Research Article

Matthias Weise*, Raphael Zender, and Ulrike Lucke

How Can I Grab That?

Solving Issues of Interaction in VR by Choosing Suitable Selection and Manipulation Techniques

<https://doi.org/10.1515/icom-2020-0011>

Abstract: The selection and manipulation of objects in Virtual Reality face application developers with a substantial challenge as they need to ensure a seamless interaction in three-dimensional space. Assessing the advantages and disadvantages of selection and manipulation techniques in specific scenarios and regarding usability and user experience is a mandatory task to find suitable forms of interaction. In this article, we take a look at the most common issues arising in the interaction with objects in VR. We present a taxonomy allowing the classification of techniques regarding multiple dimensions. The issues are then associated with these dimensions. Furthermore, we analyze the results of a study comparing multiple selection techniques and present a tool allowing developers of VR applications to search for appropriate selection and manipulation techniques and to get scenario dependent suggestions based on the data of the executed study.

Keywords: 3D Interaction, Selection/Manipulation Techniques, Virtual Reality

1 Introduction

Since the early days of computer graphics, researchers try to find convenient ways for the interaction with objects in three-dimensional virtual space (see, e.g. [48]). The progress made in the development of VR systems, particularly in the last decade, makes the technology accessible to an ever-wider target group and continually increases the possibilities for interaction with the virtual world. It is more than ever necessary to provide users with suitable ways to interact with virtual objects. Each VR application comes with its unique interaction requirements, based on which developers need to choose appropriate selection

and manipulation techniques [11]. This includes objective criteria like speed and accuracy but also ergonomic factors like the induced mental or physical workload of a technique. For example, some techniques might work better on close distances but are more prone to fatigue. To date, standards known from the desktop interaction with mouse and keyboard are missing. The interaction techniques *Simple Virtual Hand*, which maps the movements of the real hand to a virtual counterpart, or *Ray-Casting*, which uses a ray to interact with objects, can be found in many applications. However, often they are used since they are well-known and easy to implement [56]. Developers of VR applications often do not know the full design space of possible interaction techniques. To enable research results on 3D user interfaces finding their way into the development of VR applications, we have to show the advantages and disadvantages of different interaction techniques for specific scenarios [13] and need to increase the visibility of the results.

In this article, we present our steps towards such a systematization and recommendation. We start with an introduction to the major issues of interaction in VR. Then an adapted version of an existing taxonomy for selection and manipulation techniques [54] is presented, which allows classifying techniques regarding different dimensions. We also discuss the effects of the different dimensions on the presented issues of interaction in VR. We used the taxonomy to select techniques for an exploratory user study where we compare these techniques regarding their performance in different application scenarios as well as in terms of usability and user experience aspects. Subsequently, we present a tool that allows the filtering and suggestion of techniques based on the taxonomy and the objective and subjective measurements of the study. The article closes with a summary and planned future work.

2 Interaction in Virtual Reality

Technically, virtual reality can work with classic input devices like a gamepad, a keyboard, or a mouse. However,

*Corresponding author: Matthias Weise, University of Potsdam, Potsdam, Germany, e-mail: matthias.weise@uni-potsdam.de
Raphael Zender, Ulrike Lucke, University of Potsdam, Potsdam, Germany, e-mails: raphael.zender@uni-potsdam.de, ulrike.lucke@uni-potsdam.de

the possibility to interact with the virtual world by using movements in three-dimensional space drastically improves the immersive experience. In this article, we focus on this type of 3D interaction based on real spatial input [32]. A 3D interaction technique maps such input to the actions in the virtual 3D space. Bowman refers to the three universal tasks navigation, selection, and manipulation [8]. These tasks have to deal with similar issues. However, most of the existing interaction techniques incorporate selection and manipulation, whereas there is a high number of dedicated techniques for locomotion in VR. Therefore, in this article, we focus on selection and manipulation techniques, but our results are not limited to them. In the following, when we speak about interaction techniques, we refer to selection and manipulation techniques. We investigate techniques that support one of the common tasks: selection, positioning, rotation, and scaling. We do not consider specialized techniques, which are, for example, only used to delete an object or to change particular properties like the color. Furthermore, we do not consider cooperative interaction or multi-selection/manipulation of objects.

The research in this field led to a high number of different 3D interaction techniques since capable input devices exist. Over several years we found over 110 different 3D selection and manipulation techniques from literature research and experience with VR applications but raise no claim for completeness. However, this comprehensive set of interaction techniques reflects the common (and several uncommon) forms of interaction in VR applications.

It is important first to understand the problems we have to deal with in virtual reality interaction to help developers of VR application selecting suitable interaction techniques. That is why the next section describes the most common and examined issues of interaction in VR in literature. We then introduce an adapted version of our previously developed taxonomy [54] and link the interaction issues to the dimensions of the taxonomy.

2.1 Problems

Interaction in VR mainly relies on input devices that track the movements of the user in three-dimensional space. Unfortunately, midair interaction always causes noise because of the *natural hand tremor* [30]. Without a stable rest, it is impossible to fully maintain a position with the hands, which inevitably leads to imprecise interactions. Even with perfect tremor compensation, human *precision* cannot be perfect because of the limited hand-eye coordination, restricted motor precision, and the unachievable

fine control of the muscle groups [30]. Furthermore, even though the input devices are quickly getting better, tracking jitter and lag are still issues and negatively influence the possible precision.

Additionally to movements in 3D space, VR interaction relies on discrete inputs (e. g., buttons) for sub-tasks like confirming a selection or releasing an object. However, this comes along with a problem called the *Heisenberg effect* [15]. Pressing a button on a tracked input device inevitably affects the position of the device, which can, for example, lead to a missed selection when the virtual selection tool consequently leaves the object.

Fatigue is another problem that can arise in midair interaction [26]. This issue is also known as the gorilla arm effect, which is a well-known problem when using vertical touch screens for a longer time. Notably, techniques, where the user needs to extend their arm for a longer time, are tiring.

Besides the physical demands of an interaction technique, the induced *cognitive load* has a significant impact on usability. The cognitive load increases with the complexity of an interaction technique and directly connects to the ease of use [8]. For example, the usage of multiple buttons or the need to frequently switch between different modes can have an impact on the complexity. It is not advisable to increase the performance of a technique by considerably increasing its complexity [3] as this can lead to stress and frustration on the user side.

A problem that mainly affects pointing techniques that do not originate from the eyes is the *eye-hand visibility mismatch* [4]. The objects which are fully visible from the eyes point of view are not necessarily unoccluded when seen from the pointing origin. Especially in dense environments, this allows the selection of objects which are actually not visible and respectively hinders the selection of objects which are visible to the user.

The *Midas Touch effect* [28] is another issue that can appear during the selection of objects. For example, with an eye tracker interface, the user may select unwanted objects if the technique does not use dwell time or if it is too short. Hand-based techniques are also prone to the Midas Touch effect if they use hovering to select objects [20].

Several environment parameters can hamper the interaction in VR. Many studies showed that the *distance* and the *size* of an object have a significant impact on the interaction performance (e. g. [11]). Greater distances and smaller sizes of an object decrease the visual size, which makes it harder to interact with an object if the interaction technique does not account for that. Furthermore, some techniques, like the *Simple Virtual Hand*, only work on

short distances. Environments with a high number of objects can show a high *density*, object *occlusion*, and *movements*. These issues can lead to unwanted selections, or the user needs to change the perspective for a better angle if the interaction technique does not provide a mechanism to solve these issues.

A problem that can arise in manipulation tasks is *clutching*. For instance, if a technique directly maps the wrist rotation to the object rotation, it is often necessary to execute the task stepwise. If the user wants to rotate the object by a high amount, she often has to release the object, then rotate the hand back and grab the object again. This movement overhead can be very time consuming and unsatisfying for the user [59].

In VR, we can overcome limitations of reality like gravity or the reach of the arms. This freedom enables a high *expressiveness* of interaction in VR and allows us to do things we are not able to do in reality [3]. For example, it is possible to enable users not only to rotate and translate objects but also to scale them. It is also possible to add disambiguation mechanisms to support the selection of objects in high-density environments. The disadvantage is that such interaction techniques can quickly get complex and harder to learn.

2.2 Taxonomy for Selection and Manipulation Techniques

As already mentioned, there is no perfect interaction technique that can deal with all of the described problems, but over time, a lot of different techniques were developed. They target one or multiple issues but can also have disadvantages in other scenarios. VR developers have to define the interaction requirements of an application to find suitable interaction techniques [11]. A taxonomy for interaction techniques can guide the derivation of such requirements. Furthermore, a taxonomy can help to group techniques that share similar approaches to support users with the difficulties in VR interaction. The dimensions of the taxonomy can give a first hint on whether techniques tackle a problem or not.

