From Research to Practice: Survey and Taxonomy of Object Selection in Consumer VR Applications

Mykola Maslych [*]
University of Central Florida
Difeng Yu [†]
University of Copenhagen
Amirpouya Ghasemaghaei
University of Central Florida
Yahya Hmaiti
University of Central Florida
Esteban Segarra Martinez
University of Central Florida
Dominic Simon
University of Central Florida
Eugene M. Taranta II
University of Central Florida
Joanna Bergström [‡]
University of Copenhagen
Joseph J. LaViola Jr. [§]
University of Central Florida

Abstract

Object selection has been explored extensively in the VR research literature. However, the research is typically conducted in constrained experimental setups. It remains unclear whether the designed selection techniques fit the prevalent practical uses and whether the experimental tasks represent important challenges in real applications. To identify and help bridge these gaps, we surveyed current consumer VR applications, containing 206 popular VR game and 3D modeling applications. We extracted 1300+ selection scenarios based on video analyses of these applications and derived a taxonomy to understand common patterns on *where* and *how* selections occur. Our findings reveal significant gaps in selection tasks and techniques between research and consumer applications. We also present an interactive visualization tool to help researchers explore the VR object selection scenarios. Finally, we discuss how our work can help researchers and developers evaluate techniques in meaningful tasks and drive the design of techniques.

Keywords: 3D user interfaces, consumer applications, database, object selection, target selection, video games, virtual reality.

1 Introduction

Object selection is a fundamental interaction in virtual reality (VR). It is also the most common interaction and works as a basis for other interactions in VR. For example, object selection often precedes manipulation, such as when selecting an object to translate it to another location, and initiates system commands, like selecting virtual buttons for menu navigation. The research literature is burgeoning new selection techniques in VR and task scenarios that evaluate them. However, we lack an overview of the *where* object selection occurs in consumer VR applications and *how* techniques are used in these applications. This has two undesirable outcomes for research and practice.

First, assessing the potential impact of new object selection techniques is difficult as the evaluations of the techniques in research are typically conducted in constrained experimental setups. Bergström et al. [7] reviewed two decades of research in VR selection and manipulation studies. They found that the diversity of the tasks used in the studies hampers direct comparison of the techniques and thereby the accumulation of knowledge about their strengths and weaknesses. This suggests that the study setups might be biased for testing the specific advantages that the novel technique designs provide. And even if the field would have standardized tests, it is unclear how the tasks would generalize to, and represent the tasks of the applications in practice.

Second, accumulating knowledge of good designs is bounded in our respective domains. The lack of overlap between consumer and research spheres is highlighted by a recent paper by Steed et al. [45], that discusses the novel trends in 3D user interface (3DUI) design across a variety of contemporary VR applications. For example, the research designs are rarely adapted in consumer applications, and even if they would, it is unclear if they would improve object selection in them, because the tasks in practice might not leverage the particular strengths of the state-of-the-art techniques in research.

The purpose of this work is to identify and thus help bridge the gaps between research and practice on VR object selection. We first created a dataset of 1300+ selection scenarios across 206 contemporary consumer VR game and 3D modeling applications based on video analyses. We then iteratively defined two taxonomies of 'where' (the tasks) and 'how' (the techniques) selections occur in these scenarios. The taxonomies created through this bottom-up approach present new categories that were previously overlooked in research. We then analyzed the frequencies of different categories of object selection tasks and techniques and discussed how VR research can be inspired by our findings from consumer applications. Finally, we provide an interactive visualization tool for exploring the dataset.

This work can help bridge the gap from research to practice through the following contributions:

- Taxonomies of where and how selection occurs in consumer VR applications. The taxonomies provide a systematic overview of object selection tasks and techniques in applications. These can help to consider the validity of evaluations (e.g., generalizability and representativeness of tasks) and drive ideas for design (e.g., motivating new attributes of the techniques in practice).
- An initial analysis of the collected dataset which can help identify commonalities but also to direct attention to fringe cases in both research and practice.
- An open-source dataset of 1300+ interaction scenarios from 206 consumer VR applications which can be explored with a web-based visualization tool. For practitioners, it offers a tool to view other applications. For researchers and students, it provides a resource to explore consumer space developments and guide new techniques and study methods.

2 Related Work

This section presents an overview of 'where' selection occurs (i.e., the task) and 'how' selection occurs (i.e., the technique) within the research literature. We also refer the readers to several taxonomies and literature reviews for a more comprehensive landscape about VR object selection tasks and techniques in research [2,7,27,51,58], while emphasizing the need to construct new overviews of those in consumer VR applications.

[*]e-mail: maslychm@gmail.com
[†]e-mail: diyu@di.ku.dk
[‡]e-mail: joanna@di.ku.dk
[§]e-mail: jlaviola@ucf.edu

2.1 VR selection tasks

The VR research community uses a vast amount of object selection tasks to evaluate interaction techniques and to gain insights into their design [7,58]. Some studies emphasize accumulation of knowledge and comparison of the results across studies by using more controlled or even standardized task setups. Others emphasize representativeness of the tasks in expected applications of VR to inform practice.

One predominant, controlled task paradigm is based on an extension of the ISO 9241 standard, also known as the Fitts's ring [42]. Typically in that, the users need to select multiple targets (mostly spheres) distributed on a circular ring sequentially in a pre-defined order. Such a task design facilitates the precise control of experimental variables, including factors like target distance and width. It also considers the impact of target direction on the selection performance results, given that users need to traverse through the entire circular layout [2, 7]. Similar task configurations may incorporate reciprocal selections between targets [24, 60]. Although the ISO 9421 can be used in 3D VR, it is a 2D task. Some 3D spiral-shaped versions exist, but are not standardized (e.g., as used in [29, 34] to incorporate depth). Many selections in VR are truly 3D, and in those 2D tasks that exist (e.g., menus), fast sequential selection is rarely a represented feature. An exception to this is the selection occurring on a QWERTY keyboard layout (e.g., [43,55]).

Another, often highly controlled and abstract task setup is to arrange the targets within a predetermined 3D space in front of the user. One common approach is to place the targets in a grid layout [32, 59]. Another is to randomize the targets within the defined space. For example, targets could be randomly positioned within the user's arm reach [47,54] or dispersed at a greater distance in a larger space [6, 39]. In these scenarios, the targets could also be moving [11,26] or get occluded by other objects [33].

In addition to the more abstract task settings, the research community has also investigated the practicability of the selection techniques in more realistic application scenarios. The users may need to acquire targets located on a shelf [16, 54], in a molecule structure [35, 54], and within a complex 3D car engine model [3]. The shapes of targets can also vary widely, ranging from basic geometric shapes (e.g., spheres, cubes, pyramids) to complex 3D objects like a 3D bunny model [61], buildings in a VR city [41], industrial pipes found in factories [49], and everyday items such as chairs [50].

In contrast to the preceding categories, where targets are typically positioned within the virtual environment (VE), some tasks require users to select targets that are anchored around/on their own virtual body. For instance, users may be required to acquire targets that are positioned above their wrist [30, 56], attached to their body surfaces [4, 63], and distributed around their body [57, 65].

These types of task setups suggest that the VR community has striven to both establish controlled and standard tasks [7, 58], and to develop task scenarios representative of actual VR applications, in order to evaluate selection techniques. Furthermore, their underlying characteristics have been captured in many taxonomies for object selection tasks (e.g., [15, 27]).

