46 students, 22 employed, 11 pupils, 3 retired, 1 unemployed, 2 unknown). 23 participants used the gaze mode, 21 the laser mode, 21 the controller mode, and 20 the hand mode.

We did not do a specific participant screening. This allowed us to assess whether our interaction modes can be used by a broad variety of different user groups. This is of major importance for our overall goal of allowing for usability evaluations of technical device prototypes in VR. The reason is that also for such evaluations a broad variety of test persons may need to be recruited.

A majority of 65 participants had no experience with VR and 14 only heard about it. None of our participants considered him- or herself a VR expert. 45 participants first used the coffee scene and 40 first used the copy scene. The logging of actions and the recording of the screen did not work for the first participant (gaze mode, coffee scene first) resulting in one recording less.

As we were at central locations, we had to ensure that our participants were not biased by observing other participants. This is important for correctly measuring if there is a learning effect between the interaction in the first and the

second scene. For this, we did not project anything of what the participants saw during a test to a larger screen. We also informed new audience that if they wanted to participate, they must not look at our computer screen. In addition, if participants had observed previous test participants, we informed them that their own sessions will be different and that their observations will be of no help or relevance. We also ensured that subsequent participants used different interaction modes.

All participants were introduced and observed by the same two evaluators. The evaluators split their tasks. The first evaluator did the introduction and the interviews of the participants and ensured that the participants were physically safe while interacting with the scenes. The other evaluator started the VR scenes and took the notes. For test and data consistency, these roles stayed the same for all test sessions.

After the case study execution, we performed a data post-processing and analysis. This included a card-sorting for grouping the detected usability issues. For this, both evaluators were asked separately to define categories of user issues and statements and to assign the notes to these categories. The categories of both evaluators were then matched to each other and harmonized. Through an analysis of the screen casts and the log files, we determined additional data, such as the duration of individual actions of the participants.

5.4   Findings

From the case study data, we derive different types of usability issues. Some issues concern the interaction in general and occurred for both scenes (*general issues*). Other issues are specific to an interaction mode (*mode issues*) or to one of the virtual prototypes (*device issues*). In addition, we took note of other user comments. All issues and comments as well as our duration measurements are listed in the following.

General issues: Table 2 shows an overview of the general issues and their number for each interaction mode as well as in total. In the first row the table also lists the number of participants that used a certain interaction mode for

	Gaze	Laser	Controller	Hand	Total
Number of participants	23	21	21	20	85
Genera issues					
Detailed help	-	9	18	2	29
Object detached	5	4	-	12	21
Too far away	12	4	3	-	19
Snap Drop Zone	5	2	4	4	15
Reset	-	-	1	12	13
Mode unintuitive	2	1	-	-	3

Table 2. General issues with the total and mode specific number of occurences.

better reference. The most prominent issue (29 times) concerns the participants' need for a more detailed help of how to interact with the VR. For the controller mode (18 times), this includes a second demonstration of the two pressure points of the trigger button. Another problem is that the cup or the paper fell down unwillingly. This happened 21 times, most often for the hand mode (12 times). 19 participants stood too far away from the technical device to be able to interact properly, 12 of them using the gaze mode. It was difficult for some participants to use the snap drop zones correctly. In total, 15 struggled to understand the concept of this feature and tried to place the object exactly without letting this be taken over by the snap drop zones.

Mode issues: 7 participants using the gaze mode did not look long enough at a tangible object to trigger an action, 3 did not recognize the cross hair, and 2 grabbed an object unintentionally. With the controller mode, the participant had two possibilities to grab an object: 14 grabbed it by touching it with the controller, the other ones used the laser. 8 participants of this mode tried to trigger the use action by touching a button with the controller. For the laser mode, 1 participant would have liked to have the controller tooltips in the field of view and another did not see them initially. With the hand mode, 6 participants had difficulties to grab an object and 4 mentioned that the hardware needs improvement for this mode, especially due to tracking and grabbing problems.