In this section, we present an adapted version (see Figure 1) of the classification of Weise et al. [54]. The taxonomy is based on existing classifications and taxonomies, which we refer to in the corresponding subsection. They usually cover a small number of characteristics or consider a single type of interaction or specific properties. Therefore we decided to develop a comprehensive taxonomy that does not incorporate all possible characteristics of interaction techniques but is limited to the most important

Figure 1: Overview of the 13 dimensions of the taxonomy.

and meaningful aspects based on literature. There is no claim to completeness, and the taxonomy can be easily extended. For a more detailed delineation of the taxonomy, we refer to Weise et al. [55].

In comparison to the previews version, we moved *Degree of Freedom* to *Input Device Requirements* as a subcategory and reworked the *CD-Ratio* dimension. The resulting *Mapping* dimension breaks the used transfer function down for each possible base task. Furthermore, we removed the dimensions *Reference Frame* and *Spatial Compliance* as they had less additional value and added *Visual Feedback* and *Interaction Termination*, which are essential distinctive features of interaction techniques.

In the following, we present the dimensions of the taxonomy to be able to discuss how they relate to the already mentioned VR interaction issues. Table 1 visualizes the connections. If a cell contains an x, there is a possible influence of the corresponding dimension on a VR interaction issue. That means if techniques differ in a dimension, the severity of the connected issue may be different for the techniques. Vice versa, if there is no x, we could not find evidence for an influence of the dimension on the interaction issue in the literature.

2.2.1 Metaphor

The underlying real-world metaphor can give a first hint on how a technique works and what the possibilities and limitations are [32]. We reduced the classification from LaViola et al. [32] and differentiate between the three classes *grasping*, *pointing*, and *hybrid*. Grasping incorporates all techniques that share the metaphor of picking up an object by hand. The subclasses of grasping are *hand-based* and *finger-based* techniques. The well-known *Simple Vir-*

Table 1: Possible influences of the dimensions of the taxonomy on major issues in VR interaction.

VR interaction issues	Taxonomy dimensions												
	Metaphor	Executable Tasks	Input Device Requirements	Reach	Two-Handedness	Mapping	Transformation Separation	Visual Feedback	Interaction Termination	Directness	Disambiguation	Constraints	Interaction Fidelity
Natural Hand Tremor	x		x			x	x		x	x	x	x	
Imprecision	x		x		x	x	x		x	x	x	x	
Heisenberg Effect			x		x	x			x	x	x	x	
Gorilla-Arm-Effect	x				x	x	x						
Cognitive Workload		x	x		x	x	x	x		x	x	x	x
Eye-Hand Visibility Mismatch	x							x					
Midas Touch Effect									x				
Clutching						x							
Expressiveness	x	x	x	x	x								
Distance	x			x		x		x		x	x	x	
Object Size	x					x			x		x	x	
Moving Objects									x		x	x	
Density	x					x					x	x	
Occlusion											x	x	

tual Hand technique is a representative of the hand-based class, where a virtual proxy imitates the movements of the real hand. Usually, a button press grabs an object if the virtual hand touches it and enables the rotation and translation of the object. Finger-based techniques additionally utilize the fingers of the user. The pointing metaphor has the subclasses *vector-based* and *volume-based*. *Ray-Casting* [39] is a vector-based technique that is often used for selection but rarely for manipulation.

Volume-based techniques like *Flashlight* [34] use a selection volume. Here, a cone originates from the hand. Since multiple objects can fall into such selection volumes, a disambiguation mechanism is needed to ensure the selection of only one object (see Section 2.2.11). Hybrid techniques incorporate components of both grasping and pointing techniques. For example, *HOMER* [10] uses a ray for selection and a virtual hand for manipulation to combine the strengths of both.

The natural hand tremor negatively influences all interaction techniques but has a particularly strong effect on pointing techniques because the impact increases if the target is far away [16]. Similarly, the possible pointing precision decreases on greater distances [30]. On the other hand, pointing techniques are often used with a resting arm position. Therefore, they are less vulnerable to fatigue [26] in comparison to grasping techniques, which usually

incorporate more physical hand movements [4]. The eye-hand visibility mismatch is mainly a problem of pointing techniques as the selection point is closer to the position of the input device in grasping based techniques. Multiple studies showed that pointing techniques work better for selection on greater distances (e. g. [12]), whereas grasping techniques are better suitable for manipulation tasks improving their expressiveness [32]. Volume-based techniques can improve the selection of small objects as they cover a greater selection space but have problems with high-density environments because multiple objects fall into the selection volume [52]. For the same reason, volume-based techniques can support the selection of moving objects [24].

2.2.2 Executable Tasks

There is an infinite number of actions someone can do in virtual reality, like turning a crank or throwing an object. It is not possible to analyze interaction techniques regarding every possible task, and we need to focus on the most common ones. We use the classification by LaViola et al. [32] defining *selection*, *positioning*, *rotation*, and *scaling* as basic tasks. Some techniques can execute only one task, like the selection technique *Flashlight* [34]. *Spindle* [35], on

the other hand, is capable of all four task types. The center between the two hands serves as the selection point. Pressing a button on both controllers simultaneously grabs the object. Moving the hands in the same direction positions the object, rotating the hands around each other rotates the objects, and moving the hands apart or together scales the object. The executable tasks inherently influence the cognitive load and expressiveness of an interaction technique. The more tasks are supported, the more complex the technique gets, and the more powerful a technique is.

2.2.3 Input Device Requirements

It is essential to know the requirements of an interaction technique to ensure that it is supported by the VR system. We define the three subclasses *tracked body parts*, *Degrees of Freedom* (DoFs) and *1D Input*, to characterize the requirements on the input device. The tracked body parts can be *hand* (one or two), *forearm*, *upper arm*, *fingers*, *head*, or *eyes*. Most of the interaction techniques only rely on one or two hands. The *Push* technique [20] is one of the few techniques that use the forearm and the upper arm to detect whether the arm is stretched out, which triggers the selection. The category DoFs states if the position (*x*-, *y*-, and *z*-axis) and the rotation (*roll*, *pitch*, and *yaw*) of a body part need to be tracked. We differentiate between the minimum DoFs needed to be able to use a technique and the maximum DoFs supported by the technique. If an input device supports all maximum DoFs, the technique can be potentially more powerful. For example, it is possible to use the *Ray-Casting* technique by devices, which only support rotational tracking like the controller of the Oculus Go.¹ However, positional tracking additionally allows changing the origin of the ray, which gives the user better control. The 1D Input class defines whether and how many *buttons* or *scroll wheels* (or touchpads) are needed.

The natural hand tremor is inevitable if tracking of the hand is needed. Furthermore, Accot and Zhai [1] found out that the smaller the tracked muscle group, the higher the possible interaction precision. It is possible to avoid the Heisenberg effect if no buttons are needed, but the problem can also occur when finger-based gestures are used for the selection [58]. As already discussed, more buttons can also induce more cognitive load [8]. However, the number of used buttons and the tracked DoFs can give a hint on the expressiveness of a technique.

2.2.4 Reach

Reach is a category that can help developers to identify whether a technique is specialized on a distance, or it supports multiple distances. Mendes et al. [38] divide the reach of an interaction technique in *arm-length*, *scaled*, and *infinite*. Grasping-based techniques with isomorph mapping like the *Simple Virtual Hand* only allow interaction in the natural working space. Techniques with a scaled reach allow the interaction with objects up to some meters away from the user. For example, the *Go-Go* technique [45] extends the virtual arm when the user reaches a specified distance away from her body. Pointing techniques like *Ray-Casting* [39] theoretically have an infinite reach.

The reach of a technique inherently indicates on which distance it is usable. Respectively, techniques that can interact with objects on multiple distances are more expressive. Most of the techniques with a greater reach are prone to accuracy problems arising from the natural hand tremor, the human imprecision, and the Heisenberg effect because they use *Ray-Casting* or a scaled control-display ratio [30]. However, it is not possible to conclude that the reach is directly connected to these accuracy problems because several techniques can overcome these problems. For instance, the *Scaled Scrolling World in Miniature* technique [9] extends the original *World in Miniature* technique [39] by the possibilities to zoom into the world with the help of a scroll wheel in the secondary hand and to move the model by dragging it with the primary hand. These extensions allow the manipulation of distant objects comfortably from the working space of the user.