However, it is unclear how the controlled tasks characteristics generalize to application tasks, and how representative the practice-oriented setups are in VR applications. It is also unclear whether the existing taxonomies reflect a similar set of attributes that occur in current VR applications. This paper aims to help in uncovering these connections and through that to design better task setups in VR studies. We do that by giving an overview of the tasks in a form of a taxonomy, that is, 'where' the selection occurs when deploying the techniques in consumer VR applications.

2.2 VR selection techniques

Selection techniques in VR typically fall into two primary categories [2]. The *virtual hand* technique creates a virtual replica of a user's physical hand in VR to interact with nearby objects. The *virtual pointing* technique utilizes a ray projected from the user's hand position into the environment to select targets that may be located at a distance. The research community has proposed a range of techniques aimed at enhancing the usability of both these methods. We highlight a few examples to provide readers with a sense of the scope and diversity in this field [2, 27, 58].

Since the virtual hand technique often limits users to select objects within arm's reach, methods have been developed to address this constraint. For instance, techniques like Go-Go [36] and PRISM [19] enable users to extend their effective arm reach by scaling the speed or distance of their movements. Users may also extend their reach with a very long virtual arm [34], a virtual ray accompanied by a 3D cursor [6], with a portal [1, 23], or by employing multiple virtual hands distributed throughout the entire VE [39].

Input inaccuracies resulting from hand tremors during 3D object selection are also a notable concern. Consequently, numerous techniques have been developed to mitigate this problem. A technique may involve dynamically altering the size of the cursor or highlighting the closest target to more accurately capture the user's intended selection [6, 32]. Computational models can also play a role in predicting a user's target of interest [24, 60].

Virtual hand and virtual pointing techniques have also been adapted to address more complex selection tasks. For example, users can employ a volume-based cursor to select a group of objects simultaneously [46, 54]. They can iteratively refine their choices or utilize a mini-map (i.e., proximity of the VE) to facilitate the selection of an object that might be occluded from the view [5, 33, 64]. Crossing-based selection techniques, wherein users can pinpoint a target by crossing its boundaries, have also been investigated in the research literature [48]. Furthermore, a user may utilize gaze input along with hand input for target selection [41].

These examples highlight how the techniques are designed for particular goals in the task space, and selection techniques have also been characterized in many taxonomies, often related to the tasks (e.g., [2, 8, 37, 38, 52]). Although a large number of selection techniques are designed and introduced in research every year, it is unclear if the techniques have found practical usage in consumer VR applications and whether the developed techniques fulfill the needs of such applications. This paper aims to help in answering these questions by providing an overview of the techniques in a form of another taxonomy, that is, 'how' the selection occurs in consumer VR applications.

3 Scenario Identification and Classification

We define the scope of this research as follows. We consider object selection as "the task of acquiring or identifying a particular object or subset of objects from the entire set of objects available" [27]. We do not consider object manipulation (e.g., translation, rotation, or scaling of an object) or navigation (e.g., teleportation that would transport the user's viewpoint). The spatial position/orientation of the input device should also be involved in the selection process. Lastly, we focus on first-person view HMD-based VR applications.

As an initial step of understanding *where* object selection occurs in consumer VR applications and *how* different techniques have been applied, we identified 1300+ applications scenarios, based on video analysis of 206 consumer VR applications, primarily games (184) and modeling applications (22). Note we took a bottom-up approach by starting with the scenarios—an instance of a user completing a task of object selection within a specific time frame in a video—rather than predefined task types in the literature. This allowed us to identify new scenarios without being restricted by earlier work.

During this process, we encountered numerous instances that fall under the general definition of object selection, but did not fit into categories of the existing taxonomies [2, 9, 17, 18, 27, 37, 38, 44, 52]. Some examples of such common scenarios include aiming (for shooting and throwing), manipulation of sliders, levers, or knobs for the purpose of selection, and selecting storage units in inventory systems, as well as an explicit distinction between acquisition and system control as different outcomes of the same selection process. As a result, we decided to create new taxonomies (*where* and *how*) that better encompass attributes of the selection process in VR applications. In the following, we present the five steps that we followed during the dataset curation process.

3.1 Phase 1: Identify relevant VR applications

We first identified highly used VR applications across multiple major platforms, including Meta Quest Store, Steam Store, and PlayStation Store. From each platform, we collected a list of top selling and featured applications (as of January 2024), and determined 264 candidate applications after removing duplicates. We filtered out applications with the following exclusion criteria: (1) applications with less than 100 user reviews on a platform, (2) applications without native VR support, (3) applications that require specific hardware to operate (computer mice, keyboards, car steering wheels). This process left us with 206 applications for further processing.

3.2 Phase 2: Identify factors that describe ‘where’ and ‘how’ selection occurs

We started building our coding manual to identify the ‘where’ and ‘how’ based on existing taxonomies, classifications, and attributes. Four authors first collated attributes relevant to VR object selection from the taxonomies in [2,9,17,18,37,38,44,52]. These taxonomies focus on the ‘how’, that is, on the techniques. The variables that were commonly manipulated in user studies listed in Bergström et al. [7], were added to cover attributes of ‘where’. Finally, the same four authors discussed the collated attributes for the coding manual and added attributes that identified based on their experience and initial explorations of the applications in Phase 1, and that were still missing from the initial literature-based list. We used these attributes to both decide on what counts as selection in including scenarios in Phase 3, and as initial labels in the coding manual in Phase 4.

3.3 Phase 3: Extract unique selection scenarios

The goal of this step was to extract as many unique selection scenarios as possible from online videos. A selection scenario refers to a user completing a task of object selection within a specific time frame in a video. We used the YouTube platform to search for videos with the following search terms: ‘no commentary full walkthrough’, ‘tutorial’, ‘review’, and ‘controls’ together with the application name, to find the best-fitting video for an application. We recorded the exact timestamps manually when a selection scenario is identified to directly access the selection events in the database. We also checked the store pages for possible input devices to ensure coverage. To reduce scenario redundancy in the dataset, we excluded entries that might end up with the exact same levels of each attribute on a per-application basis. For example, directly grasping a can or a bottle with a virtual hand is considered as one scenario.

Six authors coded the scenarios for inclusion in the dataset. The authors met weekly to discuss the inclusions, to align the coding of unclear cases until consensus was reached in inclusion, and to adjust the attributes in the coding manual to ensure inclusion of all selection scenarios. This phase resulted to a dataset of 1300+ unique selection scenarios across the applications.

3.4 Phase 4: Identify the attributes of ‘where’ and ‘how’ to establish taxonomies

In this phase, we coded the selection scenarios to detect underlying patterns, and subsequently establish taxonomies that are particularly fitted for VR consumer applications.

Six authors assigned the attributes to the selection data, meeting weekly to adjust the coding manual where needed. The criteria that we applied for arriving to the final taxonomies was, that each category within the taxonomies should capture distinct attributes of a set of selection scenarios that are unavailable in other categories, but the categories do not need to be fully mutually exclusive. This led to two taxonomies on ‘where’ and ‘how’ the selection occurs (more details in Section 4).