Device issues: For the coffee machine, 38 participants initially tried to use a wrong button (19 the buttons below the display, 17 the buttons on the top right, and 2 the display itself). When the participants used one of the three correct buttons, 56 used the middle one (strong coffee), 17 the left (light coffee), and 7 the right (two light coffees). The other five participants pressed several buttons at once (hand mode). For the copier, the most prominent issue was that 35 participants did not open it before trying to place the paper. Furthermore, 28 participants initially pressed a wrong button (16 the top left, 10 the top right, and 2 the button left of the actual copy button). 12 participants placed the paper incorrectly at first. 8 put it into the paper tray and 4 on the paper spawn zone. 5 participants already knew our printer model. 4 participants asked if the paper needed to be turned to be copied.

Additional comments: The additional comments of the participants are listed in Table 3. Most prominently, 29 participants said that they had difficulties in seeing the icons of the coffee machine or copier properly because they seemed to be diffuse. For 6 participants, the meaning of the icons was unclear even though they could see them sharply. 2 participants mentioned this for the coffee machine, 2 for the copier, and 2 for both devices. 7 claimed that, for the given task, only one button would suffice and the other ones were distracting. 7 participants stated that the interaction was easy to learn, in contrast to 5 other participants who said that the interaction would be easier for younger people.

Time measurements: For analyzing the performance and learnability of the different modes and scenes, we measured the duration of the interaction. For this, we divided the tasks into two phases. The first phase starts when the

Issue	Number
Icons diffuse	29
One button for task would be sufficient	7
Easy to learn	7
Icons partially unclear	6
Easier for younger people	5
Prefer to see own body	3
Prefer to use hands (did not use hand mode)	3
Problems with concept of VR	3
Icons too small	2
Haptic feedback missing	2
Easier for technically skilled people	2
Does not like VR	2
Copier more difficult than coffee machine	2

Table 3. Additional comments of at least two participants of the case study.

participants intentionally touched a tangible object and ends when the cup or paper are placed correctly. This starts the second phase which ends with the task completion, i.e., when the participant pressed the correct button on the technical device. The first phase of the copy scene includes the opening of the copier, the second phase includes its closing. The resulting mean durations in seconds for each interaction mode distributed on the VR scenes and the corresponding phases are listed in Table 4 in the columns called *total*.

The values show, that for all interaction modes, the first phase was completed faster with the coffee machine than with the copier (total values). For the coffee machine, the participants of the controller mode were fastest (6.8s), the participants of the hand mode slowest (19.4s). For the copy scene, the gaze mode participants finished fastest (24.9s) while the hand mode participants were slowest (56.8s). The second phase was about equally quick to accomplish in both scenes. In contrast to the first phase, the hand mode participants were fastest (20.1s coffee scene, 12.5s copy scene) and the controller participants slowest (34.2s coffee scene, 30.1s copy scene). These differences are signicant. We performed a pairwise two-sided Wilcoxon test [31] for every column in Table 4

| | Coffee scene | | | | | | | | Copy scene | | | | | | | |
| | Phase 1 | | | Phase 2 | | | Complete | | Phase 1 | | | Phase 2 | | | Complete | |
Mode	first	second	total	first	second	total	first	second	first	second	total	first	second	total	first	second
Gaze	13.9	10.8	12.7	23.9	19.8	22.3	37.8	30.6	24.4	25.1	24.9	24.4	26.7	25.8	48.9	51.9
Laser	21.1	8.6	15.1	26.4	15.1	21.0	47.5	23.7	46.1	27.7	36.5	19.7	19.8	19.8	65.8	47.6
Controller	8.9	4.9	6.8	43.4	25.9	34.2	52.3	30.8	45.1	32.8	39.2	34.9	24.9	30.1	80.0	57.7
Hand	24.4	14.3	19.4	21.0	19.1	20.1	45.4	33.4	63.8	49.8	56.8	16.3	8.7	12.5	80.1	58.5

Table 4. Mean durations for the individual phases and overall usage of the coffee and the copy scene in seconds.

with a Hochberg p-value adjustment [12]. We rejected the null hypothesis, that durations are equal, if the p-value is below 0.005 [2].

Table 4 also shows differences of the durations between completing the task and phases depending on whether a scene was the first or the second scene to use by the participants. These values are shown in the columns named *first* and *second* in the respective parts of the table. From the values, we can see that for the coffee scene, the participants were always faster in all aspects (phases or complete duration) when they used the scene as the second one, i.e., they used the copier scene first. This is also independent of the concrete interaction mode. For the copier scene, this is only the case for the controller and the leap mode. In the other modes, the participants using the scene as the second one were not necessarily faster. According to our Wilcoxon test, only the differences between the hand and the gaze mode, and the hand and the controller mode for phase two in column *second* for the copier are significant.