2.2.5 Two-Handedness

Adding the possibility to use both hands can increase the performance because the user can use her everyday experience on bimanual interaction [32]. Ulinski et al. [50] developed a categorization for bimanual interaction, which consists of the classes symmetric-synchronous, symmetric-asynchronous, asymmetric-synchronous, and asymmetric-asynchronous. In symmetric techniques, both hands do the same, whereas, in asymmetric techniques, both hands do different things. Synchronous and asynchronous refers to the simultaneousness of the actions of both hands. The previously described technique *Spindle* [35] is a symmetric-synchronous technique as both hands synchronously do the same. Lévesque et al. [33] developed an *Asymmetric Bimanual Gestural Interface*, which is asymmetric-asynchronous. The left hand determines the

¹ <https://www.oculus.com/go>

manipulation type (rotation, translation, or scaling), and the right hand executes the manipulation. It should be mentioned that a few interaction techniques do not need the hands at all like the *Head Based Selection* [40], which uses the view direction of the head and dwell time (see Section 2.2.9) as selection indication.

Asymmetric interaction techniques can produce less fatigue and cognitive load in comparison to symmetric bimanual techniques [51]. However, bimanual techniques, in general, can produce more fatigue than single-hand interaction because both hands have to be used [37]. Bossavit et al. [7] found out that bimanual interaction can overcome tracking jitter induced by the input device. Furthermore, using two hands increases the expressiveness of an interaction technique. For example, scaling is often implemented by using the distance between the two hands [8]. Putting the selection action and the manipulation action on different hands can also prevent the Heisenberg effect [15].

2.2.6 Mapping

One of the main tasks of an interaction technique is the mapping of the real movements tracked by the input device on movements in the virtual world. We consider the transfer functions for selection, positioning, rotation, and scaling separately. For selection, we refer to the mapping of the input on the selection tool. According to Mendes et al. [37], for selection, positioning, and rotation, we distinguish between *isomorph*, *scaled*, and *remapped* transfer functions. The mapping type for a scaling task can be *distance* or *remapped*. A technique may support multiple mapping types for a single task.

An isomorph interaction technique maps the real movements 1-to-1 on a virtual representation like the *Virtual Hand* technique. According to König et al. [30], scaled mapping can be further divided into *target-oriented*, *velocity-oriented*, and *manual switching*. We extended this classification by the *area-oriented* approach to cover techniques that use a predefined area around the user to adapt the mapping. For example, the *Go-Go* technique [45] uses an isomorph mapping near the user, but beyond this area, the reach extends non-linearly. Target-oriented techniques adapt the mapping according to the distance to the target, which simulates stickiness. A velocity-oriented mapping further reduces the movements on slow velocity for higher precision or increases the movements on higher velocity to, for example, reduce clutching. *Scaled HOMER* [55] adds such a velocity-oriented translation mapping to the *HOMER* technique [10] when manipulating an object.

Manual switching gives the user control of the control-display ratio. For instance, in the *ARM* technique, the user can press a button to switch between a normal and a precision mode. The execution of a task can also be remapped entirely by using a different transformation (e. g., using translation for rotation in the *Crank Handle* technique [7]) or buttons. Scaling techniques often use the distance between the hands to manipulate the size of an object. Some techniques incorporate multiple mappings. For example, *Go-Go + PRISM* [5] uses the velocity-oriented mapping of the *Go-Go* technique [45] to increase the range and adds the velocity-oriented mapping of *PRISM* [22] to allow higher precision on slow movements even on greater distances.

A scaled mapping that further slows down slow movements can increase the accuracy and therefore deals with problems like the natural hand tremor, imprecision, and the Heisenberg effect [30]. The increased precision also helps to select small objects [30]. On the other hand, reducing the control-display ratio can negatively affect the possible precision but enables the user to bridge greater distances [42] and can reduce clutching [32]. This can also have an effect on fatigue as the user can reduce the needed movements. A non-isomorph mapping can furthermore increase the cognitive load as the users need to understand the mapping [43] and may be irritated by the different positions of the real hand and the virtual cursor [30]. Manual switching between scaling modes can also increase the cognitive load [30]. Furthermore, target-oriented techniques do not work well in high-density environments as the areas of influence of the object interfere with each other [30].

2.2.7 Transformation Separation

Most of the available interaction techniques allow multiple transformations simultaneously. For example, they often support translation and rotation without switching between these two manipulation types. However, users frequently rotate and translate objects separately, and for some tasks, separated manipulation is more suitable [7]. The category transformation separation describes whether tasks can be executed separately and on which axes the manipulation is separated [37]. For example, the *Simple Virtual Hand* technique has no transformation separation as it simultaneously allows the translation and rotation of an object. In contrast, the *Asymmetric Bimanual Gestural Interface* [33] has three modes allowing translation and rotation simultaneously, only rotation or only scaling, and therefore it uses a partial transformation separation.

In the *Crank Handle* technique [7], the user can translate an object after she closes her hand. Opening and closing the hand again switches to the rotation mode, where rotations on the three primary axes are possible by moving the hand around the particular axis. Therefore, *Crank Handle* has a full transformation separation.

The possibility to manipulate objects on specific axes can increase precision [37]. Furthermore, in tasks where only one manipulation type is needed, transformation separation can simplify the task and therefore reduce the cognitive load [27]. However, transformation separation usually comes with the need to switch modes, which can also increase the cognitive load [37].

2.2.8 Visual Feedback

Feedback is an essential tool in human-computer interaction to inform the user about the current state of the interaction and the actions the user has to do to finish the task successfully. We concentrate on visual feedback because other types like haptic and auditive feedback are rarely considered so far in the design of 3D interaction techniques. Based on the analyzed interaction techniques, we consider the following classes: *3D cursor*, *target highlighting*, *adopting cursor*, *additional cursor*, *widgets*, and *proxy objects*. A technique can use multiple kinds of visual feedback. A 3D cursor is used by almost every interaction technique in the form of a virtual hand (or controller), a point cursor, a virtual ray, or a selection volume. For instance, *Head-based selection* [40] uses a point cursor indicating the gaze direction. Target highlighting can, for example, be done by changing the color or by using a shadow object. In selection tasks, target highlighting indicates typically the object that will be selected if the user triggers the selection. Cursors can adapt during interaction by, for instance, changing the size, color, or form or by bending to the target. Some techniques use additional cursor like a second ray or a second hand. *IntenSelect* [24] generates a second ray that snaps to the object the user probably wants to select based on heuristic calculations. (3D) widgets often provide supportive information or are used in the form of interactive menus. For example, in *Stretch Go-Go* [10], a gauge widget always indicates in which area the user has placed her hand, controlling the length of the arm. Proxy objects usually are miniature models of the real virtual objects with which the user interacts instead of the real objects. For instance, the *World In Miniature* technique [39] offers a miniature model of the real world in which the user can interact with the objects.

The changes to the proxy objects are mapped to the actual objects.

Providing feedback is mandatory as users are not able to interact with the virtual world without it [58]. Therefore, the given feedback positively influences the cognitive load. However, on easy tasks, additional feedback can even reduce performance by increasing the cognitive load due to additional information to process [57]. Furthermore, Argelaguet and Andujar [2] showed that adapting visual feedback can reduce the effect of eye-hand visibility mismatch. On greater distances, visual feedback can improve the selection of objects [46]. One would expect that feedback supports the interaction in challenging scenarios with, for example, small or occluded objects. However, multiple studies did not find significant effects of feedback like target highlighting (e. g. [23] or [47]). Further research is needed here.

2.2.9 Interaction Termination

There are multiple ways to indicate the selection and release of an object. For selection, Argelaguet Sanz [21] differentiates between *on button press*, *on button release*, *dwell*, and *gesture*. The analyzed techniques use the same mechanisms for the release of an object, so we decided to use the classification for both. Most of the techniques select an object the moment a button is pressed. However, the selection on button release could be less sensitive to the Heisenberg effect [21]. The release of objects is often done on the release of the button as the user holds down the button during manipulation. If dwell time is used, the user needs to place the cursor in a specific area for the specified amount of time. If this time is too short, the technique is prone to the Midas touch effect [3]. Furthermore, with dwell time, it is challenging to select small objects, and the natural hand tremor and imprecision can play a role as the cursor needs to be placed in a possibly small area for a longer time [49]. These problems can be transferred to moving objects. On the other hand, with dwell time, no Heisenberg effect can arise. Finger-based techniques typically use gestures to select or release objects like the closing/opening hand gesture. Unfortunately, gestures cause a positional change of the hand, which can influence the intended selection/release position [18].