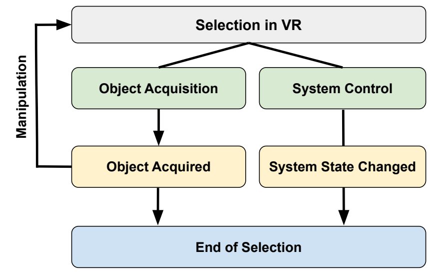

Figure 1: VR selection considers object acquisition (obtaining a virtual object) and system control (non-acquisition changes in the system state). Some selection scenarios involve first acquiring an object and then manipulating that object for system control.

1. The ‘where’ taxonomy captures the spatial characteristics of selections. It addresses questions related to the reference frame of selections, the proximity at which selections take place, and the type of the selectable objects.
2. The ‘how’ taxonomy delves into aspects related to interaction techniques, dissecting the selection process into its indication and confirmation stages, highlighting the interaction method, and picturing the selection outcome.

3.5 Phase 5: Validate the taxonomies

To assess inter-rater reliability (IRR), we performed Fleiss’ kappa analysis. We randomly selected 50 scenarios from the database, ensuring that all taxonomy categories were present. The blank entries for these scenarios were distributed in separate spreadsheets to six authors, including only the timestamped link and a short scenario description. The six authors filled-out eight selected attributes per scenario based on the taxonomies. The resulting values were encoded as categorical labels and organized into a table with 400 subjects (50×8). Fleiss’ kappa analysis revealed a substantial agreement among the six authors ($\kappa = 0.830, z = 125.163, p < 0.001$), indicating that the observed level of agreement is not likely due to chance. This result gives us confidence in the consistency of the assigned attributes in our taxonomy.

4 Taxonomy

We constructed two taxonomies based on the phases outlined in the Methodology (Section 3). The primary goal of these taxonomies is to identify and describe the key patterns in ‘where’ and ‘how’ selection occurs in consumer VR applications. As discussed in Section 3, earlier taxonomies did not many capture use cases in consumer applications.

4.1 Forms of VR Selection

Based on the scope of object selection, we consider selection can be in the form of either *object acquisition*, obtaining the possession of virtual objects in the environment, or *system control*, where selection is used to change the system state to achieve desired outcomes without possessing the target object. A simple example of object acquisition includes grabbing a virtual bottle, while system control could involve selecting buttons on 2D user interfaces.

However, object selection could be more complex in the application scenarios—a user can first acquire a slider, knob, or lever in the VE and manipulate it to a certain configuration to control the

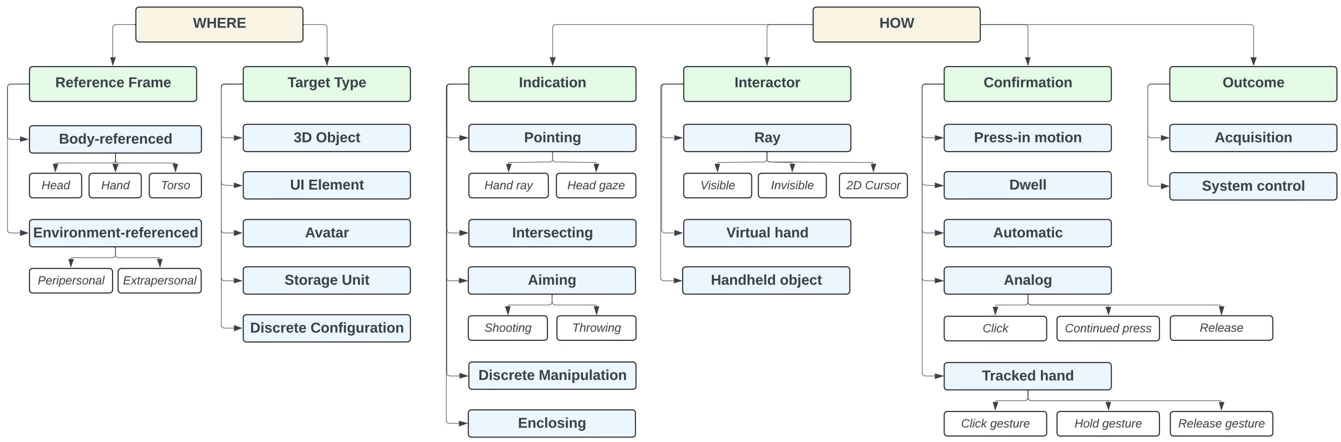

Figure 2: The derived taxonomies: left — **'where'** selection occurs; right — **'how'** selection occurs.

state of the system. In such cases, system control cannot be achieved without first acquiring an object, and both object acquisition and system control are indispensable for target selection (i.e., a desired system state). Similarly, an acquired object can be used to acquire other objects (e.g., grab a fishing net to catch fish).

Figure 1 shows a flowchart of the selection process. In more straightforward scenarios, a selection action ends after objects are acquired or the system state is changed. In the more complex scenarios, acquired (e.g., handheld) objects can be manipulated to achieve additional selection. Such manipulation could include carefully positioning a handheld object to a specific configuration (e.g., position and orientation). It could also be releasing or throwing the handheld object to the target [20, 21]. While such mechanics are prevalent in consumer VR applications and fall under the general definition of selection, they are less represented in the research literature.

4.2 The 'where' taxonomy

The 'where' taxonomy (Figure 2, left) captures the spatial attributes of the selectable objects, covering the reference frame in which the interaction occurs and the target types (i.e., the forms of the selectable objects).

4.2.1 Reference Frame

Reference frame refers to the origin of the coordinate system in which selectable target objects are situated at the time of selection. Inspired by previous works [2, 27, 45], we categorize selectable objects into *body-referenced* and *environment-referenced*. *Body-referenced* objects follow the movement of a user's body part and are divided into:

- *Hand-referenced* objects are attached to one or both hands of the user [2, 45]. An example of this are 2D UI elements on a menu that follows the user's non-dominant hand, selectable with a ray controlled by the user's dominant hand.
- *Head-referenced* objects follow the position and orientation of the user's head [25, 27]. A menu UI that always follows the center of the user's view is an example of head-referenced 2D UI.
- *Torso-referenced* objects follow the estimated position of the user's body. In many applications, a user can look down to see a utility belt with selectable items in an inventory system, attached to the user's torso.

Objects situated in the virtual world's coordinate system are *environment-referenced*, as they do not follow the reference frame of a user's body part. While previous work has provided granular classification of the human interaction space [22], in practice it is difficult to quantify the exact distance from the user's view to objects through video analysis [25], so we divide the environment reference frame into two categories based on the proximity to the user [13]:

- *Peripersonal* objects are within a user's arm reach.
- *Extrapersonal* objects are beyond a user's arm reach, requiring selection techniques designed for interacting with distant targets.

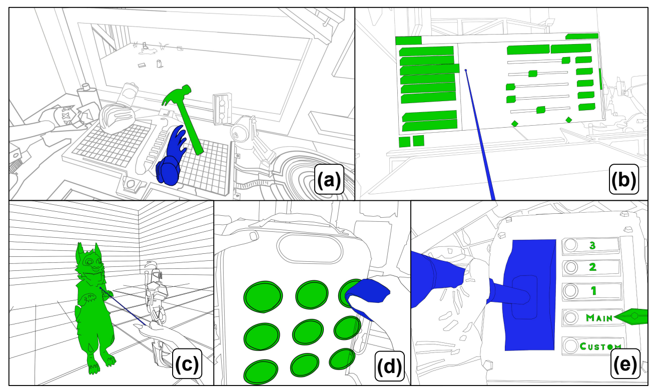

Figure 3: Target Types (green) in consumer VR applications: (a) 3D Object in *Rick and Morty: Virtual Rick-ality*, (b) UI Element in *Breachers*, (c) Avatar in *VRChat*, (d) Storage Unit in *The Walking Dead: Saints & Sinners*, (e) Discrete Configuration in *GORN*.