5.5   Counter Evaluation

As mentioned, our virtual devices were based on real technical devices. To be able to evaluate, whether we correctly identify usability issues of these devices in VR, we also performed usability evaluations of the real devices. These evaluations were structured as the second phase of the VR evaluations described above. This means, we asked the participants to use the device for the same task as in VR. In the meantime, we observed them and took notes of comments and issues. Afterwards, we interviewed them using the same guiding statements as for the device assessment in VR. The test participants we recruited for this evaluation were different from those using our VR scenes.

We had 10 participants for the coffee machine evaluation. 9 of them had misconceptions or verbalized an unsureness about the three buttons for brewing the coffee. 7 participants used the middle button. 2 of them explicitly mentioned that the icon on this button looks like a full cup matching the goal of their task. These results are similar to a counter evaluation for the same device described by Holderied [13].

Also for the copier, we had 10 participants. 3 mentioned an unsureness about the copy button, but used it correctly. 2 users used the input tray, the others opened the top and put the paper there.

5.6   Discussion

Based on our results, we answer our research questions. RQ 1 asks to what extent current consumer VR systems and corresponding interaction modes can be used for interacting with virtual prototypes of technical devices in VR. Our results show that all but one participants were able to accomplish all given tasks. Hence, we conclude:

The identified four categories of consumer VR systems and corresponding interaction modes can all be used for interacting with virtual prototypes of technical devices.

But there are differences between the interaction modes resulting in different user efficiency. The values below one minute for task completion seem, at least for us, in an acceptable range for all interaction modes considering the application area of usability testing. But due to different results for phase 1 and phase 2 of the scenes and also our list of interaction mode specific user issues, we derive that for grabbing and using objects, different interaction modes work best. This means in detail:

Using buttons works best with the hand mode.
Grabbing objects is easiest with the controller mode.

For the usage of VR for usability evaluations of technical device, it is also important that users can easily learn the interaction modes. For assessing this learnability, we compared the different times the participants needed using a certain scene either as first or as second scene. The reduced time that the users needed with a specific mode when used in the second scene indicates that:

The learnability of all interaction modes is generally high.

This applies although the interaction times within the VR were relatively short. For the laser and the controller mode, the learnability was strongest. The gaze mode instead does not show an efficiency improvement. However, it has a low error rate except for virtually standing too far away. This can be derived from the mode specific issues we detected. Hence, we derive:

The initial position of test participants in VR when using the gaze mode must be tested and selected with care.

The required resets and the number of detached objects let us conclude that the hand mode needs improvement for actions like grabbing. In contrast, the need for a more detailed help is highest for the controller mode. This also indicates that an initial understanding of the controller itself, especially with multiple pressure points of a button, is difficult.

Based on our results, we also provide answers for RQ 2 focusing on the applicability of current consumer VR system for usability evaluations of technical device prototypes. Our results show that independent of the interaction modes:

It is possible to detect real usability issues of the technical devices using virtual prototypes in VR.

For example, we found that for brewing coffee, most users used the button for a strong coffee. This result correlates with the findings of our counter evaluation and the ones in [13]. A similar problem was detected for the copier. For some participants, it was unclear which button triggers the copy function. Considering our setup, we also conclude that:

For the virtual prototypes, all affordances related to an evaluated user tasks must be implemented.

This stems from the fact that some participants tried to put the paper in the tray on the copier top instead of opening the copier. This means, they tried to use an affordance for which we did not provide an implementation. Hence, if a

technical device has multiple affordances for the same task, all of them must be simulated. In contrast, we also saw that neither displays of devices need to show anything nor additional functionality needs to be implemented. We derive this from the fact that the virtual copier did not show anything on its display but only one user asked if the copier needed to be switched on.