2.2.10 Directness

LaViola et al. [32] define interaction techniques as either *direct* or *indirect*. Most of the grasping-based techniques

are direct as the user directly touches the virtual object. Indirect techniques use an intermediate or an additional tool to select or manipulate an object. If buttons are used to manipulate an object (for example, increasing or decreasing the size of an object with two buttons), the technique is also considered indirect. In addition to direct and indirect techniques, we define *semi-direct* techniques according to Jerald [29]. Semi-direct interaction also uses an intermediate but quickly feels like direct interaction. For instance, as already mentioned, the *World In Miniature* technique allows the manipulation of the virtual object in the miniature model. Therefore, there is no direct interaction with the objects, but manipulating the miniature objects feels direct.

Direct interaction techniques can only be used on short distances if they do not provide an appropriate mapping (see Section 2.2.6). Indirect techniques can select or manipulate objects at high distances. However, using an intermediate object to interact with an object always introduces an offset, which makes the technique prone to unprecise input. Furthermore, using buttons for manipulation tasks feels unnatural and therefore increases the cognitive load [8], but over time this type of control can increase the accuracy, especially on greater distances [8].

2.2.11 Disambiguation

Many interaction techniques use selection volumes, which improve the selection of small objects but comes along with potentially multiple selectable objects. To ensure that only one object is selected, disambiguation is needed. This involves the two components *progressive refinement* [38] and *disambiguation mechanism* [3]. Progressive refinement describes the process of reducing a set of potentially selectable objects until only the target object is left [38]. The progressive refinement strategy can be either *continuous* or *discrete*. For example, *Intend Driven Selection* [44] uses a scalable sphere to continuously reduce the number of objects until the target object is selected. The discrete refinement can be either done in a *single step* or with *multiple steps*. For instance, the *Flashlight* technique [34] selects the object which is closest to the center of the cone when the user presses the selection button and therefore uses a single step strategy. On the other hand, the *Expand* technique [19] uses a cone allowing the user to select multiple objects in the first phase. In the second phase, these objects are arranged as a grid in front of the user enabling the selection of the desired object.

The disambiguation mechanism describes how the technique detects the target object and can be *manual*, *heuristic*, or *behavioral* [3]. For example, during the second phase of the *Expand* technique, the user manually chooses the target object. The heuristic approach ranks the objects according to heuristic calculations and selects the objects with the highest rank. For instance, *Flashlight* uses the distance of the objects to the center of the ray to determine the ranks. The behavioral approach takes into account the actions of the user before the selection. For example, the pointing technique *IntenSelect* [24] uses an invisible cone and calculates scores for the objects according to the distance to the center of the cone and the retention time in the cone.

In general, disambiguation can increase selection performance in difficult environments with small, distant, or moving objects [24, 37]. Techniques with manual multi-step disambiguation can also work well in cluttered environments [31]. However, manual approaches can cause a higher cognitive load due to multiple steps the user has to execute [21]. Overall, techniques with disambiguation are less frail to imprecision due to the used disambiguation mechanisms [42].

2.2.12 Constraints

Constraints restrict the action space to simplify the interaction and to allow a higher precision [39]. We distinguish between the three constraint types *DoF reduction*, *snap to position*, and *snap to object*. DoF reduction ensures that in manipulation tasks, the position or rotation of an object only changes on one or two axes. For example, the *Knob* technique [18] allows only rotations around one axis, which derives from the hand movements. *Knob* also incorporates a snap to position constraint, which drags the object to a specific position. Here, the technique repositions the object to a slightly earlier position after release to overcome the Heisenberg Effect. Snap to object constraints are often used in selection techniques to connect the selection tool to a target object. For Instance, *IntenSelect* [24] uses a second bendable ray that snaps to the object with the highest score.

DoF reduction can increase the possible precision as it reduces the movements to the needed axis [32]. As already discussed, this often comes along with the requirement to change modes, which can lead to more cognitive load [37]. Snap to constraints can help for selection in difficult environments with small, distant, moving, cluttered, or occluded objects as the cursor can snap to the intended target [24].

2.2.13 Interaction Fidelity

The interaction fidelity describes how natural an interaction technique feels. Naturalism can improve the learnability and performance of an interaction technique, but at the same time, it can limit the possibilities of the technique [14]. McMahan [36] developed the Framework for Interaction Fidelity Analysis (FIFA), which consists of the three categories *biomechanical symmetry*, *control symmetry*, and *input veracity*. The categories are further divided into subcategories, which we will not go into detail in this article. The biomechanical symmetry describes how much the movements during the task match the movements during a comparable real-world interaction. The control symmetry characterizes how the interaction technique maps the actions in the real world on actions in the virtual world. The input veracity describes how precise the input devices capture and measure the user's actions. The input veracity is not considered in our classification as we analyze the techniques independently from specific input devices. An interaction technique with a high interaction fidelity is the finger-based *Virtual Hand*. The mapping of the finger and hand movements are isomorph and it corresponds to a real-world task. Less natural is a technique used in the application Engage,² which we named *Bimanual Fishing Reel + Scale*. After selecting an object with a ray it is attached to the ray which allows controlling the position of the object. Four buttons allow to change the distance and the size of the object. Additionally, while a button on the second controller is pressed its rotation is transferred to the rotation of the object.

Most of the presented categories influence the fidelity of an interaction technique. For example, a non-isomorphic mapping or the usage of disambiguation typically reduces the fidelity of a technique. Therefore the interaction fidelity is indirectly connected to all of the described problems. However, we can only directly connect fidelity to the cognitive load as the learnability and ease of use of a technique increases if the user knows a similar form of interaction from the real world [32].

3 Finding Suitable Interaction Techniques

The dimensions of the taxonomy described in the last section can give a first indication of whether an interac-

tion technique is suitable for an application scenario and which issues it tackles. However, for a reliable assessment of which advantages and disadvantages a technique has, more in-depth knowledge is required. In this section, we discuss a user study where we compare multiple selection techniques. Furthermore, we present a tool allowing developers to find interaction techniques according to the dimensions of the presented taxonomy and to suggest techniques based on the data of the study.

3.1 User Study

There are already several user studies comparing different interaction techniques. However, they can only compare a small number of techniques in specific scenarios and are often used to identify the advantages of newly developed techniques. To the best of our knowledge, only Bowman et al. [12] compared a high number of interaction techniques to explore the design space. Furthermore, we wanted to find indications on how the dimensions of the taxonomy influence the suitability of techniques for different application scenarios as well as usability and user experience aspects of the techniques. Therefore, we conducted an exploratory user study enabling us to compare a high number of techniques in various scenarios. To cover as many of the dimensions of the taxonomy as possible, we tried to ensure that there is at least one technique for each possible value of a dimension (and sub-dimension). However, that was not always possible. For example, in pretests, we observed a noticeable negative impact of input devices capable of tracking the fingers or the eyes on the performance of interaction techniques because of their unprecise tracking. To avoid such influences of the tracking devices, we decided to discard corresponding techniques. Therefore, there are no representatives for the corresponding values of the dimensions Metaphor (finger-based) and Tracked Body Parts (fingers and eyes). We divided the study into two sub-studies, where the participants either needed to select or manipulate an object. In the following, we will focus on the sub-study on the selection task. We are preparing another article for the manipulation sub-study.

Ten of the selected interaction techniques support the selection of objects: Bimanual Fishing Reel (BFR), Expand [19], Flashlight [34], Go-Go + PRISM (GP) [5], Head Based Selection (HBS) [40], IntenSelect [24], Scaled HOMER (SH) [55], Simple Virtual Hand (SVH), Spindle [35], and Scaled Scrolling World in Miniature (SSWIM) [9]. For reasons of space, we refrain from describing the techniques again and refer to Section 2.2. The implementation details of the

² <https://engagevr.io>

Figure 2: (left) Activation sphere starting the time on selection. (right) Arrangement of spheres with a green target sphere.

techniques can be found in the tool we present in Section 3.2. To cover a broad range of possible application scenarios we decided to use three of the most important and influential environment parameter [32, 53] as independent variables: distance (0.6, 3 or 6 m), size (15, 10 or 5 cm), and density (single object, 10 or 5 cm between objects). These variables also overlap with some of the interaction issues presented in Section 2.1. However, it is not possible to cover all issues.