4.2.2 Target Type

Target types are the forms of the selectable objects presented in VEs (Figure 3). The most common target types are 3D Objects and UI Element. In addition, we identified three main target types (Avatar, Storage Unit, and Discrete Configuration) that are not mutually exclusive from the previous categories but deserve separate categories because they aid in understanding how selection technique is implemented.

- *3D Object* represents three-dimensional virtual entities with width, depth, and height. Many of them mimic objects in the physical world, like bottles, fruit, and furniture.
- *UI Element* is an interactive component that is typically used for system control. Common ones include buttons, menus, and 2D virtual keyboards. These elements can be observed in traditional 2D window layouts inside VR [45].

- *Avatar* refers to the digital representation of living beings in virtual environments. Avatars can be in the form of foes, wildlife, humanoids, or other users in a multi-user application. These type of targets are normally moving in the virtual space.
- *Storage Unit* is an empty unit space that is most frequently observed in inventory systems. They can be 2D and 3D. Users can select a storage unit to store or retrieve items like 3D objects.
- *Discrete Configuration* is position and orientation of a proxy object (e.g., slider, lever or knob) that corresponds to a single state from a set of states [14, 15]. The proxy object is limited in degrees of freedom and is always set to one of the discrete positions. The user manipulates a proxy object and sets it to a discrete position to control the system.

4.3 The 'how' taxonomy

The 'how' taxonomy (Figure 2, right) concerns the interaction techniques that lead to selection, breaking up the process into two stages, and highlighting the interactor (i.e., the selection "tool") and selection outcome. *Indication* and *confirmation* [8, 27, 45] are pivotal for understanding the selection process. The indication stage determines a specific object from a set of selectable objects within the VR environment, commonly achieved through pointing or direct touch (intersection). Once an object is identified, the confirmation stage validates the user's actual intent to make a selection. Physical switch presses, finger pinches, and dwell-based confirmation are widely used in research, so we use them as basis and augment them with additional categories found in consumer VR applications.

4.3.1 Indication

In the indication stage, the user sends their selection intent to the system with specific methods. In our classification, the primary methods include pointing, intersecting, aiming, discrete manipulation, and enclosing. We consider indication being achieved through spatial input modalities such as hand-held controllers, bare hands, or head-mounted displays (HMDs).

- *Pointing* requires the user to control a virtual ray or an extended volume, such as in the shape of a cone, emanating from the input device and following the device's orientation. The virtual ray could also be curved (e.g., parabolic shape). *Hand ray*-based pointing originates at the position of a hand or handheld controller. *Head gaze*-based pointing emanates from the headset position.
- *Intersecting* represents colliding (e.g., "touching") the user's hand or a handheld object with the target.
- *Aiming* includes gamified and physics-based indication methods mostly used for in-game weapons. We identified two approaches. *Shooting* requires aiming an invisible ray while holding an in-game weapon at eye level, partially occluding the target in the process. *Throwing* requires aiming at a target and performing a throwing motion with the intention of releasing a handheld object in that direction with speed.
- *Discrete manipulation* requires physical manipulation of an acquired proxy object (e.g., slider, lever or knob), setting it to one of the available selectable options.
- *Enclosing* requires the user to specify a volume (e.g., a bounding box) that likely includes more than one selectable object.

4.3.2 Interactor

Spatial indication implies that a user-controlled ray, a volume, or a point intersects the volume or area of a selectable object (Figure 4). This interactor follows the user's hand or HMD and can be visible or invisible. In addition to hand and ray, which are commonly researched, we include a handheld interactor category for cases where an object that was previously picked up is used for indication.

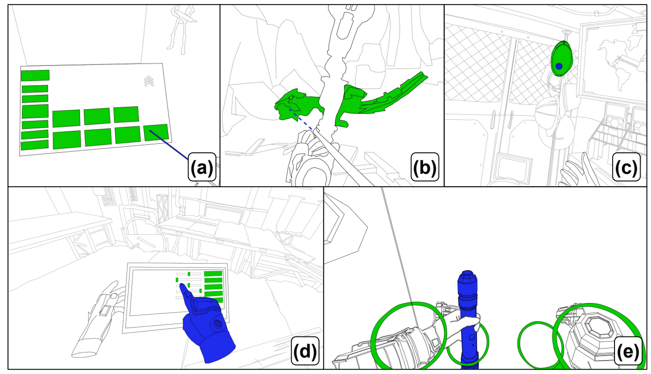

Figure 4: Interactor types (blue) for indication stage: (a) Ray - Visible in *Guardians Frontline*, (b) Ray - Invisible in *Horizon Call of the Mountain*, (c) Ray - 2D Cursor in *I Expect You to Die 2: The Spy and the Liar*, (d) Virtual hand in *STACK*, (e) Handheld object in *Sweet Surrender*.

- *Ray* (or an extended volume) can serve for target selection, usually involving pointing or aiming at objects. We have discerned three types of rays. *Visible* ray extends from its origin point following a user-input direction. *Invisible* ray acts similar to a visible ray, except the user can not see its rendered representation. *2D Cursor* leverages an invisible ray but provides feedback by rendering the point of intersection between the ray and the environment.
- *Virtual hand* is a virtual representation of a user's hand in VR. It is created through handheld controllers or other trackers. Virtual hand is often used to interact with virtual objects directly.
- *Handheld object* refers to a virtual item obtained for selection purposes, like a pen for color dipping or a knob for state selection. In these scenarios, the selection process is dependent on the presence and use of these handheld objects; without them, the selection cannot be completed.

4.3.3 Confirmation

Confirmation delineates the process of triggering the selection the target that is under indication. In the surveyed VR applications, a confirmation command is often issued through analog switch input, hand gestures, press-in motion, and dwelling. The confirmation can also be triggered automatically.

- *Analog* confirmation demands a controller with physical switch (e.g., button, trigger). The switch can be operated in three different forms. *Click* refers to pressing and instantly releasing a physical switch. *Continued press* refers to pressing and holding a physical switch for acquisition or a specific time period. *Release* refers to un-pressing a pressed physical button.
- *Tracked hand* confirmation requires input devices that track the user's hand gestures in the physical environment. Similar to an analog input, there are three different confirmation methods. *Click gesture* refers to performing a gesture momentarily (e.g., pinch, grip) and returning the hand to a non-gesture pose. *Hold gesture* refers to holding a gesture for acquisition or a specific time period. *Release gesture* refers to returning the hand to a non-gesture pose from a predefined gesture (e.g., releasing a pinch to drop an object into a storage unit).
- *Press-in motion* necessitates a physics-based interaction that involves the user's hand trajectory. It refers to the user moving their hand forward while pushing a virtual 3D button until its actuation point, mimicking a real-world button press. In contrast to tracked hand confirmation, specific hand gestures are not necessary; instead, the emphasis lies on the motion itself.