From the study, we can also draw conclusions considering the differences between local and remote usability evaluation. In local usability evaluation, the evaluators can support the participants. Hence, the challenges for the participants to interact with the VR can be a bit higher. For remote evaluation, the evaluators are separated from the participants. Hence, the VR systems and interaction modes in this scenario should be as simple as possible and easy to learn. Our results show, that this is given best for the gaze and the laser mode. Hence, we propose:

> *For remote usability evaluations, use the gaze or the laser mode.*
> *For local usability evaluations, use the controller or hand mode.*

This would also match the fact that the corresponding VR system categories for remote usability evaluation would be less complex and cheaper than for local evaluations reducing the overall burden for test participation. Our participants partially mentioned problems with a diffuse view leading to a bad recognition of icons on buttons. Hence, when performing usability tests with current consumer VR systems, it needs to be considered that:

> *The resolution of the VR system may influence the evaluation results negatively.*

Finally, considering the technical setup of the VR scenes and the snap drop zones, we conclude that:

> *Snap drop zones must be implemented as realistic as possible.*

Otherwise, users may become distracted as partially happened in our study. The evaluators of usability tests also need to keep in mind that technical issues like an overheating of the VR system or tracking problems may occur. We did not observe typical VR issues as cyber sickness [8]. But this may be caused by the relatively short usage times.

5.7   Threats to Validity

The validity of our case study may have been affected by several threats. For example, due to the setup of our VR area, some participants may have watched the interaction of their predecessors and might have experienced a learning effect, even though they used another interaction mode and although we actively tried to prevent this. Since we executed one of our sessions in the university, half of our participants were students and, therefore, our test sample might be too homogeneous. On top of that, many potential participants in the mall did not want to join the case study for different reasons like being afraid or having heard bad things about VR. So we might have missed a relevant user group.

Furthermore, we still had only a small number of participants per mode and first scene combination.

During the case study, many participants seemed rather impressed by VR. Our demographic data also shows that 65 of our participants did not have any previous experience with VR. This may have positively influenced their assessment of our interaction modes or the virtual technical devices.

During the case study, the same two evaluators were responsible for executing the whole case study. This might have caused an evaluator effect. Furthermore, we did not switch off the room scale tracking of the HTC Vive for the gaze and the laser mode although the envisaged VR systems categories for these modes only provide orientation tracking for the head set. We did this to allow for position corrections of the participant during the evaluation. But this may have influenced the results for these two modes.

6   Conclusion and Outlook

In our work, we assessed how state-of-the-art consumer VR technology can be used for performing usability testing based on virtual prototypes of technical devices. For this, we first identified four different categories of consumer VR systems. Then we implemented one interaction mode per category and tested in a large case study if they can be used for interacting with the virtual technical devices. Overall, we found that a gaze mode and a laser mode currently work best. In addition, a mode where users can use their real hands for interaction has quite some potential as long as the required tracking techniques improve. We also showed that usability issues of technical devices can be found by using VR. In addition, we uncovered some issues that may occur when performing this type of usability evaluation, such as difficulties in correctly seeing details of the technical devices.

For future research, similar case studies should be executed with other VR systems and further user groups to have an enhanced validation of our results. In these evaluations, a first study with remote usability evaluation should be performed. This would include recording and analyzing the VR usage without creating screen casts. For this, our already used logging mechanism may be an option. The logged actions may be used for replaying the VR usage. In addition, we will consider how our hand-based interaction can be improved so that it becomes an option for our intended scenarios, as it seems to be most intuitive but technically challenging.

7   Replication Kit

All the data we recorded in the case study, the performed statistical tests, as well as the VR scenes have been published in a replication kit available at https://doi.org/10.5281/zenodo.894173.