3.1.1 Procedure

Before the experiments, each of the participants filled out a questionnaire asking for some personal information like gender, age, handedness, and experience with relevant technologies and applications. Then a document was handed out explaining the procedure and the following tasks. The selection task consists of two phases, which are visible in Figure 2. The task was developed to allow for a simple variation of the used independent variables and is based on tasks used in similar studies (e. g. [6]). A red sphere, an arm-length away from the user, had to be selected first to start the time measurement and to spawn an arrangement of one or multiple spheres where the target sphere was colored green. Objects hit by the selection tool of an interaction technique were highlighted by a yellow outline to increase the comparability of the techniques. The tasks were generated randomly once, and each participant executed the same tasks in random order. However, because two of the techniques do not support greater distances, we ensured that all short distance tasks came first. This ensures that the techniques are comparable on short distance and minders the effect of training. Furthermore,

we generated additional dummy tasks for the two techniques replacing the unsupported tasks to ensure that the same number of tasks were executed with each technique. For each possible variable combination, three tasks were generated. Accordingly, a participant executed 81 tasks per technique. For each task execution, we measured the time and the missed selections (accuracy), which are most relevant for the objective performance of a technique [8].

Each participant evaluated five of the ten interaction techniques. For each technique, the participant got a short explanation and had up to 5 minutes of training time. The time limit per task was 30 seconds. After finishing all the tasks with a single technique, the participant had to fill out questionnaires. We were interested in the usability of the interaction technique as well as user experience aspects. Therefore, the participants answered two standardized questionnaires on usability (System Usability Scale) [17] and workload (NASA Task Load Index) [25]. Additionally, we asked five custom questions regarding naturalness, fun, precision, speed, and motion sickness. The whole procedure was repeated until all five techniques were evaluated.

A well-known issue in user studies comparing interaction techniques is the speed-accuracy tradeoff. It is difficult for participants to optimize both speed and accuracy while solving a task. To tackle this problem, we used a similar approach as Wingrave and Bowman [57] to represent the real world speed-accuracy tradeoff. For each task, the participant got two scores (one for the speed and one for the number of misses) based on their performance in comparison to the other participants. The scores were summed up for each technique, and the participant with the highest score got a 10 € Amazon voucher. The participants were

Figure 3: Boxplots for the task completion times and bar charts for the average number of misses separated according to the used variables distance, density, and object size. No density means that there was only one object (the target object). The y-axis is logarithmically scaled.

informed about the procedure beforehand. However, the speed-accuracy tradeoff needs to be kept in mind when analyzing the results. The subjective ratings on speed and accuracy can be used for further verification.

Twenty participants (ten females, one left-handed, age between 18 and 42) took part in this experiment. Four participants never used VR, fourteen participants used VR less than 10 hours per year in the last three years, and two participants used VR less than 50 hours per year. The Hardware Setup consisted of an HTC Vive Pro and an Alienware 17 R4 (Intel i7-7700HQ, NVIDIA GTX 1070, 16 GB Ram). Unity was used to implement the test environment.

3.1.2 Discussion of the Results

Because of the exploratory nature of the study, we refrain from showing statistical significance in the data and focus on a descriptive analysis of the results. Therefore, this study can find indications for the advantages or disadvantages of techniques or dimensions of the taxonomy in specific scenarios. However, dedicated studies are necessary to verify the results. Figure 3 shows the measured times and the number of misses for each tested scenario separately. The results of the questionnaires are presented in

Figures 4 and 5. Motion sickness was no factor during the experiment and will not be further discussed. We also removed the question asking for the temporal affordance from the NASA questionnaire for the analysis of the results because we noticed that most of the participants did not understand the question correctly or referred to the temporal affordance of the task and not the technique. This does not affect the expressiveness of the NASA questionnaire but can limit the comparability with other studies where the questionnaire was used.

The hand-based techniques *SVH* and *Spindle* could only be tested for close range. The mean times of both techniques range among the fastest techniques, whereas *Spindle* is a little bit slower. However, on close range, the mean times of the best techniques mostly only differ less than 0.25 seconds. Therefore, it is hard to draw reliable conclusions. Both techniques show a small number of misses, but they continuously arise in contrast to other techniques. The participants often moved their hand back to early to prepare for the next task while pressing the trigger, resulting in misses. We assume that this happened because of the unsophisticated interaction allowed by the isomorph mapping, and the already mentioned speed-accuracy tradeoff. An earlier selection recognition or a simple disambiguation mechanism could help. However, the number of misses cannot lead to the conclusion

Figure 4: Boxplots for the results of the System Usability Scale (SuS) and the NASA-TLX and its sub-categories. The results are mapped on a 0–100 % scale. For the NASA-TLX and its sub-categories a lower value is better.

Figure 5: Results of the custom questions.

that the techniques are unprecise as the number is small, and the participants rated the techniques as rather precise. Accordingly, the speed and fun ratings for the SVH are one of the highest of all techniques as the isomorph transfer function leads to predictable movements. The participants also perceived the technique as very natural as it is based on a well-known metaphor from the real world. However, *Spindle* was rated less favorable in terms of fun, speed, and naturalism. Some participants noted that it felt unnatural to reach out with both hands only to select an object. This may also lead to the lower usability rating in comparison to most other techniques. In contrast, the usability rating of the single-handed SVH technique was rather good. The overall workload for SVH was low in comparison to most of the other techniques, while *Spindle* got a slightly higher rating. For both techniques, the physical demand and the effort were rated worse than for the best pointing techniques. This was expected as the participants had to move the hand(s) more often and extend the arm(s) more frequently in comparison to the pointing techniques. *Spindle* was even more demanding. Therefore, at least for selection tasks, we cannot confirm that the perceived fatigue is higher for the single-handed technique in comparison to the bimanual technique, as Bossavit et al. [7] found out for manipulation tasks.

IntenSelect and *Flashlight*, on average, accomplished the fastest times. Especially for single objects, the techniques are very fast as the user only needs to point roughly in the right direction to select the target. For single objects, there is no noticeable impact of the object distance and size. The object density only had a noticeable impact on the most challenging scenario (high distance, high density, and small object size). Here, because of the higher number of objects, small movements sometimes changed the predicted target shortly before the selection was triggered, causing slower times and more misses. This effect was also visible more often, with higher distance and object size. Therefore, the participants needed to move the cursor with more caution resulting in slightly slower times. The simple handling and learnability of the techniques lead to high usability values. The naturalism of the techniques was rated positively by only a few participants, as the secondary ray and the cone felt unusual. The perceived speed was rated very positive, which presumably lead to high ratings in terms of fun. The precision of *IntenSelect* was rated best among all techniques. Therefore, the secondary ray sticking to objects which cannot be controlled directly did not influence the perceived precision. However, *Flashlight* got less positive ratings here. A reason may be that it is hard to see which object is closest to the

center of the cone. An additional ray in the center may help. The workload was rated rather low for both techniques in comparison to the other techniques. Especially for the physical demand and the effort, this was expected, as the user can hold the hand in a comfortable position during the interaction.

Expand behaves like *Flashlight* for single objects and accordingly accomplishes similar fast times in these scenarios. This is also visible for short distances where it is possible to include only one object in the cone. If this is not possible anymore, the times get noticeable slower, as now two steps are necessary to select the object. This causes the slower times on mid and high distances and for small objects with high density on a short distance. However, for these cases, the times are relatively constant, and the object size and the density do have low impact. Interestingly, the participants had different approaches to solve the tasks. They either roughly pointed in the direction of the target object in the first phase accepting more unwanted objects in the second phase or tried to reduce the number of objects in the second phase by including as few objects as possible in the cone in the first phase. The second approach sometimes caused misses as the target object was often located at the edge of the cone. Because of the Heisenberg effect, sometimes, the target object was missing in the second phase. Overall the precision was rather good at the expense of the slower times confirming the results of other studies (for example [31]). This is also visible in the questionnaire results. The precision was rated higher, and the speed was rated lower in comparison to the *Flashlight* technique due to the two-phase progressive refinement. The participants confirmed the high precision and had fun using the technique. As expected, most of the participants rated the naturalism rather low as it is hard to find a corresponding real-world metaphor. The usability and the workload were perceived as similar positive as for the flashlight technique, which shows that the multi-step progressive refinement strategy had no significant negative influence.