- *Dwell* mandates the user to maintain a ray (or an extended volume) or virtual hand intersected with a target for a predetermined duration. No specific hand gesture is necessary; the emphasis lies in maintaining the intersection.
- *Automatic* confirmation needs no explicit command from the user following the indication; the selection is triggered automatically. For example, touching and intersecting an object often requires no explicit confirmation and are considered as *Automatic*.

4.3.4 Outcome

A selection action could result in an object being acquired by the user (*object acquisition*) and the system state being changed (*system control*), as mentioned in Section 4.1.

5 Visualization Tool

We have systematically gathered and examined 1355 scenarios from 206 consumer VR applications. With researchers and practitioners in mind, we provide an interactive online visualization tool via https://www.eecs.ucf.edu/isuelab/research/vr_selections/ with four main features: (1) the ability to explore and download the open-source selection scenarios database; (2) built-in filters to aid with exploring the database online; (3) the ability to view each scenario through timestamped YouTube video links; (4) contributions and feedback instructions.

The filtering feature allows to narrow a search down to very specific selection scenarios of interest. As an example, to search for occurrences of pointing to a handheld 2D UI menu, user can set the filter to *Dimensions*: 2D, *Reference Frame*: Hand, *Interactor*: Ray - Visible. Filtering also provides information about prevalence of specific selection attribute combinations over others. For example, through our dataset we discovered that on-hand 2D UIs (targets), combined with intersection (indication) occur more frequently (46 entries) than on-hand 2D UIs with pointing (35 entries). Another example is that virtual hand combined with Automatic confirmation is more frequent (116 entries) than its combination with Dwell confirmation (31 entries). Overall, we believe the tool can help researchers and students to explore the landscape of object selection in consumer applications, which will shape the directions for new techniques and study methods.

6 Dataset Analysis and Discussion

We set out to investigate where and how object selection occurs in VR applications to identify and bridge the gaps between research and practice. We took a bottom-up approach by extracting selection scenarios and deriving new taxonomies based on the scenarios.

In this section, we first discuss the identified gaps in 'where' (the task) and 'how' (the technique) between research and practice. The discussions are centered around Figure 5, which shows the occurrences of the 'where' and 'how' taxonomy labels. These numbers represent the percentages of the surveyed consumer VR applications that include interactions associated with specific label types at least once (e.g., 81.07% of applications have pointing with a ray that extends from the user's hand). In addition, Figure 6 shows how Indication-Confirmation and Indication-Target Type are often associated, which provides further insights into the interaction within and between taxonomies. Finally, we discuss the implications of our taxonomies for research and practice.

6.1 The differences in 'where' selection occurs

Both in research and in practice, object selections occur mostly in the Environment on 3D objects (97.57%) and on UI elements (89.87%). However, the data reveals two frequently occurring task attributes that are less studied in research.

6.1.1 Body as a Reference Frame

Body-referenced targets are largely neglected in the VR research literature and taxonomies as compared to environment-referenced targets, save some sparse works, such as investigating how to position targets around the wrists [30, 56]. For example, a review of object selection studies by Bergström et al. [7] had no body-based targets in their sample. In contrast, our data reveals that 59.22% of the applications included body-referenced selection tasks.

Among the collected scenarios, hand-referenced objects (45.63%) predominantly involve one hand grasping a collection of items (such as an inventory backpack or a book with menus), while the other hand is used for pointing or touching to make selections (like in [31]). In certain instances, the targets are directly attached to the hand (e.g., hand-anchored UI like in [28], buttons on the arm). While on-body input and on-skin input as well as target layouts for those are widely studied in non-VR settings, they represent a niche in VR object selection studies. Similarly, head-referenced targets (9.22%) commonly involve placing specific food items in the mouth position for consumption, and a notable number of torso-anchored objects (26.21%) are featured in storage units attached to the user's body, but the VR research field provides only a few examples (e.g., in [63]).

The diverse settings of body-referenced objects (e.g., target number and layout) in the applications necessitate further research to assess their usability. For instance, one from research community could imagine it to be quite cumbersome to target objects accurately on the opposite hand with a pointing ray due to the limited distance. Utilizing one hand to interact with a closely-placed inventory on the other hand may pose issues with potential collisions between handheld controllers. Furthermore, exploring how to position objects around or on the torso warrants additional research.

6.1.2 Avatars and Storage Units as Targets

While target types such as 3D objects and UI elements are widespread as tasks attributes in evaluating object selection in research, in consumer apps a discernible proportion of applications incorporate avatars (50.49%) and storage units (44.17%) as targets.

Avatars are predominantly featured in shooting or combat games. While these selection scenarios may not necessarily prioritize efficiency-maximizing techniques, the distinctive shapes, such as humanoid forms, and dynamic movements presented in this category can be encountered in various conditions. For example, in a multiplayer application like *VRChat*, a user can employ a virtual ray to point and select other players in the VE, and Figure 6-Left shows that aiming and direct intersection are also common to select avatars across applications. In research, targets of avatar size and form are rare, and temporal or moving target selection tasks are represented much less than static targets (with a few examples, like [62]).

Storage units are commonly featured in inventory systems, requiring specific selection methods for retrieving or storing objects, often achieved through intersection. These targets are frequently organized in a grid layout and are located on-body or within the user's arm reach, most often requiring direct intersection to select them (see Figure 6). Such targets demand extra attention from the research community. Additionally, discrete configurations, such as knobs and levers, which require physics-based interactions, also warrant attention in the selection literature (e.g., the ergonomics of such an interaction [40]).

6.2 The differences in 'how' selection occurs

Both in research and in practice, object selections are mostly performed by indicating the target with Intersecting (92.72%) or Pointing with a Hand Ray (81.07%), and confirming the selection by an Analog Click (88.35%). Similarly, the interactors of virtual hand and visible ray are prevalent both in research and in practice. However, the data reveals two frequently occurring technique attributes that are less studied in research. We discuss those next.

Figure 5: Percentage of applications where taxonomy labels have occurred: Reference Frames (a), Target Types (b), Interactors (c), Outcomes (d), Indication stages (e), Confirmation stages (f).

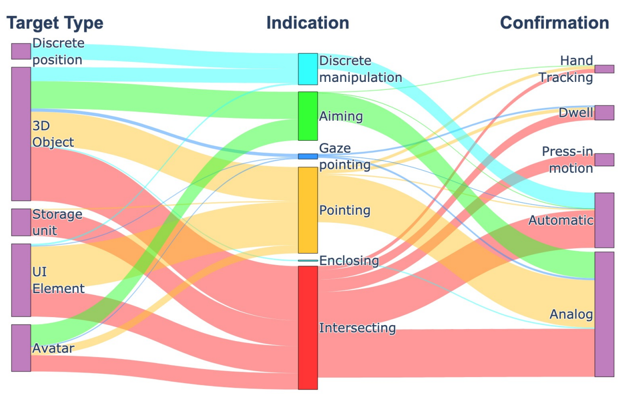

Figure 6: Sankey diagrams show occurrence frequencies between combinations of (Left) Target Types and Indication and (Right) Indication and Confirmation. Note target types are not mutually exclusive.