References

1. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability (ISO 9241-11:1998), 1998.
2. Daniel J Benjamin, James O Berger, Magnus Johannesson, Brian A Nosek, Eric-Jan Wagenmakers, Richard Berk, Kenneth A Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, et al. Redefine statistical significance. *Nature Human Behaviour*, 2017.
3. Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. Questioning naturalism in 3d user interfaces. *Commun. ACM*, 55(9):78–88, September 2012.
4. Fabio Bruno and Maurizio Muzzupappa. Product interface design: A participatory approach based on virtual reality. *International Journal of Human-Computer Studies*, 68(5):254 – 269, 2010.
5. Marina Carulli, Monica Bordegoni, and Umberto Cugini. An approach for capturing the voice of the customer based on virtual prototyping. *Journal of Intelligent Manufacturing*, 24(5):887–903, Oct 2013.
6. HTC Corporation. Vive, 2017. retrieved 9/2017.
7. G. Di Gironimo, G. Matrone, A. Tarallo, M. Trotta, and A. Lanzotti. A virtual reality approach for usability assessment: case study on a wheelchair-mounted robot manipulator. *Engineering with Computers*, 29(3):359–373, 2013.
8. Ralf Dörner, Wolfgang Broll, Paul Grimm, and Bernhard Jung, editors. *Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
9. Christianne Soares Falcão and Marcelo Márcio Soares. Application of virtual reality technologies in consumer product usability. In Aaron Marcus, editor, *Design, User Experience, and Usability. Web, Mobile, and Product Design*, pages 342–351, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
10. Blender Foundation. Blender, 2017. retrieved 9/2017.
11. Marcus Hegner. *Methoden zur Evaluation von Software*. Arbeitsbericht. IZ, InformationsZentrum Sozialwiss., 2003.
12. Yosef Hochberg. A sharper bonferroni procedure for multiple tests of significance. 75:800–802, 12 1988.
13. Hanna Holderied. Evaluation of Interaction Concepts in Virtual Reality. In Maximilian Eibl and Martin Gaedke, editors, *INFORMATIK 2017*, LNI, pages 2511–2523. Gesellschaft für Informatik, Bonn, September 2017.
14. Jason Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA, 2016.
15. Satoshi Kanai, Taiga Higuchi, and Yukiaki Kikuta. 3d digital prototyping and usability enhancement of information appliances based on usixml. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 3(3):201–222, Aug 2009.
16. K. Kuutti, K. Battarbee, S. Sade, T. Mattelmaki, T. Keinonen, T. Teirikko, and A. M. Tornberg. Virtual prototypes in usability testing. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pages 7 pp.–, Jan 2001.
17. Glyn Lawson, Davide Salanitri, and Brian Waterfield. Future directions for the development of virtual reality within an automotive manufacturer. *Applied Ergonomics*, 53:323–330, 2016. Transport in the 21st Century: The Application of Human Factors to Future User Needs.

18. Inc. Leap Motion. Unity, 2017. retrieved 9/2017.
19. Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
20. Oculus. Oculus go, 2018. retrieved 8/2018.
21. Hyungjun Park, Hee-Cheol Moon, and Jae Yeol Lee. Tangible augmented prototyping of digital handheld products. *Computers in Industry*, 60(2):114–125, 2009.
22. Fabio Paternò. Tools for remote web usability evaluation. In *HCI International 2003. Proceedings of the 10th International Conference on Human-Computer Interaction. Vol.1*, volume 1, pages 828–832. Erlbaum, 2003. retrieved 9/2017.
23. Fabio Paternò, Andrea Russino, and Carmen Santoro. Remote evaluation of mobile applications. In Marco Winckler, Hilary Johnson, and Philippe Palanque, editors, *Task Models and Diagrams for User Interface Design*, volume 4849 of *Lecture Notes in Computer Science*, pages 155–169. Springer Berlin Heidelberg, 2007.
24. Diana Reich, Christian Buchholz, and Rainer Stark. *Methods to Validate Automotive User Interfaces Within Immersive Driving Environments*, pages 429–454. Springer International Publishing, 2017.
25. Michael Richter and Markus D. Flückiger. *Usability Engineering kompakt: Benutzbare Software gezielt entwickeln*. It Kompakt. Springer Berlin, Heidelberg, 2013.
26. Alistair Sutcliffe, Brian Gault, Terence Fernando, and Kevin Tan. Investigating interaction in cave virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 13(2):235–267, June 2006.
27. Alistair Sutcliffe, Brian Gault, and Neil Maiden. Isre: immersive scenario-based requirements engineering with virtual prototypes. *Requirements Engineering*, 10(2):95–111, May 2005.
28. Unity Technologies. Unity 3d, 2017. retrieved 9/2017.
29. Jouke Verlinden, Willem Van Den Esker, Liesbet Wind, and Imre Horváth. Qualitative comparison of virtual and augmented prototyping of handheld products. In D. Marjanovic, editor, *DS 32: Proceedings of DESIGN 2004, the 8th International Design Conference, Dubrovnik, Croatia*, pages 533–538, 2004.
30. Google VR. Google cardboard, 2017. retrieved 9/2017.
31. Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.