SSWIM accomplished average results in comparison to the other techniques. Unexpectedly, the distance had a significant influence on the times, although the distances are much shorter in the miniature model. However, the initial sphere was always on a short distance, causing differences in the amounts of needed movement to reach the target object. Furthermore, it was possible to set a comfortable zoom level for the tasks on short distances as they all came first. For mid and long distances, the participants needed to scale down the environment increasing the difficulty of the task. The object size and density further negatively influence the execution time as they make

it harder to hit the object with the cursor. This also had an impact on the number of misses, which is low overall but increases especially for smaller objects and a higher density. Scaling was possible by moving the thumb up and down on the touchpad. Several participants found that rather cumbersome and tried to scale as little as possible. This is also visible in the usability rating, which is rather low in comparison to the other techniques. Similar to the *Spindle* technique, the participants said that the technique felt unusual for the selection of objects, and they had long practice times. This impacts the naturalness rating. Even though the technique is based on two well-known constructs from the real world (miniature models and grasping metaphor), the combination does not feel natural for the selection of objects for most of the participants. However, despite the lower usability values, the fun factor, precision, and speed were rated quite positive. It is also noticeable that the additional effort for scaling and moving the miniature model did not have a negative influence on these subjective ratings. However, that is not the case for the workload. Especially the mental demand was rated rather high because of the complexity of the technique. Furthermore, the physical demand and overall effort were rated higher in comparison to most of the other techniques as both hands needed to be used.

SH and *BFR* have very similar mean times. The only difference in the implementation of the techniques is that the ray of *BFR* gets transparent at the end. In simple scenarios, the techniques accomplish similar times than the best techniques. However, for mid and high distances, the negative impact of decreasing visual object size is noticeable. On high distances with small objects, the techniques accomplished the slowest times. Accordingly, there is a high number of misses in these scenarios. The reasons are the already mentioned natural tremor [30] and the Heisenberg effect [15]. The high variance in times originates from the different capabilities of the participants in keeping their hand calm. *BFR* accomplished even worse results as it is harder to align the transparent end of the ray with the target object. The usability of both techniques was rated primarily positive but worse than most of the other techniques and again with a high variance among the participants. Interestingly the precision and speed for *SH* were still rated positively by the majority of the participants. However, for *BFR*, the precision was rated rather conservative, presumably, due to the problematic ray representation. The problems in challenging scenarios also condense in less positive ratings for the perceived fun and naturalness. The workload was rated higher than for most of the

other techniques. Notably, physical demand, effort, frustration, and their performance was rated negatively due to the mentioned problems. Some participants said that it was fatiguing to hold the arm in a steady position over a long time in these scenarios. The mental demand for *SH* was quite high in comparison to *BFR*. We assume that the high frustration of the participants had an impact on this rating.

Throughout all scenarios, *HBS* is one of the slowest techniques. The object density did not influence the times. A higher distance did only matter in combination with smaller objects. A smaller visual size made it difficult for the participants to place the cursor on the object for the needed amount of time. This also resulted in a less positive rating for the perceived precision, although there are no misses in any scenario. The dwell time makes it hard to choose the wrong object. However, this results in slower selection times and less positive ratings for the perceived speed. The participants suggested a shorter dwell time, but this can lead to the already mentioned Midas Touch effect [28]. The naturalism was rated rather high even though no hands were used. Presumably, the technique was compared with focusing an object of interest with the eyes. The question related to the perceived fun was answered with some neutral scores, which corresponds to the comments of the participants on the long dwell time. The usability and workload range between the other techniques. Especially the overall effort was rated rather high.

GP is usually among the slowest techniques. The object size and density have a small influence on the times. However, with higher distances, the technique gets slower. The combination of velocity-based and area-oriented transfer function lead to unexpected movements of the hand, like König et al. [30] already described. This is also visible in the lower ratings for fun, naturalism, precision, and speed. Our results support the findings of another study comparing *GP* with the *Go-Go* technique, where the participants said they had less control compared to the *Go-Go* technique. We also noticed that the participants were mainly forced to align the visual feedbacks (moving the cursor to the object until it is highlighted) and were not able to rely on proprioception because of the non-isomorph mapping. The usability accordingly was rating as the lowest of all tested techniques. The workload was also rated as the worst of all techniques. Especially the effort and the physical demand was rated very high. Often the participants needed to stretch out their arm for a long time to place the cursor on the object leading to fatigue.

3.1.3 Conclusion

We compared a high number of interaction techniques and varied several environment parameters, which are crucial in real VR scenarios. We found indications for the advantages and disadvantages of techniques in specific scenarios and were able to show some of the relations between the dimensions of the taxonomy and major interaction issues presented in Section 2. However, the used task was artificial, which has to be kept in mind when transferring the results.

The volume-based techniques *Flashlight* and *IntenSelect* accomplished the best results regarding objective and subjective measurements due to the usage of heuristic and behavioral disambiguation mechanisms. The poor results in difficult scenarios of the pointing techniques *SH* and *BFR*, which work like *Ray Casting* for selection, also emphasize the importance of disambiguation. We observed the effect described by Bossavit et al. [31], where such immediate techniques rely on a single high-precision spatial selection with the consequence of higher execution times and a higher error probability. Bossavit et al. [31] alternatively suggest manual disambiguation techniques like *Expand*. In our experiment, *Expand* drastically reduced the number of misses because of the multi-step disambiguation at the expense of the needed time. Techniques relying on heuristic and behavioral disambiguation seem to be a good tradeoff but are prone to misses in challenging scenarios. We need further research on when it is useful to improve the speed and accuracy of basic techniques at the expense of their simplicity. *BFR* also shows the importance of proper feedback, as it is difficult to align a ray with a transparent end with small objects.

Unexpectedly, the described problems of the pointing techniques lead to a high perceived workload as it was fatiguing, trying to place the ray on the object precisely. The higher physical demand for the hand-based techniques, on the other hand, was expected. The bimanual techniques *Spindle* and *SSWIM* were even more fatiguing. Furthermore, these two techniques got lower usability ratings compared to the other techniques, as they felt unusual for a selection task. Further research is needed on whether the reason for this is the usage of two hands.

GP, which incorporates velocity-oriented and area-oriented mapping, was the worst technique in terms of objective as well as subjective measurements due to the unexpected behavior of the cursor. This cannot be transferred to all techniques using one or multiple non-isomorph mappings. However, the mapping needs do be comprehensible, and considering the results of techniques with an iso-

Figure 6: Main page of S3DIT (Suggestions for 3D User Interfaces).

morph mapping, this demonstrates the advantage of proprioceptive feedback [41].

We could not find an influence of the perceived naturalism on the usability or the measured times. Techniques with lower ratings for naturalism like *IntenSelect* and *Expand* accomplished similar good results than the most natural technique *SVH*.

Head-based selection, in combination with dwell time, did not work well in our test scenario. Lower dwell time can counter the slow selection times but need to account for the Midas Touch effect. The use of a disambiguation mechanism could also avoid the need for precise placement of the cursor on the object. However, we could not find evidence that the head is less convenient for the selection of objects than techniques relying on the hands.

3.2 Suggesting Appropriate Interaction Techniques

The study shows that interaction techniques differ in terms of usability, user experience, and their suitability for specific application scenarios. Developers of VR applications need to be able to identify appropriate techniques. To support this decision, we developed a web-based interactive tool that can rank the evaluated techniques according to a

given scenario called S3DIT³ (Suggestions for 3D User Interfaces). However, the main goal of the tool is to allow VR developers to search through a comprehensive filterable database of interaction techniques. All over 110 techniques considered in this article are available. Figure 6 shows the main page of the tool. Here, the user can filter the techniques based on the dimensions of the taxonomy presented in Section 2.2. For example, it is possible to display only techniques that work on greater distances or to show only single-handed techniques. This allows matching the techniques with the demands of the target VR system and the application scenario. However, the presented techniques should not be taken as complete and implemented as they are. Instead, the techniques should be considered as an impulse on how to solve specific issues arising in the interaction in VR. Subcomponents of the techniques may be suitable for the application and should be carefully integrated with existing interaction forms of the application. The details of each interaction technique are visible on a separate page. Here a description of the techniques functionalities is given, and the technique is classified according to the dimensions of the taxonomy. Additionally, images visualize the technique. If the technique was evaluated in the presented study, implementation details and the study results are listed.

³ <https://s3dit.cs.uni-potsdam.de>

Under the tab Suggestions, the techniques evaluated in the studies (see Section 3.1) can be sorted considering a given scenario or subjective measurements. To get suggestions for a scenario, the user needs to execute two steps of a wizard. In the first step, the user chooses the target task, which can either be selection or manipulation. Furthermore, she decides whether she wants to rank the techniques according to objective or subjective measurements. If the first option was chosen, the user defines a scenario in the second step. For selection tasks, it is possible to select the distance, object size, and density and for manipulation, the subtask, distance, and manipulation amount. Furthermore, the user can prioritize speed or precision. If the subjective option was chosen, the user can decide whether she wants to rank the techniques concerning the usability, naturalness, fun, precision, speed, motion sickness, workload, mental demands, physical demands, performance, effort, or frustration.