6.2.1 Automatic and Analog Confirmations

Automatic selection as a confirmation is common in both research and practice (71.84%) when it relates to 3D objects or UI elements as a target type, such as selecting a button on the UI or a virtual object by directly touching it (i.e., Intersecting). However, Automatic selection is common in applications for also dynamic, moving objects (e.g., slicing the cubes in *Beat Saber*) and for Avatar targets (e.g., stabbing an enemy with a knife or hitting with a axe like in *Superhot VR* and *BONELAB*). In contrast, research offers only a few examples on crossing-based selection (e.g., [48]), and even less for moving targets. As Figure 6-Right shows, a considerable proportion of Automatic confirmations are also linked to discrete manipulations, where the selection is automatically triggered by configuring a lever or knob to a specific state; these too are less represented in research.

Analog release (59.22%) is also a prevalent Confirmation technique in consumer applications, with many instances linked to throwing actions. Hence, exploring throwing mechanics in VR might require more attention in research (e.g., [20, 21, 66]). Analog click and analog continued press actions may result in hand tremors, causing the pointing direction or position to shift and potentially miss the target (i.e., the Heisenberg Effect [53]). Therefore, it is important to consider the entire selection process in research, in addition to more common focus on solely Indication.

6.2.2 Invisible Rays and Handheld Objects as Interactors

Invisible rays are frequently used Interactors in practice (47.09%), whereas in research visible rays or Rays with a 2D Cursor are much more common. Invisible rays appear for instance in shooting, where an indication for aiming at the target might be lacking completely, or as a kind of a spotlight, where a potential target is highlighted once the invisible ray hits it (e.g., an interactable object) and otherwise not (e.g., if the ray simply hits an unselectable environmental surface like a wall). Therefore, research should invest in techniques supporting invisible rays as interactors.

The data also reveals significant use of handheld objects for target selection in practice (67.48%)—in these scenarios, users are often required to acquire a handheld object before selecting the target of interest. While one could argue that this process resembles the repetitive selection of multiple objects, we emphasize a crucial distinction: the acquired object may significantly influence the shape or pointing direction of the interactor and the interaction itself, in contrast to virtual hands or rays. For instance, slicing an object with a sword differs from using a hand to cut (it is more similar to cutting objects with an extended arm). Similarly, shooting an enemy with a weapon may alter the pointing direction of the ray compared to emanating the ray from the hand position. In addition, actions such as throwing an object or performing discrete manipulations often necessitate a handheld object. In essence, handheld objects offer new forms of selection that are frequently overlooked in research.

6.3 Implications from Research to Practice

In this work, we have attempted to identify the gaps between research and practice on VR object selection. The above sections highlighted some important gaps both in selection tasks and in selection techniques. However, the tasks from practice cannot be directly adopted in research in order to validate the generalizability of the used tasks or increase their representativeness in practice. Nor can we assume the effectiveness of the research-based designs by directly applying them in practice. Here, we discuss how this work can help bridge the gaps between research and practice, and what are the important challenges in doing so.

6.3.1 Implications for Selection Technique Evaluation

When conceptualizing or assessing new selection techniques, there are typically two types of objectives. Firstly, the research community may delve into pioneering selection techniques and scenarios not yet prevalent in current consumer applications (e.g., selecting objects that are fully occluded [33,64]). Secondly, the focus is on enhancing or optimizing a selection technique for a specific application scenario. A thorough understanding of selection scenarios in consumer applications can greatly benefit the latter type of research.

In approaching an application scenario, it is crucial for researchers and practitioners to initially comprehend its inherent objectives and constraints. For example, if the primary goal is entertainment and the selection was intentionally made to be difficult (e.g., in shooting games), additional enhancements may not be necessary. On the other hand, if usability is the priority, understanding the constraints within the game scenarios becomes essential. Applications must incorporate selection mechanisms into a broader context, such as storytelling, styles, and game mechanics, which introduces new constraints to the design [45]. For instance, if an application requires a fixed number of items (say, 20) to be placed on-hand, it will be useful if there has been research on how to optimally deal with such cases already. Integrating such constraints into the design process, instead of solely focusing on determining the optimal placement positions in an ideal condition without constraints, can make research designs more applicable and pragmatic.

The application scenarios can also inspire how we evaluate the designed interaction techniques. A researcher could aggregate all the intended applications from our dataset for a specific technique and construct a testbed around these applications, rather than relying solely on abstract tasks. For example, object layout could follow a grid pattern, similar to the arrangement of storage units, rather than randomly initializing their locations or placing them on a Fitts' ring.

6.3.2 Implications for Selection Technique Design

Many selection techniques proposed in the research community are under-utilized in consumer applications. While extremely rare, we did find some interesting examples in consumer applications. In *Another Fisherman's Tale*, a user could stretch their arm like springs to grab fruit hanging on a tall tree, like many techniques that aim to extend the reach of virtual hands in research (e.g., Go-Go [36]) but with a different implementation. In *Rec Room VR*, a user can select a cloth in the drawer with a two-step process. The user first chooses the drawer that arranges clothes on a grid and then selects the target cloth, similar to techniques like SQUAD [5] and Expand [10] in research. In *Guardians Frontline*, a user could draw a lasso on the ground to select a group of robots, like in some research papers [46, 64]. We also observed two-handed grab that allows the users to translate, rotate, and scale the selected objects or the world around the user. It was similar to Spindle+Wheel [12], and we observed it in 16 of the modeling applications. This technique is powerful for dealing with distant, occluded, and differently scaled objects, providing an example of a research-originated technique making its way into consumer VR applications.

Considering that the adoption of research output in consumer applications is limited, it may be essential to reconsider how to design techniques in research—given the constraints (e.g., storytelling, game style) in consumer applications, it can be challenging to directly apply some techniques. Additionally, efforts should be directed towards enhancing accessibility of research outputs to the developer community. This could involve creating cheat sheets summarizing selection techniques, providing developers with an overview of current state-of-the-art in research. We believe that collaborative efforts between the research and developer communities are indispensable for the advancement of the field.

6.4 Limitations and Future Work

While we used a rather comprehensive approach to gather data by analyzing video footage from various applications, we acknowledge the inherent limitations of this approach. The videos may not encompass all possible selection scenarios, and we acknowledge that certain instances could be overlooked during the coding and classification process.

We also did not distinguish whether an application was designed for room-scale, sitting, or standing interactions. The current 'where' taxonomy focuses on describing where the selection happens virtually in space without considering physical settings (e.g., obstacles) or previous interactions (e.g., navigation). The physical interaction space will influence the design of techniques. We have included this information in our dataset to motivate future analysis on this issue. Extracting the types of user feedback that follows indication and confirmation stages also has some difficulties. We recorded the presence of Audio and Visual feedback for all scenarios, however, information about Tactile feedback is not available through video footage analysis. Also, features like object sizes are hard to determine through video analysis because they depend on perspectives.

Furthermore, it is crucial to highlight the evolving nature of games. For example, integration of hand and eye tracking in VR has become more prevalent, leading to new incorporation of this feature into existing games (13 of the 206 surveyed applications state eye tracking support on their store pages). Although the goal of this work is to provide a snapshot of the state of VR selection in consumer applications as of spring of 2024, we open-source our dataset and encourage community contributions of newly found selection scenarios to the dataset. This will foster continued exploration and refinement, through an up-to-date online database for future research to build upon our findings. For instance, future research could consider a more systematic comparison of the research output and applications to investigate the delay of research and application transfer, as well as investigate the quality of user experience in applications.

7 Conclusion

In conclusion, our investigation into object selection in virtual reality (VR) has provided valuable insights into the practical selection scenarios and usage of selection techniques in consumer applications. Through a comprehensive survey of over 1300 selection scenarios in games and 3D modeling applications, we derived taxonomies that provide dedicated perspectives on 'where' and 'how' selection occurs in these applications. The taxonomies and our further analysis of the database highlighted scenarios that have been overlooked within the research community and also revealed a slow adaptation of advanced techniques developed in research into consumer applications. Our findings prompt a reconsideration of the design and evaluation of object selection techniques in VR. By understanding the discrepancies between research and consumer applications, researchers and developers can calibrate their methods to create more effective interaction techniques.

Acknowledgments

The authors wish to thank the anonymous reviewers for their feedback. This work was supported in part by a grant from US Army Research Lab (DEVCOM Award W912CG2320004) and by the Pioneer Centre for AI, DNRF grant number P1.

References

[1] D. Ablett, A. Cunningham, G. A. Lee, and B. H. Thomas. Point & portal: A new action at a distance technique for virtual reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 119–128, 2023. doi: 10.1109/ISMAR59233.2023.00026 2

[2] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, May 2013. doi: 10.1016/j.cag.2012.12.003 1, 2, 3, 4

[3] F. Argelaguet, A. Kunert, A. Kulik, and B. Froehlich. Improving co-located collaboration with show-through techniques. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 55–62. IEEE, 2010. doi: 10.1109/3DUI.2010.5444719 2

[4] T. Azai, S. Ushiro, J. Li, M. Otsuki, F. Shibata, and A. Kimura. Tap-tap menu: body touching for virtual interactive menus. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2, 2018. doi: 10.1145/3281505.3281561 2

[5] F. Bacim, R. Kopper, and D. A. Bowman. Design and evaluation of 3d selection techniques based on progressive refinement. *International Journal of Human-Computer Studies*, 71(7-8):785–802, 2013. doi: 10.1016/j.ijhcs.2013.03.003 2, 8

[6] M. Baloup, T. Pietrzak, and G. Casiez. Raycursor: A 3d pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019. doi: 10.1145/3290605.3300331 2

[7] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk. How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, article no. 533, 20 pages. ACM, New York, NY, USA, 2021. doi: 10.1145/3411764.3445193 1, 2, 3, 6

[8] D. A. Bowman, D. B. Johnson, and L. F. Hodges. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '99, 8 pages, p. 26–33. ACM, New York, NY, USA, 1999. doi: 10.1145/323663.323667 2, 5

[9] D. A. Bowman, D. B. Johnson, and L. F. Hodges. Testbed evaluation of virtual environment interaction techniques. *Presence: Teleoperators and Virtual Environments*, 10(1):75–95, Feb. 2001. doi: 10.1162/105474601750182333 2, 3

[10] J. Cashion, C. Wingrave, and J. J. LaViola Jr. Dense and dynamic 3d selection for game-based virtual environments. *IEEE transactions on visualization and computer graphics*, 18(4):634–642, 2012. doi: 10.1109/TVCG.2012.40 8

[11] Y. Chen, J. Sun, Q. Xu, E. Lank, P. Irani, and W. Li. Empirical evaluation of moving target selection in virtual reality using egocentric metaphors. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part IV 18*, pp. 29–50. Springer, 2021. doi: 10.1007/978-3-030-85610-6_3 2

[12] I. Cho and Z. Wartell. Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments. In *2015 IEEE symposium on 3D User Interfaces (3DUI)*, pp. 133–136. IEEE, 2015. doi: 10.1109/3DUI.2015.7131738 8

[13] J. Cléry, O. Guipponi, C. Wardak, and S. B. Hamed. Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: knowns and unknowns. *Neuropsychologia*, 70:313–326, 2015. doi: 10.1016/j.neuropsychologia.2014.10.022 4

[14] R. Dachselt and M. Hinz. Three-dimensional widgets revisited - towards future standardization. 2008. 5

[15] R. Dachselt and A. Hübner. Three-dimensional menus: A survey and taxonomy. *Computers & Graphics*, 31(1):53–65, Jan. 2007. doi: 10.1016/j.cag.2006.09.006 2, 5

[16] H. G. Debarba, S. Perrin, B. Herbelin, and R. Boulic. Embodied interaction using non-planar projections in immersive virtual reality. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, pp. 125–128, 2015. doi: 10.1145/2821592.2821603 2

[17] K. R. Dillman, T. T. H. Mok, A. Tang, L. Oehlberg, and A. Mitchell. A visual interaction cue framework from video game environments for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Montreal QC Canada, Apr. 2018. doi: 10.1145/3173574.3173714 2, 3

[18] N. Elmqvist and P. Tsigas. A taxonomy of 3D occlusion management techniques. In *2007 IEEE Virtual Reality Conference*, pp. 51–58, Mar. 2007. ISSN: 2375-5334. doi: 10.1109/VR.2007.352463 2, 3

[19] S. Frees, G. D. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(1):2–es, 2007. doi: 10.1145/1229855.1229857 2

[20] A. Ghasemaghaei, Y. Hmaiti, M. Maslych, E. S. Martinez, and J. J. LaViola. Throwing in virtual reality: Performance and preferences across input device configurations. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 897–898, 2024. doi: 10.1109/VRW62533.2024.00245 4, 7

[21] A. Ghasemaghaei, M. Maslych, Y. Hmaiti, E. S. Martinez, and J. LaViola. Towards better throwing: A comparison of performance and preferences across point of release mechanics in virtual reality. In *Graphics Interface 2024 Second Deadline*, 2024. 4, 7

[22] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold Jr, M. Durbin, M. S. Edmonson, J. Fischer, D. Hymes, S. T. Kimball, et al. Proxemics. *Current anthropology*, 9(2/3):83–108, 1968. 4

[23] D. Han, D. Kim, and I. Cho. Portal: Portal widget for remote target acquisition and control in immersive virtual environments. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, VRST '22, article no. 16, 11 pages. ACM, New York, NY, USA, 2022. doi: 10.1145/3562939.3565639 2

[24] R. Henrikson, T. Grossman, S. Trowbridge, D. Wigdor, and H. Benko. Head-coupled kinematic template matching: A prediction model for ray pointing in vr. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. doi: 10.1145/3313831.3376489 2

[25] Y. Hmaiti, M. Maslych, E. M. Taranta, and J. J. LaViola. An exploration of the effects of head-centric rest frames on egocentric distance judgments in vr. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 263–272, 2023. doi: 10.1109/ISMAR59233.2023.00041 4

[26] M. Khamis, C. Oechsner, F. Alt, and A. Bulling. Vrpursuits: Interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 international conference on advanced visual interfaces*, pp. 1–8, 2018. doi: 10.1145/3206505.3206522 2

[27] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017. 1, 2, 4, 5

[28] I. Lediaeva and J. LaViola. Evaluation of body-referenced graphical menus in virtual environments. In *Graphics Interface 2020*, 2020. 6

[29] J. Li, I. Cho, and Z. Wartell. Evaluation of cursor offset on 3d selection in vr. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*, SUI '18, 10 pages, p. 120–129. ACM, New York, NY, USA, 2018. doi: 10.1145/3267782.3267797 2

[30] Z. Li, J. Chan, J. Walton, H. Benko, D. Wigdor, and M. Glueck. Armstrong: An empirical examination of pointing at non-dominant arm-anchored uis in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. doi: 10.1145/3411764.3445064 2, 6