4 Summary and Further Research

In this article, we discussed common problems related to interaction with objects in VR and associated these problems with the dimensions of a taxonomy for interaction techniques. We discussed the effects of the dimensions of the taxonomy based on a user study where we compared several interaction techniques in terms of their performance in different scenarios and according to usability and user experience aspects. We presented a tool that allows to find techniques considering the dimensions of the presented taxonomy, and that can recommend suitable techniques based on the data of the presented user study.

The next steps include the field evaluation of S3DIT. We plan to carry out a qualitative evaluation in which we conduct interviews with experts. These experts should be composed of VR developers who will use the tool to find suitable interaction techniques for an application to be developed. The evaluation should target whether a recommendation of techniques based on limited user studies is reasonable and whether all dimensions of the taxonomy are useful for developers when searching for interaction techniques. A possible outcome could be that only a few basic dimensions are important to find suitable techniques, and every additional dimension unnecessarily increases the complexity. The tool could also be extended to guide developers without suggesting specific interaction techniques. Existing guidelines and results of other

studies could be consolidated to give more general hints on what needs to be considered in the design of user interfaces for specific application scenarios. Furthermore, in the future, dedicated user studies should be conducted to verify the results of the presented study.

References

- [1] Johnny Accot and Shumin Zhai, Scale effects in steering law tasks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 2001, pp. 1–8. doi:10.1145/365024.365027.
- [2] Ferran Argelaguet and Carlos Andujar, Efficient 3D pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6), 34–43, 2009.
- [3] Ferran Argelaguet and Carlos Andujar, A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3), 121–136, 2013.
- [4] Ferran Argelaguet, Carlos Andujar and Ramon Trueba, Overcoming eye-hand visibility mismatch in 3D pointing selection. In: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*. ACM, New York, NY, 2008, p. 43. doi:10.1145/1450579.1450588.
- [5] Chris Auteri, Mark Guerra and Scott Frees, Increasing precision for extended reach 3D manipulation. *International Journal of Virtual Reality (IJVR)*, 12(1), 66–73, 2015.
- [6] Marc Baloup, Thomas Pietrzak and Géry Casiez, RayCursor: a 3D pointing facilitation technique based on raycasting. In: *2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2019, pp. 1–12. doi:10.1145/3290605.3300331.
- [7] Benoît Bossavit, Asier Marzo, Oscar Ardaiz, Luis D. de Cerio and Alfredo Pina, Design choices and their implications for 3D mid-air manipulation techniques. *Presence: Teleoperators and Virtual Environments*, 23(4), 377–392, 2014.
- [8] Doug Bowman, Principles for the design of performance-oriented interaction techniques. In: *Handbook of Virtual Environments*, 2002.
- [9] Doug A. Bowman, Jian Chen, Chadwick A. Wingrave, John F. Lucas, Andrew Ray, Nicholas F. Polys, Qing Li, Yonca Hacıahmetoglu, Ji-Sun Kim, Seonho Kim, Robert Boehringer and Tao Ni, New directions in 3D user interfaces. *IJVR*, 5(2), 3–14, 2006.
- [10] Doug A. Bowman and Larry F. Hodges, An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: A. van Dam (Ed.), *Proceedings of the 1997 Symposium on Interactive 3D Graphics*. ACM, New York, NY, 1997.
- [11] Doug A. Bowman and Larry F. Hodges, Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, 10(1), 37–53, 1999.
- [12] Doug A. Bowman, Donald B. Johnson and Larry F. Hodges, Testbed evaluation of virtual environment interaction techniques. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. ACM, New York, NY, 1999, pp. 26–33. doi:10.1145/323663.323667.

- [13] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola and Ivan Poupyrev, An introduction to 3-D user interface design. *Presence: Teleoperators and Virtual Environments*, 10(1), 96–108, 2001.
- [14] Doug A. Bowman, Ryan P. McMahan and Eric D. Ragan, Questioning naturalism in 3D user interfaces. *Communications of the ACM*, 55(9), 78–88, 2012.
- [15] Doug A. Bowman, Chadwick A. Wingrave, Joshua M. Campbell, V. Q. Ly and C. J. Rhoton, Novel uses of pinch gloves™ for virtual environment interaction techniques. *Virtual Reality*, 6(3), 122–129, 2002.
- [16] Douglas A. Bowman, 1999. *Interaction techniques for common tasks in immersive virtual environments*, Citeseer.
- [17] John Brooke, SUS – a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7, 1996.
- [18] Fabio M. Caputo, Marco Emporio and Andrea Giachetti, Single-handed vs. two handed manipulation in virtual reality: a novel metaphor and experimental comparisons. In: *Proceedings of the Conference on Smart Tools and Applications in Computer Graphics, STAG '17*. Eurographics Association, Goslar Germany, Germany, 2017, pp. 39–45. doi:10.2312/STAG.20171225.
- [19] Jeffrey Cashion, Chadwick Wingrave and Joseph J. LaViola, Dense and dynamic 3D selection for game-based virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 634–642, 2012.
- [20] Peter Dam, Priscilla Braz and Alberto Raposo, A study of navigation and selection techniques in virtual environments using Microsoft Kinect®. In: D. Hutchison, T. Kanade and J. Kittler (Eds.) *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments. Proceedings of the International Conference on Virtual, Augmented and Mixed Reality 2013*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin/Heidelberg, 2013, pp. 139–148. doi:10.1007/978-3-642-39405-8_17.
- [21] Fernando Argelaguet Sanz, Informàtica Gràfica, *Pointing Facilitation Techniques for 3D Object Selection in Virtual Environments*, Doctoral Thesis, 2011.
- [22] Scott Frees, G. D. Kessler and Edwin Kay, PRISM interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 14, 2-es, 2007.
- [23] Tovi Grossman and Ravin Balakrishnan, The design and evaluation of selection techniques for 3D volumetric displays. In: *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, UIST '06*. ACM, New York, NY, USA, 2006, pp. 3–12. doi:10.1145/1166253.1166257.
- [24] Gerwin d. Haan, Michal Koutek and Frits H. Post, IntenSelect: using dynamic object rating for assisting 3D object selection. In: E. Kjems (Ed.) *Proceedings of the 11th Eurographics Conference on Virtual Environments*. Eurographics Association, Aire-la-Ville, Switzerland, 2005. doi:10.2312/EGVE/IPT_EGVE2005/201-209.
- [25] Sandra G. Hart and Lowell E. Staveland, Development of NASA-TLX (task load index): results of empirical and theoretical research. In: P. A. Hancock and N. Meshkati (Eds.) *Advances in Psychology: Human Mental Workload*. North-Holland, 1988, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9.
- [26] Juan D. Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian and Pourang Irani, Consumed endurance. A metric to quantify arm fatigue of mid-air interactions. In: *32nd Annual ACM Conference on Human Factors in Computing Systems*. Assoc. for Computing Machinery, New York, NY, 2014, pp. 1063–1072. doi:10.1145/2556288.2557130.
- [27] Ken Hinckley, Randy Pausch, John C. Goble and Neal F. Kassell, A survey of design issues in spatial input. In: P. Szekely (Ed.) *Proceedings of the ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, 1994, pp. 213–222. doi:10.1145/192426.192501.
- [28] Robert J. K. Jacob, John J. Leggett, Brad A. Myers and Randy Pausch, Interaction styles and input/output devices. *Behaviour & Information Technology*, 12(2), 69–79, 1993.
- [29] Jerald, Jason, *The VR Book: Human-Centered Design for Virtual Reality*. ACM Books, vol. 8. Morgan & Claypool, San Rafael, 2016.
- [30] Werner A. König, Jens Gerken, Stefan Dierdorf and Harald Reiterer, Adaptive pointing – design and evaluation of a precision enhancing technique for absolute pointing devices. In: T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates and M. Winckler (Eds.) *Human-Computer Interaction – INTERACT 2009. 12th IFIP TC 13 International Conference*. Lecture Notes in Computer Science, vol. 5726. Springer, Berlin, 2009, pp. 658–671. doi:10.1007/978-3-642-03655-2_73.
- [31] Regis Kopper, Felipe Bacim and Doug A. Bowman, Rapid and accurate 3D selection by progressive refinement. In: *IEEE Symposium on 3D User Interfaces (3DUI), 2011*. IEEE, Piscataway, NJ, 2011, pp. 67–74. doi:10.1109/3DUI.2011.5759219.
- [32] LaViola, Joseph J., Kruijff, Ernst, McMahan, Ryan P., Bowman, Doug, and Poupyrev, Ivan P., *3D User Interfaces. Theory and Practice*. Addison-Wesley Professional, 2017.
- [33] Julien-Charles Lévesque, Denis Laurendeau and Marielle Mokhtari, An asymmetric bimanual gestural interface for immersive virtual environments. In: R. Shumaker (Ed.) *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments. 5th International Conference, VAMR 2013*. Lecture Notes in Computer Science, vol. 8021. Springer, Berlin-Heidelberg, 2013.
- [34] Jiandong Liang and Mark Green, JDCAD. A highly interactive 3D modeling system. *Computers & Graphics*, 18(4), 499–506, 1994.
- [35] Daniel P. Mapes and J. M. Moshell, A two-handed interface for object manipulation in virtual environments. *Presence: Teleoperators and Virtual Environments*, 4(4), 403–416, 1995.
- [36] Ryan P. McMahan, *Exploring the effects of higher-fidelity display and interaction for virtual reality games*, 2011.
- [37] Daniel Mendes, Ariel Caputo, Andrea Giachetti, Alfredo Ferreira and Joaquim A. P. Jorge, A survey on 3D virtual object manipulation: from the desktop to immersive virtual environments. *Computer Graphics Forum*, 38(1), 21–45, 2019.
- [38] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Eduardo Cordeiro, Alfredo Ferreira and Joaquim A. Jorge, Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. *Computers & Graphics*, 67, 95–102, 2017.
- [39] Mark Mine, ISAAC: A virtual environment tool for the interactive construction of virtual worlds. *UNC Chapel Hill Computer Science Technical Report TR95-020*, 1995.

- [40] Mark R. Mine, Virtual environment interaction techniques. *UNC Chapel Hill computer science technical report TR95-018*, 1995.
- [41] Mark R. Mine, Frederick P. Brooks Jr. and Carlo H. Sequin, Moving objects in space: exploiting proprioception in virtual-environment interaction. In: G. S. Owen (Ed.) *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 19–26. doi:10.1145/258734.258747.
- [42] Alec G. Moore, John G. Hatch, Stephen Kuehl and Ryan P. McMahan, VOTE. A ray-casting study of vote-oriented technique enhancements. *International Journal of Human-Computer Studies*, 120, 36–48, 2018.
- [43] Noritaka Osawa, Automatic adjustments for efficient and precise positioning and release of virtual objects. In: *Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications*. ACM, New York, NY, 2006, p. 121. doi:10.1145/1128923.1128943.
- [44] Frol Periverzov and Horea Ilies, IDS. The intent driven selection method for natural user interfaces. In: *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, 23–24 March 2015, Arles, France. IEEE, Piscataway, NJ, 2015, pp. 121–128. doi:10.1109/3DUI.2015.7131736.
- [45] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst and Tadao Ichikawa, The go-go interaction technique: non-linear mapping for direct manipulation in VR. In: *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, 1996, pp. 79–80. doi:10.1145/237091.237102.
- [46] Ivan Poupyrev and Tadao Ichikawa, Manipulating objects in virtual worlds. Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1), 19–35, 1999.
- [47] Ivan Poupyrev, Tadao Ichikawa, Suzane Weghorst and Mark Billinghurst, Egocentric object manipulation in virtual environments. Empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17(3), 41–52, 1998.
- [48] Ivan Poupyrev and Ernst Kruijff, 2000. *20th Century 3DUI Bib. Annotated Bibliography of 3D User Interfaces of the 20th Century*.
- [49] Anthony Steed, Towards a general model for selection in virtual environments. In: *3DUI 2006. IEEE Symposium on 3D User Interfaces 2006: proceedings*, Alexandria, Virginia, USA, March 25–26, 2006. IEEE, Piscataway, NJ, 2006, pp. 103–110. doi:10.1109/VR.2006.134.
- [50] Amy C. Ulinski, Zachary Wartell, Paula Goolkasian, Evan A. Suma and Larry F. Hodges, Selection performance based on classes of bimanual actions. In: *IEEE Symposium on 3D User Interfaces, 2009. 3DUI 2009*, Lafayette, Louisiana, USA, March 14–15, 2009. IEEE, Piscataway, NJ, 2009, pp. 51–58. doi:10.1109/3DUI.2009.4811205.
- [51] Amy C. Ulinski, Catherine Zambaka, Zachary Wartell, Paula Goolkasian and Larry F. Hodges, Two handed selection techniques for volumetric data. In: *Proceedings/3DUI, IEEE Symposium on 3D User Interfaces 2007*, Charlotte, North Carolina, USA, March 10–11, 2007. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2007. doi:10.1109/3DUI.2007.340782.
- [52] Lode Vanackén, Tovi Grossman and Karin Coninx, Exploring the effects of environment density and target visibility on object selection in 3D virtual environments. In: *Proceedings/3DUI, IEEE Symposium on 3D User Interfaces 2007*, Charlotte, North Carolina, USA, March 10–11, 2007. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2007. doi:10.1109/3DUI.2007.340783.
- [53] Matthias Weise and Raphael Zender, A systematic literature review on independent variables used in user studies on 3D selection and manipulation techniques. In: P. Grimm, Y. Jung and R. Dörner (Eds.) *16. Workshop der GI-Fachgruppe VR/AR. Virtuelle und erweiterte Realität*, 2019.
- [54] Matthias Weise, Raphael Zender and Ulrike Lucke, A comprehensive classification of 3D selection and manipulation techniques. In: *Mensch und Computer 2019*. ACM, New York, NY, USA, 2019.
- [55] Curtis Wilkes and Doug A. Bowman, Advantages of velocity-based scaling for distant 3D manipulation. In: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*. ACM, New York, NY, 2008, p. 23. doi:10.1145/1450579.1450585.
- [56] Chadwick A. Wingrave, 3DUI flavors beyond vanilla. In: *SEARIS 2009 2th Workshop on IEEE VR*, 2009.
- [57] Chadwick A. Wingrave and Doug A. Bowman, Baseline factors for raycasting selection. In: *Proceedings of Virtual Reality International*, 2005.
- [58] Chadwick A. Wingrave, Doug A. Bowman and Naren Ramakrishnan, Towards preferences in virtual environment interfaces. In: W. Stürzlinger and S. Müller (Eds.) *Proceedings of the Workshop on Virtual Environments 2002*, Barcelona, Spain, May 30–31, 2002. Association for Computing Machinery, New York, 2002.
- [59] Shumin Zhai, Paul Milgram and William Buxton, The influence of muscle groups on performance of multiple degree-of-freedom input. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1996, pp. 308–315. doi:10.1145/238386.238534.

Bionotes

Matthias Weise

University of Potsdam, Potsdam, Germany
matthias.weise@uni-potsdam.de

Matthias Weise studied Computational Science at the University of Potsdam. He is working at the chair for Complex Multimedia Application Architectures at the University of Potsdam where he is involved in a project on action-oriented learning with a Virtual Reality painting simulator. His main research goal is the improvement of interaction in Virtual Reality. He is also spokesman of the working group VR/AR-Learning of the Gesellschaft für Informatik.

Raphael Zender
University of Potsdam, Potsdam, Germany
raphael.zender@uni-potsdam.de

Dr. Raphael Zender studied Information Technology/Computer Engineering and obtained his doctorate for innovative e-learning infrastructures. As a research associate at the chair for Complex Multimedia Application Architectures at the University of Potsdam, he performs various tasks in the field of teaching and research. He is currently researching the use of virtual reality in education. He is also spokesman of the working group VR/AR-Learning of the Gesellschaft für Informatik.

Ulrike Lucke
University of Potsdam, Potsdam, Germany
ulrike.lucke@uni-potsdam.de

Prof. Dr.-Ing. habil. Ulrike Lucke holds the chair for Complex Multimedia Application Architectures at the University of Potsdam since 2010. She studied computer science at the University of Rostock and also completed her PhD and habilitation in this discipline. Her areas of scientific interest include complex applications and infrastructures for media-based education and research. In the German Informatics society she was speaker of the SIG E-Learning from 2008 to 2014, and she is active as a member of the executive board since 2014 and of the SIG E-Science since 2015. Moreover, she was Chief Information Officer of the University of Potsdam from 2010 to 2018 and is vice chair-person of the German association of CIOs at universities.