[31] R. W. Lindeman, J. L. Sibert, and J. K. Hahn. Hand-held windows: towards effective 2d interaction in immersive virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pp. 205–212. IEEE, 1999. doi: 10.1109/VR.1999.756952 6

[32] Y. Lu, C. Yu, and Y. Shi. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality. In *2020 IEEE Conference on virtual reality and 3D user interfaces (VR)*, pp. 35–43. IEEE, 2020. doi: 10.1109/VR46266.2020.00021 2

[33] M. Maslych, Y. Hmaiti, R. Ghamandi, P. Leber, R. K. Kattoju, J. Belga, and J. J. LaViola. Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 460–470. IEEE, 2023. doi: 10.1109/VR55154.2023.00061 2, 8

[34] J. McIntosh, H. D. Zajac, A. N. Stefan, J. Bergström, and K. Hornbæk. Iteratively adapting avatars using task-integrated optimisation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 709–721, 2020. doi: 10.1145/3379337.3415832 2

[35] D. Mendes, M. Sousa, R. Lorena, A. Ferreira, and J. Jorge. Using custom transformation axes for mid-air manipulation of 3d virtual objects. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–8, 2017. doi: 10.1145/3139131.3139157 2

[36] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. 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*, UIST '96, 2 pages, p. 79–80. ACM, New York, NY, USA, 1996. doi: 10.1145/237091.237102 2, 8

[37] I. Poupyrev and T. Ichikawa. Manipulating objects in virtual worlds:

categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1):19–35, Feb. 1999. doi: 10.1006/jvlc.1998.0112 2, 3

[38] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. A framework and testbed for studying manipulation techniques for immersive VR. In *Proceedings of the ACM symposium on virtual reality software and technology*, VRST '97, pp. 21–28. ACM, New York, NY, USA, 1997. doi: 10.1145/261135.261141 2, 3

[39] J. Schjerlund, K. Hornbæk, and J. Bergström. Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021. doi: 10.1145/3411764.3445759 2

[40] D. Schön, T. Kosch, F. Müller, M. Schmitz, S. Günther, L. Bommhardt, and M. Mühlhäuser. Tailor twist: Assessing rotational mid-air interactions for augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023. doi: 10.1145/3544548.3581461 6

[41] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality. In *Proceedings of the 2020 chi conference on human factors in computing systems*, pp. 1–13, 2020. doi: 10.1145/3313831.3376438 2

[42] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International journal of human-computer studies*, 61(6):751–789, 2004. doi: 10.1016/j.ijhcs.2004.09.001 2

[43] M. Speicher, A. M. Feit, P. Ziegler, and A. Krüger. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3174221 2

[44] A. Steed. Towards a general model for selection in virtual environments. In *3D User Interfaces (3DUI'06)*, pp. 103–110, Mar. 2006. doi: 10.1109/VR.2006.134 2, 3

[45] A. Steed, T. M. Takala, D. Archer, W. Lages, and R. W. Lindeman. Directions for 3d user interface research from consumer vr games. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4171–4182, Nov. 2021. doi: 10.1109/TVCG.2021.3106431 1, 4, 5, 8

[46] R. Stenholt. Efficient selection of multiple objects on a large scale. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, pp. 105–112, 2012. doi: 10.1145/2407336.2407357 2, 8

[47] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–9, 2017. doi: 10.1145/3139131.3139149 2

[48] H. Tu, S. Huang, J. Yuan, X. Ren, and F. Tian. Crossing-based selection with virtual reality head-mounted displays. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–14, 2019. doi: 10.1145/3290605.3300848 2, 7

[49] L. Wang, X. Liu, and X. Li. Vr collaborative object manipulation based on viewpoint quality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 60–68. IEEE, 2021. doi: 10.1109/ISMAR52148.2021.00020 2

[50] M. Wang, Z.-M. Ye, J.-C. Shi, and Y.-L. Yang. Scene-context-aware indoor object selection and movement in vr. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 235–244. IEEE, 2021. doi: 10.1109/VR50410.2021.00045 2

[51] M. Weise, R. Zender, and U. Lucke. A comprehensive classification of 3d selection and manipulation techniques. In *Proceedings of Mensch Und Computer 2019*, MuC '19, 12 pages, p. 321–332. ACM, New York, NY, USA, 2019. doi: 10.1145/3340764.3340777 1

[52] M. Weise, R. Zender, and U. Lucke. A comprehensive classification of 3D selection and manipulation techniques. In *Proceedings of Mensch und Computer 2019*, MuC'19, pp. 321–332. ACM, New York, NY, USA, Sept. 2019. doi: 10.1145/3340764.3340777 2, 3

[53] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–10, 2020. doi: 10.1145/3313831.3376876 7

[54] Z. Wu, D. Yu, and J. Goncalves. Point-and volume-based multi-object acquisition in vr. In *IFIP Conference on Human-Computer Interaction*, pp. 20–42. Springer, 2023. doi: 10.1007/978-3-031-42280-5_2 2

[55] W. Xu, H.-N. Liang, A. He, and Z. Wang. Pointing and selection methods for text entry in augmented reality head mounted displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 279–288. IEEE, 2019. doi: 10.1109/ISMAR.2019.00026 2

[56] X. Xu, A. Dancu, P. Maes, and S. Nanayakkara. Hand range interface: Information always at hand with a body-centric mid-air input surface. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–12, 2018. doi: 10.1145/3229434.3229449 2, 6

[57] Y. Yan, C. Yu, X. Ma, S. Huang, H. Iqbal, and Y. Shi. Eyes-free target acquisition in interaction space around the body for virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3173616 2

[58] D. Yu. *Enhancing Mid-Air Selection and Manipulation for Complex Virtual Reality Interaction*. The University of Melbourne, 2023. 1, 2

[59] D. Yu, R. Desai, T. Zhang, H. Benko, T. R. Jonker, and A. Gupta. Optimizing the timing of intelligent suggestion in virtual reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–20, 2022. doi: 10.1145/3526113.3545632 2

[60] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. doi: 10.1145/3355089.3356544 2

[61] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021. doi: 10.1145/3411764.3445343 2

[62] D. Yu, B. V. Syiem, A. Irlitti, T. Dingler, E. Velloso, and J. Goncalves. Modeling temporal target selection: A perspective from its spatial correspondence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023. doi: 10.1145/3544548.3581011 6

[63] D. Yu, Q. Zhou, T. Dingler, E. Velloso, and J. Goncalves. Blending on-body and mid-air interaction in virtual reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 637–646. IEEE, 2022. doi: 10.1109/ISMAR55827.2022.00081 2, 6

[64] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves. Fully-occluded target selection in virtual reality. *IEEE transactions on visualization and computer graphics*, 26(12):3402–3413, 2020. doi: 10.1109/TVCG.2020.3023606 2, 8

[65] Q. Zhou, D. Yu, M. N. Reinoso, J. Newn, J. Goncalves, and E. Velloso. Eyes-free target acquisition during walking in immersive mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3423–3433, 2020. doi: 10.1109/TVCG.2020.3023570 2

[66] T. Zindulka, M. Bachynskyi, and J. Müller. Performance and experience of throwing in virtual reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–8, 2020. doi: 10.1145/3313831.3376639 7